



Enzymatic Conversion of Red Blood Cell Antigens to the Universal Blood Group O

Jensen, Mathias

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Jensen, M. (2023). *Enzymatic Conversion of Red Blood Cell Antigens to the Universal Blood Group O*. DTU Bioengineering.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Enzymatic Conversion of Red Blood Cell Antigens to the Universal Blood Group O

PhD thesis

Mathias Jensen

Protein Glycoscience and Biotechnology

Department of Biotechnology and Biomedicine

DTU Bioengineering

October 2023

Preface

This PhD thesis presents and summarizes my research, which was carried out in the Protein Glycoscience and Biotechnology group at the Department of Biotechnology and Biomedicine, DTU Bioengineering, under the supervision of Professor Maher Abou Hachem, and co-supervision of Professor Jens Preben Morth. This project was performed in collaboration with the Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University, headed by Professor Martin L. Olsson, and in collaboration with the DTU Structural Biology Core, headed by Professor Jens Preben Morth. The project was funded by the Independent Research Fund Denmark, under the Technology and Production Research Project proposal number 0136-00086B. Enzymes described herein were cloned by Bashar Shuokar in a previous PhD project supported by the Ministry of Higher Education and Scientific Research of Iraq.

The Project has resulted in the following manuscript, which has been included as a chapter in this thesis:

Mathias Jensen, Linn Stenfelt, Jennifer Ricci Hagman, Jakob Michael Pichler, Julia Weikum, Tine Sofie Nielsen, Annika Hult, Jens Preben Morth, Martin L. Olsson, Maher Abou Hachem. *Targeting extended blood group antigens by Akkermansia muciniphila exoglycosidases unveils a missing link for generating ABO-universal donor blood*. Currently in revision. (2023).

Results from the project have also resulted in 5 poster presentations and 2 oral flash presentations at various conferences as indicated in the Appendix.

Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my main supervisor, Professor Maher Abou Hachem, for employing me onto this extremely exciting project. It has been an incredible journey, not the least because of your motivating enthusiasm and great support. I want to thank you for pushing me and holding me to a high standard, I have learned and grown tremendously from this experience. You're guidance has been instrumental for everything I have achieved during this PhD project. I sincerely appreciate you always making time to give feedback or have discussions with me, despite your busy schedule. Thank you for your ideas, your patience, and for sharing your admirable scientific knowledge. You have been truly inspiring.

I would also like to thank everyone at the Division of Hematology and Transfusion Medicine in Lund for a great collaboration. In particular, I want to thank Dr Linn Stenfelt for the fantastic work on the clinical parts of the project. There is no doubt you have poured a lot of time and effort into this project, contributing to the high impact of the work. I would also like to thank Jennifer Ricci Hagman and Annika Hult for their additional contributions to this project, both at the very start and at the end. Finally, I would like to thank Professor Martin L. Olsson for the plethora of insightful and thought-provoking discussions we have had throughout our meetings over these 3 years, which highlighted the strengths of doing cross-disciplinary research projects.

I would also like to thank my co-supervisor, Professor Jens Preben Morth, for teaching me the art of crystallography. Your joy and enthusiasm for the field is contagious, and I have no doubt learned a lot. Structure refinement seemed daunting at a first glance, but your help swiftly gave me the boost I needed to get the hang of it, and more importantly to enjoy it as well.

Throughout this project I have received invaluable help from our laboratory technician, Tine. I want to thank you for always stepping in when I become too busy to do it all. Additionally, thank you for introducing me to the laboratory and the group when I was first starting, you were a great help. I would also like to thank the laboratory trainees, Suzana and Nina, who worked with me during parts of the project and helped boost productivity.

Of course, my time here would not have been nearly as enjoyable without all the members of both the Protein Glycoscience and Biotechnology group, and our neighbor group, Protein Chemistry and Enzyme Technology, headed by Professor Birte Svensson. You are all wonderful colleagues and great friends. Thank you for the many great discussions, funny stories, and for always being supportive and helpful. I want to specially thank my office mate, Michi, for giving me a warm welcome and for sharing your expertise throughout these years. I want to thank Yun and Simone for reminding me to also have some fun every now and then, and for an especially great time in Boston. I want to thank Sanchari for being helpful earlier this year when I was submitting my very first structures to the PDB. I also want to thank Ásta for bringing in new and exciting energy to the office, and for being great company during our stay in Göteborg.

Last but certainly not least, I would like to thank my family for their everlasting love and support, and for always believing in me throughout my whole life. I would also like to thank my incredible girlfriend, Chelina, for always having my back. Thank you for understanding when things got very busy, and for always believing in me. Pursuing a PhD would not have gone this smoothly without you.

Abstract

Every year, millions of blood transfusions are performed worldwide. Miss-matching of donor blood type to the recipient can result in hemolytic transfusion reactions, which can be deadly. These reactions occur when recipient plasma antibodies recognize non-self ABO antigens on the surfaces of donor Red Blood Cells (RBCs). Therefore, correct matching of donor and recipient ABO blood group is essential for safe transfusions. Increasing blood demand for the ageing populations, and limited donor numbers presents a blood supply challenge. An attractive solution encompasses the use of exo-glycosidases to convert the A and B antigens into the universal type, O, in order to maintain stocks of only universal blood. This would alleviate issues related to keeping stocks of different blood groups and minimize spill of unused units. Efficient enzymes targeting the A and B antigens have been reported, but unexplained incompatibility issues between recipient plasma and the Enzyme Converted O (ECO) RBCs, raise questions about the safe implementation of the method. Strikingly, the clinical significance of four extensions of type A and B antigens on transfusion compatibility has remained unexplored.

In this thesis, we discover and employ two cocktails of 5 enzymes, collectively targeting both the canonical and the extended A and B antigens. We show for the first time that targeting the extended A and B antigens significantly decreases the level and severity of positive crossmatches, as a proxy for transfusion compatibility, between ECO RBCs and recipient plasma, in comparison to only targeting the canonical A and B antigens. We highlight the complex nature of the glycan phenotype that decorate surface glycoconjugates of human RBCs and the matching complexity and diversity in plasma antibody repertoires. This work contributes to changing the current paradigm in regards to what constitutes universal donor blood. Additionally, we provide evidence for the evolutionary adaptation of enzymes from the mucin-associated human gut microbiota symbiont, *Akkermansia muciniphila*, to efficiently target human glycoconjugates. We report the crystal structures of three novel enzymes that exhibit high efficacy in the conversion of ABO antigens on RBCs, allowing us to dissect enzymatic signatures that correlate with activity on RBCs. Herein, we report the first structure of a blood-active GH110 with an ancillary CBM-51 domain. We also report a structure of a GH20 β -N-acetylhexosaminidase appended to a novel putative CBM with GalNAc binding activity. Altogether, we provide novel insights into structural features that correlate to high efficacy on RBC surface glycoconjugates and unravel the hitherto unknown importance of targeting the extended A and B antigens. Here, we set the foundation for a new generation of enzymes for truly ABO-universal enzymatically generated blood.

Dansk Resumé

Hvert år udføres mere end 100 millioner livreddende blodtransfusioner, hvilket har givet det en plads på Verdenssundhedsorganisationens liste over essentielle lægemidler. Korrekt matchning af ABO-blodgruppen mellem donorer og modtagere er vigtigt, da en mis-match kan føre til alvorlige, potentielt dødelige hæmolytiske transfusionsreaktioner. Disse reaktioner opstår, når immunsystemet genkender fremmede A- eller B- blodgruppeantigener fra donorens blodlegemer. En mangel på sjældne blodtyper, samt en aldrende befolkning, har gjort det svært at opretholde en tilstrækkelig blodforsyning, og har ledet til en stigende efterspørgsel på bloddonorer. En innovativ løsning indebærer brugen af enzymer til at omdanne A- og B- antigener til den universelle blodtype O, som kan modtages af alle uden problemer. Dette ville nedsætte blodmangel, minimere spild af dyrebart doneret blod og reducere sikkerhedsrisici forbundet med mis-match. I de seneste 40 år, har forskning ført til opdagelser af enzymer, der omdanner A- og B-antigener, men uforklarlige kompatibilitetsproblemer mellem enzym-konverteret O (ECO) blodlegemer og modtagerens plasma, har rejst bekymringer om metoden. Forbavsende, er den potentielle påvirkning af fire forlængelser af A- og B- antigenerne på transfusionskompatibilitet, forblevet udforsket.

I dette projekt opdager og anvender vi en cocktail af enzymer fra menneskets tarmmikrobiota, der kollektivt omdanner både A, B og alle kendte forlængelser. Vi demonstrerer for første gang, at omdannelsen af disse forlængelser af A- og B-antigener reducerer både forekomsten og alvorligheden af inkompatibilitet mellem ECO-blod og modtagerens plasma. Vi udfordrer den nuværende forståelse af sikkert, enzymatisk konverteret blod, og giver indblik i potentialet af enzymer fra tarmmikrobiota-symbionten, *Akkermansia muciniphila*, som er udviklet til effektivt at nedbryde disse antigener. Bestemmelse af krystalstrukturer af tre nye enzymer, der er yderst effektive til at omdanne ABO-antigener på røde blodlegemer, har vist os vigtige strukturelle træk, der korrelerer med effektiv antigenkonvertering på røde blodlegemer. Vores forskning tilbyder nye indsigter i de strukturelle egenskaber ved enzymer, der gør dem yderst effektive til omdannelsen af blod antigener på blodlegemer, og understreger vigtigheden af at omdanne forlængede A- og B-antigener. Vi lægger grundlaget for en ny generation af enzymer, der bringer os tættere på ægte universelt blod, gennem enzymatisk behandling.

Abbreviations

RBCs	Red Blood Cells
WHO	World Health Organization
ECO	Enzyme Converted O
HGM	Human Gut Microbiota
CAZyme	Carbohydrate Active enZymes
GH	Glycoside Hydrolase
GT	Glycoside Transferase
PL	Polysaccharide Lyase
CE	Carbohydrate Esterase
AA	Auxiliary Activities
CBM	Carbohydrate binding Module
Gal	Galactose
Fuc	Fucose
Neu5Ac	(5-) <i>N</i> -acetylneuraminic acid
GalNAc	<i>N</i> -acetylgalactosamine
GlcNAc	<i>N</i> -acetylglucosamine
LacNAc	<i>N</i> -acetylactosamine
LNB	Lacto- <i>N</i> -Biose
GNB	Galacto- <i>N</i> -biose
GIT	Gastro Intestinal Tract
PTS	Proline-Threonine-Serine
Le ^{a/b/x/y}	Lewis A/B/X/Y
HMO	Human Milk Oligosaccharide
GLP	Glucagon-Like Peptide
ICAM	IntraCellular Adhesin Molecule
ABC	ATP-Binding Cassette
A β	β -Amyloid
PCM	Porcine Colonic Mucin
PGM	Porcine Gastric Mucin
Ig	Immunoglobulin
ISBT	International Society of Blood Transfusion

Contents

Preface	i
Acknowledgements	ii
Abstract	iii
Dansk Resumé	iv
Abbreviations	v
Chapter 1 - Introduction	1
1.1. Universal donor blood	1
1.2. Carbohydrate nomenclature	2
1.3. Carbohydrate Active Enzymes	4
1.3.1. Glycoside hydrolases	4
1.3.2. Carbohydrate Binding Modules	7
1.3.3. CBMs with human glycan affinities	8
1.4. The Human Gut Microbiota	9
1.4.1. Human gut biogeography	9
1.4.2. The mucus layer	11
1.4.3. Mucin O-glycan Structures	12
1.5. Glycan Utilization by HGM	15
1.5.1. <i>Akkermansia muciniphila</i>	15
1.5.2. Mucin foraging by <i>A. muciniphila</i>	17
1.6. Red Blood Cell Glycans	19
1.6.1. The ABO blood group	20
1.6.2. AB antigen extensions	20
1.6.3. The Lewis, SID, and I blood groups	21
1.6.4. Globoside, P1PK, NOR, and FORS blood group	22
1.6.5. Clinical Significance of blood groups	23
1.7. Enzymatic Conversion of ABO blood antigens	26
1.7.1 First generation enzymatic conversion of AB antigens	28
1.7.2 Bacterial enzymes for universal blood	30
1.7.3. Endo-active enzyme for RBC antigen conversion	31
1.7.4. Enzymes from the HGM for ABO antigen conversion	32
Chapter 2 – Manuscript: Targeting extended blood group antigens by <i>Akkermansia muciniphila</i> exoglycosidases unveils a missing link for generating ABO-universal donor blood	35

Chapter 3 – Structural analysis of <i>Akkermansia muciniphila</i> enzymes active on RBCs.....	116
3.1. Methods	116
3.1.1. Crystallization and structure determination of <i>AmGH109B</i>.....	116
3.1.2. AlphaFold model predictions.....	116
3.2. <i>Akkermansia muciniphila</i> encodes two distant GH110 enzymes.....	117
3.3. Active site loop divergence of <i>A. muciniphila</i> GH20 enzymes.....	122
3.4. <i>AmGH20E</i> shares characteristics with a 6S-GlcNAcase.....	126
3.5. <i>AmGH109B</i> shortened loops confer a more open active site compared to the previously described blood converting GH109 enzyme.....	128
3.6. Conclusion	133
Chapter 4 – Discussion	134
4.1. Enzyme source and discovery strategy.....	134
4.2. Conversion of ABO blood group antigens on RBCs	135
4.3. Compatibility of ECO RBCs with recipient plasma	137
4.4. Structural signatures of enzymes active on RBCs	138
Conclusions and perspectives	141
References	142
Appendix – Poster Abstracts.....	160

Chapter 1 - Introduction

1.1. Universal donor blood

Transfusions of Red Blood Cells (RBCs) from healthy donors are considered an essential medicine according to the World Health Organization (WHO)¹, with roughly 118 million transfusions performed yearly. Correct blood type-matching between recipient and donor is crucial in order to avoid hemolytic immune reactions, with potential fatal consequences². The ABO blood group, discovered in 1900 by the Nobel Prize winner Karl Landsteiner after observing various patterns of hemagglutination when mixing RBCs with different plasma samples^{3,4}, is considered the most critical blood group to match⁵. The ABO blood group is defined by carbohydrates coating the surfaces of RBCs. Group O blood contains the H antigen, precursor to both the A and B antigens, and can be received by any recipient. Group A and B individuals express the A and B antigens, resulting from the addition of an α -N-acetylgalactosamine or α -galactose to the H antigen, respectively⁶. Furthermore, 3 extensions of the A antigen⁷⁻⁹ and 1 extension of the B antigen¹⁰, are currently known, and these will be detailed below. The A and B antigens are recognized by circulating antibodies in the plasma of individuals lacking the corresponding antigen¹¹. Hemolytic transfusion reactions occur when antibodies recognize their antigens and initiate a complement cascade, lysing the RBCs in circulation¹². In 1963, mentions of enzymes with activity against the A and B antigens surfaced¹³, followed by a report in 1975 of a Coffee bean α -galactosidase able to convert the B antigen on RBCs into the underlying H antigen¹⁴. It was proposed that exo-glycosidases could potentially be used to convert both blood group A and B antigens into the universal group O. Since then, several enzymes have been identified with activity against the A and B antigens, respectively, resulting in seemingly complete conversion of A and B RBCs into type O. Despite these successes, limited clinical data show that residual crossmatch reactivity remains between Enzyme Converted O (ECO) RBCs and type O recipient plasma¹⁵, and the practice has not been implemented in a clinical setting.

Recent studies by Rahfeld *et al.*¹⁶ and Anso *et al.*¹⁷, showcase the high efficiency of enzymes from members of the Human Gut Microbiota (HGM) against ABO blood group antigens. The ABO antigens are present in mucins lining the gut¹⁸, which is utilized as a carbon source for certain members of the HGM¹⁹. This thesis reports the discovery of enzymes that convert the A and B antigens, as well as all known extensions thereof. Hitherto, the significance of these extended structures for transfusion crossmatch reactions has remained unknown²⁰. Chapter 1 of this thesis starts with a basic introduction to carbohydrates and carbohydrate-active enzymes, covering essential background knowledge. A brief description of the HGM with a focus on the mucus layer and associated glycan convergence by the commensal member, *Akkermansia muciniphila*, will follow. Finally, a description of the carbohydrate-based blood groups of RBCs and their clinical significance will set the stage, followed by an up-to-date review of enzymatic conversion of ABO antigens. The enzymatic treatment with the discovered enzymes will be described in Chapter 2. Chapter 3 includes a structural analysis of the discovered enzymes to elucidate signatures of enzymatic activity on RBCs. Finally, chapter 4 will entail a discussion of the research in this thesis.

1.2. Carbohydrate nomenclature

Carbohydrates are the most abundant organic molecules in nature. In their basic form, carbohydrates are polyhydroxy aldehydes or ketones (Figure 1a), classified according to the number of carbon atoms they contain²¹. The simplest carbohydrates are trioses (3 carbons), followed by tetroses (4 carbons), pentoses (5 carbons), hexoses (6 carbons), and so on. Collectively, these are referred to as “monosaccharides”, and can be linked to one another by glycosidic bonds to form carbohydrate oligomers or polymers. In solution, pentoses and larger monosaccharides exist in either cyclic or open forms. In their open form, monosaccharides contain a carbonyl group present as part of either the aldehyde (in aldoses) or ketone (in ketoses). Following organic chemistry nomenclature, the carbonyl carbon is referred to as C1 in aldoses or C2 in ketoses, with the carbon number increasing along the hydroxymethyl chain. In the cyclic form, the carbonyl carbon will react with one of the hydroxyl groups, most commonly on the C4 or C5, forming a hemiacetal group. The resulting five- and six-membered rings are termed furanoses and pyranoses, respectively. The absolute configuration of a monosaccharide is determined by the furthest stereogenic carbon from the C1 in a Fischer projection²² (Figure 1b). If the hydroxyl group is located to the right of the last stereogenic carbon, the absolute configuration is termed D (from the Latin “dexter”, meaning “right”). If the hydroxyl group is located to the left, the absolute configuration is termed L (from the Latin “laevus” meaning “left”). D and L sugars are enantiomers, meaning they are mirror images of one another.

In the cyclic form, the carbonyl carbon is anomeric, and the position of the substituent hydroxyl group compared to the C5 in furanoses, or C6 in pyranoses, dictates the anomeric center configuration²³ (Figure 1c). If the hydroxyl group at the anomeric center and the C5/ C6 are located on opposite sides of the ring, the configuration is termed “ α ”. If they are on the same face of the ring, the configuration is termed “ β ”. The anomeric carbon forms glycosidic linkages to other monosaccharides in the case of carbohydrate polymers. The nomenclature of these bonds refers to the configuration of the anomeric carbon, and the specific carbons involved in the linkage. For example, a β -1,3 linkage refers to a linkage in the β orientation, from the C1 of one monosaccharide to the C3 of another monosaccharide.

Carbohydrates consisting of 2 linked monosaccharides are termed disaccharides. Thereafter, carbohydrate chains linked together in a linear or branched polymer are referred to as oligosaccharides (usually <12 monosaccharides) or polysaccharides (usually \geq 12 monosaccharides)²⁴. Carbohydrates may be covalently attached to other organic molecules, resulting in glycoconjugates. The most common glycoconjugates are glycoproteins and glycolipids, but carbohydrates can also be attached to small organic moieties, termed glycosides (e.g. some plant secondary metabolites and toxins). The carbohydrate portion of glycoconjugates is termed glycans. Glycoproteins and glycolipids play a structural and modulatory role in biological systems, including physical protection^{25,26}, solubility^{27,28}, glycoprotein folding^{29,30}, cell communication³¹, and more³². Of key interest for this thesis, is the role of glycoconjugates in the process of intrinsic and extrinsic recognition by antibodies and enzymes.

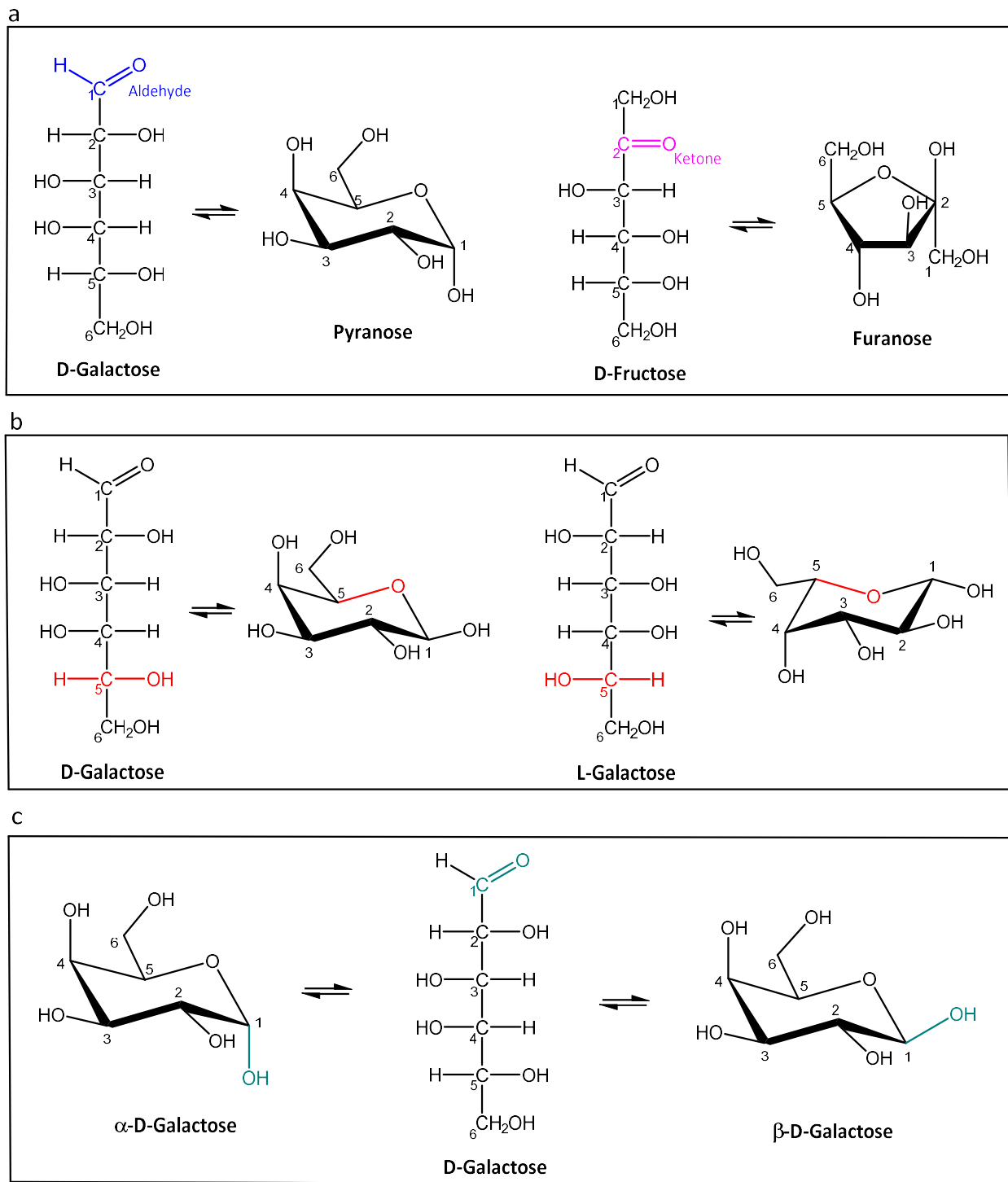


Figure 1. Monosaccharide structure and nomenclature. **a.** Open and cyclic forms of hexoses D-galactose and D-fructose. Monosaccharides contain a carbonyl carbon as part of either an aldehyde (blue) or ketone (pink) group, resulting in the six-membered pyranose or five membered furanose cyclic structures, respectively. **b.** The C5 (red) position in a Fischer projection determines absolute configuration, resulting in either D- or L-galactose. D- and L-configurations are enantiomers to each other. **c.** In the cyclic form, the position of the hydroxyl group (teal) on the C1, in relation to the C6 determined the configuration of the carbonyl carbon as either α - or β -D-galactose. This figure was drawn using ChemDraw v.18.1.

1.3. Carbohydrate Active Enzymes

Enzymes are protein catalysts for specific biochemical reactions. Enzymes that catalyze reactions involved in the breakage, formation, or modification of glycosidic bonds, are termed Carbohydrate Active enzymes (CAZymes). There are five classes of CAZymes annotated in the CAZy database³³ (<http://www.cazy.org/>); glycosyl hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary active (AA) enzymes (Table 1). GHs are enzymes that facilitate the breakage or remodeling of glycosidic linkages by hydrolysis, phosphorolysis, or transglycosylation^{34,35}. GTs catalyze the formation of glycosidic linkages³⁶. CEs catalyze the removal of *O*- or *N*-acetyl substitutions from saccharides³⁷. PLs are a smaller class of carbohydrate degrading enzymes, classified by their ability to cleave substrates containing uronic acid by a β -elimination mechanism³⁸. AA enzymes are the smallest category in CAZy, containing redox enzymes that act on carbohydrates, carbohydrate-associated, or non-carbohydrate substrates in conjunction with other CAZymes³⁹.

Table 1: Enzyme classes and activity in the CAZy database

CAZy classes	Activity	Number of families*	Modules in present families*
Glycoside Hydrolases (GHs)	Hydrolysis/rearrangement of glycosidic bonds	186	1,398,336
Glycosyl Transferases (GTs)	Formation of glycosidic bonds	116	1,177,447
Polysaccharide Lyases (PLs)	Glycosidic bond cleavage by β -elimination	43	48,702
Carbohydrate Esterases (CEs)	Saccharide <i>O</i> - or <i>N</i> -deacetylation	20	140,347
Auxiliary Activities (AAs)	Redox enzymes	17	25,174

*Data taken from the CAZy database (<http://www.cazy.org/>) in October, 2023

CAZyme classes are further divided in families, and in some cases subfamilies, assigned according to amino acid sequence similarities and stereochemical catalytic mechanism³³. Enzymes belonging to the same family typically also share the same 3-dimensional fold of the catalytic domain^{34,40}. Many CAZymes are modular, meaning they consist of two or more appended domains of autonomous functions⁴¹. Frequently, these modules include smaller domains with carbohydrate-binding abilities attached to the catalytic domains. These ancillary domains are termed carbohydrate binding modules (CBMs), which is a class of CAZyme-associated domains with carbohydrate-binding activity⁴².

1.3.1. Glycoside hydrolases

Glycoside hydrolases form the largest and most well-studied class of CAZymes, and play a vital role in energy harvest and carbon cycling³⁶. Currently, there are 186 GH families defined in CAZy ([CAZy - GH](#)). The mechanism of hydrolysis employed by GHs is usually described as one of two distinct mechanisms first proposed by Daniel Koshland in 1953⁴³, both involving a nucleophilic attack by a water molecule (or a phosphate⁴⁴) facilitated by catalytic amino acid residues in the active site of the enzyme.

One mechanism of hydrolysis proceeds via a single displacement reaction, in which a general acid residue of the enzyme protonates the oxygen of the glycosidic bond, while a general base residue deprotonates a water molecule that can attack the anomeric carbon and cleave the glycosidic bond, releasing two products (Figure 2b). This mechanism is termed inverting, due to the inversion of the stereochemistry of the anomeric carbon at the end of the catalytic cycle. The other mechanism is termed retaining, because the stereochemistry of the anomeric carbon is retained. The retaining mechanism proceeds via a double-displacement (Figure 2a). A catalytic nucleophile and corresponding acid/base residue (usually an aspartic acid or glutamic acid) are responsible for the cleavage of the glycosidic bond. The nucleophile residue attacks the anomeric center of the substrate, while the acid/base donates a proton to glycosidic oxygen to assist the departure of the leaving group in the first half reaction (glycosylation). This breaks the glycosidic bond and results in the covalent intermediate between the anomeric carbon of the cleaved bond and the catalytic nucleophile. The acid/base residue deprotonates a catalytic water molecule, which can attack the anomeric carbon to release the covalent intermediate at the glycone region of the active site (deglycosylation half reaction). Transglycosylation can occur in some cases, when a nucleophile group from a carbohydrate acceptor attacks the glycosyl-enzyme intermediate, leading to the formation of a new glycosidic bond between the carbohydrate acceptor and the product (donor)^{35,45}.

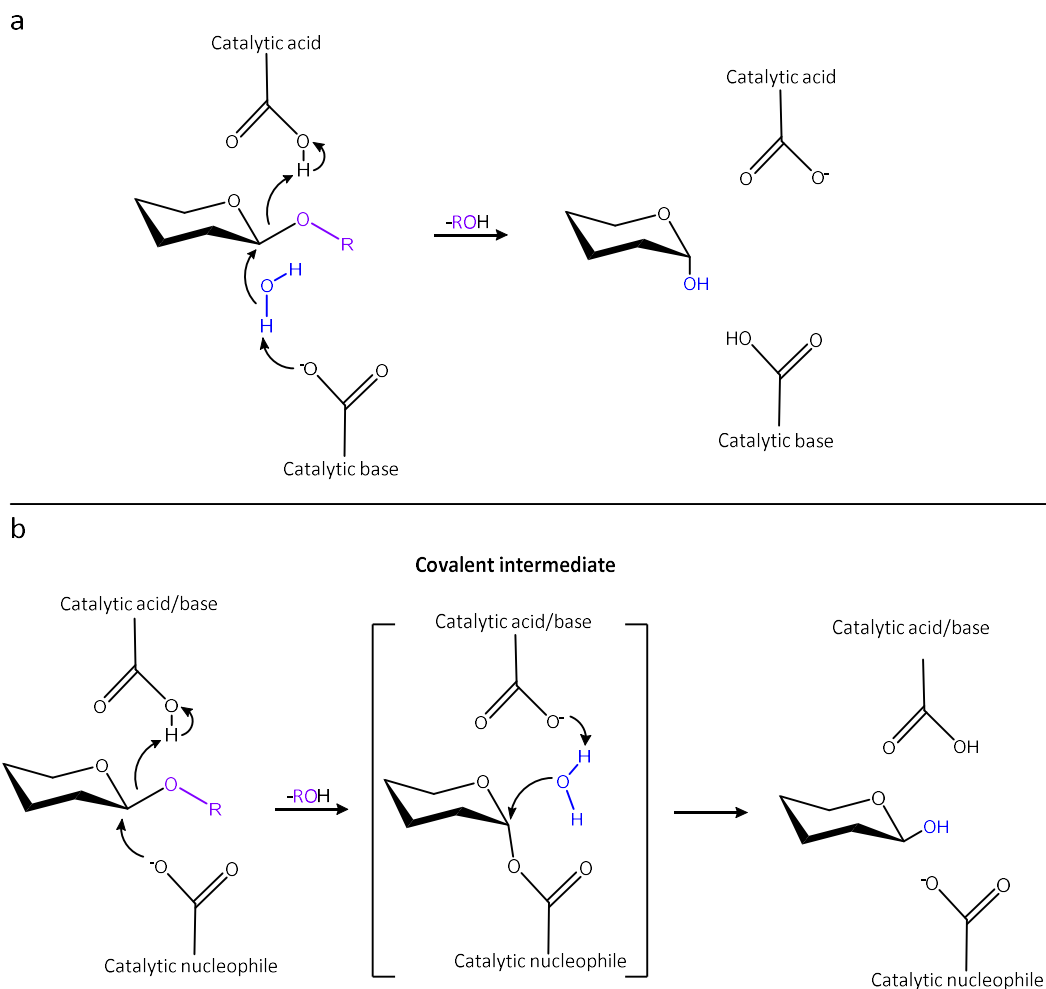


Figure 2. The classical Koshland mechanisms of glycoside hydrolases. **a.** The inverting mechanism proceeds via a single displacement reaction, resulting in the inversion of the carbonyl configuration of the saccharide product. **b.** The retaining mechanism proceeds via a double-displacement mechanism with a covalent intermediate in which the catalytic nucleophile is bound to the anomeric carbon of the product. Hydrolysis occurs when the second reaction proceeds via a nucleophilic water molecule, whereas transglycosylation occurs when a nucleophilic carbohydrate attacks the enzyme-product intermediate instead. Nucleophilic water is shown in blue, and the leaving group is shown in purple. This figure was drawn using ChemDraw v.18.1.

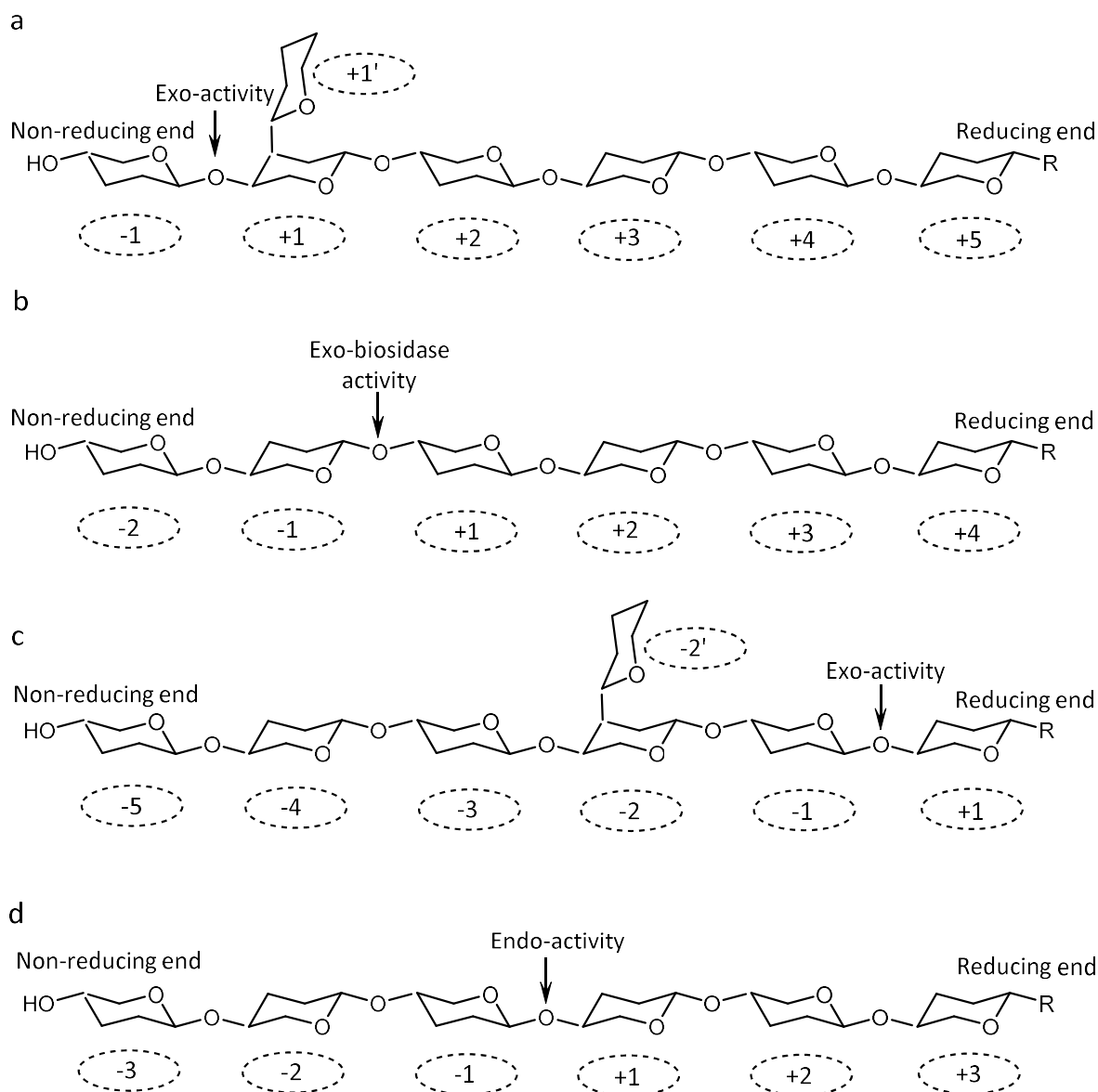


Figure 3. Exo- and endo- activities of glycoside hydrolases and sub-site nomenclature. **a-b.** Exo-enzymes cleave a carbohydrate chain from the reducing, **c** or non-reducing ends. **d** Endo-active enzymes cleave their substrate from within the chain. Arrows denote the targeted linkage for cleaving, and dashed bubbles indicate the corresponding enzyme sub-site. This figure was drawn using ChemDraw v.18.1.

Subsite nomenclature is used regularly in structural biology, describing the binding subsites of individual monosaccharide units of the substrate. The nomenclature used in this thesis is as proposed by Davies *et al.* in 1996⁴⁶. In this system, binding subsites are labeled from $-n$ to $+n$, where n are ascending Arabic integers and the $-n$ and $+n$ designate the non-reducing end and reducing subsites, respectively (Figure 3). The enzymatic cleavage of glycosidic linkages occurs between the -1 and $+1$ subsites according to this nomenclature. In the case of branching substrates, the subsites are denoted as $-n'$ or $+n'$, wherein the n is denoted according to the adjacent main-chain subsite. GHs can also be assigned as “exo” or “endo” – depending on their mode of action. Exo-enzymes cleave a glycan chain from its ends (Figure 3a-c), most often from the non-reducing end, and typically the active site resembles a pocket which form the -1 subsite. In some cases, larger glycone subsites can lead to exo-biosidases or larger exo-cleavages (Figure 3b). Endo-active enzyme are characterized by their ability to cleave a glycan randomly within the chain (Figure 3d), and the active site of these enzymes typically resembles a cleft with several subsites at either side of the scissile bond, typically no less than -2 substrate binding sites at either side of the scissile bond.

1.3.2. Carbohydrate Binding Modules

Carbohydrate binding modules (CBMs) have carbohydrate binding, but not catalytic activities, and can promote the association of the appended catalytic domain of the cognate enzymes to their substrates. The first report of a CBM was made in 1986, when Van Tilbeurgh *et al.*⁴⁷ showed that they were able to proteolytically remove a small domain of approximately 40 amino acid with cellulose-binding activity, from the catalytically active cellulase domain from *Trichoderma reesei*⁴⁷. Based on the observed binding to crystalline cellulose, these domains were initially called cellulose binding domains (CBDs), however this was later changed to CBMs to be more accommodating to other binding specificities. Like glycoside hydrolases, CBMs are also divided into families in the CAZy database based on amino acid sequence similarity. At present, there are 97 different CBM families. Furthermore, CBMs are classified according to the structural topology of their binding sites, into one of three types: the surface binding CBMs (type A), glycan chain binding CBM (type B), and finally the glycan termini binding CBMs (Type C)⁴² (Figure 4). Type A CBMs are characterized by their flat binding site, resembling a platform along the surface lined with exposed aromatic amino acid residues. Type A CBMs display affinity towards crystalline surface sites mainly through aromatic stacking, and to a lesser extent, water-mediated hydrogen bonding^{48,49} (binding is entropically driven), whereas type B CBMs harbor a cleft-like binding site. These types of CBMs are usually specific towards free glycan chains rather than crystalline surfaces, and polar interactions between amino acids and the carbohydrate substrates play an important role in the binding specificity^{50,51}, which is associated to an enthalpy-driven binding signature. It has been observed that binding specificity is reduced on shorter carbohydrates of three units or less⁴². Lastly, type C CBMs are characterized by a smaller binding site, featuring more or less open pockets, optimal for binding the termini of oligo- or polysaccharides, which explains their affinity for smaller carbohydrate molecules like mono-, di-, or tri-saccharides. Type C CBMs are described as having lectin-like properties⁵², meaning direct hydrogen bonding between amino acids and the carbohydrate substrate is more extensive in the binding specificity than for type B CBMs^{53,54}. In other words, there is relatively low hydrogen-binding density/ saccharide unit in type A CBMs, which increase for B and C type CBMs.

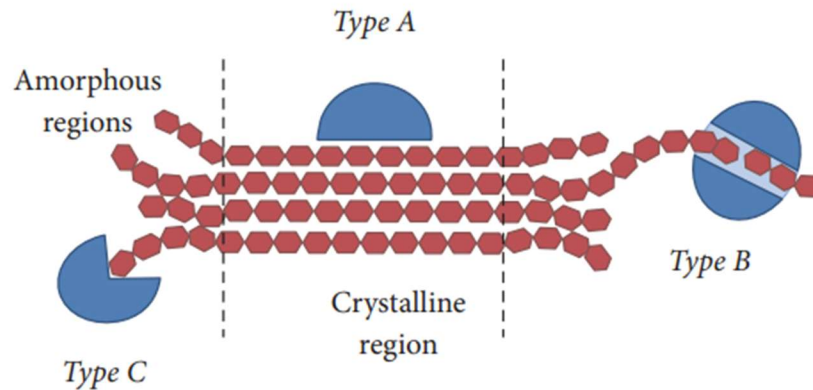


Figure 4. Overview of CBM classifications. Type A CBMs contain a flat binding surface, typically showing affinity to crystalline substrates. Type B CBMs contain a cleft-like binding site and recognized amorphous glycan chains. Finally, type C CBMs contain a smaller binding site pocket, typically showing affinity to terminal glycan mono-, di-, or tri-saccharides. Figure adapted from⁵⁵.

1.3.3. CBMs with human glycan affinities

Most characterized CBMs are specific towards plant polysaccharides like cellulose, xylan, pectins, etc. However, there are a few CBM families which show characterized binding towards mammalian glycans⁵⁶. Binding specificity towards human glycans have been reported for members of CBM 32, 40, 47, 51, and 71 (Table 2). CBMs belonging to these families share the general fold of a stranded β -sandwich and resemble mainly type C CBMs, with specificity for terminal glycan motifs. Of these families, CBM32 is the most well-studied, having members that recognize plant biomass polysaccharides in addition to those recognizing human glycan structures. CBM32 modules with mammalian glycan affinity typically recognize and bind to terminal Galactose (Gal), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), or in some cases *N*-acetyllactosamine (LacNAc)^{57–59}. The binding site architecture of CBM32 modules is not conserved, which makes it difficult to predict specificity based on amino acid sequence analysis⁶⁰, and the binding diversity in the family is likely much greater than what is currently reported. CBMs belonging to family 40 are characterized sialic acid binders, typically found in modular sialidases, such as those from GH family 33⁶¹. The binding of CBM40 modules share a conserved pattern of an ionic interaction between a basic group of an amino acid and the carboxylate of the sialic acid. The location and loop topology of the binding site vary heavily in the family⁶². CBMs belonging to family 47 and 51 are of particular interest for enzymatic conversion of blood antigens, as members from these families have been shown to bind the Lewis⁶³ and A/B antigen epitopes⁶⁴, respectively. Characterized members of these families have been observed appended to endo-acting enzymes of the GH98 family, known to cleave the internal LacNAc of the A and B type 2 antigens⁶⁵. For binding modules belonging to the CBM51 family, a conserved metal ion site, typically Ca^{2+} , is located adjacent to the binding site, and proposed to plays a role in the positioning of binding site loops⁵⁶. Metal ion binding sites are also observed in CBMs belonging to families 32 and 47, but they are distant from the binding site and do not play a role in the binding mechanism. Lastly, CBM71 have been found conjoined with a β -galactosidase from the human pathogen, *Streptococcus pneumoniae*⁶⁶. CBM71 CBMs displayed a strong preference towards the Gal β -1,4- motif in lactose and LacNAc, and play a role in the adherence of *S. pneumoniae* to the surfaces of host epithelial cells.

Table 2: CBM families with binding specificity to host glycans

CAZy classes	Specificities	Number of sequences*
CBM32	Gal, GalNAc, GlcNAc, LacNAc	18,426
CBM40	Neu5Ac	897
CBM47	Lewis antigens	1,439
CBM51	Gal, GlcNAc, A/B antigens	3,028
CBM71	Lactose, LacNAc	443

*Data taken from the CAZy database (<http://www.cazy.org/>) in October, 2023

1.4. The Human Gut Microbiota

The human gastrointestinal tract is home to a complex consortium of microorganisms including archaea, eukarya, and especially bacteria which make up the most abundant group of organisms⁶⁷. It has been estimated that the average number of bacteria present in the gut matches the number of human cells in the body 1:1⁶⁸, and this community has co-evolved with humans over thousands of years to form a mutually beneficial relationship⁶⁹. The composition of the HGM has critical implications in metabolic health and disease⁷⁰, as well as the immune system and gut inflammation⁷¹. The bacterial community of the HGM are primarily dominated by 5 phyla, accounting for more than 90% of the gut microbial populations: Bacteroidota, Bacillota, Actinomycetota, Pseudomonadota, and Verrucomicrobiota (previously called Bacteroidetes, Firmicutes, Actinobacteria, Proteobacteria, and Verrucomicrobia, respectively)⁷². The diversity of bacterial species identified at lower taxonomic levels continues to grow, thanks to the extensive efforts of HGM metagenomic sequencing, enabling the establishment of databases containing >200,000 metagenomes⁷³. The HGM is exposed to a diverse range of complex dietary and host-derived carbohydrates, which are essential for the establishment and flourishing of this community⁷⁴.

The specific make-up of the HGM varies between individuals, owing to many factors that play a role in establishment, maturation, and maintenance of bacteria in the gut. Factors such as the microbiome of the mother and mode of delivery⁷⁵, local environment⁷⁶, diseases and antibiotic use⁷⁷, lifestyle^{78,79} (smoking, drinking, exercise, etc.), and most importantly, diet^{76,80,81}. Humans encode a limited number of digestive saccharolytic enzymes, mostly targeting sucrose, lactose, and the α -1,4-Glucose linkages of starch. Other non-digestible dietary carbohydrates, termed dietary fibers, provide one of the main carbon sources for members of the HGM endowed with the ability to degrade such polysaccharides^{82,83}. Dietary fibers are fermented by the HGM to the short chain fatty acids (SCFA), butyrate, propionate, or acetate in the ratios of 1:1:3⁸⁴. These SCFAs contribute to cellular proliferation, gene expressions, and recruitment of immune cells^{85,86}. It has been shown extensively that western “low-fiber” diets can lead to the erosion of the microbiota with regard to diversity, with concomitant deterioration in gut health, further underlying the importance of dietary fibers on the maintenance of the HGM^{87,88}.

1.4.1. Human gut biogeography

The human gut is typically divided into the upper and lower gastrointestinal (GI) tract, each comprised of several organs and collectively responsible for the digestion of food and nutrient uptake⁸⁹. The upper GI tract consists of the mouth, esophagus, stomach, and part of the small intestine (duodenum). Following,

the lower GI tract consisting of the rest of the small intestine (jejunum and ileum), and the large intestine regions; the colon, rectum, and anus (Figure 5a). The local environment and the composition of the microbiota varies along the GI tract, presenting several biogeographical niches⁷². The oral microbiota is diverse and encompasses several ecological niche's, but is mostly dominated by members of *Lactobacillaceae* from Bacillota, *Prevotellaceae* from Bacteroidota, and *Actinomycetaceae* from Actinomycetota^{90,91}. The microbiota of the esophagus is not as well defined, however, data suggests it is closely related to the oral microbiota, dominated by members of *Veillonellaceae* (Bacillota), in addition to *Lactobacillaceae* and *Prevotellaceae*⁹². The stomach presents a hostile and highly acidic environment, mostly colonized by the bacteria *Helicobacter pylori*, from Pseudomonadota, which has adapted to this environment⁹³.

Compared to the large intestine, the small intestine is more acidic (low pH), has higher oxygen levels, and a higher concentration of antimicrobial peptides and bile acids (Figure 5b). Furthermore, the transit time of food through the small intestine (averaging 5 hours) is faster than the large intestine (averaging 21 hours), despite the larger travel distance^{94,95}. Simple nutrients, like monosaccharides and disaccharides, are absorbed rapidly, while more complex nutrients, like fibers, move to the large intestine. Due to these conditions, the bacteria of the small intestine are dominated by facultative anaerobic and fast growing bacteria including *Enterobacteriaceae* from the Pseudomonadota phylum, and *Lactobacillaceae* from the Bacillota phylum⁹⁶ (Figure 5a).

The longer transit time in the large intestine allows for water and electrolyte uptake, which is a slower process than simple nutrient absorption in the small intestine⁹⁴. In addition, pH increases while oxygen level and antimicrobial peptides decrease in the large intestine. The bacteria of the large intestine are strict anaerobic and include members able to ferment plant polysaccharides, and members able to utilize human glycans. Common members from the Bacteroidota phylum include *Bacteroidaceae*, *Prevotellaceae*, and *Rikenellaceae*, and from Pseudomonadota include *Lachnospiraceae* and *Ruminococcaceae*^{72,96}, and the single member from the Verrucomicrobiota phylum, *Akkermansia muciniphila*⁹⁷ (Figure 5b). Due to the higher pH, slower nutrient transit time, and decrease in antimicrobial peptides, the abundance and diversity of bacteria is greater in the large intestine compared to the small intestine. Interestingly, there is also a difference in the composition of bacteria cross-sectionally from the lumen towards the epithelial cells. This can be attributed to a small oxygen gradient selecting for more oxygen tolerant anaerobes closer to the epithelium⁹⁸, and more significantly to the ability of certain bacteria to interact with the mucrosal layer covering the epithelium.

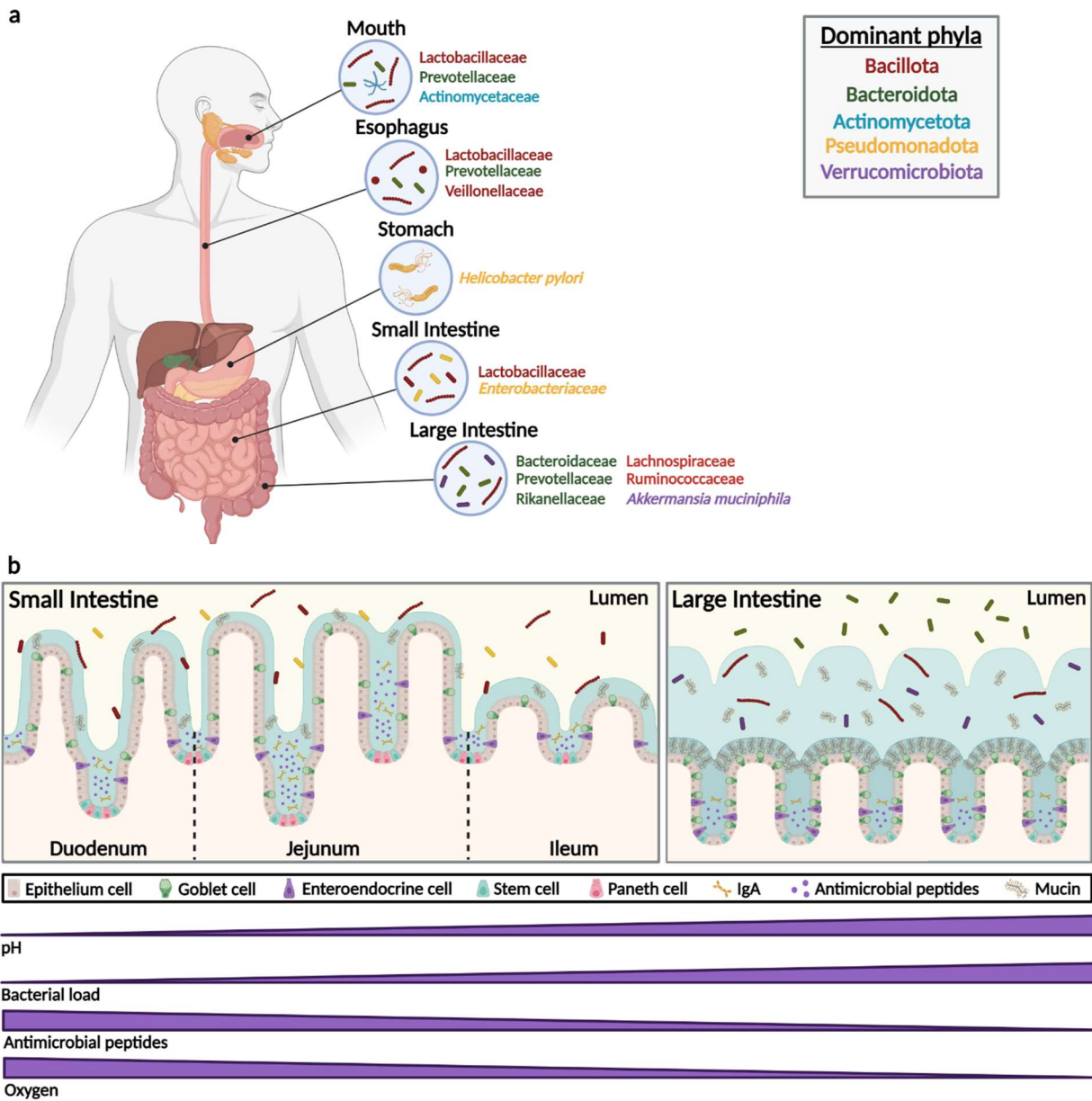


Figure 5. Gut biogeography and mucus layer. **a.** Overview of the dominant bacterial families present in various parts of the gastrointestinal tract. **b.** The progression of the mucus layer from the small intestine to the large intestine, alongside changes in pH, bacterial load, antimicrobial peptides, and oxygen levels. In the large intestine, the mucus layer is split into two parts: the outer and inner mucus layer. This figure was made using Biorender.

1.4.2. The mucus layer

In order to protect the host against physical and microbial insult, a mucosal barrier separates the epithelial cells lining the gut from the lumen⁹⁹. This barrier is termed the mucus layer, and is a dense gel-like layer, increasing in thickness longitudinally from the small intestine to the large intestine, alongside the increased abundance of microorganisms⁷² (Figure 5b). It is critical for gut health that the mucus layer be

maintained, as a collapse of this layer can lead to inflammation in the gut^{25,100,101}. This is seen in the inflammatory bowel diseases, ulcerative colitis, when the permeability of the mucus layer is increased and bacteria can be found in close proximity to epithelial cells during inflammation^{102,103}, or at worst, when the gut barrier is breached and the bacteria can translocate to the blood circulation.

The principle component of the mucus layer are heavily glycosylated glycoproteins termed mucins, which are synthesized by goblet cells lining the wall of the intestine¹⁰⁴ (Figure 6a). The human gut expresses >20 genes encoding mucin glycoproteins. Mucins are grouped as either membrane-bound or secreted¹⁸. Common to all mucins are long repeats of proline-threonine-serine (PTS domain), which are decorated with over 100 different *O*-glycan structures¹⁰⁵. The most abundant mucin in the human gut is MUCIN2, or MUC2¹⁰⁶, and up to 80 % w/w of MUC2 is estimated to be glycans. In the small intestine, the mucus layer is thin and loosely attached in order to allow for nutrient absorption¹⁰⁷. To counter this, the higher abundance of antimicrobial peptides synthesized by paneth cells in the crypt of the small intestine aid in minimizing infections¹⁰⁸. Furthermore, the mucus layer is steadily replenished by MUC2 initially anchored to the surfaces of goblet cells, then released by the protease meprin β , which is expressed under bacterial exposure^{109,110}. In the large intestine, the mucus layer is thicker and divided into two separate parts; the inner and outer mucus layer. The inner mucus layer is anchored on the cells in a tightly packed lamellar structure, and the outer mucus layer can be up to 4 times larger in volume, and consists of mucin released from the inner layer by endogenous proteases¹¹¹. In contrast to the inner mucus layer, the outer mucus layer contains pores that are permeable to bacteria, allowing for various commensal members of the HGM to colonize¹¹².

1.4.3. Mucin *O*-glycan Structures

The *O*-glycosylation of MUC2 takes place in the Golgi apparatus, initiated by the addition of an *N*-acetylgalactosamine (GalNAc) to serine or threonine residues of the PTS domain, by peptidyl-acetylgalactosamine transferases¹¹³. This results in the Tn antigen, which can then be elongated by the addition of Galactose (Gal) and/or *N*-acetylglucosamine (GlcNAc) to the C3 or C6, resulting in the most common mucin core structures (core 1-4). The core 1 structure consists of a Gal β -1,3 linkage to the Tn antigen by the glycosyltransferase, C1GALT1, forming the T antigen¹¹⁴. The addition of an GlcNAc β -1,6-linkage to the core 1 GalNAc by the transferases G2CNT1 or C2GNT3, results in the branched core 2 structure, GlcNAc β -1,6-(Gal β -1,3-)GalNAc¹¹⁵. Cores 1 and 2 are most commonly found in gastric and duodenal mucins¹¹⁶. The core 3 structure consists of the addition of GlcNAc β -1,3 to the Tn antigen by the β 3Gn-T6 transferase¹¹⁷, whereas the core 4 structure is comprised of an additional GlcNAc β -1,6- linkage by the G2CNT2 transferase, resulting in the branched GlcNAc β -1,6-(GlcNAc β -1,3-)GalNAc¹¹⁵. The core 3 and 4 structures are predominant in colonic mucin^{116,118,119} (Figure 6b).

These core structures are frequently further elongated with additional repeating Gal and GlcNAc units via either β -1,3- or β -1,4- linkages, forming the backbone of the *O*-glycan chains. Type 1 chains are terminal and composed of Gal β -1,3-GlcNAc linkages, forming Lacto-*N*-Biose (LNB) units. Type 2 chains consist of Gal β -1,4-GlcNAc linkages, forming a LacNAc backbone, which can be elongated to form poly *N*-acetylgalactosamine units. The glycan chains can be branched from the Gal units by the addition of GlcNAc β -1,6-Gal linkages within the chain. Elongation of the backbone chains can continue until the glycan is capped by the addition of fucose (Fuc), *N*-acetylneuraminic acid (Neu5Ac), or α -linked GalNAc, Gal, or GlcNAc¹¹². In the human colon, an ST6-sialyltransferase can add an Neu5Ac to the C6 of the Tn antigen

GalNAc, resulting in the sialyl-Tn antigen (sTn)¹²⁰. Gal residues in mucin can additionally be capped by sulfate, which can occur *O*-linked to the C3, C4, or C6, resulting in 3S-, 4S-, or 6S-Gal, respectively. Sulfation can also occur to the C6 of GlcNAc, resulting in 6-S-GlcNAc, which occurs both as a terminal but also on internal GlcNAc residues of mucin glycans. Levels of sulfation appear to increase towards the distal end of the GI tract, with colonic mucin harboring increased levels of sulfation compared to gastric mucin^{121,122}. The addition of caps blocks the ability of the transferases to elongate the glycan chain, and furthermore serve the purpose of shielding the internal chain from degradation by members of HGM, which lack a decapping machinery^{19,105,123}.

The diversity of mucin *O*-glycan structures is attributed mainly to the terminal regions, as the common cores and backbone structures are widely present. The terminal epitopes vary depending on the encoded glycosyltransferases, and their expression levels, in each individual. Fuc can be added to the terminal Gal of both type 1 and type 2 chains via an α -1,2 linkage by the fucosyltransferases FUT1 and FUT2, resulting in the blood group antigens, H type 2 and H type 1, respectively¹²⁴ (Figure 6c). The addition of either a GalNAc or a Gal by an α -1,3 linkage, to the Gal of the H antigens results in the blood A and B antigens, respectively, depending on the encoded activity of the ABO transferase gene of the host^{125,126}. Additional fucosyltransferases (FUT3, FUT4, FUT5, FUT6, FUT7) can add a Fuc unit to internal GlcNAc via an α -1,4 or α -1,3 linkage to the type 1 and type 2 chains, respectively¹²⁷. If the α -1,3/4- Fuc is added to terminal LNB/LacNAc units, the resulting epitopes is the Lewis a (Le^a) and Lewis x (Le^x) antigens, respectively. If the fucose is added to the H antigens, the resulting double-fucosylated epitopes are termed Lewis b (Le^b) and Lewis y (Le^y), for the type 1 and type 2 chains, respectively. Lastly, the A/B transferase can act on either of the Le^b or Le^y, resulting in A-Le^{b/y} or B-Le^{b/y}. Neu5Ac can be added to mucin glycans via either an α -2,3 or an α -2,6 linkage by the ST6GalNAc2/3/4 and ST3Gal1/2 transferases¹²⁸ (Figure 6d). Some of the known sialylated epitopes, besides the sialyl-Tn antigen, include the sialyl-T antigen, sialyllactosamine, Sd^a, and the sialylated Le^a and Le^x structures. In humans, the degree of fucosylation on mucin epitopes will decrease from the gastric mucin to the colonic mucin, whereas sialylation will increase¹¹⁶. Both the Gal and GlcNAc of the Le^a and Le^x epitopes can additionally be *O*-sulfated, resulting in 3'-,6'- or 6-*O*-sulfate Le^{a/x} (Figure 6e).

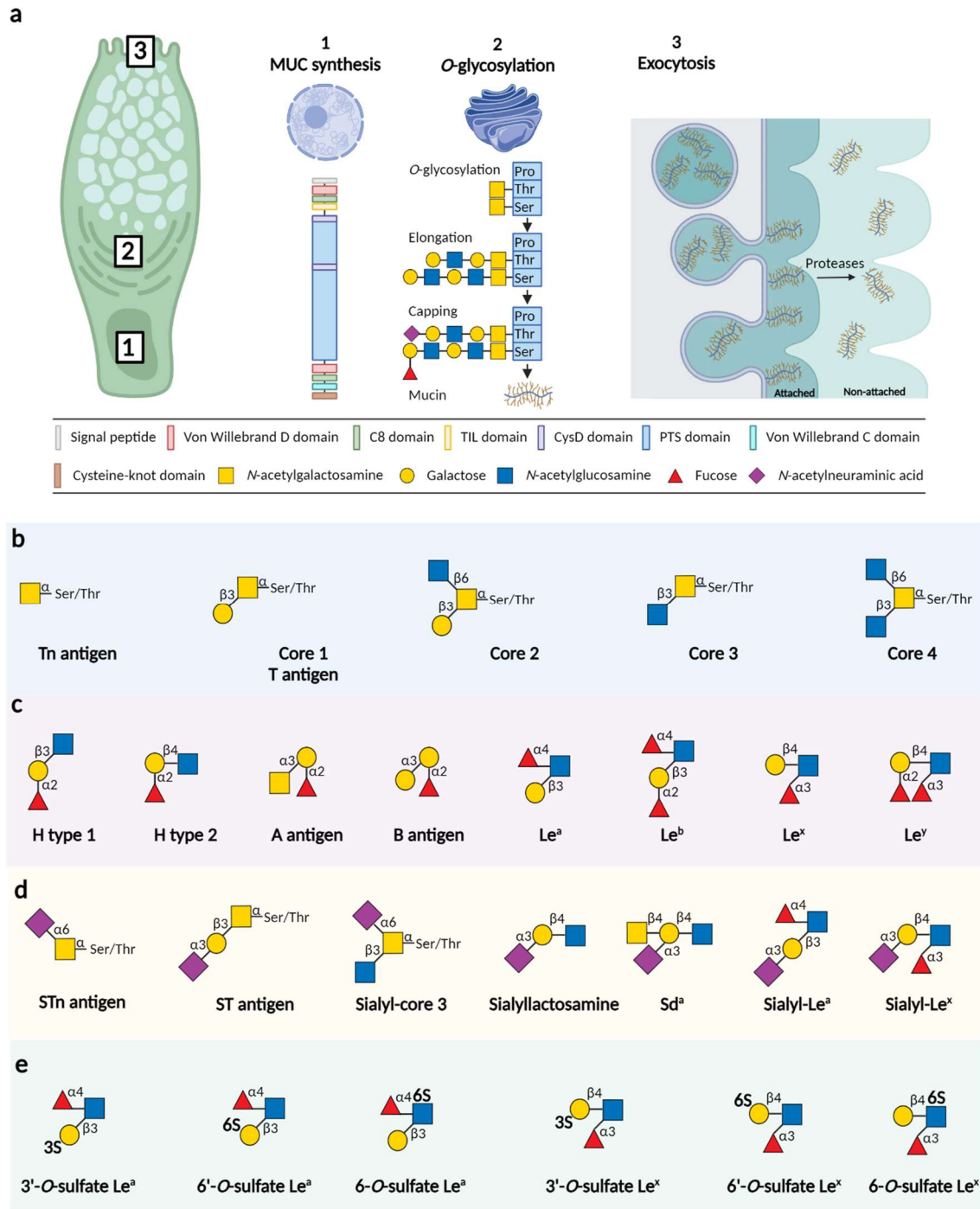


Figure 6. Synthesis and secretion of mucin glycoproteins and common glycan structures. **a.** Mucin is synthesized by goblet cells. Synthesis of the MUC protein occurs in the nucleus, and O-glycosylation occurs in the golgi apparatus. Threonine and serine residues are initially glycosylated with GalNAc, which is then elongated by Gal and GlcNAc, and terminated with fucose, sialic acid, or sulfate capping. Mature mucin is secreted as membrane-bound glycoproteins, which are released by endogenous proteases. **b.** Common core structures of mucin O-glycans. The known terminal structures of mucin include **c** ABO and Lewis antigens, **d** sialylated structures, and **e** sulfated epitopes. This figure was made using Biorender.

1.5. Glycan Utilization by HGM

It is crucial for bacteria of the HGM to adapt to the competitive environment in the gut niche. Competition for carbon and nitrogen sources has led to the evolution of a huge library of CAZymes¹²⁹ and carbohydrate transporters^{130–132} among members of the HGM. A study looking at metagenomic analyses of CAZyme distribution across 13 different broadly defined microbial communities, show that the HGM has the highest abundance of glycoside hydrolases¹³³. These glycoside hydrolases are adapted to target a diverse range of carbohydrates, including animal and host derived glycans from mucin or epithelial cells⁸². Similar to other ecological niches, bacteria may be classified as generalists or specialists with regard to their carbohydrate utilization profiles. Generalist, like most of the gut Bacteroidota species, usually encode a large amount of CAZymes allowing them to adapt their catabolism to target a variety of different glycan sources, depending on the availability. This is an evolutionary adaptation that allows generalists to survive large variations in the type of diet. Specialists, such as *A. muciniphila*, have evolved to efficiently target one or a few glycans, often giving them a competitive advantage over generalists for the particular substrate they specialize in utilizing.

1.5.1. *Akkermansia muciniphila*

A. muciniphila, the sole member of the phylum Verrucomicrobiota in the HGM known to date, is a specialist targeting host glycans, particularly mucin. Since its discovery in 2004¹³⁴, *A. muciniphila* was shown to utilize mucin as a sole carbon and nitrogen source¹³⁵. Colonization by *A. muciniphila* starts early in life⁹⁷, and develops within the first year to reach the levels observed in healthy adults¹³⁶. *A. muciniphila* has been shown to inefficiently degrade certain Human Milk Oligosaccharides (HMOs), which bear structural overlaps with terminal mucin epitopes, based on poor growth on HMOs in a synthetic medium as compared to mucin¹³⁷. This may be attributed to the lack of the efficient ATP-binding cassette (ABC) transporters for HMO uptake systems observed in bifidobacteria¹³⁸, and the different strategy of internalizing mucin oligomers into the cytoplasm¹³⁹, and potentially the adaptation of the enzymes to more complex and conjugated glycans found in mucin.

Since *A. muciniphila* is a specialized mucin degrader, and the integrity of the mucus layer is critical for a healthy gut, the role that *A. muciniphila* plays in gut health has been studied extensively. Interestingly, many investigations found *A. muciniphila* abundance correlates positively with a variety of metabolic and immunological benefits^{140,141} (Figure 7). In mice studies, it was observed that administration of *A. muciniphila* contributed to restoration of the mucus layer thickness in the gut, which can be altered due to metabolic disorders or obesity¹⁴². *A. muciniphila* produces the short-chain fatty acid, propionate, as a product of mucin fermentation¹⁴³. Propionate can interact with surface receptors on epithelial cells, including goblet cells, stimulating the release of glucagon-like peptide-1/2 (GLP1/2) to regulate glucose metabolism¹⁴¹. Furthermore, propionate enhances the secretion of mucus by goblet cells and stimulates proliferation of goblet cells, hereby contributing to the replenishment of the gut mucus barrier^{111,144}. In addition, other studies show that a protein secreted by *A. muciniphila* termed P9¹⁴⁵, and the cell-bound protein Amuc_1100^{146,147}, are both able to bind to Inter Cellular Adhesion Molecule 2 (ICAM2) receptors on surfaces of enterocytes and endocrine L cells, which in turn increases the production of GLP1/2 and Interleukin-6. This leads to a decrease in inflammation and glycaemia, increasing insulin sensitivity and fatty acid oxidation. In cancer related research, Amuc_1100 has been correlated with reduced tumor development and an increase in the number of cytotoxic and regulatory T cells¹⁴⁷. Furthermore,

supplementing *A. muciniphila* in mouse models revealed a protective effect on the central nervous system (CNS), by reducing the amount of β -amyloid (A β) protein deposition in the hippocampus¹⁴⁸. A β protein deposition plays a role in the pathogenesis of Alzheimer disease¹⁴⁹. Despite these finding, the use of *A. muciniphila* to treat inflammatory bowel disease show some contradictory effects¹⁵⁰. It appears that *A. muciniphila* may contribute with both anti-inflammatory, but also pro-inflammatory effects in mice with colitis depending on several factors, including specific strain usage, the sex of the host, and the coexistence of other pathogenic bacteria.

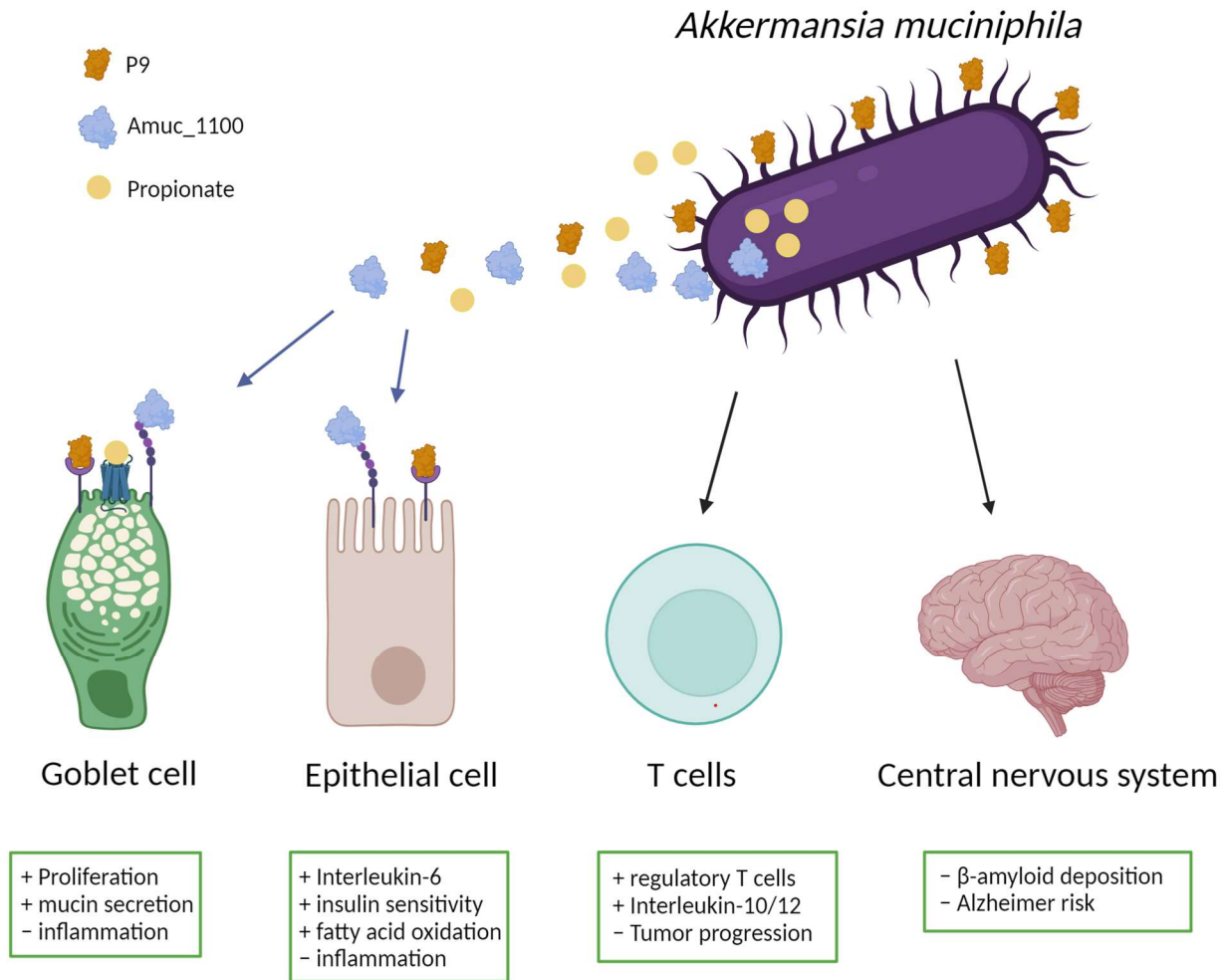


Figure 7. Overview of health benefits correlated with *A. muciniphila* abundance in the gut. *A. muciniphila* produces the short-chain fatty acid, propionate. Propionate, along with two expressed proteins, P9 and Amuc_1100, can interact with receptors on goblet cells and enterocytes to promote a healthy gut. Furthermore, *A. muciniphila* abundance has been correlated with upregulation of T cells, leading to decreased tumor progression, as well as having protective effects on the CNS, including decrease β -amyloid deposition in the brain.

1.5.2. Mucin foraging by *A. muciniphila*

Certain members of the HGM, including *A. muciniphila*, which colonize the outer mucus layer in the large intestine, encode a repertoire of CAZymes active against mucin *O*-glycans¹⁹. Mucin is capped with sulfate, sialic acid, and fucose, the two latter being present on ABO and Lewis antigens. Mucolytic bacteria that are able to grow on mucin require a repertoire of different GHs, CEs, and sulfatases to deconstruct mucin *O*-glycans. The cleavage of the mucin backbone requires specialized glycopeptidases that recognize different *O*-glycan motifs.

The annotated CAZyme of *A. muciniphila* comprises 59 GHs belonging to 26 different families. Although only few have been characterized on mucin glycoconjugates to date, enzymes from *A. muciniphila* have shown various activities related to mucin breakdown. In a recent study, fucosidases belonging to GH29 and GH95, and sialidases belonging to GH33 and the newly defined family GH181, were shown to be able to decap all known fucosyl and sialyl mucin *O*-glycan motifs¹⁵¹. *AmGH29C*, and to a lesser extent *AmGH29D*, are able to decap the α -1,3/4-fucose of all Lewis antigen structures on mucin, whereas *AmGH95B* is able to cleave the α -1,2-fucose of H antigens on type 1, 2, and 3 chains. Likewise, members of GH33 and the defining member of GH181 efficiently de-sialylated all sialyl- α -2,3/6- epitopes, including the sialyl T antigen, sialyl Tn antigen, sialylactosamine, and Sd^a. Another highlight of the study shed light on the specificity of *A. muciniphila* enzymes, showing that some enzymes exhibit broad selectivity, while others display very strict preferences to a single *O*-glycan. For example, as opposed to *AmGH95B*, *AmGH95A* was specific towards the H type 2 antigen, displaying no activity on either type 1 or type 3 chains. *AmGH181* also showed a strict specificity towards the α -2,3 sialylated T antigen, whereas both GH33 enzymes displayed a broader activity against both α -2,3 and α -2,6 linkages of *O*-glycans and *N*-glycans, albeit with different selectivities. *AmGH33A* displayed Neu5Ac α -2,3-Gal activity in both intact and released porcine colonic mucin (PCM) glycans, whereas *AmGH33B* was inactive on the intact mucin epitopes. In contrast, *AmGH33B* was active against the Sd^a core 1 of the released PCM glycans.

A. muciniphila also encodes *O*-glycopeptidases, which have been shown to cleave mucin backbone, at sites adjacent to T or Tn antigens^{152–155}. The sialylation of the recognition motif impairs the activity of these glycopeptidases, and removal of the Neu5Ac caps by the aforementioned GH33 and GH181 enzymes would be a pre-requisite for the cleavage of the mucin backbone. It has also been proposed that *A. muciniphila* can take up mucin fragments into the periplasm, in order to utilize the amino acid backbone as a nitrogen source¹⁵⁶. In addition to the *O*-glycopeptidases, a family of endo- β -1,4-galactosidases (GH16) have also been shown to cleave the abundant poly LacNAc structures within mucin glycan chains, also requiring an initial de-sialylation step¹⁵⁷. Furthermore, 3 β -galactosidases, 1 from family GH35 and 2 from GH2, were able to target the common mucin chain galactose linkages of porcine gastric mucin (PGM), including Gal β -1,3-GlcNAc, Gal β -1,4-GlcNAc, and Gal β -1,3-GalNAc¹⁵⁸.

Besides the aforementioned activities, other enzymatic activities have been shown to target additional mucin glycan structures (Figure 8). Enzymes from GH20 have been shown to target GlcNAc β -1,3-Gal substrates, an abundant linkage in mucin chains¹⁵⁹. Some GH20 enzymes from *Bifidobacterium bifidum* have β -*N*-acetyl-6-sulfoglucosaminidase activity, targeting one of the sulfated mucin epitopes¹⁶⁰, in the place of sulfatases utilized by other members of the HGM, including *Bacteroides thetaiomicron*¹⁶¹. Blood group A and B epitopes have been shown to be targeted by members of GH109 and GH110, respectively¹⁶². GlcNAc α -1,4-Gal is a common terminal epitope in gastric mucins¹⁹, and has been shown to be a substrate for enzymes belonging to family GH89 from *B. bifidum*¹⁶³. Finally, enzymes belonging to

the GH101 and GH129 families have been shown to target mucin core 1 and the Tn antigen, respectively^{164,165}. The genome of *A. muciniphila* encodes many GHs yet to be characterized, including some belonging to families GH20, GH89, GH109, GH110, with potential activity for mucin linkages (Table 3). Due to a structural link between mucin glycans and carbohydrate antigens on Red Blood Cells (RBCs), which will be described in the coming section, these may also be candidates for the generation of universal blood.

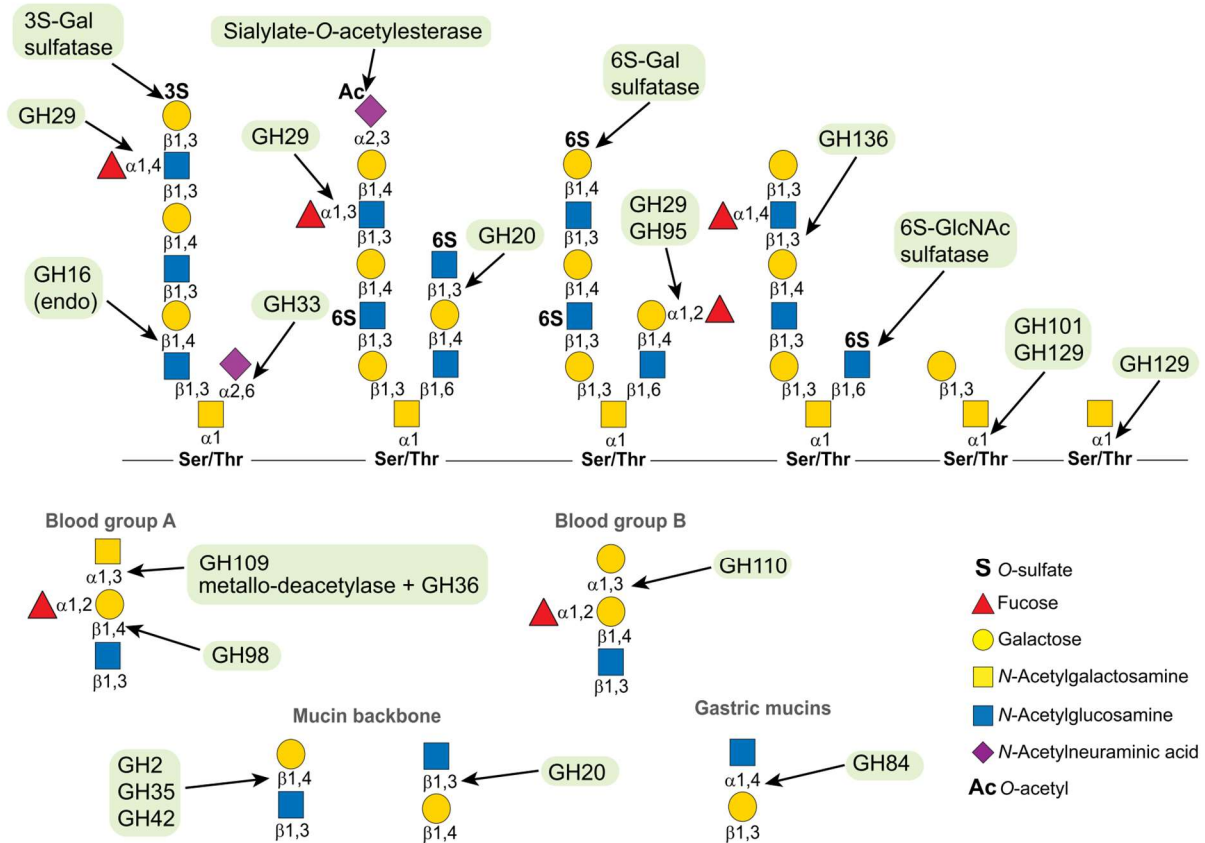


Fig. 8. GH families with specificity for mucin terminal epitopes, backbone chains, and core structures. GH family with activity against the mucin O-glycan linkages are indicated. A few errors in the figure include the mention of a GH136, which has not been reported so far. Additionally, the GH84 for gastric mucin degradation should be a GH89. Figure adapted from¹⁹

Table 3. Potential enzymatic activities for mucin degradation encoded by *A. muciniphila*

GH families	Sequences	Potential activity	Potential target
GH2	6	β -Galactosidases	Backbone
GH16	3	Endo- β -Galactosidases	Backbone
GH18	1	Endo- β - <i>N</i> -Acetylglucosaminidase	Backbone
GH20	11	β - <i>N</i> -Acetylgalactosaminidase, β - <i>N</i> -Acetylglucosaminidase	Backbone
GH27	1	α -Galactosidases	Terminal α -Gal
GH29	4	α -1,3/4-Fucosidases	Lewis antigen Fuc
GH33	2	α -Sialidases	<i>N</i> -acetylneuraminic acid caps
GH35	2	β -Galactosidases	Backbone
GH36	3	α -Galactosidases, α - <i>N</i> -Acetylgalactosaminidase	Terminal α -Gal, α -GalNAc
GH89	2	α - <i>N</i> -Acetylglucosaminidase	Gastric mucin terminal α -GalNAc
GH95	2	α -1,2-Fucosidases	H antigen Fuc
GH109	2	α - <i>N</i> -Acetylgalactosaminidase, β - <i>N</i> -Acetylgalactosaminidase	Terminal A antigen
GH110	2	α -Galactosidases	Terminal B antigen
GH177	1	α -Sialidases	<i>N</i> -acetylneuraminic acid caps
GH181	1	α -Sialidases	Sialyl-T antigen

*Data taken from the CAZy database ([CAZy - Bacteria](#))

Green indicates enzyme families whose activity has been verified on mucin substrates.

1.6. Red Blood Cell Glycans

Red blood cells (RBCs) make up nearly half of the circulating blood volume in our bodies⁶⁸, and carry out the essential function of transporting O₂ from lungs to peripheral tissues, and transporting the aerobic metabolism end product, CO₂¹⁶⁶. RBCs are small bio-concave cells lacking a nucleus, and are coated with a dense layer of glycoconjugates, which collectively forms the RBC glycome¹⁶⁷. Naturally occurring antibodies (immunoglobulins) in the body can recognize specific glycan epitopes from pathogens, or other foreign cells, for the purpose of launching an immune response¹⁶⁸. The glycan patterns on different cell types, including RBCs, depend on the expression of encoded glycosyltransferases¹⁶⁹, meaning specific glycan patterns are genetically governed and can vary between individuals. Currently there are 44 different blood groups recognized by the International Society of Blood Transfusion (ISBT), collectively containing 345 different antigens as of December 2022¹⁷⁰. Among these, clinically relevant carbohydrate based blood groups include the ABO, Lewis, Globosides, P1Pk, FORS, and SID blood¹⁷¹.

1.6.1. The ABO blood group

The ABO blood group is comprised of 4 primary phenotypes, termed A, B, AB, and O. The phenotypes are defined by the terminal carbohydrate structure of glycolipids and *N*-glycoproteins coating the surfaces of RBCs¹²⁶. ABO antigens are derived from a precursor of either type 1, 2, or 4 chains⁶ (Figure 9a). The type 2 chain is the most abundant on the surfaces of RBCs¹⁷², composed of a terminal Gal β -1,4-GlcNAc (LacNAc). The type 1 chain is composed of a terminal Gal β -1,3-GlcNAc (LNB) unit, whereas the type 4 chains are composed of a Gal β -1,3-GalNAc (Galacto-*N*-biose, GNB) unit exclusively found on glycolipids. Recent glycomic studies have shown that a majority of the ABO antigens may be present on longer poly-LacNAc chains^{167,173}. The addition of a fucosyl unit to any of the precursors by an α -1,2- linkage, results in the H antigen of the O type, the common precursor to both the A and B antigens (Figure 9a). The addition of a terminal α -1,3-GalNAc onto the non-reducing galactosyl of the H antigen results in the A antigen, while the addition of a terminal α -1,3-Gal unit instead would result in the B antigen (Figure 9b). The A, B, and AB phenotypes contain the A antigen, B antigen, or a combination of both for type AB, respectively. Interestingly, it has been suggested that type O blood has evolved from type A, as a mutation in the A allele resulting in an inactive transferase caused increased resistance towards malaria caused by *Plasmodium falciparum*¹⁷⁴. More recently, it has also been demonstrated that type O individuals also exhibit an increased resistance towards SARS-COV-2^{175,176}.

As mentioned, the addition of carbohydrate units to the LacNAc, LNB, or GNB precursor structures is catalyzed by glycoside transferases expressed by the host. The type of antigen presented on RBCs is determined by the specific activity, or inactivity, of the transferases. The synthesis of the H antigen, as mentioned earlier, is determined by the fucosyltransferases FUT1 or FUT2^{177,178}, while the ABO group is determined by the activity of the ABO allele, which can encode either an α -1,3-*N*-acetylgalactosaminyltransferase (A), an α -1,3-galactosyltransferase (B), or an inactive transferase (O)^{179,180}. The FUT1 fucosyltransferase is responsible for fucosylation specific to the type 2 and 4 chains. In contrast to the type 2 chain, the type 1 chain is a soluble molecule, first and foremost present in secretions like saliva or mucins, but is also found on glycosphingolipids in the plasma, where it gets incorporated into the membrane of RBCs. The FUT2 fucosyltransferase is active on the type 1 chains, but is not specific, meaning it may also add a fucose to other chain types as well. Individuals with an active FUT2 are termed “secretors”¹⁸¹. In rare cases, individuals do not express the FUT1 or FUT2 gene, and in these cases fucosylation of the core does not occur, resulting in the rare Bombay phenotype¹⁸²⁻¹⁸⁴.

After the synthesis of the H antigen, the A or B transferase can add the terminal GalNAc or Gal unit, respectively, resulting in the A and B antigens. In the same manner as the H antigen, both A and B antigens are predominantly expressed on the surfaces of RBCs on the type 2 chain, but also exist on type 1 and 4 chains. In addition, the A phenotype exists in two subtypes (A_1 and A_2), the main difference being the extent of expression of the A antigen. A_1 individuals express a more active A transferase resulting in higher abundance of A antigen compared to A_2 individuals¹⁸⁵. In particular, the A type 4 glycolipid is present at moderate levels in A_1 individuals, but virtually undetectable in A_2 individuals¹⁸⁶.

1.6.2. AB antigen extensions

Both the A and B antigens can be further elongated with one or more additional monosaccharide units, resulting in extended A and B epitopes (Figure 9c). In 1985-1986, Clausen *et al.* showed that the A antigen could be extended with a Gal β -1,3- unit, resulting in the Gal-A epitope, which can be fucosylated by a Fuc α -1,2- linkage, resulting in the H type 3 epitope^{8,9}. Finally, a GalNAc α -1,3- linkage can be made to the

Gal of the H type 3, resulting in the repetitive A-antigen structure, termed A type 3⁷. More recently, Hagman *et al.* discovered the extended version of the B antigen, in which a GalNAc unit is added via a β -1,3 linkage to the B antigen, resulting in the GalNAc-B (ExtB) antigen^{10,20}.

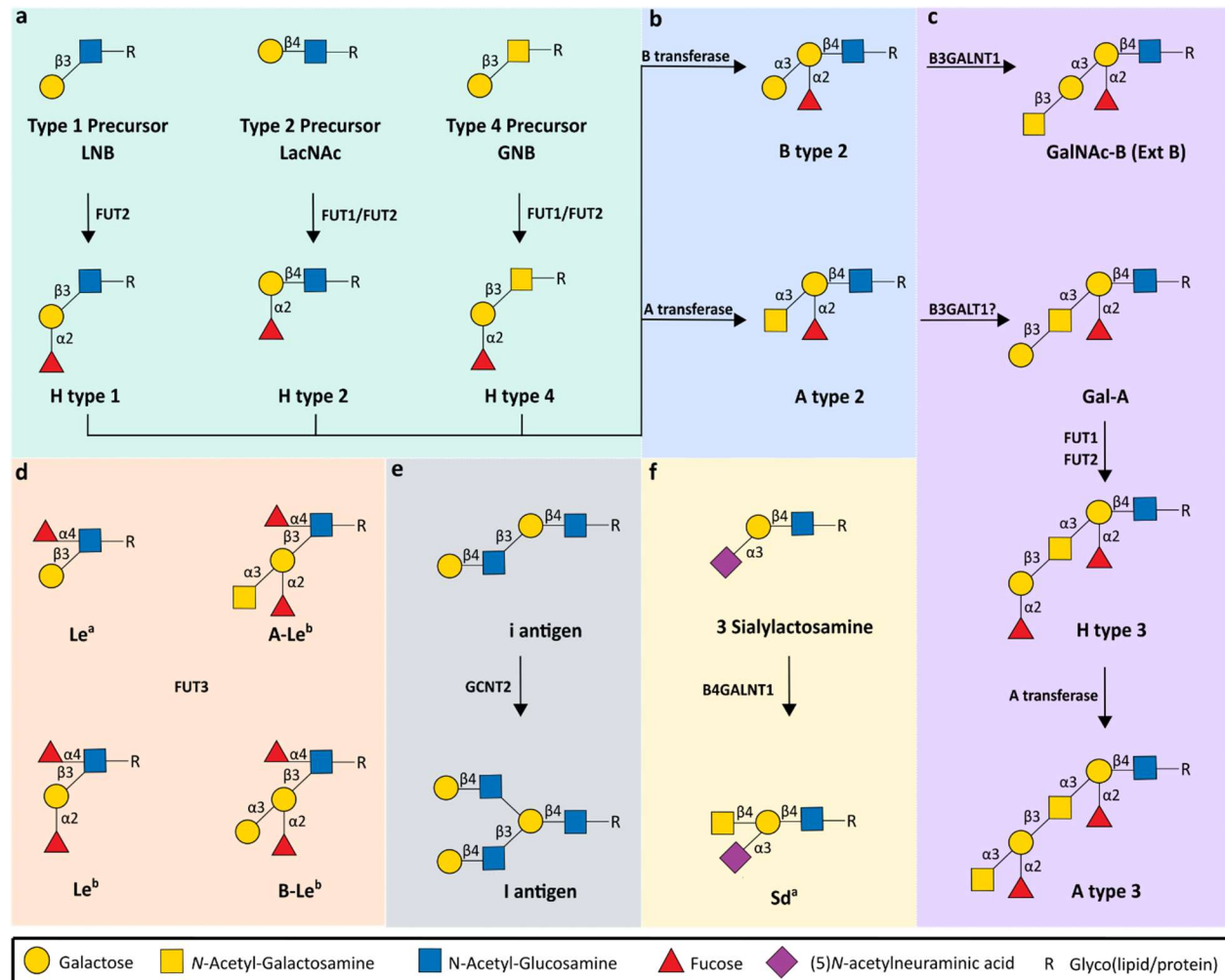


Figure 9. Overview of RBC antigen structures and transferases involved in their synthesis. **a.** Type 1, 2, and 4 precursors and H antigens, synthesized by the fucosyltransferases FUT1 and FUT2. **b.** The A or B antigens, synthesized from the H antigen by the A or B transferase. For simplicity, only the type 2 chain is depicted, as it is the abundant. **c.** Structures of the know extended B and A antigens, and the transferases involved in the synthesis. **d.** The Lewis antigen structures, resulting from the FUT3 transferase acting on either the type 1 or 2 precursor, or H type 1 or 2 structures. **e.** Antigens using the same precursor as the ABO antigens include **f** The I blood group antigens from the i antigens and the **f** The SID blood group antigens. Carbohydrates are depicted according to the Symbol Nomenclature for Glycans (SNFG) (<https://www.ncbi.nlm.nih.gov/glycans/snfg.html>).

1.6.3. The Lewis, SID, and I blood groups

The Lewis blood group is closely related to the ABO group, and are also synthesized from the common type 1 and 2 LNB/ LacNAc precursors^{127,171,187} (Figure 9d). The Lewis antigens on RBCs are a result of fucosylation by FUT3, an α -1,(3/4)-fucosyltransferase¹⁸⁸. This transferase is primarily active on type 1

precursors, resulting in the formation of Le^a antigen, but can also act on type 2 precursors, making the Le^x antigen. Once a Lewis antigen is synthesized, it can no longer be targeted by the FUT2 or A/B transferases. However, since FUT2 is active in the proximal Golgi, whereas FUT3 resides mostly in the distal Golgi, most of the type 1 or 2 precursors will be targeted by FUT2, resulting in the H antigen. As opposed to the FUT2 transferase, the FUT3 transferase can still act on the fucosylated H antigens, adding an additional fucose unit to the GlcNAc of the H antigen¹⁸⁹, resulting in the double fucosylated Le^b and Le^y antigens, respectively¹⁹⁰. Furthermore, FUT3 can act on either the A or B antigens of both type 1 and type 2, resulting in the A-Le^{b/y} or B-Le^{b/y} antigens¹⁷¹.

The Lewis antigens present on RBCs are not synthesized *in situ*, but are instead acquired from plasma glycosphingolipids passively embedded into the RBCs membrane¹⁵⁶. Therefore, all Lewis antigens on RBCs are present on secretory lipids of type 1 precursors (Le^a and Le^b), with no Lewis glycoproteins. The Le^x and Le^y of the type 2 precursors are not considered RBC epitopes, although they have been found to be present in small amounts in the plasma of O non-secretors¹²⁷.

The I blood group is characterized only by the I antigen, which is generated from unbranched repeats of LacNAc precursor, termed the i antigen (small i)^{171,192} (Figure 9e). A glucosaminyltransferase, termed GCNT2, acts on the i antigen by adding a branching GlcNAc β -1,6 linkage to the internal galactose of the i antigen, which is further elongated with the addition of a Gal β -1,4- linkage. The resulting I antigen is a branched LacNAc structure¹⁹³. This branched version is referred to as the I antigen, and is present in >99% of adults¹⁷¹. In very rare cases, a frameshift mutation in the GCNT2 gene can inactivate the transferase, resulting in the adult i phenotype, which has been correlated with the development of congenital cataract¹⁹⁴.

Finally, the SID blood group is made up of the Sd^a histo-antigen, which is a high frequency antigen occurring when a β -1,4-*N*-acetylgalactosaminyltransferase (B4GALNT2) adds a GalNAc residue to a sialylated LacNAc (Neu5Ac α -2,3-Gal β -1,4-GlcNAc)¹⁹⁵ (Figure 9e). Like the Lewis antigen group, the Sd^a antigen is mostly present in secretions that can adhere onto the surfaces of RBCs¹⁹⁶.

1.6.4. Globoside, P1PK, NOR, and FORS blood group

The globoside and P1PK blood groups are exclusively expressed on glycolipids^{197,198}. These antigens are unique, as they arise from a common lactosyl-ceramide core (Gb2), instead of the LacNAc/ LNB motifs. The addition of a Gal α -1,4- linkage by the A4GALT α -galactosyltransferase, onto the Gb2 core results in the P^k antigen. The P^k antigen is then further extended with a GlcNAc β -1,3-linkage by the β -*N*-acetylgalactosaminyltransferase, B3GALNT1, resulting in the P antigen, also called Globoside (Figure 10a). Notably, the B3GALNT1 transferase is the same enzyme that is responsible for generating the extended B antigen, GalNAc-B²⁰. P antigens are typically very abundant on the surfaces of RBCs, with up to 15 million antigens per cell¹⁹⁹.

The P1PK blood group is closely related, and is made up of the Aforementioned P^k antigen, along with the P1 and PX2 antigens²⁰⁰ (Figure 10b). The latter two are both synthesized from paragloboside, which occurs when the Gb2 precursor is elongated with a GlcNAc β -1,3-(Lc3) and a Gal β -1,4-linkage. The A4GALT transferase is responsible for the addition of the Gal α -1,4- linkage onto the paragloboside, resulting in the P1 antigen, which is mostly found on RBCs²⁰¹. Alternatively, the B3GALNT1 transferase can also add a GalNAc β -1,3- linkage onto paragloboside, resulting in the PX2 antigen²⁰².

A genetic variant of the A4GALT transferase can result in the addition of the Gal α -1,4- onto globoside instead of paragloboside, resulting in the NOR1 antigen^{203,204} (Figure 10c). The NOR1 antigen can be further extended with a GalNAc β -1,3-linkage to form the NOR_{int} antigen, which in turn can be further glycosylated with an additional Gal α -1,4-linkage by the A4GALT variant, forming the NOR2 antigen. These three antigens collectively make up the NOR blood group. Finally, an α -N-acetylgalactosaminyltransferase (GBGT1), can also add a GalNAc α -1,3-linkage to the P antigen (Globoside), resulting in the rare Forssman antigen of the FORS blood group^{205,206} (Figure 10d). Interestingly, the Forssman antigen was initially considered to be a subgroup of the A antigen, due to the high similarities between these two structures²⁰⁷. The addition of a Gal β -1,3- linkage to the P antigen results in the Globo5 (Gb5) structure, which can in turn be sialylated by a Neu5Ac α -2,3-linkage, resulting in the Luke antigen (LKE) (Figure 10e).

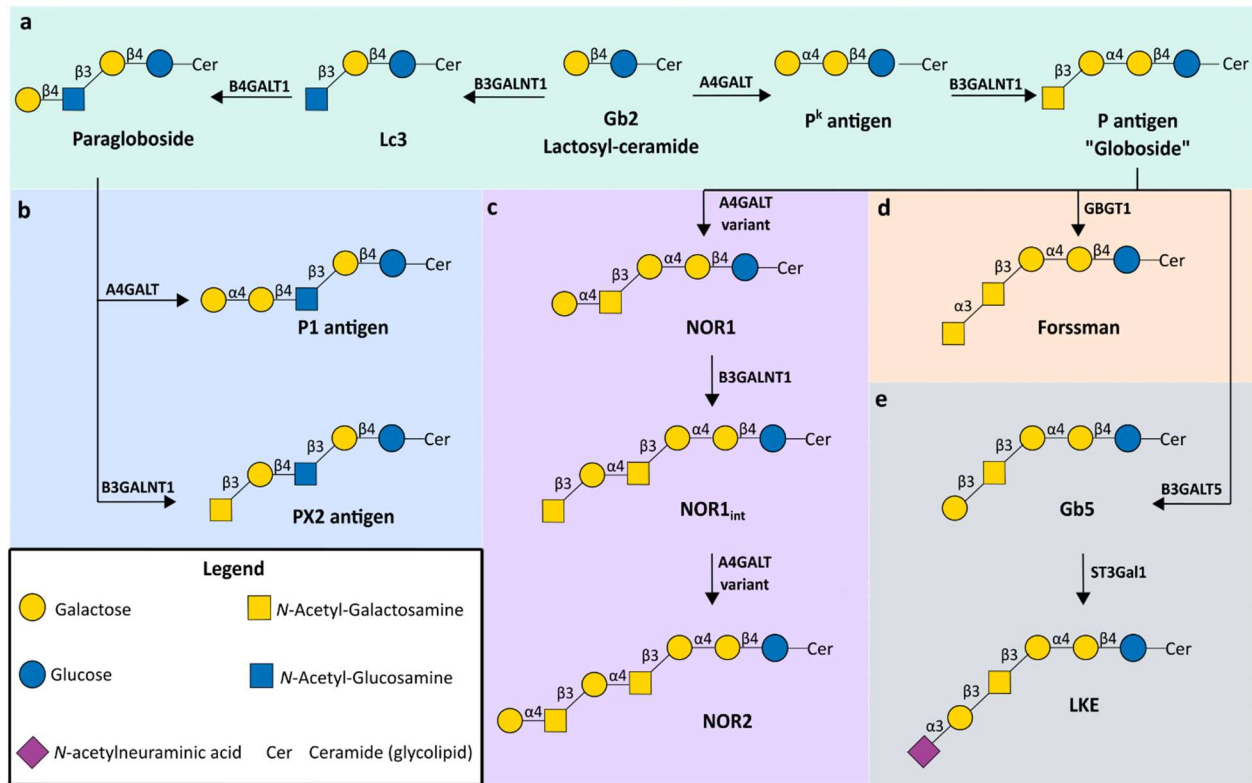


Figure 10. Overview of ceramide antigen structures and transferases involved in their synthesis. a. The Gb2 lactosyl-ceramide is the precursors to both paragloboside and the P antigen (Globoside) b. Synthesis of P1 antigen or PX2 antigen from paragloboside. c. Synthetic pathway of the NOR blood group from the P antigen. d. The Forssman antigen, synthesized from the P antigen. e. Synthetic pathway of the Luke (LKE) antigen from the P antigen. Carbohydrates are depicted according to the Symbol Nomenclature for Glycans (SNFG) (<https://www.ncbi.nlm.nih.gov/glycans/snfg.html>).

1.6.5. Clinical Significance of blood groups

Glycan antigens can be recognized by specialized antigen-binding heterodimeric proteins called immunoglobulins (Igs), which are synthesized by B cells¹⁶⁸. Immunoglobulins consist of four chains, two light chains and two heavy chains, with different variable regions determining their antigen binding

specificity. There are five classes of immunoglobulins, determined by amino acid sequence and structural commonalities between their respective heavy chains. These types are termed IgM, IgG, IgA, IgD, and IgE. IgM antibodies can be secreted by B cells as large pentamers²⁰⁸, and they function in primary immune response by coating target antigens, labeling them for destruction²⁰⁹. IgM antibodies generally have low affinity, and are not very reactive at physiological temperatures, and typically do not result in the development of autoimmune disease¹⁶⁸. IgG are the most dominant form of immunoglobulins found in the blood stream, and also have the longest half-life²¹⁰. The main function of IgGs is to recognize and protect from pathogens and xenobiotic molecules²¹¹. Therefore, antigens that are recognized by IgG antibodies are typically critical in regards to transfusion medicine, as these are known to cause severe immune responses^{168,171}. IgA antibodies are the predominant antibody type found at mucosal surfaces, where they play a role in protection from pathogens and toxins by either preventing binding to mucosal surfaces, or by direct neutralization of pathogens²¹². IgD antibodies are present only in very small amounts in the bloodstream (0.25% of immunoglobulins in serum)²¹³, and their specific function remains poorly understood, but may play a direct role in immune defense²¹⁴. Finally, IgEs are very potent antibodies associated with hypersensitivity and allergic reactions²¹⁵. The concentration of IgE antibodies in serum is, however, the lowest among all the immunoglobulin types, with normal concentrations reaching only 0.05% of IgG levels²¹⁶.

Naturally occurring immunoglobulins with specificity for blood group antigens can cause hemolytic transfusion reactions when blood transfusions to non-matching blood type occur. The abundance or types of specific antibodies varies from individual to individual, based on their respective phenotype¹⁷¹. Of the different carbohydrate-based blood groups, ABO is the most well-studied. Anti-A and anti-B antibodies are synthesized by blood type B and type A individuals, respectively. Type O individuals synthesize both the anti-A and anti-B antibodies, and also tend to express higher amounts of these antibodies²¹⁷. These antibodies were found to be primarily IgM type, but IgG and IgA isotypes also exist²¹⁸. In cases of the rare Bombay phenotype, those individuals also produce anti-H IgM and IgG antibodies²¹⁹. The anti-ABH antibodies have been shown to cause hemolytic transfusion reactions (HTRs) upon antigen recognition, which can have potential fatal consequences for blood recipients in cases of a miss-match^{220,221}.

Antibodies against the other carbohydrate-based blood groups also occur, but appear to be less critical in regards to safe transfusion. Anti-Lewis antibodies occur as IgM isotypes for individuals who do not express the FUC3 active enzyme¹⁸⁹. These antibodies are not typically reactive at physiological temperature and therefore do not cause HTRs, although recent evidence suggests that some anti-Lewis IgM antibodies are active at 37° C, and may cause HTRs^{222,223}. Anti-I antibodies are also predominantly IgM isotypes active at cold temperature, and typically do not pose a threat of HTR¹⁷¹. A potential problem arises in the rare instance of adults with the i_{adult} phenotype, who carry Allo-anti-I antibodies. A study by Chaplin *et al.*²²⁴ in 1986 reported that these antibodies may be clinically significant, although the available data on this remains scarce, perhaps since this is a very rare phenotype appearing in less than 1% of adults worldwide.

Antibodies against the P antigen occur as IgM and IgG isotypes¹². The anti-P IgG can cause potentially severe HTRs, thus the P antigen has a clinical significance¹⁷¹. Anti-P1 antibodies are typically IgM isotypes reactive at room temperature and do not usually cause any HTRs¹⁹⁸. The Anti-Px2 antibody, like the anti-P, also occurs in both IgM and IgG isotypes and may also cause HTRs¹⁹⁷. However, in the same regards as the Allo-anti-I antibodies, there is very limited data available on the significance of the different P1PK blood group antibodies.

Both the NOR and FORS blood groups are extremely rare, occurring in about 0.1 and 0.01 % of the populations, respectively^{198,207}. Consequently, the clinical significance of the anti-NOR and anti-FORS IgM antibodies is not fully understood. There is some evidence to suggest that anti-FORS antibodies may cause intravascular hemolysis²²⁵.

Finally, anti-Sd^a IgM antibodies also occur naturally in individuals with a deficiency in the B4GALNT2 transferase¹⁷¹. These antibodies are typically not expected to cause HTRs, however, some odd mixed-field hemagglutination is observed when mixing these antibodies with RBCs¹¹.

Due to their rare occurrence, and the low reactivity of IgM antibody isotypes, antibodies of non-ABO blood groups are thought to play an insignificant role in clinical blood transfusion safety, despite the lack of compelling data. The significance of the ABO blood group antibodies is undeniable, and therefore the ABO blood group has since come to be recognized as the most important for safe blood transfusions (Figure 11). When performing blood transfusions, the ABO phenotype must be taken into account to avoid crossmatches between donor RBCs and recipient plasma antibodies. In this system, O type blood lacks both the A and B antigen epitopes and can be donated to any recipient. However, type O individuals synthesize both the anti-A and anti-B antibodies, and can only receive blood from other type O donors. Type A and type B individuals can receive RBCs from either O type donors or donors of the same respective type. Some individuals carry the AB type, lacking the anti-A and anti-B antibodies, and they can receive RBCs from any donor, but can only donate to other type AB individuals.

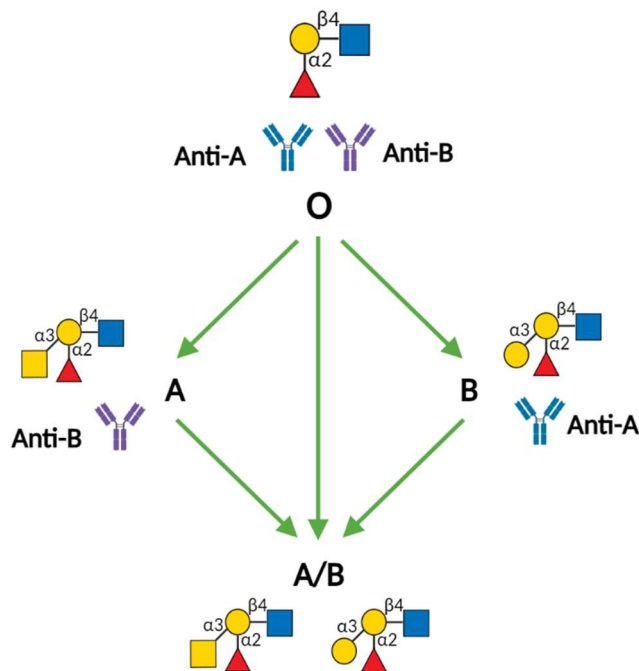


Figure 11. The classical ABO blood group system for transfusions. The antigens and respective antibodies related to the ABO blood groups are depicted. Safe transfusions are depicted as green arrow. This figure was made using Biorender.

1.7. Enzymatic Conversion of ABO blood antigens

Discordance of ABO blood groups between donors and recipients and the short lifetime of donor RBCs contribute to a high demand for donor blood. This is particularly true for type O donors, as it is frequently used in emergency situations, in which there is limited time for correct typing of RBCs. A potential solution to this issue is the use of enzymes to cleave the terminal non-reducing ends of the A and B antigen, respectively, resulting in the H antigen of the O type. Early reports indicating that naturally occurring enzymes existed with capabilities to cleave the A and B antigens was reported by a Japanese group in 1963¹³. Following this, a report describing an α -galactosidase from Coffee bean able to remove the B antigen from RBCs in 1975¹⁴, and a report of an α -*N*-acetylgalactosaminidase from *Clostridium perfringens* with apparent A antigen degrading activity in 1980²²⁶, further solidified the potential of this strategy. Since then, the field has progressed steadily over the course the next 43 years, resulting in the discovery of efficient enzymes from the HGM to target RBC antigens. A timeline summarizing the work on Enzyme Converted O (ECO) blood is shown in Figure 12.

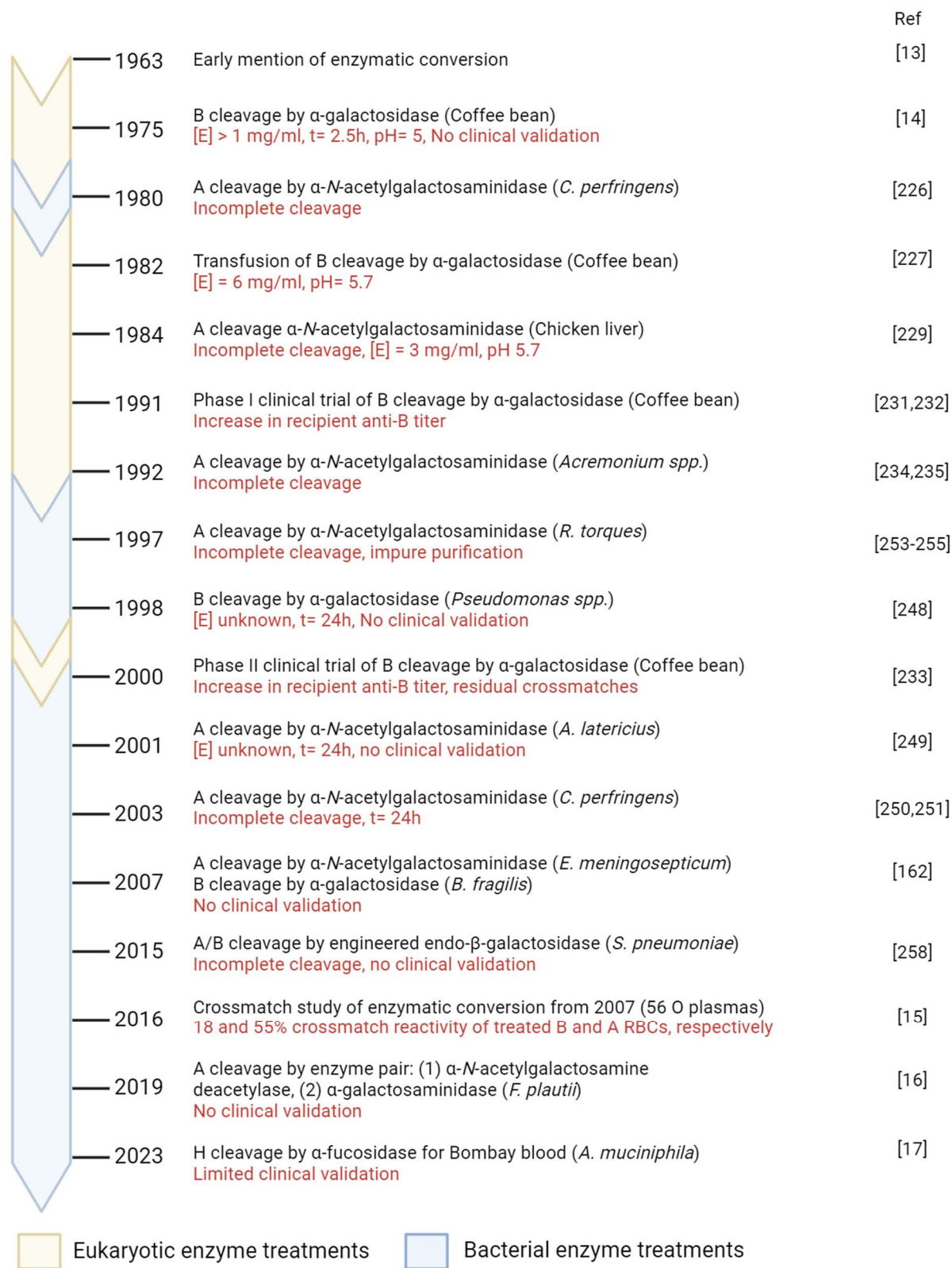


Figure 12. Timeline of research for enzymatic conversion of blood. An overview of ECO studies since 1963, highlighting the activities that were researched and the host organisms of the enzymes. Bottlenecks in each report is shown in red text. This figure was made using Biorender.

1.7.1 First generation enzymatic conversion of AB antigens

The first demonstration of a safe transfusion using ECO-B RBCs was reported by Goldstein *et al.* in 1982²²⁷. They utilized the α -galactosidase from green coffee beans to cleave the terminal galactosyl unit of the B antigen on RBCs. After verifying donor RBC viability post enzymatic treatment, initial clinical tests were performed using Gibbons. RBCs from Gibbons were treated with the enzyme and returned to the donor to monitor for potential reactions of the ECO RBCs by anti-enzyme antibodies. No adverse effects were reported, and initial clinical tests were then performed with healthy human volunteers. A volunteer of type A, B, and O were recruited, and sample of RBCs was taken from the B type donor and treated with the α -galactosidase. Following this, the enzymatically ECO RBCs were mixed with plasma samples from all three participants to test hemagglutination. Hemagglutination is the clumping of particles, and in the case of RBCs, occurs when blood group antibodies bind to their respective antigens, leading to RBC coagulation (Figure 13). Initial tests of the group B-ECO cells showed no visible hemagglutination when mixed with type A and type O plasma, verifying the loss of the antigen. It should be noted, the limit of detection for hemagglutination experiments at this time was estimated to be about 2000 antigenic sites per cell. Finally, a small volume of 5 mL of ECO cells were labeled with ⁵¹Cr and administered to all three participants to monitor the survival of the ECO cells after transfusion. Blood test were taken at 15, 30, and 60 minutes after transfusion, and over a period of 49 days. The radioactivity of the samples was used to measure the survivability of ECO RBCs in the recipients. The ECO RBCs entered the circulation successfully, and were not targeted for rapid destruction.

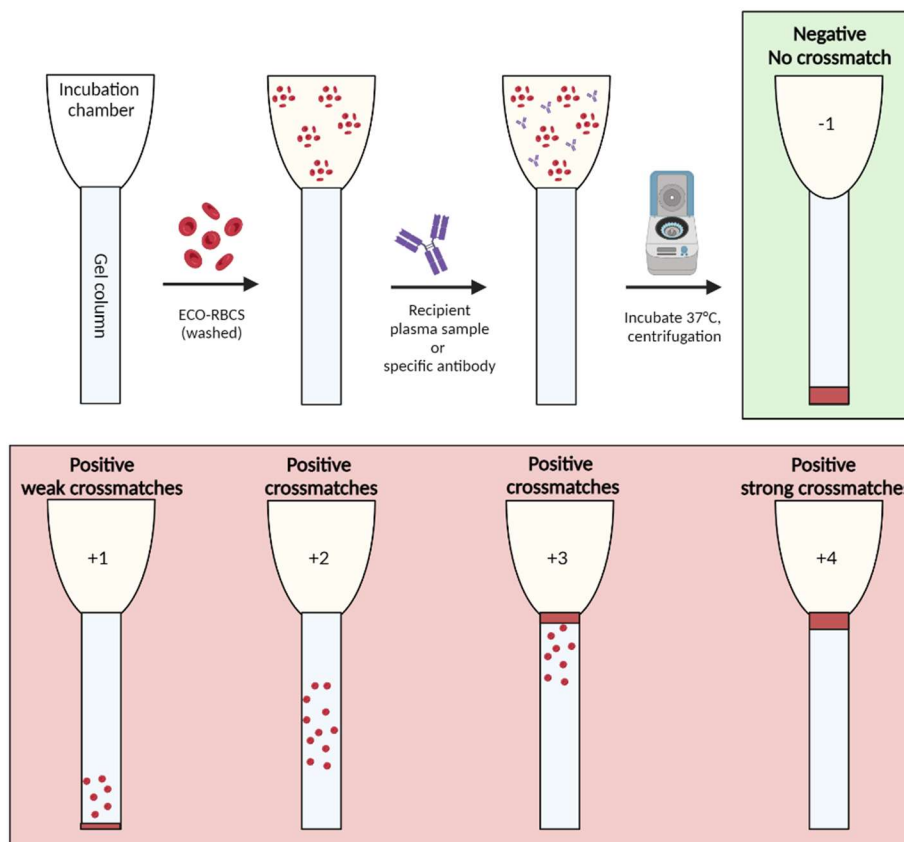


Figure 13. Hemagglutination experimental setup. Red blood cells are treated with the corresponding enzyme and washed to remove residual enzyme. The ECO RBCs are then mixed with plasma samples or specific antibodies at 37 °C for various time intervals. If antibodies bind to their corresponding antigens, RBCs start to coagulate. After incubations, the cells are spun in a centrifuge through a gel column. If no coagulation occurs, all ECO-RBCs will accumulate at the bottom (-1). If crossmatch reactions have occurred, resulting in coagulation of ECO-RBCs, traversal through the gel column is hindered. Positive crossmatch reactions are scored from +1 to +4 to indicate the weak or strong crossmatch reactivity. This figure was made using Biorender.

Although successful cleavage of the B antigen was observed, the treatment required a concentration of 6 mg mL⁻¹ at pH 5.5, to convert one full RBC unit (200 mL). The high quantities of enzyme required for conversion would be an economic barrier for implementation into clinical practice. Furthermore, enzymatic conversion at neutral pH is preferred for clinical implementation²²⁸. Goldstein and colleagues followed up in 1984 with another enzyme, an α -N-acetylgalactosaminidase from chicken liver, for the enzymatic conversion of the A antigen into O²²⁹. This enzyme was reported to convert the A antigen on RBCs on the A₂ subgroup, however experimental evidence was not published. Apparently, treatment of the A₁ subgroup required “stringent” conditions, which rendered them unsuitable for transfusion. The conversion of A RBCs was performed with 3 mg mL⁻¹ for conversion of 1 blood unit. The A enzyme was also acidic with a pH optimum of 3.65²³⁰, and a pH of 5.7 was used for conversion of RBCs.

Phase I clinical trials were initiated by Lenny *et al.* in 1991 for ECO type B RBCs, using the green coffee bean α -galactosidase^{231,232}. Recipients were administered with one or more whole ECO blood units. The administration proved to be “safe”, as no ill effects were observed on the recipients, and the ECO-RBCs entered circulation without rapid destruction (monitored by radioactivity of ⁵¹Cr labeled ECO-RBCs). In spite of this, a slight increase in Anti-B titer was observed in type O recipients, although the reason for this was not clear. In 2000, Kruskall *et al.* continued the work on the type B ECO in a larger phase II clinical trial²³³. This time, a recombinant form of the α -galactosidase was used to convert type B blood units, and 21 volunteers of type A or O received the ECO-RBC transfusions. Similar to previous results, no adverse effects were observed on the recipients, however, a slight increase in B-titer was observed once again. Crossmatch studies were also performed using the ECO-RBCs and recipient plasma samples, and strikingly, hemagglutination was observed in 20% of type A plasma samples and 40% of type O plasma samples. Ultimately, due to the unexplained crossmatch reactivity, the excessive amount of enzyme that would be necessary in order to fully convert a unit of RBCs, and the low pH optimum, clinical use of these enzymes was not deemed feasible.

An α -N-acetylgalactosaminidase from *Acremonium* spp. (Ascomycota phylum of fungi), was identified and isolated by Tsuji *et al.* in 1989²³⁴. They validated the α -GalNAcase activity against mucin substrates, including porcine, horse gastric mucin and human salivary mucin, resulting in the removal of the A antigen. The enzyme was tested on type A RBCs using hemagglutination test, and they observed a decrease in crossmatch reactivity. In 1992, they followed up with a more in depth serological study of the enzyme against RBCs membranes, as well as glycoproteins and glycolipids extracted from RBC membranes²³⁵. A antigen cleavage was observed, however complete conversion of A antigens was not achieved. Furthermore, the enzyme carried some of the same limitations seen previously, having a pH optimum of 4 and requiring long incubations times of 12 to 24 hours.

The enzymes used by Goldstein *et al.* and Tsuji *et al.* for the conversion of RBCs were all members of GH27. Characterized members from this family are typically α -galactosidases, but some eukaryotic members

have α -N-acetylgalactosamine activity^{236,237}. Other reported bacterial activities in this family include isomalto-dextranases²³⁸ and β -L-arabinopyranosidases^{239–241}. Enzymes of the GH27 family have been demonstrated to be exo-glycosidases employing the retaining double-displacement mechanism^{242,243}. Crystal structures of GH27 enzymes from *Gallus gallus* (chicken)²⁴⁴, *Oryza sativa* (rice)²⁴⁵, and *Homo sapiens* (human)²⁴⁶ revealed the structural organization of GH27 enzymes, composed of an N-terminal $(\beta/\alpha)_8$ -TIM barrel, linked to a C-terminal antiparallel β -jellyroll domain. The catalytic nucleophile and corresponding acid/base residue are a pair of highly conserved aspartic acid residues situated above and below the anomeric carbon of the (-1) subsite²⁴⁷.

1.7.2 Bacterial enzymes for universal blood

After the success of antigen cleavage by exoglycosidases by Goldstein and colleagues, other research groups sought to discover improved enzymes for the conversion of A and B group RBCs. In 1998, Bakunina *et al* employed an α -galactosidase from the marine bacterium, *Pseudomonas* spp. for conversion of group B RBCs²⁴⁸. As opposed to the eukaryotic enzymes used previously, this enzyme had a neutral pH optimum. The conversion conditions reported in the study are not clear, and a 24 hour treatment with unknown concentration of the enzyme was employed. Hemagglutination tests showed the disappearance of anti-B reactivity to ECO-RBCs, alongside an increase in anti-H reactivity, indicating B antigen cleavage. Later, In 2001, Bakunina *et al.* identified an α -N-acetylgalactosaminidase from another marine bacterium, *Arenibacter latericius*, with a pH optimum between 7 and 8²⁴⁹. Treating type A RBCs with the enzyme for 24 hours resulted in decreased hemagglutination titer similar to that reported for the B converting enzyme. Ultimately, the enzymes were not suitable for use in clinical practice due to the long treatment time required.

Smith, *et al.* purified the *C. perfringens* α -N-acetylgalactosaminidase with previously described A antigen activity²²⁶, to further validate the potential of the enzyme for conversion of the A antigen of A₂ membranes using ELISA assays^{250,251}. The enzyme had a pH optimum of 6.5–7, and was active at pH ranges from 6–8²⁵². Enzymatic treatment was tested on the A₂ membranes, and 93% of the A antigens were depleted following a 24 hour incubation time with the enzyme. Conversion of the A antigen on intact RBCs was never reported. Another group identified an α -N-acetylgalactosaminidase from a *Ruminococcus torques* strain (IX-70) from human feces, and assayed the enzyme for activity against type A RBCs using hemagglutination test²⁵³. Antigen levels were reduced, but full conversion was not achieved. Notably, the enzyme solution was not pure, including sialidases and other contaminants^{254,255}.

In 2007, Liu *et al.* reported the discovery of bacterial enzymes against both A and B antigens, with improved activity¹⁶². A library of bacterial lysates and fungal supernatants were screened for activity against fluorescently-conjugated A and B tetrasaccharides at pH 7. Enzymes able to convert the A and B oligosaccharides were discovered from Gram negative pathogenic *Elizabethkingia meningosepticum* (NagA) and mucolytic human gut member *Bacteroides fragilis* (FragA), respectively. Based on sequence analysis, these enzymes defined two new families of glycoside hydrolases, family GH109 for A converting enzymes and family GH110 for B converting enzymes. The enzymatic cleavage of the A and B antigens was monitored with flow cytometry, using fluorescently labeled anti-A and anti-B. The enzymes showed great improvements in activity compared to previous enzymes. NagA was able to fully convert a unit of A RBCs at 300 $\mu\text{g mL}^{-1}$ at pH 6.8, while FragA was able to convert a unit of type B RBCs at only 10 $\mu\text{g mL}^{-1}$. These

enzymes worked using a low ionic strength glycine buffer, which allowed the use of PBS (higher ionic strength, higher pH buffer) to wash the enzymes off the cells.

The *E. meningosepticum* enzyme, NagA, became the founding member of the GH109, characterized by their activity against the blood A antigen. A crystal structure of this enzyme revealed a unique fold among GHs, consisting of a Rossmann fold with a tightly bound NAD⁺ cofactor. The NAD⁺ contributes to the enzyme mechanism through a redox-assisted elimination. The NAD⁺ co-factor oxidizes the C3 of the substrate, which in turn acidifies the C2, similar to the previously reported mechanism in GH4. The catalytic base was determined to be a tyrosine residue attacking the C2 of the substrate, containing the *N*-acetyl group. The catalytic acid/base residue was recently identified by Teze *et al.* based the structure of AmGH109A from *A. muciniphila*²⁵⁶. A highly conserved histidine located in a flexible GGHGG loop was responsible for donating a proton to the leaving group. A dual mechanism for the enzymes was reported in this study, consisting of both α -retaining and β -inverting mechanism, although a general preference for the α -retaining mechanism is observed.

The *B. fragilis* enzyme, FragA, became the founding member of family GH110, the first family of α -galactosidases to employ the inverting mechanism. A follow-up paper from Liu *et al.* reported that enzymes of this family fall into one of two different sub-families²⁵⁷. One of the subfamilies is specific for the branched B antigen structure (GH110A), while the other subfamily is able to cleave the linear α -1,3-galactose structures in addition to the fucosylated B antigen. The first structure of a GH110 enzyme from the marine bacterium *Pseudoalteromonas distincta* was reported by McGuire *et al.* in 2020 and revealed a catalytic domain with a parallel β -helix fold, and two small inserted β -helix domains on either side of the active site pocket. A phylogenetic analysis showed that this enzyme (*PdGH110B*) belonged to the B sub-family, in accordance with the measured activity against the linear α -1,3-galactobiose. Interestingly, this enzyme belongs to a predominantly marine cluster in the phylogenetic tree and displayed no activity against B antigens. The structure of *PdGH110B* also allowed the identification of the general acid, a highly conserved aspartic acid residue. A water molecule was also located close to the C1 of the substrate, with two additional aspartic acid residues as putative general bases, within distance to be able to activate the water molecule as the nucleophile.

Crossmatch reactions using the GH109 and GH110 enzymes were performed later by Gao *et al.* in 2016¹⁵. In this study, ECO-RBC units were mixed with plasma from 56 different type O donors. Strikingly, 18% of the plasmas showed positive crossmatch reactivity with the ECO-B cells, and 55% of plasmas showed positive crossmatches with the ECO-A cells. These data suggest further underlying issues for correct blood matching besides the conversion of the A and B antigens may exist, as opposed to the previous hypothesis that enzyme efficiency was the issue.

1.7.3. Endo-active enzyme for RBC antigen conversion

A different approach by Kwan *et al.* in 2015²⁵⁸ explored the use of bacterial endo- β -galactosidases to target the core LNB and LacNAc of the ABO blood group type 1 and 2 chains, to convert both the A and B antigens using a single enzyme. Enzymes from GH98, whose founding member had been reported to cleave the A and B type 2 chains in 2005⁶⁵, were chosen for the strategy. GH98 is a small family of endo- β -1,4-galactosidases. Most of the characterized members are active against the A, B, or Lewis antigens, apart from a single enzyme from *Bacteroides ovatus* with reported endo- β -1,4-xylanase activity. The

enzymes of the GH98 family act using the inverting mechanism, and the catalytic residues were identified by structural analysis of two GH98 enzymes from the human pathogen, *Streptococcus pneumoniae* (*Sp3GH98* and *Sp4GH98*)²⁵⁹. A highly conserved glutamic acid residue acts as the general acid, while the general base consists of an aspartate and glutamate diad. The structure of GH98 enzymes consists of a $(\alpha/\beta)_8$ barrel catalytic domain, and a β -sandwich domain.

All characterized GH98 enzymes are specific towards type 2 β -1,4-linkages. To achieve complete conversion of A and B antigens, the enzyme would need to also target the type 1, 3, and 4 chains. Using structure-guided mutagenesis, Kwan *et al.*²⁵⁸ were able to generate a variant of the *S. pneumoniae* GH98 with increased activity against the type 1 LNB chain, albeit at the cost of decreased activity towards the type 2 linkage. The mutant enzyme was tested on a commercial cell strain (A431) expressing the A type 1 antigen and the A-Le^b type 1 antigen. The mutant showed improved conversion compared to the wildtype, however, complete conversion was still not achieved. An additional test was performed using type B RBCs, validated by hemagglutination tests. ECO-RBCs showed no hemagglutination, indicating removal of the B antigen was achieved, albeit at higher enzyme concentrations, consistent with the low catalytic turnover of the enzyme. Although the approach highlighted the potential of structural engineering in improving enzymatic conversion of AB antigens, the activity of the enzyme remains too low for clinical implementation. More importantly, the strategy of using an endo-enzyme to cleave the antigen chains removes the H antigen, and uncovered epitopes may be recognized by antibodies and targeted for destruction. The safety of the method still remains to be validated, and the use of exo-glycosidases to convert A and B into O type blood remains the most suitable approach for generating universal O type blood.

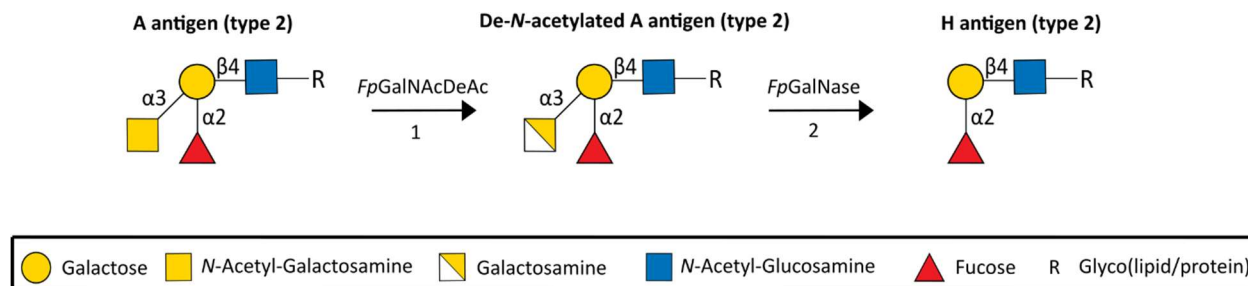


Figure 14. Two-step mechanism of *FpGalNAcDeAc* and *FpGalNase* to convert A antigen to H antigen. The reaction involves the initial De-N-acetylation of the terminal non-reducing GalNAc by the carbohydrate esterase, *FpGalNAcDeAc*, resulting in terminal galactosamine. In a second step, *FpGalNase* can cleave the terminal galactosamine, uncovering the H antigen.

1.7.4. Enzymes from the HGM for ABO antigen conversion

In the search for catalytically superior exo-glycosidases for the conversion of A and B antigens, Rahfeld *et al.* employed a metagenomic approach to screen enzymes from the HGM of an AB blood type donor¹⁶. This led to the discovery of a pair of enzymes from *Flavonifractor plautii* that, when used conjointly, were able to efficiently convert the A antigen into the H antigen (Figure 14). One of these enzymes is an esterase showing specificity towards the fucosylated A antigen, based on the lack of activity measured against linear non-fucosylated A antigen structures and linear GalNAc α -1,4-linkage chains. The enzyme, termed

FpGalNAcDeAc, de-acetylates the terminal non-reducing GalNAc of the A antigen into galactosamine. The second enzyme is a GH36 α -galactosaminidase, active on the galactosamine product of the first enzyme. These two enzymes used together worked much more efficiently than any previous A antigen converting candidate, able to convert a unit of blood using only $0.5 \mu\text{g mL}^{-1}$ of both enzymes in the presence of the molecular crowder, dextran. A molecular crowder is a large molecule that, when introduced in high concentrations, occupy a significant fraction of the available volume, effectively increasing the concentration of other biomolecules²⁶⁰. In the absence of dextran, $5 \mu\text{g mL}^{-1}$ of each enzyme is needed for conversion, still presenting a 30-fold improvement in catalytic turn-over compared to NagA from *E. meningosepticum*. This work highlights the adaptation of some members of the HGM towards human O-glycoconjugates. A potential risk using this approach, is residual galactosamine left on the RBCs potentially causing acquired B syndrome as described by Gerbal *et al.* in 1975²⁶¹. Here, antibodies in the plasma may recognize the galactosaminyl as a galactosyl residue, and would react as if the B antigen was present. Using an excess of the enzyme pair for conversion may be necessary in order to fully alleviate this risk. Furthermore, clinical validation has not been published, so the question remains whether the crossmatch reactivity of recipient plasma against ECO-RBCs is still an issue.

The first step of the conversion involves *FpGalNAcDeAc*, a carbohydrate esterase. *FpGalNAcDeAc* is the first reported carbohydrate esterase with specificity towards the blood A antigen. A crystal structure of the enzyme in complex with the B antigen revealed the mechanism and determinants of specificity. The active site involves a divalent metal binding site important for the reactions. A significant reduction in activity was observed after mutating the histidine and aspartic acid ligands of the metal ion. A glutamic acid residue was proposed to activate a nearby catalytic water molecule as the nucleophile in the deacetylation mechanism. While no hydrogen bonds were formed to the Gal of the B antigen trisaccharide, polar interactions are observed between the fucosyl unit and a serine and aspartic acid residues, explaining the selectivity for fucosylated structures.

Enzymes from GH36 share evolutionary relation to GH27, belonging to the same clan (GH-D). This means the enzymes of these families share a $(\beta/\alpha)_8$ fold and retaining catalytic mechanism. Similar to GH27, GH36 members contains both α -galactosidases and α -N-acetylgalactosaminidases, in addition to some members from plants including stachyose or raffinose synthases. Structural and mutagenic studies have confirmed the catalytic residues, consisting of a pair of aspartic acid residues acting as the general acid and general base^{262,263}. As opposed to GH27 enzymes, GH36 contains a larger N-terminal β -sheet domain, which in some cases contributes key substrate binding residues. The enzyme (*FpGalNase*) reported by Rahfeld *et al.* is the first with α -galactosamine activity in the family, and furthermore, the first instance of a GH36 enzyme with activity against blood group A antigens.

Recently, two independent studies showed the high activity of an *A. muciniphila* GH95 α -L-fucosidase, *AmGH95B*, towards H antigens. GH95 is a family of inverting α -L-1,2-fucosidases²⁶⁴, but also include members with α -L-1,2-galactosidase activity²⁶⁵. Reported fucosidases of this family are strictly active against Fuca-1,2-Gal linkages, which are found on xyloglucans, as well as human-derived glycans including the ABO and Lewis antigens (Figure 9), and human milk oligosaccharides (HMOs)²⁶⁶. This family of enzymes use a unique variation of the inverting reaction mechanism, in which the general base (typically an asparagine) residue has to be activated by a neighboring aspartate²⁶⁷. Members of GH95 harbor an $(\alpha/\alpha)_6$ barrel catalytic domain, and typically include a large N-terminal β -sandwich domain and a small C-terminal β -domain. The fold of the catalytic domain is similar to that seen in CAZy families belonging to the clan GH-L (GH15, GH65, and GH125), which all act on α -linkages. The structure of a GH95 α -L-galactosidase

from *B. ovatus* reported the determinant of the specificity²⁶⁵. A threonine was observed acting with the C6-OH group of the bound α -galactose ligand. In the corresponding fucosidase structures, this residue is a histidine.

As mentioned previously, Shuoker *et al.* reported activity of this enzyme towards the H type 1, 2, and 3 epitopes on mucin, as part of the mucin decapping strategy of *A. muciniphila*¹⁵¹. Additionally, a study by Anso *et al.* reported the activity of the same enzyme against the H antigens on RBCs, for the conversion of universal O blood into the rare Bombay type¹⁷. In this study, hemagglutination experiments were performed using 20 type O donor RBCs, together with the plasma of a Bombay type donor, or an anti-H lectin. Treatment of the RBCs with wildtype *AmGH95B* at concentrations of 5 $\mu\text{g mL}^{-1}$ for 1 hour at 37 °C resulted in no observed hemagglutination, whereas a catalytically impaired mutant of *AmGH95B* showed clear hemagglutination. To validate the results, flow cytometry analysis using fluorescent anti-H antibodies and anti-H lectin was conducted, which showed a complete loss of the H antigen after incubation. These studies highlight the efficiency of mucolytic members of the HGM as a source for efficient glycoside hydrolases.

Table 4: GH families with known ABO antigen activity

CAZy classes	Activity	Antigen target	References
GH27	α -1,3-galactosidase	B	[14,227,248]
	α -1,3- <i>N</i> -acetylgalactosaminidase	A	[234,235,249]
GH36	α -1,3- <i>N</i> -acetylgalactosaminidase	A	[226,250,251]
	α -1,3-galactosaminidase	De-acetylated A	[16]
GH95	α -1,2-L-fucosidase	H antigen	[17]
GH109	α -1,3- <i>N</i> -acetylgalactosaminidase	A	[162]
GH110	α -1,3-galactosidase	B	[162]

In conclusion, research since 1963 has resulted in the discovery of enzymatic activity against A, B and H antigens from several families of GHs (Table 6). Initially, enzymatic treatment of RBCs required high enzyme load and long incubation times. Recently, more efficient enzymes discovered from the HGM niche has resulted in a more feasible enzyme load for conversion of A and H antigens. However, in depth crossmatch analysis and clinical validation are lacking for these, and the question remains whether the high crossmatch reactivity observed in previous studies by Kruskall *et al.*²³³ and Gao *et al.*¹⁵, remain a bottleneck in the implementation of the method to blood clinics. As it has been highlighted in this introductory chapter, there are many similarities between RBC glycans and human mucin *O*-glycans, including the ABO and Lewis antigens. Recent efforts have highlighted enzymes from the HGM for RBC antigen conversion, and we aim to further test this theory by mining the genome of the mucolytic specialist, *A. muciniphila*, for enzymes active against the A, B, and extended A/B antigen structures. Additionally, we aim to test the significance of the extended antigens on ECO compatibility, using a large-scale crossmatch study. Finally, structural analysis of enzymes with high efficacy to convert RBC antigens may help further our understanding of the molecular signatures that drive enzyme activity against human glyco-conjugates. Our overarching aim, is to push the concept of ECO blood forward, and hopefully pave the way for safer, universal donor blood.

Chapter 2 – Manuscript: Targeting extended blood group antigens by *Akkermansia muciniphila* exoglycosidases unveils a missing link for generating ABO-universal donor blood

Targeting extended blood group antigens by *Akkermansia muciniphila* exoglycosidases unveils a missing link for generating ABO-universal donor blood

Authors & affiliations

Mathias Jensen^{1,4}, Linn Stenfelt,^{1,2,4} Jennifer Ricci Hagman^{2,3}, Jakob Michael Pichler¹, Julia Weikum¹, Tine Sofie Nielsen¹, Annika Hult³, Jens Preben Morth¹, Martin L. Olsson^{2,3,5}, Maher Abou Hachem^{1,5}

¹Department of Biotechnology & Biomedicine, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

²Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University, Lund, Sweden

³Department of Clinical Immunology and Transfusion Medicine, Office for Medical Services, Region Skåne, Sweden

⁴These authors contributed equally: Mathias Jensen, Linn Stenfelt

⁵Correspondence: martin.l.olsson@med.lu.se; maha@bio.dtu.dk. These authors jointly supervised this work.

Abstract

Extended versions of glycosphingolipid ABO-antigens on erythrocytes are biochemically described, but their impact on blood transfusion compatibility remains unexplored. We evaluated 21 enzymes from the dedicated mucin-degrader *Akkermansia muciniphila*, harnessing its adaptation to deconstruct ABO-epitopes on mucin *O*-glycans. This led to the discovery and optimisation of exoglycosidase blends able to transform both canonical A and B antigens and four extensions thereof with excellent efficacies. Structural analyses revealed signatures underpinning the high enzymatic activity on human cell-surface glycoconjugates. Strikingly, targeting both the canonical and extended ABO antigens on erythrocytes significantly improved compatibility with group O plasmas, compared to only converting A or B antigens. Our findings give a long-sought rationale to the involvement of extended A/B antigens in persistent crossmatch reactivities after the apparent removal of A/B antigens. This reopens avenues towards generating enzyme-converted ABO-universal blood, and highlights mucolytic microbiota specialists as sources of superb biocatalysts to achieve this goal.

Introduction

More than 118 million life-saving blood transfusions are performed yearly¹. Matching blood groups of donor red blood cells (RBCs) and recipients is crucial to avoid potentially fatal transfusion haemolytic reactions². The ABO system is the prime immunological barrier in blood transfusion due to potential positive crossmatch reactions between naturally-occurring antibodies against non-self ABO antigens on donor RBCs³. Immuno-dominant terminal glycans in glycolipids⁴ and *N*-glycoproteins⁵ on RBC surfaces define the A, B and H antigens (Fig. 1a). The relative overconsumption of universal group O is attributed to medical emergencies with unknown patient blood group, paediatric transfusions, and ambiguous ABO-typing due to previous transfusion or acquired/inherited variants⁶. A universal group O blood inventory would offer an attractive solution to blood shortages, reduce outdated blood units, and eliminate the risk of ABO-dependent haemolytic adverse events. The enzymatic conversion of group A/B RBCs to the ABO-universal group O (ECO-blood) was pioneered in 1982 using coffee-bean α -galactosidase to convert B antigens to H⁷. Later, higher-efficacy α -*N*-acetylgalactosaminidases (α -GalNAcase) and α -galactosidase, discovered from a bacterial library⁸, defined glycoside hydrolase families GH109 and GH110, respectively. Recently, mining a genomic library from human gut microbiota (HGM) identified a sequential deacetylase/ α -galactosaminidase system for A antigen conversion⁹.

Despite successful group B-derived ECO-blood transfusion in early clinical trials¹⁰, safety concerns were raised due to incompatibility issues, which were observed as cryptic positive crossmatches with recipient plasmas^{11–15}, despite near-complete conversion of canonical A/B antigens. Blood group A¹⁶, is divided into A₁ and A₂ subtypes. The A₁ subtype differs quantitatively from A₂, by displaying considerably higher amount of A antigens^{17,18}. Clausen et al. Interestingly, A antigens can be extended by a β -galactosyl-1,3-unit, which can be α -1,2-fucosylated to give the Gal-A and the H type 3 structures, respectively^{19–21} (Fig. 1a). Further extension H type 3 precursors results in the A type 3 antigen, which is a repetitive A structure, uniquely detected on A₁ RBCs (Fig. 1a). Recently, an extension of the B antigen with an β -*N*-acetylgalactosamine-1,3-unit has been reported and the corresponding transferase was identified^{6,22}, (Fig. 1a). Surprisingly, the relevance of these extensions to ECO-blood compatibility remains unknown, and no enzymes targeting these extensions are available. Epithelial enterocyte surfaces²³ and *O*-glycans on mucins, the main scaffolds of the mucosal gut barrier, are decorated with ABO epitopes. We focused on the dedicated mucolytic human gut symbiont *Akkermansia muciniphila*²⁴ that possesses enzymes to defucosylate mucin *O*-glycans including the H type 3 antigen, also displayed as an extended A structure on RBCs²⁵.

Here we report the discovery of remarkably high-efficacy enzymes, not only against A and B antigens, but also against their extensions on RBCs. Extensive crossmatch analyses established the antigenicity of extended A/B-extensions and their conversion led to significant compatibility improvements of ECO-RBCs. Structural analyses allowed us to dissect the signatures associated with high catalytic efficacy on RBC surface antigens. Our work establishes the conversion of extended A and B antigens as a previously unexplored solution to ECO-blood incompatibility and highlights the potential of gut microbiota specialists for the discovery of high-efficacy human glycoconjugate-active enzymes.

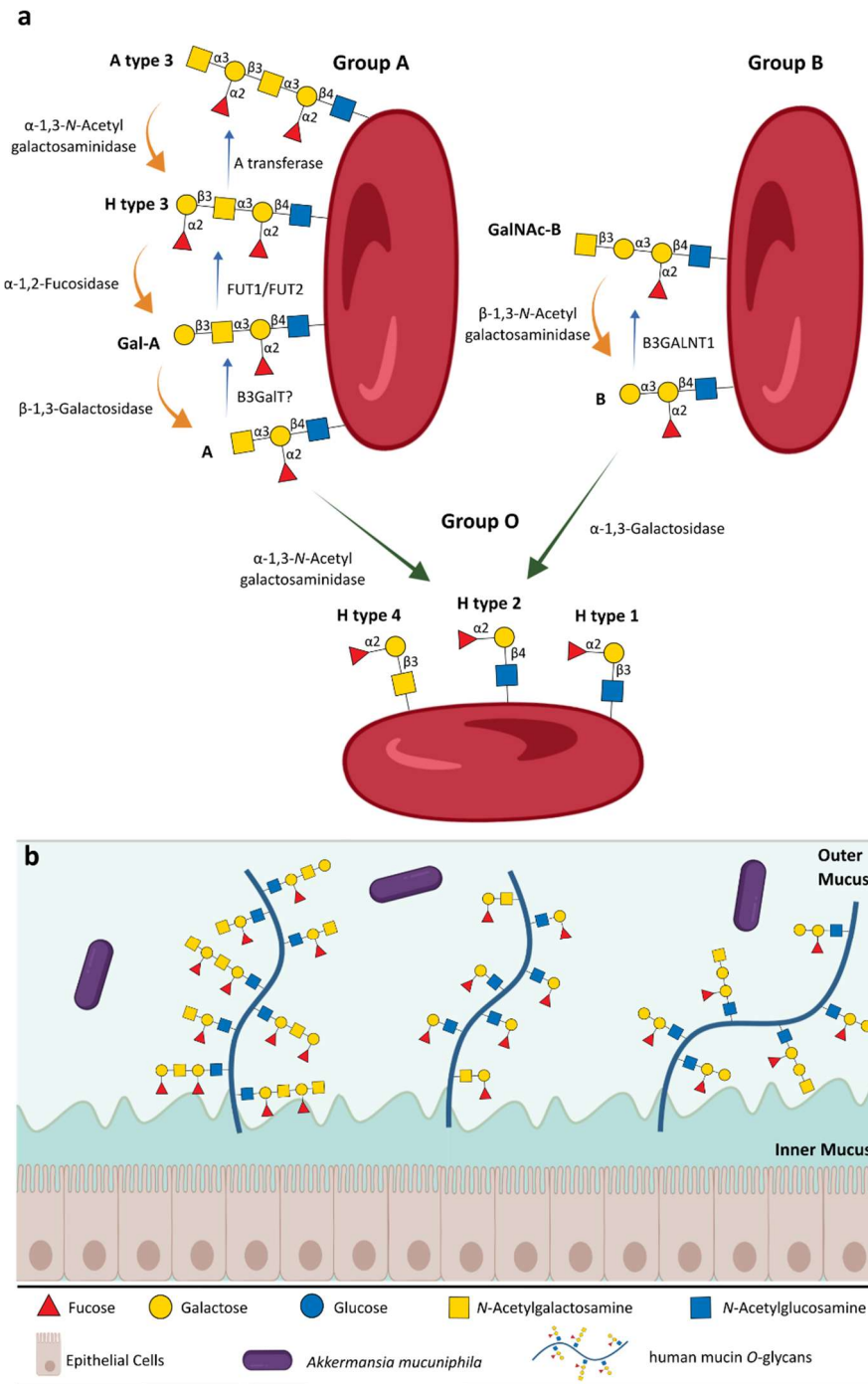


Figure 1: The ABO blood group antigens on red blood cells and mucins. **a** Cartoon showing the quantitatively predominant blood group A, B, and H type 2 antigens on red blood cells as well as the H type 1 and 4 chains. The three extensions of the A antigen (Gal-A, H type 3, and A type 3) and a single reported extension of the B antigen (GalNAc-B or ExtB) are also depicted. The previously known glycosyltransferases, which assemble these antigens from H antigens are indicated with blue arrows and the corresponding exoglycosidase activities that cleave these antigens are indicated with orange arrows. **b** The ABO antigens are present on mucin O-glycoconjugates, according to the blood group of the individual. The mucolytic specialist *Akkermansia muciniphila* colonizes the outer mucus layer and is expected to possess the enzymatic machinery to degrade mucin the ABO epitopes that are displayed on mucin O-glycans.

Results

Conversion of the B and extended GalNAc-B antigens

We started by selecting candidates for removing the B antigen. *A. muciniphila* possesses two orthologues, that populate hitherto undescribed clusters in the phylogenetic tree of GH110 (Supplementary Fig. 1), and which share <30% identity with the characterized B antigen-cleaving FragA⁸ (Supplementary Fig. 2a,b). Amuc_0480 (henceforth *AmGH110A*) has an N-terminal carbohydrate-binding module that is distantly related to A/B-antigen-binding CBM51 members²⁶. Notably, the CBM51-GH110 combination has not been reported to date. Amuc_1187 (*AmGH27*) was also selected based on the presence of A/B antigen-converting enzymes in GH27^{7,27} (Supplementary Fig. 2c,d), similarly to Amuc_0216 (*AmGH36A*), Amuc_0517 (*AmGH36B*), and Amuc_0855 (*AmGH36C*) (Supplementary Fig. 2e,f). These candidates (Supplementary Tables 1 and 2), were produced and α -galactosidase activity was established for *AmGH110A*, *AmGH27* and *AmGH36C* in contrast to *AmGH110B* and *AmGH36B*, whereas *AmGH36A* displayed only α -*N*-acetylgalactosaminidase (α -GalNAcase) activity (Supplementary Table 3). Assays against the B-antigen type 2 hexasaccharide (Supplementary Table 4) demonstrated that only *AmGH110A* was active (Supplementary Fig. 3a-f), displaying a specific activity=89.6 \pm 2.3 U mg⁻¹ (Supplementary Fig. 3g-h, Supplementary Table 5), which was >13-fold higher than the FragA benchmark⁸. Then, we aimed at identifying enzymes targeting the extended GalNAc-B motif (Fig. 1a). Strikingly, GH20 has the highest multiplicity in *A. muciniphila*, with 11 diverse putative β -*N*-acetylhexosaminidases (Supplementary Tables 1,2, Supplementary Figs. 4 and 5). These enzymes were produced, and all but one (*AmGH20K*) exhibited the desired β -*N*-acetylgalactosaminidase activity (Supplementary Table 3). However, the evaluation of the selectivity of these enzymes towards the GalNAc-B motif was not possible due to the unavailability of an oligosaccharide substrates. Therefore, we went on to analyse the removal of the B and GalNAc-B motifs on group B RBCs (Fig. 2a,b).

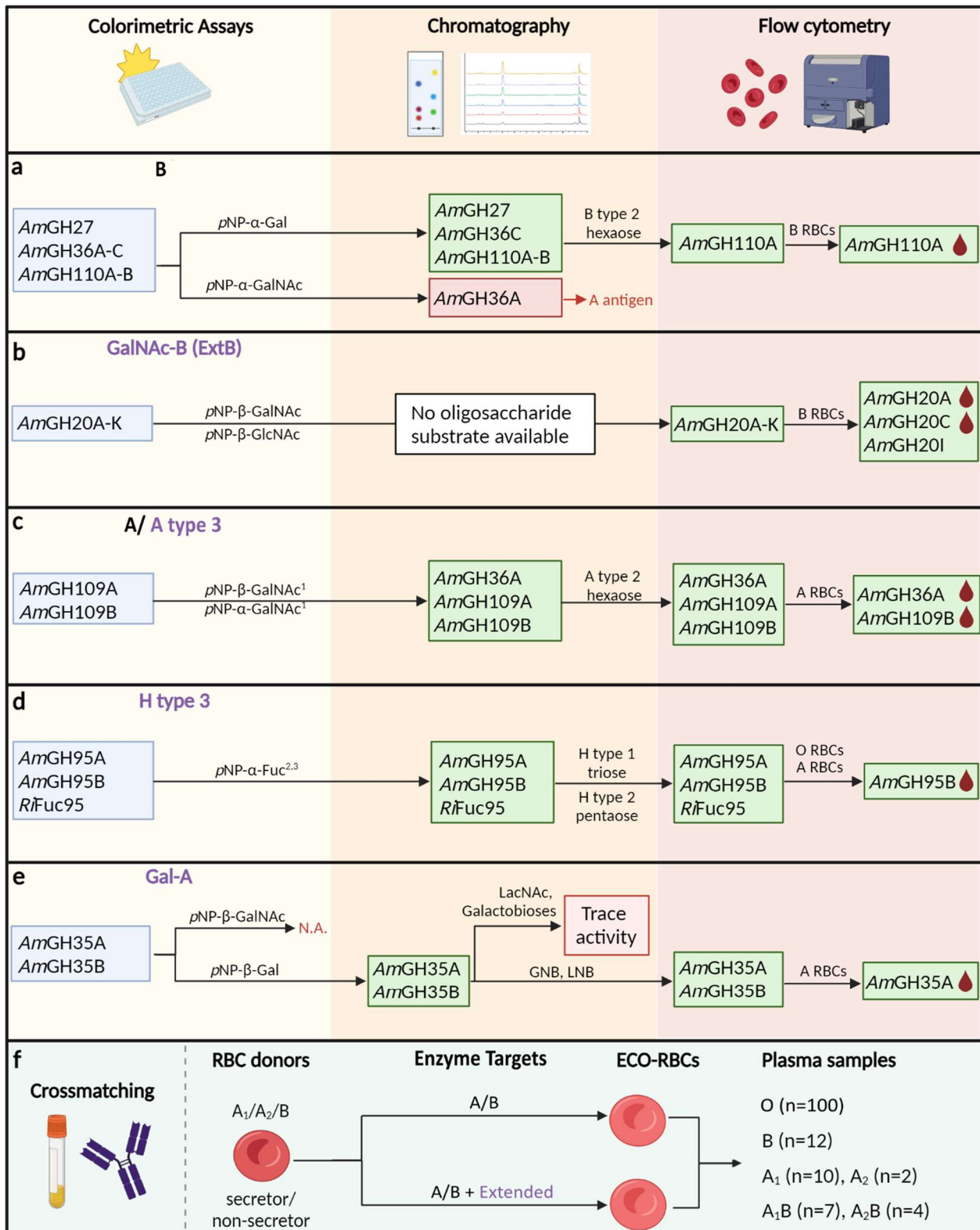


Figure legend on next page

Figure 2: Overview of the enzyme selection strategy leading to development of enzyme blend for conversion of A and B group donor RBCs. Overview of the discovery and selection of exoglycosidases for the conversion of: **a** B-antigen, **b** GalNAc-B antigen (Ext-B), **c** A and A type 3 antigens, **d** Gal-A antigen, and **e** H type 3 antigen. An initial screening assay against aryl model substrates, guided the selection of enzymes, which were further assayed against model free saccharides using thin layer chromatography. The normalised rates of the best performing enzymes were determined using high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and these enzymes were then evaluated for their ability to convert target antigens on RBCs using flow cytometry. The blood drop symbol denotes the top-performers, which were selected for crossmatch reactivity analyses as a proxy for their transfusion compatibility. **f** Crossmatch reactivity analyses were performed by enzymatic conversion of RBCs from A₁, A₂ and B donors. Single enzyme treatments targeting A or B antigens as well as sequential and one pot treatments with enzymes cocktails targeting both A/B antigens and their respective extended forms were performed. The converted RBCs were crossmatched with 100 group O plasmas per treatment type (single enzyme/enzyme cocktail), blood group (A/B), and secretor status (secretor/ non-secretor) which amounts to 800 crossmatch assays, besides a lower number of crossmatch assay with plasmas from other ABO groups.

Both the B and extended GalNAc-B antigens on RBCs are efficiently removed by *A. muciniphila* enzymes

We evaluated the conversion of B antigens on RBCs by incubating 1 μ M of *AmGH110A*, *AmGH36A*, *AmGH36C* or *AmGH27* with RBCs from three randomly-selected donors (30min, RT) in conversion buffer (200mM glycine, 3mM NaCl, pH6.8)⁸. Then, we analysed B antigen levels by immunostaining the RBCs with a monoclonal anti-B, followed by a Phycoerythrin (PE)-conjugated secondary antibody and measured PE-fluorescence using flow cytometry, which allows detection of low antigen amounts²⁸. Anti-B staining resulted in high signal from B RBCs, but not O RBCs negative controls (Supplementary Fig. 6a). Treatment with *AmGH110A*, but not other candidates, depleted B antigens on B RBCs to negative control levels (Supplementary Fig. 6b-f). The efficacy of *AmGH110A* in PBS was reduced by 50% relative to conversion buffer (Supplementary Fig. 6g-i). *AmGH110A* displayed an impressive efficacy, depleting B antigens on RBCs within 30min in conversion buffer (Supplementary Fig. 6j,k).

Next, we assayed the extended GalNAc-B epitope removal by the β -HexNAcases *AmGH20A-K*. To overcome the lack of a commercial monoclonal antibody, we used naturally-occurring anti-GalNAc antibodies, which were captured from archived reagent plasmas from GalNAc-B-deficient group A₁B and B individuals of P₁^k type. The signal from the natural antibodies was adequate to monitor the GalNAc-B conversion, albeit considerably lower than typical levels from optimized monoclonal antibodies. Only *AmGH20A*, *AmGH20C* and *AmGH20I* removed GalNAc-B in the screening (1 μ M enzyme, 30min) (Supplementary Fig. 7). At lower enzyme dosing, *AmGH20C* displayed highest efficacy, followed by *AmGH20A*, justifying further analyses using these enzymes (Fig. 2b, Supplementary Fig. 8a-e). Both *AmGH20A* and *AmGH20C* failed to remove the GalNAc-B antigen on RBCs in PBS buffer (Supplementary Fig. 8f,g).

We investigated how GalNAc-B antigen influences the removal of B-antigens. Treatment with *AmGH110A* depleted B antigens, but was hindered by GalNAc capping as expected (Fig. 1a, Fig. 3a-f). Hence, a sequential (Fig. 3g-l) or a one-pot treatment (Fig. 3m-q) with *AmGH110A* and *AmGH20A*, was required to deplete both B and GalNAc-B antigens. These findings indicate that removing the GalNAc-B antigen is crucial for the conversion of B RBCs antigens.

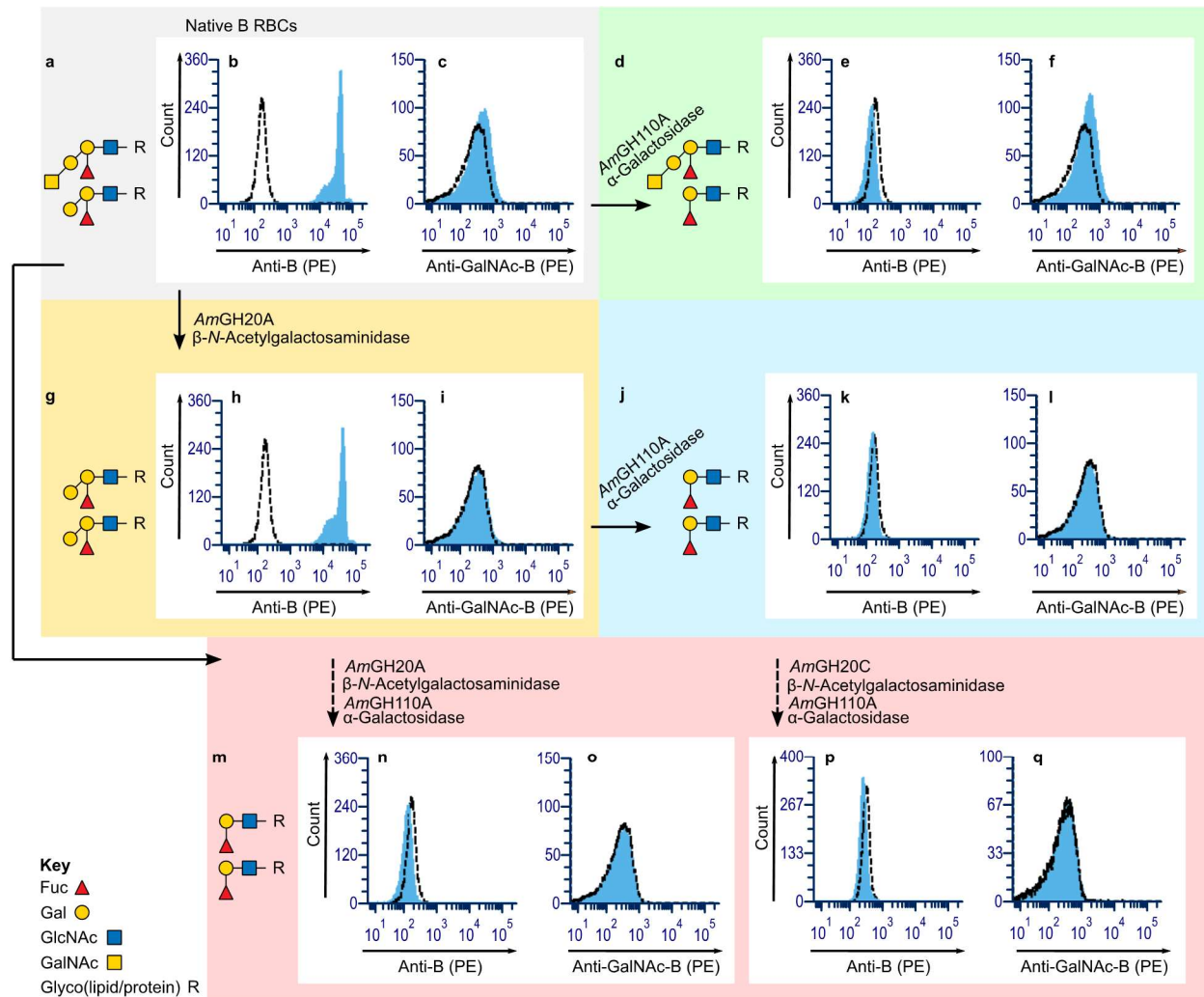


Figure 3: The removal of the extended GalNAc-B and B antigens on red blood cells by *A. muciniphila* enzymes. **a** Glycan structures of the B and GalNAc-B antigens, displayed on surfaces of group B RBCs. **b** and **c** show the levels of the B and GalNAc-B antigens on the RBCs in **a**, respectively. **d-f** The same as **a-c**, post treatment with the α -galactosidase *AmGH110A* ($0.5\mu\text{M}$). **g-i** Same as **a-c** post treatment with the β -*N*-acetylgalactosaminidase *AmGH20A* ($0.2\mu\text{M}$) and **j-l** Same as **j-l** post treatment with the α -galactosidase (*AmGH110A*). **m-q** The same as **a-f** post treatment with a one-pot blend of *AmGH110A* ($0.5\mu\text{M}$) and *AmGH20A* ($0.2\mu\text{M}$). Blue histograms represent group B RBCs stained with anti-B (clone 9621A8) and anti-GalNAc-B from human plasma, respectively. Dotted line histograms are representative native group O RBC negative controls. The analysis was performed using Phycoerythrin (PE)-conjugated secondary antibodies followed by flow cytometry. Only the H type 2 chain is depicted for simplicity. The depletion of antigens was performed successfully on RBCs from 25, 26 and 18 donors, respectively, in 15, 11 and 9 experiments each (by flow cytometry) using 1-2 batches of *AmGH110A*, *AmGH20A* and *AmGH20C*, with similar results as detailed in Supplementary Information 2, Tables 1-3.

Conversion of A and extended A antigens

We sought to identify α -GalNAcases to convert A antigens and extended A type 3 antigens (Fig. 1a). Two previously reported²⁹ distant GH109 orthologues (*AmGH109A* and *AmGH109B*) of NagA⁸ (Supplementary Fig. 9a,b, Supplementary Fig. 10) were selected. Moreover, *AmGH36A* was selected, despite the poor efficacy of reported GH36 α -GalNAcases against A RBCs⁹ (Supplementary Fig. 11). Thin-layer chromatography (TLC) screening towards the A type 2 hexaose showed that *AmGH109A* was the poorest, followed by *AmGH109B*, whereas *AmGH36A* was superior (Supplementary Fig. 12a-c) and displayed a normalized rate $V_0/E=127\pm 4.1\text{ s}^{-1}$ on the same substrate (Supplementary Fig. 12d-e).

The next targets were H type 3 and Gal-A epitopes (Fig. 1a). Recently, we showed that the *A. muciniphila* α -1,2-fucosidase *AmGH95A* exclusively converted H type 2 backbones, while its orthologue *AmGH95B* was active towards H type 1, 2 and 3 on mucin²⁵. The latter enzyme was also shown to convert H-antigens on group O RBC to the Bombay type³⁰. We also considered an intracellular α -1,2-fucosidase (*RiFuc95*), which efficiently de-fucosylated H type 1 trisaccharide³¹. *AmGH95A* failed to convert 2mM H type 1 pentaose within 2h, whereas complete conversion was achieved within 20 min and 5 min by *RiFuc95* and *AmGH95B*, respectively at 1 μ M of each enzyme (Supplementary Fig. 13a-c). Only *AmGH95B* almost indiscriminate against H type 2 triose and H type 1 pentaose (Supplementary Fig. 13d-i). For the Gal-A antigen (Fig. 1a), we considered two putative GH35 β -1,3-galactosidases: *AmGH35A* and *AmGH35B* both containing a galactose binding-like domain, but *AmGH35B* also harbours a CBM32 (Supplementary Fig. 9c,d; Supplementary Fig. 14). We produced both enzymes (Supplementary Tables 1 and 6) and showed β -galactosidase activity against *para*-nitrophenyl- β -galactopyranoside (Supplementary Table 3). Both enzymes were β -(1,3)-galactosidases, with preference for β -Gal-1,3-[GlcNAc/GalNAc] (Supplementary Fig. 15), which justified testing on group A RBC s.

***A. muciniphila* enzymes efficiently remove the A antigen and its extended forms on group A RBC**

The top performing A enzymes *AmGH109A/B* and *AmGH36A* (Fig. 2b) were tested on A₁ and A₂ RBCs from 3 donors of each subtype at 1 μ M of each enzyme, RT, 30min in conversion buffer. Lowest activity was observed for *AmGH109A*, followed by *AmGH109B*, while *AmGH36A* was superior (Supplementary Fig. 16a-j). The two top-performing enzymes exhibited better efficacy in conversion buffer than PBS (Supplementary Fig. 16k-o). Activity assays at different incubations times and concentrations established that incubation with 0.5 μ M *AmGH36A* for 30min was sufficient for full-conversion of A antigens from 3 donors (Supplementary Fig. 16p,q). Interestingly, only 10nM *AmGH36A* converted A type 3 antigens on RBCs (Supplementary Fig. 16r).

To convert H type 3 antigens, we evaluated *RiFuc95*³¹ and *AmGH95B*²⁵. We chose A₁ secretor RBCs, expected to have H type 1-4 antigens as opposed to nonsecretors, who virtually lack H type 1 and 3 antigens³². To maximize detection signal, a pre-treatment with *AmGH36A* was performed (Supplementary Fig. 17a-h) to expose H antigens including H type 3 (Supplementary Fig. 17f,g). While *RiGH95* was inactive (Supplementary Fig. 17i-k), *AmGH95B*, displayed extraordinary efficacy on H type 3, judging by a large increase in the underlying Gal-A antigen (Supplementary Fig. 17l-n). The efficacy of these fucosidases was compared on RBCs from a nonsecretor and a secretor (Supplementary Fig. 17o-q), which demonstrated high efficacy of *AmGH95B* against both phenotypes. Fucosidase activity was compared in PBS and conversion buffer, including the H type 2 specific *AmGH95A*²⁵. Interestingly, *AmGH95B* was superior in both buffers, *RiGH95* preferred PBS, and activity *AmGH95A* was not observed (Supplementary Fig. 18a-i). We assayed the removal of H antigens, in presence of *AmGH36A* to maximize the exposure of said antigens. *AmGH95B* displayed exceptional efficacy on A₁ RBCs at 2nM enzyme and 30min incubation (Supplementary Fig. 18j).

Next, we evaluated the β -1,3-galactosidases against Gal-A, showing that *AmGH35A*, but not *AmGH35B*, removed this antigen on A₁ RBCs (Supplementary Fig. 19a-d). Exposing the Gal-A structure with an *AmGH36A* and *AmGH95B* co-treatment, revealed a more compelling activity by *AmGH35A* (Supplementary Fig. 19e). The sequential treatment of A₁ RBCs with *AmGH36A*, *AmGH95B*, *AmGH35A* resulted in the exposure of additional A epitopes (Fig. 4a-u), which required a second *AmGH36A* treatment to generate A-negative RBCs (Fig. 4v,w). This was also observed when *AmGH36A* was replaced with *AmGH109B* in the α -GalNAcase treatment steps (Supplementary Fig. S20).

The additional incubation and washing steps associated with the sequential approach were less ideal for feasibility. Therefore, we implemented a one-pot reaction. Indeed, successful removal of the targeted antigens after 30min with both A₁ and A₂ RBCs was shown regardless of secretor status (Fig. 2x-ac, Supplementary Fig. 21). The conversion of A, H and Gal-A antigens by *AmGH36A*, *AmGH95B* and *AmGH35A*, respectively, could be reproduced on RBCs from 41, 33 and 31 donors, respectively, in 19, 12 and 9 independent experiments for each antigen and using 2-6 batches of enzyme, as detailed in (Supplementary Information 2, Tables 4-6).

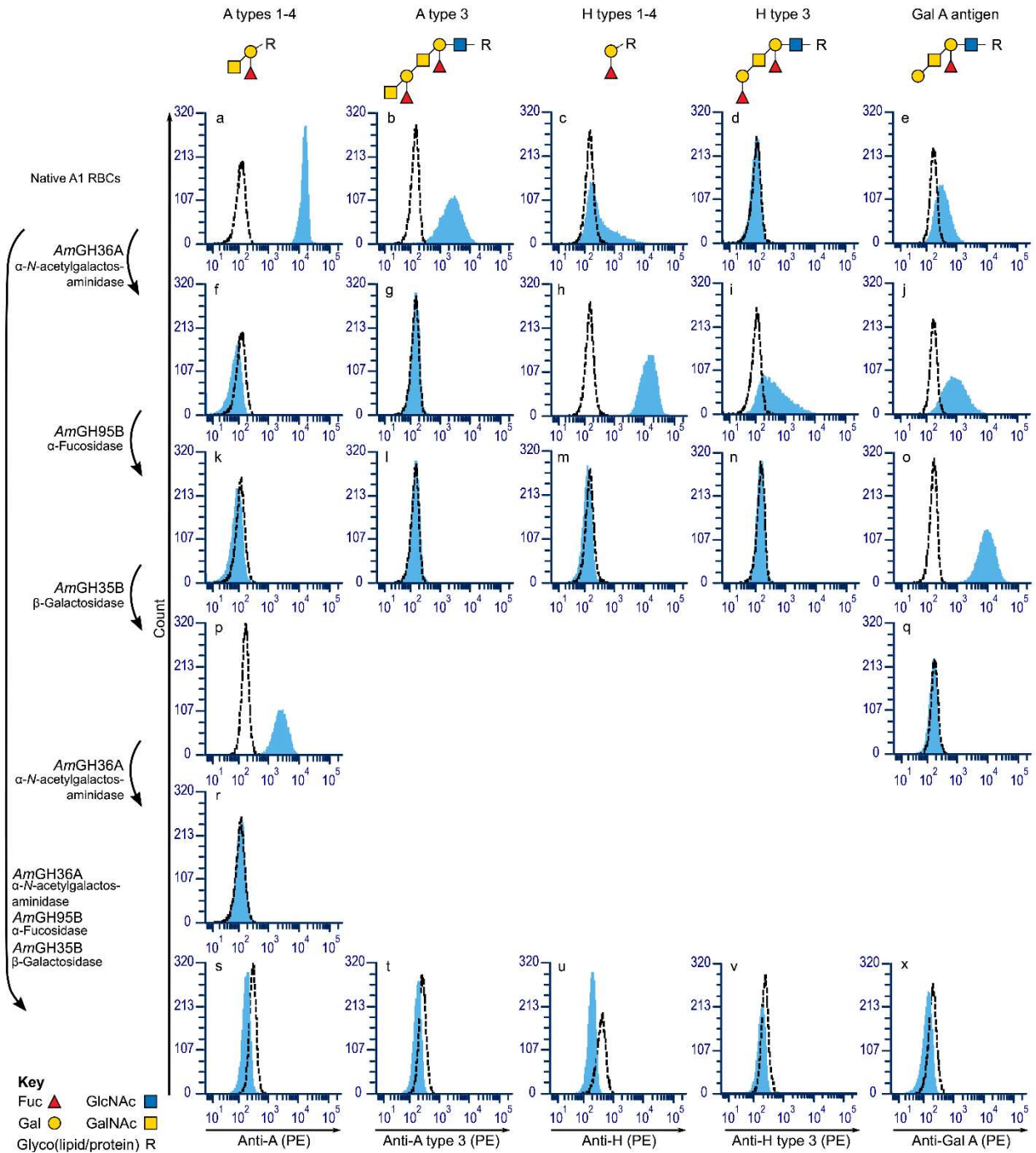


Figure 4: Enzymatic conversion of the A-core extended epitopes on RBCs. Schematic representation of the sequential and one-pot enzymatic conversion of the A type 2 epitope and its reported extensions on RBCs as monitored by immunostaining followed by flow cytometry. **a-f**, Representative histograms of the native A-core antigens on A₁ phenotype RBCs, without enzymatic treatment or after sequential treatment with *AmGH36A* (**g-l**), *AmGH95B* (**m-r**), *AmGH35B* (**s-u**), *AmGH36A* (**v-w**), and (**x-ac**) a one-pot treatment with all three enzymes for 30min in conversion buffer. Solid blue histograms depict untreated and enzyme treated A₁ RBCs, while dotted line histograms represent negative controls of each staining (O RBCs for anti-A, Anti-A type 3, Anti-H type 3 and Anti-Gal A, as well as Bombay RBCs for anti-H). The primary monoclonal antibodies used in this analysis were: Anti-A (clone

ES-15), for total which A antigen staining^{8,28-33}; Anti-H (clone BRIC231) for total H straining; TH-1, HH14 and 3C9 for detection of A type 3¹⁹, H type 3^{20,21} and Gal A³⁴ extensions, respectively.

Removing extended A and B antigens significantly improves compatibility with group O donor plasmas as compared to only removing A or B antigens

We compared the conventional removal of A and B antigens to the unprecedented targeting of extended antigens. A haemagglutination crossmatching assay used in clinical practice to avoid incompatible transfusions was employed and incubations were performed in gel card wells containing rabbit anti-human IgG gel to capture agglutinates caused by both IgG and IgM antibodies at 37°C, mimicking physiological antibody-binding activities. Group B RBCs from a secretor and a nonsecretor, were incubated with *AmGH110A* (0.5µM) or a blend of *AmGH110A* and either *AmGH20A* or *AmGH20C* (both at 0.4µM). The treated RBCs were crossmatched with 100 randomly selected group O plasmas and haemagglutination was read as negative (-), positive (from 1+ to 4+), or mixed field (m). Strong reactions of RBCs were observed with 10 of the O plasmas, whereas treatment of RBCs with *AmGH110A* markedly reduced reactivity (Fig. 5a), consistent with previous studies¹¹⁻¹³. The treatment with *AmGH110A* and *AmGH20A*, but not *AmGH20C*, led to a significant further reduction in reactivity, than only the *AmGH110A*-treatment. Generally, treatment with *AmGH110A* and *AmGH20A* increased the number of both non-reactive and less-reactive RBCs, compared to *AmGH110A* alone. A few exceptions showed similar, or rarely, increased crossmatch reactivities, and more so for *AmGH20C* treatments (Fig. 5b,c, Supplementary Fig. 22). In addition, treated B RBCs were also crossmatched with smaller cohorts of A and AB plasmas (Supplementary Fig. 23). The AB plasmas (n=11) were negative with both native and treated RBCs. The majority of A plasmas reacted strongly (3+ to 4+) with untreated B RBCs. However, all but one, became negative after the treatments. Interestingly, the exception was a weak (1+) reaction to a single *AmGH110A*-treated secretor B RBC sample. In summary, removing both the B and extended GalNAc antigens from two donors, resulted in a significant decrease in the median reactivity with O plasmas, compared to the removal of only the B antigen. The number of crossmatch-negative tests increased for secretors (from 79% to 91%) and the nonsecretors (from 80% to 96%), which establishes the importance of the extended GalNAc-B antigen for crossmatch reactivity, as a proxy for transfusion compatibility of ECO B-RBCs.

Next, we pursued a similar approach to investigate the removal of the A antigen and its extensions. First, A₁ RBCs from the sequential treatments (Fig. 4a-w) were crossmatched with two randomly selected O plasmas, in addition to a third known to be highly reactive with A/B antigens. All three plasmas reacted strongly (4+) with native A₁ RBCs, which was dependent on the individual plasma (Supplementary Fig. 24). Encouragingly, the reactivity of the highly-reactive plasma decreased (2+), while the remaining two became negative. Next we performed a larger crossmatch analysis with A₁ RBCs from a secretor and a nonsecretor, which were treated to remove either the A antigen or the latter and its extended structures. The resulting RBCs (Fig. 4g-l,x-ac; Supplementary Fig. 25) were crossmatched against group O plasmas (n=100). Despite the high efficacy of *AmGH36A*, the treatment with this enzyme decreased the crossmatch reactivity to only about 2+, and only 17-20 plasmas lost reactivity (Fig. 5d-f). The treatments of RBCs with *AmGH36A*, *AmGH95B* and *AmGH35B* further reduced the mean reactivity significantly to about 1+ (Fig. 5d). The triple enzyme-treated RBCs became negative for 48 and 53 of the secretor and the nonsecretor group O plasmas, respectively. These unprecedented findings showcase that removing the A antigen is not sufficient to generate universally compatible RBCs, likely due to the antigenicity of extended A antigens.

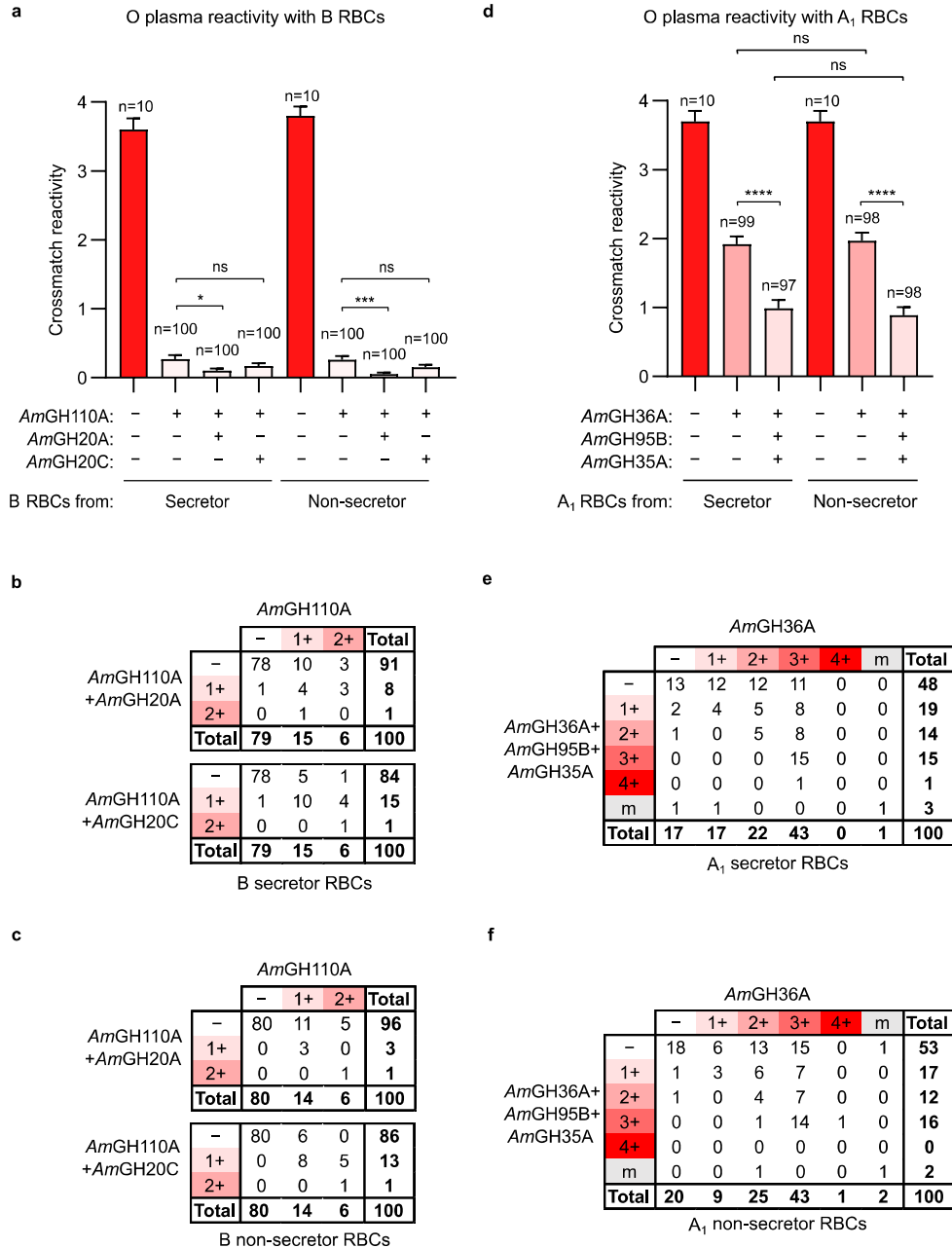


Figure 5: Removal of the extended A and B antigens significantly improves the compatibility of RBCs with group O plasmas as compared to removal of only the A and B antigens. a O plasmas mean reactivity with B RBCs from one secretor and one nonsecretor. Untreated RBCs crossmatched with a random subset of O-plasma to illustrate baseline reactivity of untreated cells, after AmGH110A (0.5µM) treatment, combined AmGH110A (0.5µM) and AmGH20A (0.4µM) or AmGH110A (0.5µM) and AmGH20C (0.4µM) treatment for 30 min in conversion buffer. **b,c** Distribution of O plasma (n=100) reactivities and comparison between the different treatments, with the B secretor RBCs and B nonsecretor RBCs, respectively. **d** O plasma mean reactivity with A₁ RBCs from one secretor and one nonsecretor with native RBCs, after AmGH36A (1 µM) treatment, or combined AmGH36A (1µM), AmGH95B (20nM) and AmGH35A (0.2µM) treatment for 30min in conversion buffer. **e,f** Distribution of the reactivities and comparison between different treatments, with A₁ secretor RBCs and A₁ nonsecretor RBCs, respectively. Reactions where reactivity level could not be determined due to mixed field are excluded from panel d. Error bars are the SEM, ****P<0.0001, ***P<0.001, *P<0.05, ns= not significant (Mann-Whitney U test).

The B-antigen enzyme *AmGH110A* reveals atypical structural signatures that correlate to high efficacy on RBCs

To date, the only structurally characterised enzyme from GH110 is *PdGH110B*³⁵ (PDB: 7JW4), implicated in breakdown of sulphated α -galactan (λ -carrageenan). We determined the structure of *AmGH110A* in complex with galactose (Fig. 6a, Supplementary Table 7). Despite the shared fold (Supplementary Table 8) with *PdGH110B*³⁵, *AmGH110A* possesses an additional N-terminal CBM51 and a less bulky β -barrel 1 than *PdGH110B* (Supplementary Figs. 26a and 27). The -1 subsite is largely conserved between *AmGH110A* and *PdGH110B* (Supplementary Fig. 26b), but the +1 subsite is markedly different (Fig. 6b,c). The active-site side facing the Gal-O2 at subsite +1 in *AmGH110A* is spacious, consistent with the accommodation of the fucosyl of the H antigen. A tyrosine (Y433) and an arginine (R544), both conserved in the phylogenetic cluster of *AmGH110A* (Supplementary Fig. 2), are candidates for recognising the fucosyl of the H antigen (Fig. 6b). By contrast, this space in *PdGH110B* is blocked by a lysine (K207) (Fig. 6c), which likely recognises the Gal-2-sulphatyl group at this position³⁵.

The CBM51 putative binding-site and the *AmGH110A* active-site are aligned (Supplementary Fig. 26a). Notably, shared ligand-binding residues with the closest structural orthologue, the *O*-glycopeptidase ZmpB-CBM51 from *Clostridium perfringens*³⁶ (Fig. 6d-e, Supplementary Fig. 28, Supplementary Table 9), suggests that the CBM51 in both enzymes may promote substrate binding. Besides the atypical presence of a CBM51 in *AmGH110*, the surface topology and charge, represent potentially important adaptations to promote catalytic efficiency on RBC densely-sialylated glycans. Thus, the active-site of *AmGH110A* is flanked by a flat and positively charged surface (Fig. 6f), consistent with the highest *pI* value (8.87) (Supplementary Table 9) amongst known B-antigen active enzymes.

The Gal-B converting enzyme *AmGH20A* harbours a previously unknown CBM

We determined the structure of *AmGH20A* in ligand-free form, lactose-bound, and GalNAc-bound (Supplementary Table 8). Similarly to *AmGH110A*, The active-site of *AmGH20A* is flanked by a flat and positively charged surface (Supplementary Fig. 29a). The previously reported structure of *AmGH20C*³⁷ (PDB: 7CBO), the closest structural orthologue to *AmGH20A* (Supplementary Table 12), shares a similar positively charged ring flanking the active-site (Supplementary Fig. 29b), but differs by the elongation of two active-site loops (Supplementary Fig. 30a). The recognition of the GalNAc bound in the active is conserved between these GH20 enzymes (Supplementary figure 30b). Strikingly, a second GalNAc unit was bound in a shallow pocket in the *AmGH20A* C-terminal β -sandwich domain (100 residues). The observed GalNAc-binding, warrants the assignment of this domain (henceforth CBM) to be a carbohydrate-binding module (Figure 6k). The CBM binding-site features a solvent accessible tyrosine (Y590), two histidine residues (H595 and H623) and an arginine (R619). The *N*-acetyl of the GalNAc is coordinated by two alanine mainchains (A594 and A622), the O3 is coordinated by the mainchains of another alanine pair (A596 and A651), while the axial O4 is recognized by R619 (Fig. 5h). This domain is structurally unique (Dali score=7.9), sharing only 8% sequence identity to the top structural orthologue (Supplementary Table 12). We constructed a phylogenetic tree from 253 homologues (20.8-94.3% sequence identity), mainly from the Bacterioidota and Verrucomicrobiota phylum that harbours *A. muciniphila*, but included sequences from Plantomycetota, Armatimonadota, and Lentisphaerota (Supplementary Fig. 31a). Interestingly, the GalNAc binding residues from 3 loops showed high conservation, especially in the *Akkermansia*-dominated cluster harbouring *AmGH20A*-CBM (Supplementary Fig. 31b).

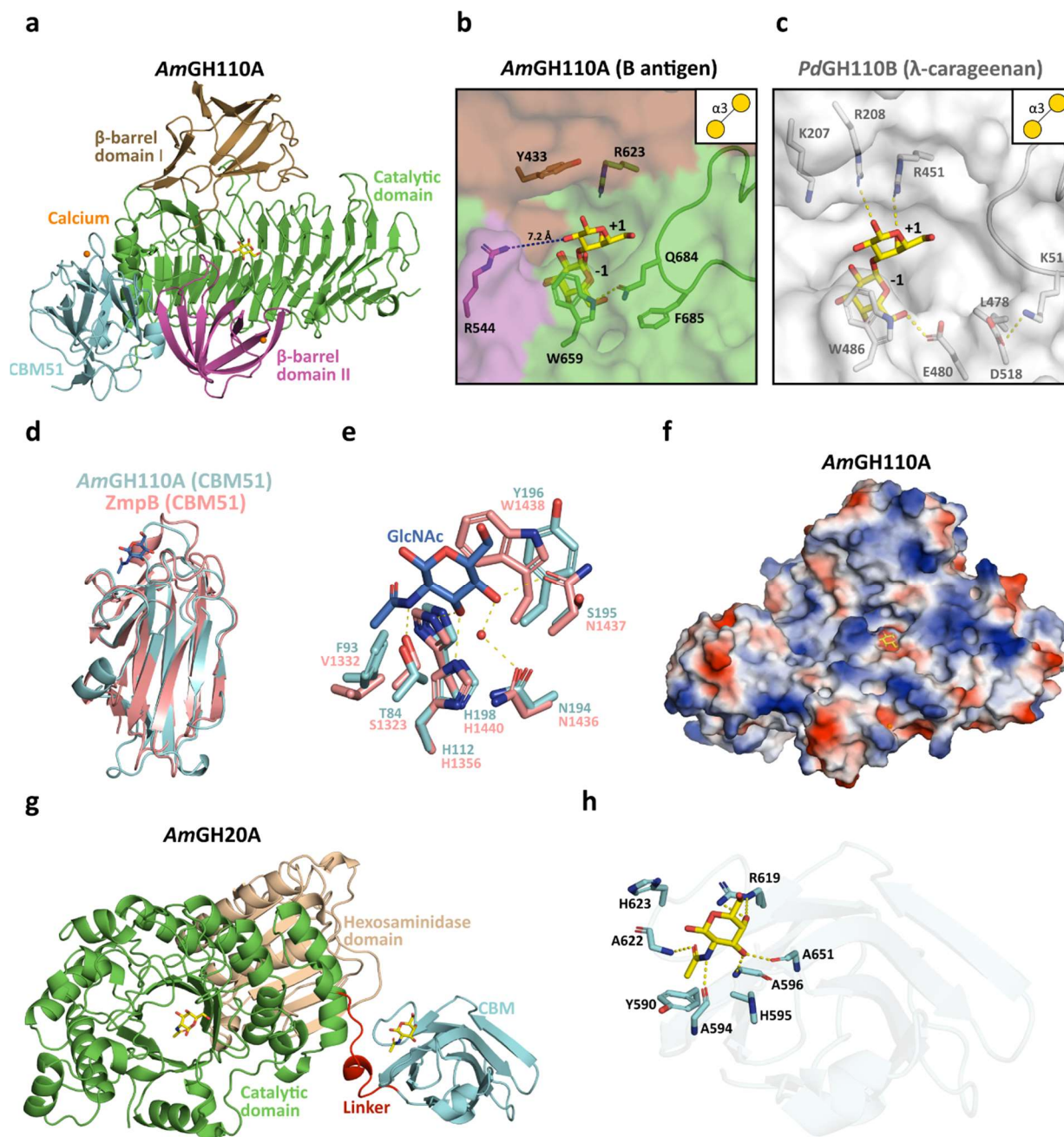


Figure 6: Structural signatures of the B-antigen converting enzymes. **a** Overall structure of *AmGH110A* in complex with α -galactose. *AmGH110A* comprises an N-terminal CBM51 (aa 64-212), a catalytic β -helix domain (aa 213-795) flanked by β -barrel 1 (aa 330-435) and β -barrel 2 (aa 525-608) domains. Bound Ca^{2+} ions are depicted as orange spheres. **b,c** Semi-transparent surface of *AmGH110A* and the λ -Carrageenan-specific *PdGH110B*, respectively. The α -galactobiose bound to *PdGH110B* is shown in both panels to highlight the -1 and +1 subsites. Polar interactions are denoted with yellow dotted lines. **d** Overlay of the CBM51 domain of the *O*-glycopeptidase *ZmpB* from *Clostridium perfringens* (PDB: 7JRL) (red) with a bound *N*-acetylglucosamine (blue) onto the N-terminal CBM51 domain of *AmGH110A* (cyan). **e** Comparison of the CBM51 binding sites between *ZmpB* and *AmGH110*. The red sphere is H_2O . **f** The surface electrostatic potential of *AmGH110A*. Positive potential is in blue, negative potential is in red, and apolar surface is in grey. The surface coating the substrate stacking residues (W659 in *AmGH110A*), has been made transparent to show the galactose at subsite -1. **g** Overall structure of *AmGH20A* in complex with β -*N*-acetylgalactosamine. *AmGH20A* comprises a GH20 N-terminal hexosaminidase domain (aa 18-164), the catalytic

(β/α)₈ barrel domain (aa 165-543), and a C-terminal CBM (aa 555-665), connected to the catalytic domain by a short linker (red, 544-554). **h** The β -*N*-acetylgalactosamine binding site of the CBM-like domain. Sidechains are not shown for mainchain interacting residues.

Discussion

Forty years after pioneering the ABO-universal ECO-blood concept⁷, clinical translation remains elusive, partly due to safety concerns related to cryptic incompatibility issues with enzymatically-converted RBCs^{7,11-15}. Here, we report the discovery of enzymes, which convert extended antigens for evaluation of their antigenicity. Our focus on mucolytic microbiota symbionts aimed at harnessing the co-evolutionary adaptation towards ABO antigens on intestinal mucins. *A. muciniphila* caught our attention due to its exclusive utilisation of mucin as a carbon and nitrogen source, in contrast to other mucolytic symbionts. Our strategy proved fruitful in the discovery of all required enzymatic specificities and remarkably-high efficacies towards all ABO-related epitopes, including extended structures on RBCs. The best performing *A. muciniphila* enzymes were highly basic ($pI \approx 9$), unlike all previously discovered RBC-active enzymes (Supplementary Table 6, Supplementary Table 10). Specifically, positively charged rings around active-sites appeared as a hallmark of adaptation to sulphated and sialylated mucin *O*-glycans³⁸, which also applies to abundant sialyl-glycoconjugates on RBCs³⁹. These findings may guide rational engineering of more potent enzymes variants to improve the safety and feasibility of ECO-RBCs. The presence of ancillary CBMs in *AmGH110A* and *AmGH20A* represents another adaptation to promote binding to mucin/cell-surface glycoconjugates^{25,40,41}. Conversely, the lack of CBMs does not exclude high efficacy on RBCs, as indicated by the exceptional activity of *AmGH95B*.

Our conversion of RBCs was uniquely performed at the highest concentration of RBCs (38% haematocrit), shortest assay time (30min) and lowest temperature (RT) as compared to previous work (Supplementary Table 14). These mild conditions without additives (*e.g.* dextran) together with the excellent enzyme efficacies are important feasibility parameters in clinical applications. Enzyme yields were 15-35 mg L⁻¹ after 2 purification steps. Based on our data, about 18 and 8 mgs of the limiting enzyme would be needed for conversion of 1 unit (200mL) of A and B RBCs, respectively, based on the enzyme concentration used in the ECO-RBCs used for the crossmatching experiments. Generally poorer enzymatic efficacies in PBS buffer, as compared to conversion buffer, may stem from lower affinities due to electrostatic screening in PBS buffer. This is an important advantage, which facilitates enzyme removal post-treatment by RBC-friendly PBS washing cycles to avoid adverse effects caused by enzyme carry-over.

To date, it has been assumed that enzyme efficacy was the crucial milestone for clinical application of ECO-RBCs. However, unexplained incompatibility issues of early clinical studies with group B ECO-RBCs, despite apparent removal of canonical B antigen, were indicative of additional underlying factors¹¹. Strikingly, another study showed that 18% of group B plasmas and 55% of group A plasmas had positive crossmatch reactions with group A₁B RBCs treated with a blend of A- and B-enzymes¹⁴, further underscoring incompatibility issues, despite depletion of canonical A/B antigens. Our work establishes a significant contribution of the extended antigens, to the crossmatch reactivity of ECO-RBCs. Thus, removing both B and extended B antigens reduced incompatible O plasmas to <9%, and importantly, the remaining positive plasmas had reduced reactivity. Similarly, A RBCs, where both A and extended A antigens were depleted, displayed increased compatibility and decreased positive crossmatch severity as compared to removal of only the A antigen. The large compatibility variations with recipient plasma highlights considerable, albeit to date underappreciated, inter-human diversity in specificities and titers of antibodies against different A-core antigens. Persistent incompatibility with distinct plasmas, post-

conversion of described A antigens requires additional work. A plausible hypothesis is that enzyme promiscuity may cause the exposure of cryptic non-ABO antigens. An O plasma and an A₂B plasma reacted mildly with group O RBCs treated with AmGH95B-AmGH35B (Supplementary Fig. 24). Additionally, a single AB plasma out of eleven was reactive with AmGH36A, while two other were mildly reactive with the one pot triple enzyme treated RBCs (Supplementary Fig. 32, Supplementary Discussion). These observations may indicate exposure of neo-antigens by the enzymes, which are recognized by a few plasmas. Alternatively, some of the crossmatches could be related to carryover of trace amounts of enzymes bound to RBCs, which theoretically could cause antibody binding. Lastly, previously-undescribed antigenic modifications of the A antigen may be resistant to our enzymes. Until the 1980s, the paradigm was that only the A antigen was specific to group A. Description of the H type 3, Gal-A and A type 3 structures by Clausen *et al.*¹⁹⁻²¹ contradicted that. Our study establishes the antigenicity of these extensions via the extensive crossmatch analysis comparison.

In conclusion, our work identified high-efficacy enzymes for conversion of all four described extensions of group A and B antigens, besides canonical A and B antigens, resulting in enzyme blends for one-pot conversion of group A and B-core antigens. This provided premiere and compelling evidence for the Antigenicity of the extended A and B antigens, thereby offering a missing link towards production of ABO-universal blood for transfusion and potentially transplantation organs^{42,43}. This work offers unprecedented insight into the extended A and B antigens as determinants of transfusion compatibility for the A and B RBC phenotypes and highlights mucolytic specialists of the gut microbiota as an important reservoir for discovery of glyco-conjugate converting enzymes.

Methods

Chemicals and reagents

Blood group A type 2 hexaose, B type 2 Hexaose, H type 1 Pentaose, and H type 2 triose were all purchased from Elicityl (Crolles, France). All other substrates and reagents were from Sigma unless otherwise stated. An overview of saccharide substrate structures and their purity is shown in Supplementary Table 4.

Cloning, mutagenesis and production of Recombinant enzymes

Gene fragments encoding enzymes lacking their signal peptides as predicted using SignalP v. 4.1⁴⁴ from *Akkermansia muciniphila* ATCC BAA-835 (DSM 22959) were amplified from genomic DNA using primers listed in Supplementary Table 1. The amplicons were inserted into the pET28a(+) vector (Novagen, Darmstadt, Germany) within the XhoI and NcoI sites using In Fusion Cloning (Takara, Shiga, Japan). The recombinant enzymes were designed as fusions with C-terminal His₆-tags and expressed in *E. coli* BL21 (DE3) (Novagen). The catalytically impaired AmGH20A mutant (E339A) construct was purchased from GenScript Biotech (Leiden, Netherlands). Cultures were grown to OD_{600} 0.6 before induction with IPTG to 0.1mM and continued growth at 18° C for 24h. Pellets were harvested by centrifugation (10.000g, 30min) and stored frozen at -20°C, until further use. Pellets were resuspended in binding buffer (20mM NaPO₄, 500mM NaCl, 10mM Imidazole, 10% glycerol, pH7.5) and lysed by one passage through a high-pressure homogenizer (Stansted 'Pressure Cell' Homogenizer SPCH Model FPG12800, Essex, UK) at 1000 bar, or ultrasound sonication (QSonica, Newtown, CT, USA) at 40% amplitude (4min; 20s on, 40s off). The cell lysates were treated with Benzonase (Novagen, Damstadt, Germany) for 1 hour at 4° C, and then centrifuged (30.000g, 30min, 4°C) to remove cell debris. Next, the recombinant enzymes were purified by His-affinity chromatography using 5 mL HisTrapTM columns (Cytiva, Uppsala, Sweden) installed on an Äkta Avant (Cytiva) using a standard protocol. The eluted fractions were pooled, concentrated using a using Amicon Ultra-15 centrifugal filter units (Merck, Burlington, MA, USA) with a molecular weight cutoff of 30kDa, or 10kDa depending on enzyme size and further purified by loading onto a size-exclusion chromatography Hiload 16/60 pg200 column (Cytiva) and eluted at 0.5mL min⁻¹. Pure elution fractions were collected and concentrated as above. Enzyme concentrations were determined using a Nanodrop (Thermo Fisher, Waltham, MA, USA) and the theoretically determined molar extinction coefficients $\epsilon_{280\text{ nm}}$ (Supplementary Table 2). The purity and size of the enzymes were assessed by SDS-PAGE analysis.

Enzyme Activity Assays

All enzymes were assayed in 20mM Sodium Phosphate, 150mM NaCl, 0.005% (v/v) Triton X-100, pH7.0, henceforth referred to as standard assay buffer. All reactions were performed in 3 independent replicates (n=3) unless otherwise stated. Specific assay conditions for all enzymes mentioned are shown in Supplementary Table 15.

Hydrolytic activity of enzymes was assayed against *para*-nitrophenol (*p*NP) conjugated substrates in 96-well microtitre plates using the standard assay buffer. The release of *p*-nitrophenol was monitored by the increase in A_{410} over time using a PowerWave XS microplate reader (BIO-TEK[®], Winooski, VT, USA). The initial reaction rates (V_0) were determined from the linear part of the A_{410} versus time and *p*NP standard curves (0.02-1mM). The reaction rates were normalized by dividing with the enzyme concentration ($\frac{V_0}{E}$) (s⁻¹). Enzyme activity on blood antigen oligosaccharides in standard assay buffer was initially screened using thin layer chromatography (TLC), and quantified using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). For analysis by HPAEC-PAD,

enzymatic reactions were quenched at various time intervals of 30s to 5min by adding 11 μ L of the reaction mixture with 99 μ L of 0.1 M NaOH. The quenched reactions were centrifuged (10.000g, 10min, 4°C) and transferred to Fisherbrand™ 9mm Short Thread Plastic Vials (0.3mL) (Thermo Fisher, Waltham, MA, USA) for analysis. The normalized rates as defined above were calculated by plotting the area underneath the product peaks (identified by comparison to standards) at different time intervals. Galactose (Gal), *N*-acetylgalactosamine (GalNAc) and Fucose (Fuc) standards (0.015-1mM) were measured and used to convert peak areas to concentrations and determine initial reactions rates (V_0), and normalized rates ($\frac{V_0}{E}$) (s^{-1}). The normalized reaction rate of *AmGH35A* against Galacto-*N*-biose was measured using an L-arabinose/D-galactose kit (Megazyme, Bray, Ireland). Reactions were heat-inactivated (98°C, 2min) at time intervals of 30s to 5min. The amount of released galactose was then measured using the kit, according to the manufacturer's protocol. Galactose (Gal) standards (0.1-0.5 mM) were measured in order to convert absorbance data into concentrations to determine normalized rates.

Thin layer chromatography and high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC) analyses of enzyme activity

Enzyme activity was assayed using TLC by spotting 1 μ L of the enzymatic reactions onto silica gel 60 F454 plates (Merck, Burlington, MA, USA) at varying time points. The sugars were separated by a mobile phase consisting of butanol: ethanol: water (5:3:3) (all v/v). For visualization, the plates were sprayed with a DPA solution (Diphenylamine (2g): Aniline (2g): MeOH (100mL): concentrated phosphoric acid (10mL)) and subsequently visualized by heating to 550 °C with an Einhell heat gun. TLC assays were performed in 3 independent replicates.

The HPAEC-PAD analysis was carried out on a Dionex ICS-2100 using a 4 X 250mm CarboPac 1 (Dionex) (*AmGH110A*; *AmGH36A*; *AmGH95B*) column in combination with a 4 X 50mm CarboPac guard columns packed with the same resin (Dionex, Sunnyvale, CA, USA). An injection volume of 10 μ L was used for all samples, and elution was carried out at 0.25mL min⁻¹ with a mobile phase of 100mM NaOH for 10min, followed by a linear increase of 0–0.4M NaOAc for 10min, which was kept for an additional 2min and then decreased back to 0mM sodium acetate for 3min. Concentrations of analyte sugars were determined from the area underneath the corresponding peaks, as a function of the concentrations of the standards using Chromeleon v. 6.80 SR15 Build 4656 (Dionex, Sunnyvale, CA, USA).

Enzyme Crystallization and Data Collection

All enzymes were buffer exchanged and concentrated into a minimal buffer containing 2mM MES, 50mM NaCl, pH6.5, and crystallization screening was performed by sitting drop vapour diffusion using crystallizations kits (Molecular Dimensions, Sheffield S1 4DP, UK). Drops of either 1:3, 1:1, or 3:1 ratios of enzyme solution: reservoir solution (200nL) were spotted on an MRC 2 Lens Crystallisation Plate (SWISSCI, Buckinghamshire, UK) using mosquito® Xtal3 handling robot (SPT Labtech, Melbourne, Australia). After the formation of protein crystals, the crystallization conditions were further optimized by incomplete factorial design, using various concentrations of the reservoir solution components, and the pH. Crystals picked for data collection were cryoprotected in either 25% (m/v) glycerol or 25% (m/v) ethylene glycol mixed with reservoir solution and frozen under liquid nitrogen. Complete X-ray diffraction data was collected at 100K at BioMAX beamline at the MAX IV synchrotron radiation facility (Lund, Sweden), or the PETRA III P13/P14 beamline at Deutsches Elektronen-Synchrotron DESY (Hamburg, Germany). An

overview of enzyme concentrations, initial crystal conditions, final crystal conditions, and cryoprotectants used listed in Supplementary Table 16.

Structure Determination

All X-ray diffraction data was initially processed and integrated with xia2⁴⁵ using the 3dii pipeline, with XDS (version: January 31, 2020). The phases were solved by molecular replacement using the Phaser version, included in Phenix package (v.1.19.2-4158). The structures of *PdGH110B* (PDB: 7JFW) combined with an AlphaFold model generated in colab^{46,47} was used to solve the structure of *AmGH110A*. The *AmGH20C* (PDB: 7CBO) structure was used as search models to solve the structure *AmGH20A*. The structures were refined using phenix.refine⁴⁸ initially, then manual rebuilding and addition of solvent content was performed in COOT⁴⁹. The models were validated using Molprobity, and the data statistics and refinement details are in Supplementary Table 7 and Supplementary Table 11 for *AmGH110A* and *AmGH20A* respectively.

Blood samples

Anonymized left-over blood samples with known ABO blood group, collected during routine blood donation from healthy blood donors and processed at the Department of Clinical Immunology and Transfusion Medicine at the Office for Medical Services in Lund, Sweden, were used in this study. Additionally, archived rare reagent RBCs and plasmas from the in-house collection were also included, *e.g.* as controls in flow cytometry and typing. According to the Swedish research law, using waste/excess fully anonymized and/or pooled biological material does not require ethical approval. Thus, no particular approval was required for this study, apart from the permission from the Dept. of Clinical Immunology and Transfusion Medicine to obtain these anonymized blood samples from waste material (ref. no. 2020:16 and 2022:19).

Crossmatch haemagglutination assays

Serologic testing was conducted according to standard blood banking practice in gel cards (Bio-Rad, Hercules, CA, US). The reactions were read as negative (-), positive, graded from one plus (1+) to four plus (4+) or mixed field (m). For confirmation or further blood group typing monoclonal anti-A (A003, Seraclone/Bio-Rad), anti-B (9621A8, Diagast, Loos, France), anti-H (lectin *Ulex europeus*, Immucor, Norcross, GA, US), anti-Le^a (GAMA701, Immucor) and anti-Le^b (GAMA704, Immucor) were used.

Enzymatic digest of carbohydrate epitopes on RBCs

One volume of packed RBCs (pRBCs) was primed in 4x volume conversion buffer⁸, unless otherwise stated, for 15min while mixing, followed by centrifugation for 5min at 1000xg. Supernatant was removed and cells diluted in 6x volume buffer and pelleted again. The enzyme was diluted in conversion buffer and the pRBCs were diluted in 1.6x enzyme solution in while mixing, pRBC volumes ≤50μL by horizontal shake on MTS 2 (IKA-Werke, Germany) at 200 rpm and greater volumes on a tube rotator (Fisherbrand, Thermo Fisher Scientific, Waltham, MA, USA). The reaction was stopped by washing the cells in PBS. The protocol was performed at room temperature and the duration 30min if not otherwise stated. The cells were finally diluted in ID-cellstab (Bio-Rad) and stored at 4°C.

Papain treatment of RBCs

Packed RBCs (pRBCs) were incubated in 2x volume 0.1% papain solution for 5min at 37°C. To confirm digest, glycine soja lectin was used in haemagglutination assay. This treatment removes selected cell surface glycoproteins, thereby decreasing the negative charge of the RBC surface and consequently increasing the sensitivity of the immunostaining and detection of ABO epitopes on glycolipids.

Anti-ExtB (GalNAc-B) eluate

Reagent plasma (1.2mL) derived from an individual with blood group B P₁P^k was incubated with 500μl pRBCs blood group O, at 37°C for 45 min, followed by 30 min on ice. This was repeated twice with O RBCs with the incubation at 37°C lowered to 30 min. The procedure was repeated a fourth time with papain-treated RBCs of blood group O P₂ and a fifth time with papain treated pRBCs of blood group B. The plasma was subjected to crossmatches with different pools of RBCs (blood group O, B, papain-treated B and papain-treated O P₂). Finally, adsorption of anti-ExtB was made by incubating the remaining antibodies with group B pRBCs, same setting as previous incubations. The RBCs were washed six times in cold PBS and the antibodies heat-eluted in 6% (w/v) BSA for 10min at 56°C⁵⁰. The eluate specificity (including absence of anti-P and -PX2) was confirmed by flow cytometric analysis by staining of appropriate control RBCs of the following blood groups: B P₁^k (negative), B pp (positive), O pp (negative) and O (negative). The extended B antigen (in this study otherwise called GalNAc-B to reflect the nature of the extension and the corresponding enzyme specificity required to convert it to B antigen, see Fig. 1a) was acknowledged by the International Society of Blood Transfusion on Dec 15, 2021 under the official name ExtB antigen.

Flow cytometry

Fixed RBCs were stained with monoclonal anti-A (clone ES-15, Quotient, Eysins, Switzerland), anti-B (clone 9621A8) and anti-H (clone BRIC231, IBGRL research products, Bristol, UK) followed by phycoerythrin (PE)-conjugated secondary rat anti-mouse-κ (RAM-κ) as described elsewhere²⁸. The same procedure was used for staining with anti-T/Gal-A (3C9), anti-A type 3 (TH1) and anti-H type 3 (HH14), although the RBCs were papain-treated before staining. The three latter antibodies were a kind gift from Dr. Ulla Mandel, Copenhagen Center for Glycomics, University of Copenhagen, Denmark)

ExtB staining was conducted on papain-treated RBCs, 4 μL 2.5% (v/v) cell suspension were incubated with 15μL anti-ExtB eluate for 20min room temp while mixing, followed by 30min at 4°C. The cells were washed twice in PBS and stained with secondary goat anti-human-IgG:PE and goat anti-human-IgM:PE (both from Jackson ImmunoResearch Europe, Cambridge, UK) diluted 1:50, incubation 10 min at room temperature while mixing. Flow cytometry data of 20,000 stained cells were recorded on BD FACS Canto II (BD, Franklin Lakes, NJ, US) and results analyzed in FCS Express v.6 (De Novo Software, Pasadena, CA, US). A representative gating strategy for the flow cytometry data is shown in Supplementary Fig. 33.

Bioinformatics.

SignalP (V.6.0), DeepTMHMM (V.1.0.20), and Jpred4 server were used for signal peptide and secondary structure predictions. Molecular weight, theoretical *pI*, and Extinction coefficients (M⁻¹ cm⁻¹) at 280 nm were predicted using ProtParam ExPASy server. InterPro and dbCAN meta servers under default settings were used to analyse modular organisation of enzymes. Sequences for phylogenetic analysis were retrieved from the CAZy database (March 2023) using NCBI Batch Entrez, or through batch BLASTP searches against the non-redundant database for a max target of 5000 sequences using the catalytic

domains of the enzymes and orthologues as described in the Supplementary information (Supplementary Fig. 1, 5, 7, 8, 11, 31). Sequences were initially aligned using MAFFT v. 7 (<https://mafft.cbrc.jp/alignment/software/>) and trimmed to only include the catalytic domain of each protein using JalView v.1.8.3, based on the region of the alignment corresponding to the catalytic domain of the query sequences, and then filtered for redundancy using CD-HIT under default settings with a 95% sequence identity cut-off. The sequences were then realigned using MAFFT server v. 7, and phylogenetic trees were generated using the neighbour-joining algorithm with bootstrap = 1000 iterations, using the MAFFT server. Phylogenetic trees were visualized and annotated using iTOL, and characterized enzymes were labelled using InkScape 1.1. Closest structural orthologues to *AmGH110A* and *AmGH20A* were identified using the Dali server⁵¹. Sequence logos were generated using the WebLogo server⁵² at default settings.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data Availability

The structure coordinate files have been submitted to the PDB data base under the accession numbers 8PVS for *AmGH110A*, and 8PXT, 8PXU, and 8PXV for *AmGH20A*-GalNAc, unliganded *AmGH20A*, and *AmGH20A*-lactose, respectively. All other data are available in the main text or supplementary materials. Source data are provided with this paper.

References

1. World Health Organization. WHO model list of essential medicines - 22nd list, 2021. *Tech. Doc.* 2021 (2021).
2. Daniels, G. The molecular definition of red cell antigens. *ISBT Sci. Ser.* **5**, 300–302 (2010).
3. Clausen, H. & Hakomori, S. -i. ABH and Related Histo-Blood Group Antigens; Immunochemical Differences in Carrier Isotypes and Their Distribution. *Vox Sang.* **56**, 1–20 (1989).
4. Hakomori, S. & D. Strycharz, G. Cellular blood-group substances. I. Isolation and chemical composition of blood-group ABH and Leb isoantigens of sphingoglycolipid nature. *Biochemistry* **7**, 1279–1286 (1968).
5. Jarnefelt, J., Rush, J., Li, Y. T. & Laine, R. A. Erythroglycan, a high molecular weight glycopeptide with the repeating structure [galactosyl-(1 leads to 4)-2-deoxy-2-acetamido-glucosyl(1 leads to 3)] comprising more than one-third of the protein-bound carbohydrate of human erythrocyte stroma. *J. Biol. Chem.* **253**, 8006–8009 (1978).
6. Clausen, H. & Olsson, M. L. Towards universally acceptable blood. *Nat. Microbiol.* **4**, 1426–1427 (2019).
7. Goldstein, J., Siviglia, G., Hurst, R., Lenny, L. & Reich, L. Group B Erythrocytes Enzymatically Converted to Group O Survive Normally in A, B, and O Individuals Published by : American Association for the Advancement of Science Stable URL : <http://www.jstor.org/stable/1687589>. *Science (80-)*. **215**, 168–170 (1982).
8. Liu, Q. P. *et al.* Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **25**, 454–464 (2007).
9. Rahfeld, P. *et al.* An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat. Microbiol.* **4**, 1475–1485 (2019).
10. Rahfeld, P. & Withers, S. G. Toward universal donor blood: Enzymatic conversion of A and B to O type. *J. Biol. Chem.* **295**, 325–334 (2020).
11. Kruskall, M. S. *et al.* Transfusion to blood group A and O patients of group B RBCs that have been enzymatically converted to group O. *Transfusion* **40**, 1290–1298 (2000).
12. Lenny, L. L., Hurst, R., Goldstein, J., Benjamin, L. J. & Jones, R. L. Single-unit transfusions of RBC enzymatically converted from group B to group O to A and O normal volunteers. *Blood* **77**, 1383–1388 (1991).
13. Lenny, L. L., Hurst, R. & Galbraith, R. A. Transfusions to group O subjects of 2 units of red cells enzymatically converted from group B to group O. *Transfusion* **34**, 209–214 (1994).
14. Gao, H. W. *et al.* Evaluation of group A1B erythrocytes converted to type as group O: Studies of markers of function and compatibility. *Blood Transfus.* **14**, 168–174 (2016).
15. Olsson, M. L. & Clausen, H. Modifying the red cell surface: Towards an ABO-universal blood supply. *Br. J. Haematol.* **140**, 3–12 (2008).
16. Goel, R. *et al.* ABO blood group and COVID-19: a review on behalf of the ISBT COVID-19 Working Group. *Vox Sang.* **116**, 849–861 (2021).
17. Jeyakanthan, M. *et al.* Chemical Basis for Qualitative and Quantitative Differences Between ABO Blood Groups and Subgroups: Implications for Organ Transplantation. *Am. J. Transplant.* **15**, 2602–2615 (2015).
18. Clausen, H. *et al.* Blood group a glycolipid (AX) with globo-series structure which is specific for blood group A1 erythrocytes: One of the chemical bases for A1 and A2 distinction. *Biochem. Biophys. Res. Commun.*

- 124**, 523–529 (1984).
19. Clausen, H., Levery, S. B., Nudelman, E., Tsuchiya, S. & Hakomori, S. Repetitive A epitope (type 3 chain A) defined by blood group A1-specific monoclonal antibody TH-1: Chemical basis of qualitative A1 and A2 distinction. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 1199–1203 (1985).
 20. Clausen, H., Holmes, E. & Hakomori, S. I. Novel blood group H glycolipid antigens exclusively expressed in blood group A and AB erythrocytes (type 3 chain H). II. Differential conversion of different H substrates by A1 and A2 enzymes, and type 3 chain H expression in relation to secretor status. *J. Biol. Chem.* **261**, 1388–1392 (1986).
 21. Clausen, H., Levery, S. B., Kannagi, R. & Hakomori, S. I. Novel blood group H glycolipid antigens exclusively expressed in blood group A and AB erythrocytes (type 3 chain H). I. Isolation and chemical characterization. *J. Biol. Chem.* **261**, 1380–1387 (1986).
 22. Ricci Hagman, J. *et al.* β 1,3GalNAc-T1-dependent extension of the human blood group B antigen results in a novel ABO-related glycolipid structure on erythrocytes. *Vox Sang.* **114**, 53 (2019).
 23. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **5**, (2015).
 24. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* **54**, 1469–1476 (2004).
 25. Shuoker, B. *et al.* Sialidases and fucosidases of *Akkermansia muciniphila* are crucial for growth on mucin and nutrient sharing with mucus-associated gut bacteria. *Nat. Commun.* **14**, 1833 (2023).
 26. Gregg, K. J., Finn, R., Abbott, D. W. & Boraston, A. B. Divergent modes of glycan recognition by a new family of carbohydrate-binding modules. *J. Biol. Chem.* **283**, 12604–12613 (2008).
 27. Ueda, T. Isolation and Characterization of a Blood Group A Substance-degrading α -TV-Acetylgalactosaminidase from tivity of blood group substances by degrading the carbohydrate portions of the glycopro-composing enzyme from *Trichomonas foetus*. and the *Clostridium*. **53**, 111–120 (1989).
 28. Hult, A. K. & Olsson, M. L. Many genetically defined ABO subgroups exhibit characteristic flow cytometric patterns. *Transfusion* **50**, 308–323 (2010).
 29. Teze, D. *et al.* The Catalytic Acid-Base in GH109 Resides in a Conserved GGHGG Loop and Allows for Comparable α -Retaining and β -Inverting Activity in an N-Acetylgalactosaminidase from *Akkermansia muciniphila*. *ACS Catal.* **10**, 3809–3819 (2020).
 30. Anso, I. *et al.* Turning universal O into rare Bombay type blood. *Nat. Commun.* **14**, 1765 (2023).
 31. Pichler, M. J. *et al.* Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. *Nat. Commun.* **11**, (2020).
 32. Kelly, R. J., Rouquier, S., Giorgi, D., Lennon, G. G. & Lowe, J. B. Sequence and expression of a candidate for the human Secretor blood group $\alpha(1,2)$ fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J. Biol. Chem.* **270**, 4640–4649 (1995).
 33. Moore, S., Chirside, A., Micklem, L. R., McClelland, D. B. L. & James, K. A Mouse Monoclonal Antibody with Anti-A,(B) Specificity Which Agglutinates Ax Cells. *Vox Sang.* **47**, 427–434 (1984).
 34. Clausen, H., Stroud, M., Parker, J., Springer, G. & Sen-Itiroh, H. Monoclonal antibodies directed to the blood group a associated structure, galactosyl-A: Specificity and relation to the thomsen-friedenreich antigen. *Mol. Immunol.* **25**, 199–204 (1988).

35. McGuire, B. E. *et al.* The structure of a family 110 glycoside hydrolase provides insight into the hydrolysis of α -1,3-galactosidic linkages in l-carrageenan and blood group antigens. *J. Biol. Chem.* **295**, 18426–18435 (2020).
36. Pluvinage, B. *et al.* Architecturally complex O-glycopeptidases are customized for mucin recognition and hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
37. Xu, W., Yang, W., Wang, Y., Wang, M. & Zhang, M. Structural and biochemical analyses of β -N-acetylhexosaminidase Am0868 from *Akkermansia muciniphila* involved in mucin degradation. *Biochem. Biophys. Res. Commun.* **529**, 876–881 (2020).
38. Robbe, C., Capon, C., Coddeville, B. & Michalski, J. C. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J.* **384**, 307–316 (2004).
39. Aoki, T. A comprehensive review of our current understanding of red blood cell (RBC) glycoproteins. *Membranes (Basel)*. **7**, 1–19 (2017).
40. Higgins, M. A., Ficko-Blean, E., Meloncelli, P. J., Lowary, T. L. & Boraston, A. B. The overall architecture and receptor binding of pneumococcal carbohydrate-antigen-hydrolyzing enzymes. *J. Mol. Biol.* **411**, 1017–1036 (2011).
41. Hobbs, J. K., Pluvinage, B., Robb, M., Smith, S. P. & Boraston, A. B. Two complementary α -fucosidases from *Streptococcus pneumoniae* promote complete degradation of host-derived carbohydrate antigens. *J. Biol. Chem.* **294**, 12670–12682 (2019).
42. Wang, A., Ribeiro, R. V. P., Ali, A., Brambate, E., Abdelnour-Berchtold, E., Michaelsen, V., Zhang, Y., Rahfeld, P., Moon, H., Gokhale, H., Gazzalle, A., Pal, P., Liu, M., Waddell, T. K., Cserti-Gazdewich, C., Tinckam, K., Kizhakkedathu J. N., West, L., Kes, M. Conversion of blood type A donor lungs into universal blood type lungs using ex vivo ABO enzymatic treatment. *Sci. Transl. Med.* (2021).
43. MacMillan, S., Hosgood, S. A. & Nicholson, M. L. Enzymatic blood group conversion of human kidneys during ex vivo normothermic machine perfusion. *Br. J. Surg.* **110**, 133–137 (2023).
44. Nielsen, H. Predicting secretory proteins with signalP. *Methods Mol. Biol.* **1611**, 59–73 (2017).
45. Gildea, R. J. & Winter, G. Determination of Patterson group symmetry from sparse multi-crystal data sets in the presence of an indexing ambiguity. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 405–410 (2018).
46. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
47. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
48. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).
49. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
50. Judd, W. J. Elution—Dissociation of antibody from red blood cells: Theoretical and practical considerations. *Transfus. Med. Rev.* **13**, 297–310 (1999).
51. Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32**, 1–18 (2023).
52. Crooks, G., Hon, G., Chandonia, J. & Brenner, S. NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190 (2004).

Acknowledgements

The study has been funded by a Research Fund Denmark Project 2 grant (0136-00086B) to M.A.H. with M.L.O. and J.P.M as co-applicants. Additional funding was from the Swedish Research Council (grant no. 2019-01683), the Knut and Alice Wallenberg Foundation (grant no. 2020.234) and the Swedish government and county councils (grant no. ALFSKANE-446521) to M.L.O. J.R.H. was supported by a research and development grant for Ph.D. students from Region Skåne. The authors would like to Acknowledge Bashar Shouker, who had cloned most of the *A. muciniphila* enzymes and produced AmGH20 enzymes in his PhD project, which was supervised by Professor Eva Nordberg Karlsson, and funded by a Scholarship from the Iraqi Ministry of Higher Education and Scientific Research. We acknowledge MAX IV Laboratory for time on the BioMax beamline under proposals 20190334 and 20200120. We thank Ana Gonzales and Uwe Muller for assistance in using the beamline and data collection. We acknowledge DESY (Hamburg, Germany), for time on the P13 beamline at Petra III under the proposal MX846, and would like to thank Isabel Bento for assistance in using the beamline and data collection.

Author contributions

M.J. and L.S. contributed equally to this work.

Conceptualisation: M.A.H and M.L.O.; Methodology: M.A.H, M.L.O., A.H., and J.P.M; Investigations: M.J., L.S, J.R.H., T.S.N., J.W. and A.H.; Analysis: M.J., L.S, J.R.H., M.J.P., Visualisation: M.J. and L.S; Funding Acquisition: M.A.H., J.P.M., M.L.O.; Project administration: M.A.H. and M.L.O. Writing original draft: M.J., L.S, M.L.O and M.A.H. Writing, review and editing: M.J., L.S., J.R.H., J.M.P., T.S.N, A.H., J.P.M., M.L.O., M.A.H.

Competing interest

A patent application has been filed based on data from this study.

Additional information

Supplementary information

The online version contains supplementary material available at xxxx.

Correspondence and requests for materials Should be addressed to Maher Abou Hachem and Martin L. Olsson.

Supplementary Information

Targeting extended blood antigens by *Akkermansia muciniphila* exoglycosidases unveils a missing link for generating ABO-universal donor blood

Supplementary Discussion

Supplementary Figures

Supplementary Fig. 1: Phylogenetic tree of GH110 sequences from the CAZy database

Supplementary Fig. 2: Modular organisation and sequence divergence of enzymes against the B antigen

Supplementary Fig. 3: Activity profiles of *A. muciniphila* enzymes against the B type 2 hexaose

Supplementary Fig. 4: Modular organisation and sequence divergence of enzymes against the extended GalNAc-B antigen

Supplementary Fig. 5: Phylogenetic tree of GH20 sequences from *A. muciniphila*

Supplementary Fig. 6: Activity of enzyme candidates towards B antigens on group B RBCs

Supplementary Fig. 7: Activity of enzymes towards extended GalNAc-B antigen on group B RBCs

Supplementary Fig. 8: Effect of enzyme concentration and reaction buffer on removal of the GalNAc-B antigen on group B RBCs

Supplementary Fig. 9: Modular organization and sequence divergence of enzymes against A and extended Gal-A antigens

Supplementary Fig. 10: Phylogenetic tree of GH109 enzymes from the CAZy database

Supplementary Fig. 11: Phylogenetic tree of GH36 enzymes

Supplementary Fig. 12: Activity of *A. muciniphila* enzymes against the A type 2 hexaose

Supplementary Fig. 13: Activity of α -1,2-fucosidases against the H type 1 pentaose and H type 2 triose

Supplementary Fig. 14: Phylogenetic tree of GH35 enzymes from the CAZy database

Supplementary Fig. 15: Activity profiles of the *A. muciniphila* GH35 enzymes towards galactobioses

Supplementary Fig. 16: Activity of *A. muciniphila* enzymes towards A antigens on group A RBCs

Supplementary Fig. 17: Activity of the α -1,2-fucosidases *RiFuc95* and *AmGH95B* towards H and H type 3 antigens on group A RBCs

Supplementary Fig. 18: Dependence of α -1,2-fucosidase activity towards H antigens on RBCs on buffer and enzyme concentration

Supplementary Fig. 19: Evaluation of GH35 β -galactosidases towards the extended Gal-A antigen

Supplementary Fig. 20: Sequential treatment of A₁ RBCs with *AmGH109B*, *AmGH95B* and *AmGH35A*.

Supplementary Fig. 21: One-pot treatment of A₁ RBCs with *AmGH36A*, *AmGH95B* and *AmGH35A*

Supplementary Fig. 22: Crossmatch reactivity of O plasmas (n=100) towards enzymatically converted B RBCs

Supplementary Fig. 23: Crossmatch reactivity of AB and A plasmas towards enzymatically converted B RBCs

Supplementary Fig. 24: Crossmatch reactivity of donor plasmas of different types towards enzymatically converted A₁ RBCs

Supplementary Fig. 25: Crossmatch reactivity of O plasmas towards enzymatically converted A₁ RBCs

Supplementary Fig. 26: Structural comparison between *AmGH110A* and *PdGH110B*

Supplementary Fig. 27: Surface topology of *AmGH110A* and *PdGH110B*

Supplementary Fig. 28: Comparison of the *AmGH110A* CBM51 domain to structural orthologues
Supplementary Fig. 29: Surface electrostatics of *AmGH20A* and *AmGH20C*
Supplementary Fig. 30: Comparison of *AmGH20A* structures in free and ligand-bound forms
Supplementary Fig. 31: Phylogenetic analysis of the *AmGH20A* CBM
Supplementary Fig. 32: Crossmatch reactivity of AB and B plasmas towards ECO-A RBCs
Supplementary Fig. 33: Representative gating strategy for flow cytometry data
Supplementary Fig. 34: Crossmatch reactions of group O plasmas towards group A RBCs from a secretors and nonsecretors after enzymatic conversion

Supplementary Tables:

Supplementary Table 1: Cloning primers of *A. muciniphila* genes encoding enzymes used in the study
Supplementary Table 2: Properties of enzymes for conversion of B and extended GalNAc-B antigens
Supplementary Table 3: Normalised rates of enzyme candidates against aryl substrate analogues
Supplementary Table 4: Oligosaccharide substrates used for enzyme activity analyses.
Supplementary Table 5: Normalised rates of enzymes against oligosaccharide substrate analogues
Supplementary Table 6: Properties of enzymes for canonical and extended A antigen conversion
Supplementary Table 7: Data collection and refinement statistics for *AmGH110A*
Supplementary Table 8: *AmGH110A* top structural orthologues
Supplementary Table 9: *AmGH110A*-CBM51 top structural orthologues
Supplementary Table 10: Isoelectric point (pI) of characterized B antigen-active GH110 enzymes
Supplementary Table 11: Data collection and refinement statistics for *AmGH20A*
Supplementary Table 12: *AmGH20A* top structural orthologues
Supplementary Table 13: *AmGH20A* CBM domain top structural orthologues
Supplementary Table 14: Comparison of conditions for enzymatic conversion of ABO antigens on RBCs
Supplementary Table 15: Enzyme assay conditions for the determination of normalized rates
Supplementary Table 16: Enzyme crystallization conditions

Supplementary Discussion

The sequentially (*AmGH36A*→*AmGH95B*→*AmGH35A*→*AmGH36A*) treated A₁ RBCs were crossmatched with AB plasmas, to get an indication if the enzymes generate neo- or cryptantigens (*i.e.*, an antigen on the surface of a red cell that is not normally visible) unrelated to the ABO system (Supplementary Fig. 24). One A₂B plasma (P140) reacted weakly with the A₁ RBCs treated with *AmGH36A*→*AmGH95B* (1+), and another (P143) reacted with the *AmGH109B*-treated RBCs. The one A₂B plasma tested, reacted with RBCs post the *AmGH95B* digests, both with the *AmGH36A*->*AmGH95B* (2+) and *AmGH109B*->*AmGH95B* treatment (1+), which could be due to antibodies recognizing remaining A₁ specific antigens, *e.g.*, Gal-A, which may be present in A₂B plasma. To further evaluate the cause of reactivity, A₁ and O cells subjected to both single and sequential treatments with the *AmGH95B* and *AmGH35A* were included in the crossmatch testing (Supplementary Fig. 24). *AmGH95B*→*AmGH35A* treated group O RBCs reacted weakly (1+) with one O plasma (P138) and one A₁B plasma (P140). Interestingly, treatments of group O RBCs with either *AmGH95B* or *AmGH35A* did not cause any reactivity with the plasmas. Similarly, A₁ RBCs treated with *AmGH95B* and *AmGH35A* (single or sequential) had no unexpected reactivities. A plausible explanation for this, could be that the removal of the fucosyl from H antigens, which is mediated by *AmGH95B*, exposes the terminal Gal-1,3-GlcNAc motif in the H type 1 antigen (Fig. 1a). This motif will be a substrate for *AmGH35A*, which would catalyse the removal of the terminal Gal unit, there by exposing the GlcNAc, which may be recognised by antibodies in the plasma. The weak reactions correlate to the dominance of the H type 2 chain and the low abundance of the chain type 1 chains in human plasma. Collectively, these data suggest that neither *AmGH95B* nor *AmGH35A* exposed neo- or cryptantigens, but their combination potentially exposed cryptantigens that were, however, mildly reactive particularly with group O and A₂B plasma.

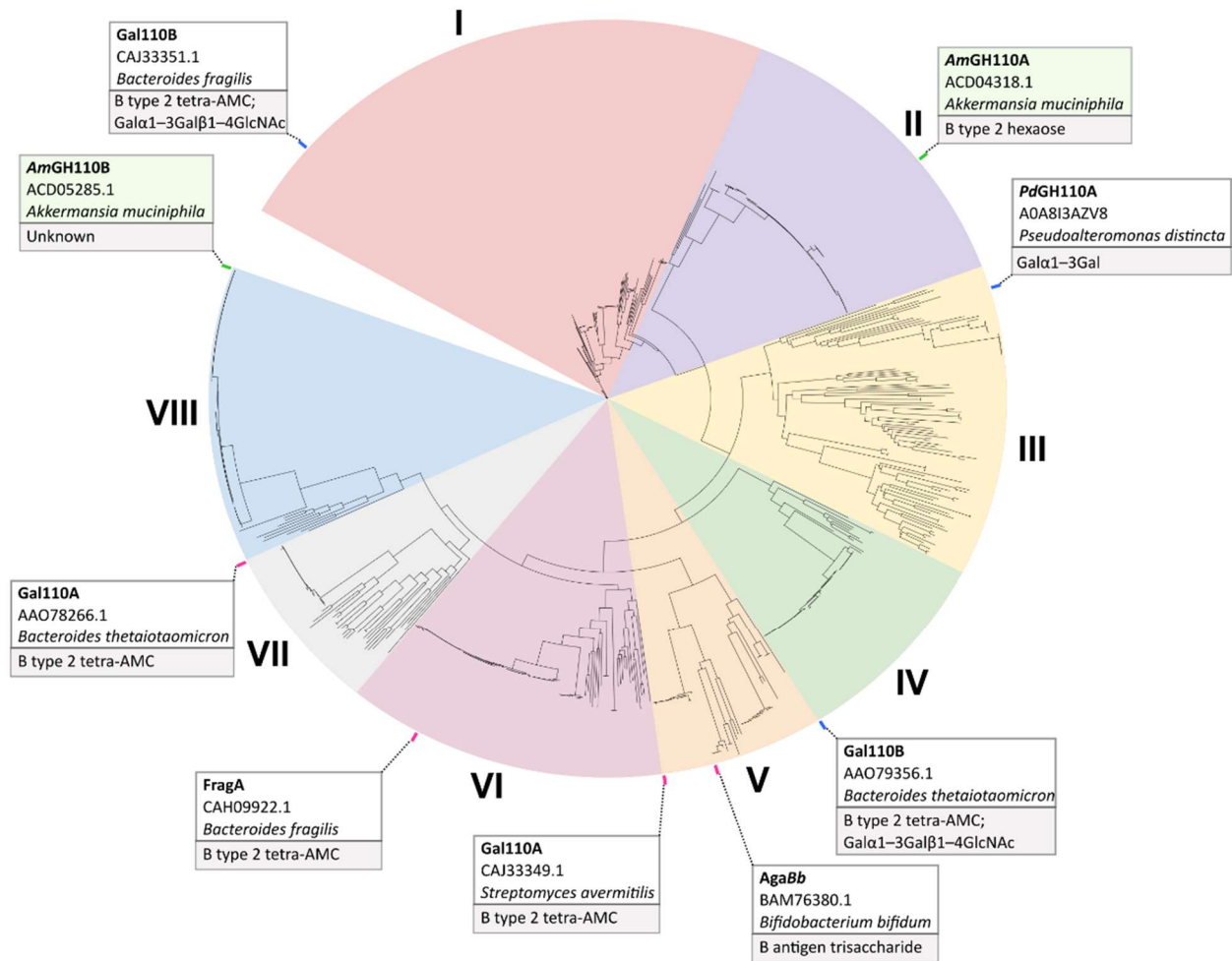
The A₁ cells treated with all four steps of digests were further crossmatched with plasma from two donors of each blood group, A₁, A₂ and B (Supplementary Fig. 24). Amongst these 6 plasmas, all but one B plasma (2+), were non-reactive, consistent with clearly enhanced compatibility associated with the removal of the extended A structures. So far, there is no obvious explanation behind the residual reactivity observed post removal of the A-core antigens. While all blood donor plasmas have been screened for irregular blood group antibodies, there is still a risk/chance that such antibodies have gone unnoticed or have reappeared to interfere with our testing. Since anonymized plasmas were used, we had no possibility to scrutinize back and re-test the same plasmas. Regarding the reactivities observed with plasmas from A₁B and A₂B donors, it is not uncommon that such individuals have naturally-occurring anti-H and anti-A₁, respectively. These are often quite weak but may in some cases have interfered with our crossmatch testing here.

In the larger crossmatch screen, we evaluated the reactivity of 100 group O plasmas with single (*AmGH36A*) or one-pot (*AmGH36A*+*AmGH95B*+*AmGH35A*) treated A₁ RBCs. The reactivity of the triple-enzyme treatments was lower or similar to the single treatment in general, which underscores the importance of the extended A structures in ECO-RBC crossmatching and shows that antigenicity of A₁ RBCs is much more complex than currently described in textbooks. However, in our crossmatch screen a few interesting exceptions were observed (Fig. 4 and Supplementary Fig. 25). One plasma (8i), which showed a strong 3+ with the single-treated RBCs from the secretor donor, reacted 4+ with the RBCs from the same donor, treated with all three enzymes. The reactivity with the same plasma (8i) was decreased from 4+ to 3+ when tested against the nonsecretor A₁-RBCs, which may point to type 1 chain involvement. Another example from the treatment of A₁ nonsecretor RBCs had a reactivity of 2+ in the single treatment with O plasma (2a), which increased to +3 in the triple-enzyme treatment. The same plasma (2a) increased its reactivity from negative to 2+ with the secretor A₁-RBCs, indicating that a factor independent of secretor/nonsecretor status may be at play.

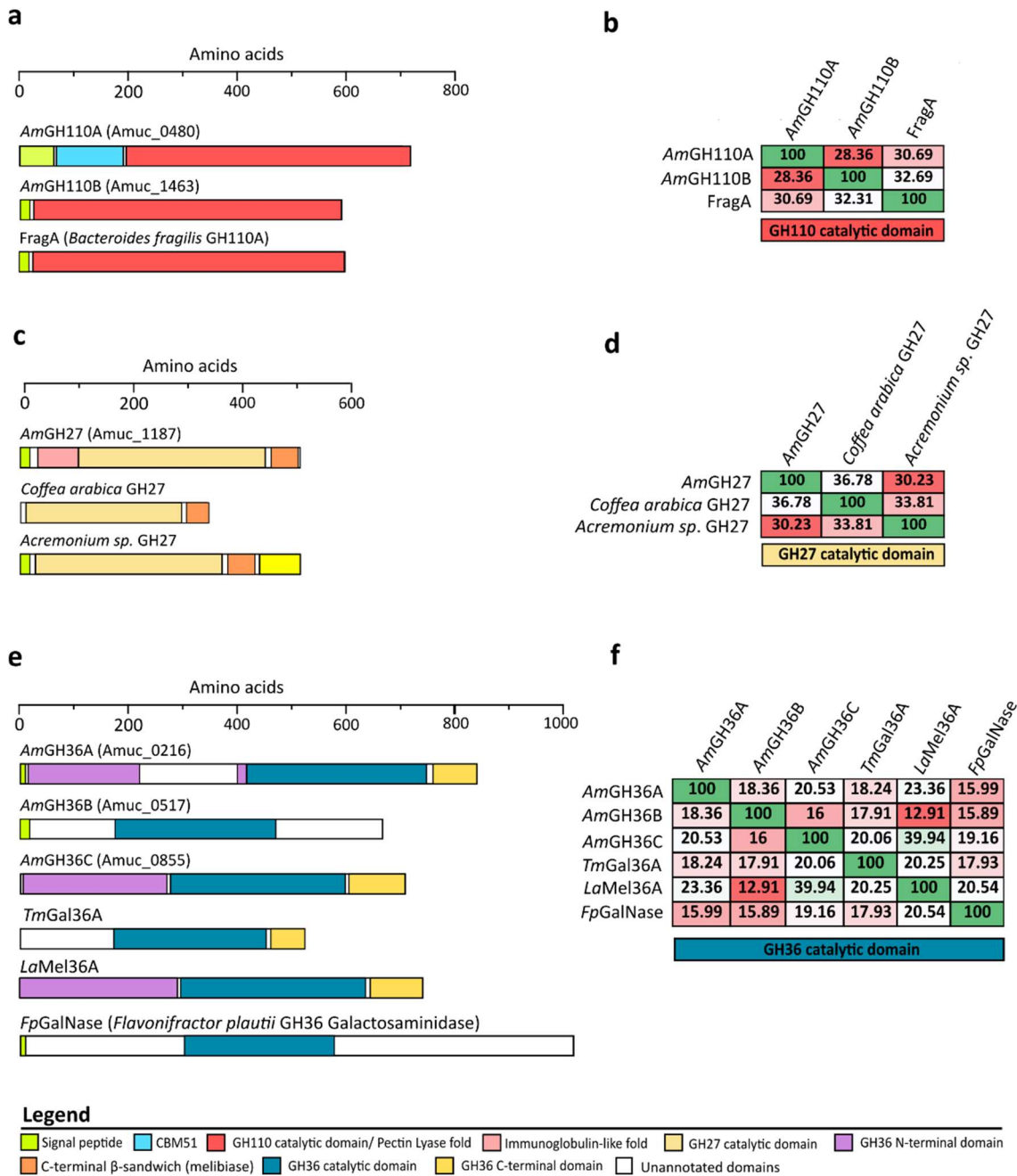
In addition to the O plasma crossmatches, the one-pot treated A₁ RBCs were subjected to crossmatch tests with 12 group B and 11 group AB plasmas (Supplementary Fig. 32). The B plasmas had a similar reaction pattern with RBCs from both donors and was clearly positive against native RBCs as expected. When the RBCs were treated only with *AmGH36A*, five B plasmas became non-reactive, while treatment with all three enzymes generated RBCs non-reactive with nine B plasmas, and the remaining reactivity was weak (<1+). One A₁B plasma (220) reacted 2+ with *AmGH36A*-treated RBCs from the secretor donor, but not at all when treatment with the triple-enzyme blend was performed. An A₂B plasma (224) reacted weakly with triple-enzyme treated RBCs. Collectively, these data highlight a compelling trend of reduced or abolished crossmatch reactivity but decoding the molecular basis for the remaining or rarely increased reactivity of ECO A₁ RBCs with different plasmas merits further investigations beyond the scope of this study.

We did not observe a significant difference between plasma reactivity towards enzyme-treated RBCs from A₁ secretors compared to A₁ nonsecretors. The median reactivity strength of 100 O plasmas was similar comparing both single (*AmGH36A*) as well as triple (*AmGH36A+AmGH95B+AmGH35A*) treated RBCs (Fig. 4). Three O plasmas were further crossmatched with an additional set of RBCs from A₁ secretors (n=5) and nonsecretors (n=7) one-pot treated with *AmGH36A*, *AmGH95B* and *AmGH35A* (Supplementary Fig. 34). The least reactive plasma (P-9i) was negative with all, but one, enzyme-treated RBC sample from a nonsecretor. The next plasma (P-9j) reacted 3+ with all A₁ secretors as compared to 2+ or 3+ with the nonsecretors, but the difference was not statistically significant. The third plasma tested (P-10a) reacted 1+ with treated RBCs from all but one donor (RBCs from the same donor that reacted with plasma no. P-9i) who showed a reactivity of 3+ with plasma P-10a. Treated A₂ RBCs were not reactive with the P-9i plasma and less reactive with the other two plasmas as compared to treated A₁ RBCs. However, only two A₂ RBCs (one secretor and one non secretor) were tested. Collectively, these data suggest that the secretor/nonsecretor background of the recipient plasmas appears to have a modest if any effect on the crossmatch results with ECO A-RBCs. It can also be noted that individual donors (like the one denoted C022 in Supplementary Fig. 27) can be more reactive with different plasmas, indicating that unrecognized inter-individual RBC differences relating to the ECO concept may exist, but have remained elusive so far.

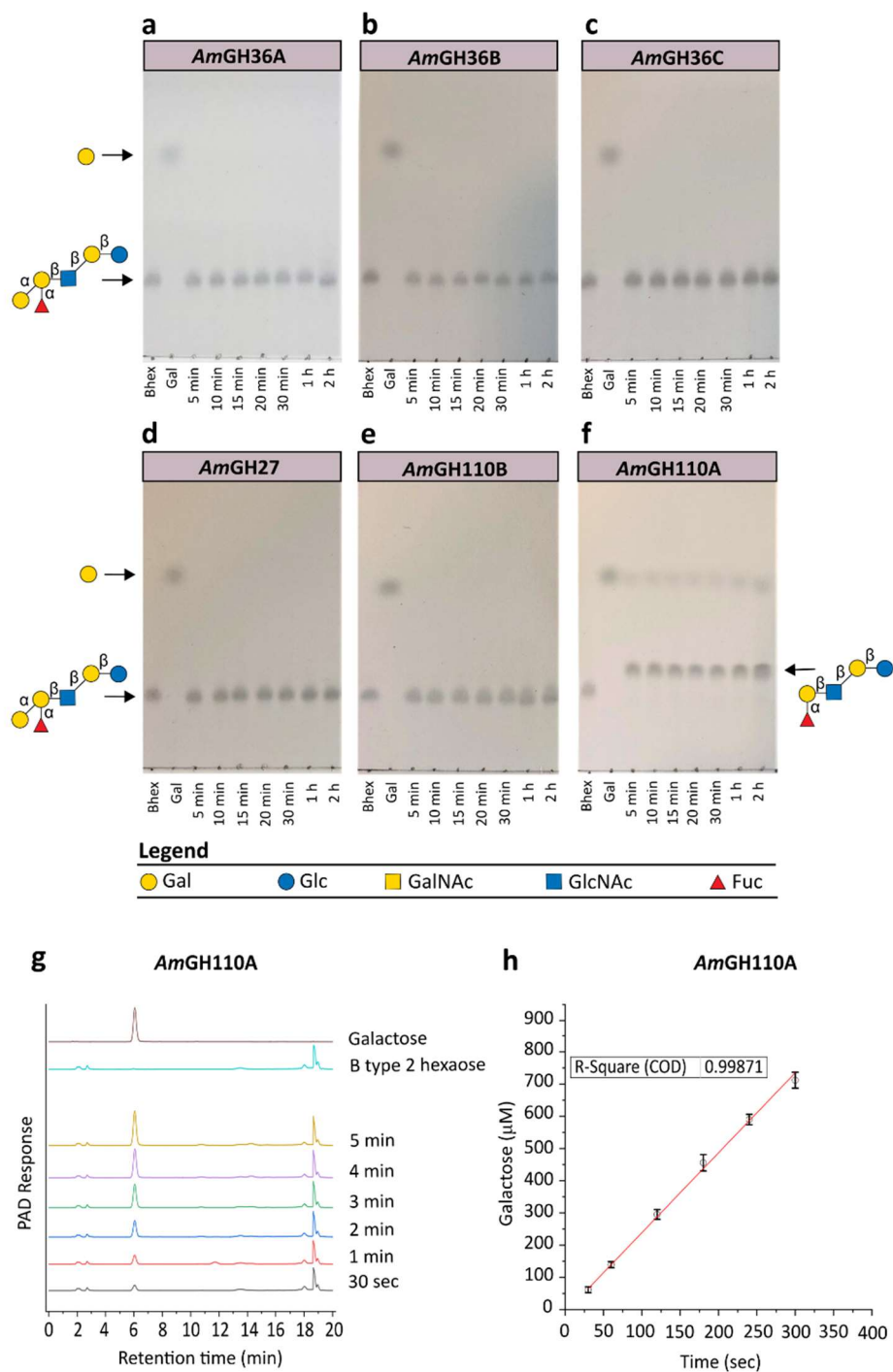
Supplementary Figures



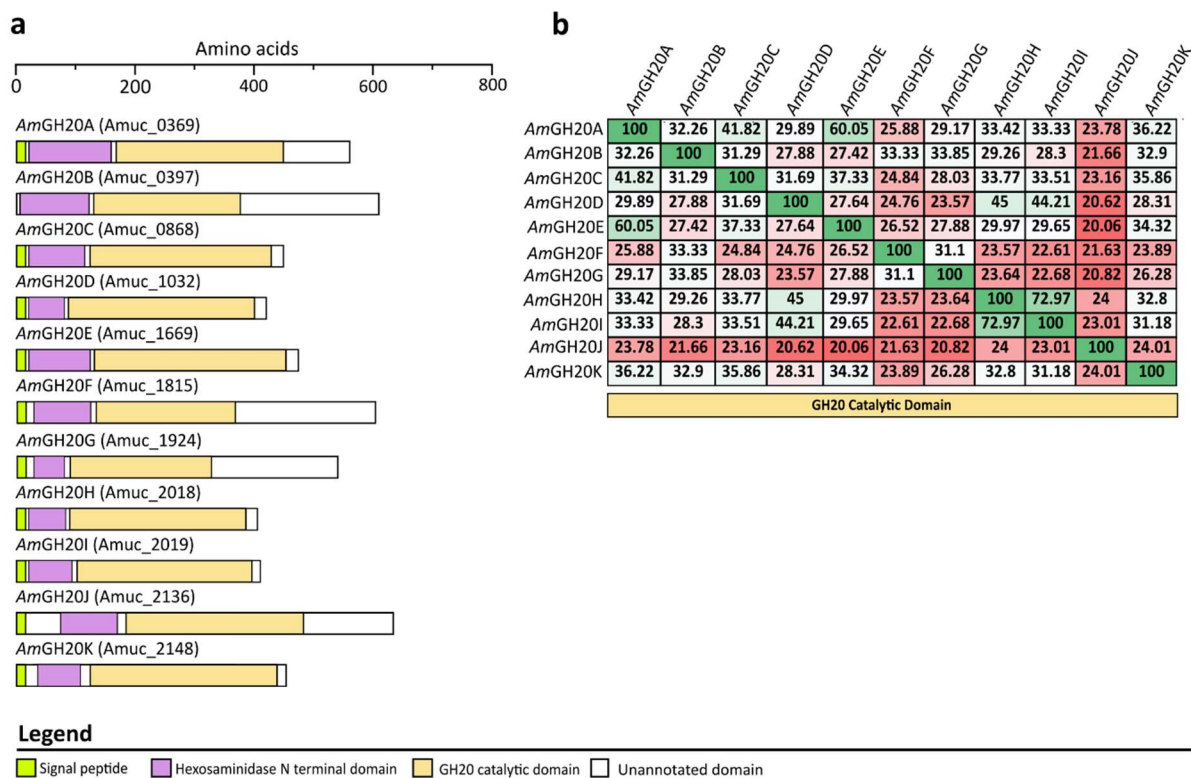
Supplementary Fig. 1: Phylogenetic tree of GH110 sequences from the CAZy database. Previously characterized enzymes that harbour B type antigen activity or λ -carrageenan activity are indicated with enzyme name, GenBank accession and substrates, where enzyme reactivity has been demonstrated. The different clusters are denoted with Roman numerals. A total of 966 annotated GH110 sequences, retrieved from CAZy, were initially aligned using MAFFT V.7 and then trimmed to include only the catalytic domains corresponding to residues 213–795 of *AmGH110A*. The sequences were filtered using MaxAlign with the default settings, and a final list of 748 sequences were subsequently realigned, and a phylogenetic tree was then made using the neighbour-joining algorithm on the MAFFT server with bootstraps performed with 1000 iterations.



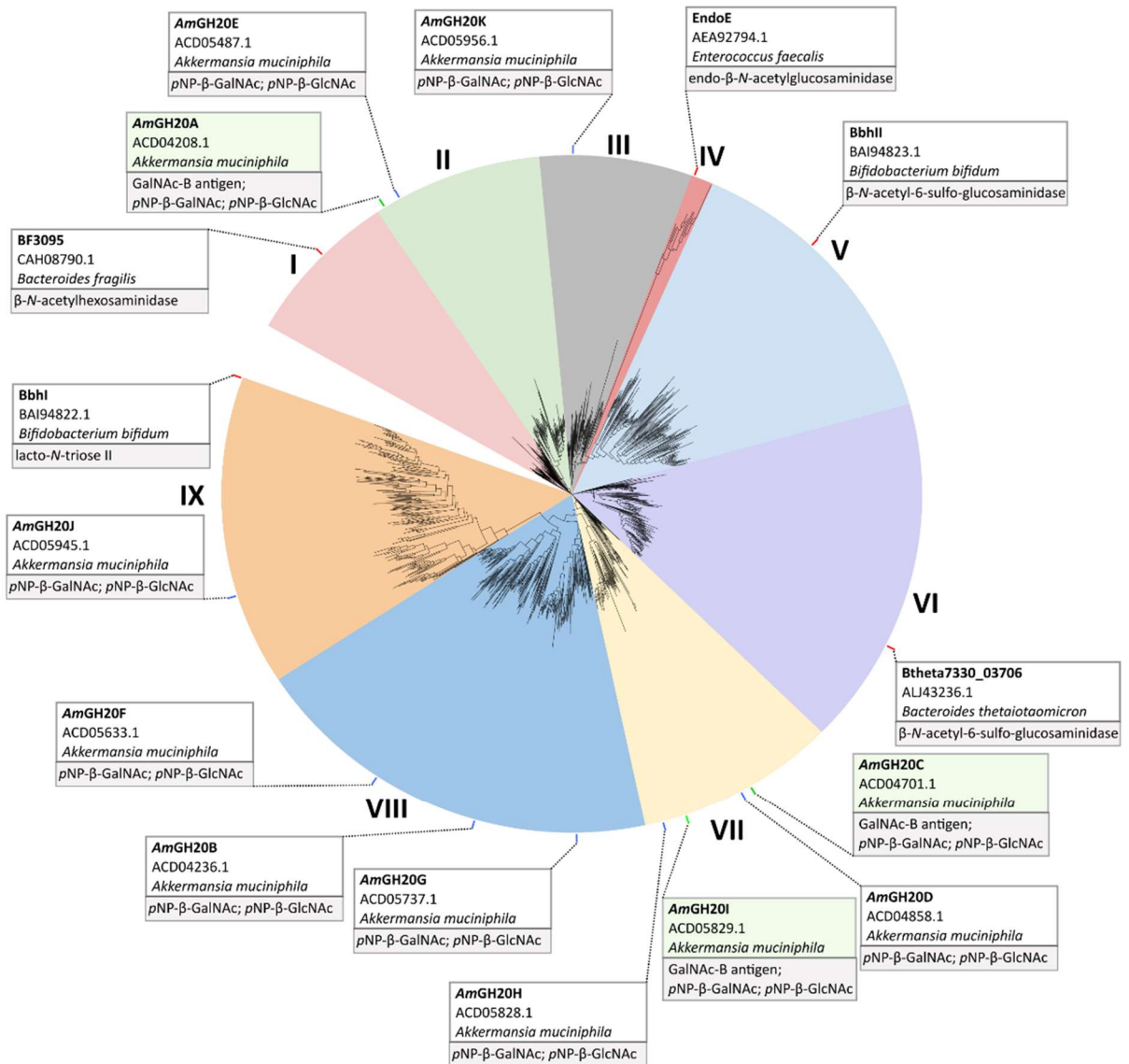
Supplementary Fig. 2: Modular organisation and sequence divergence of enzymes against the B antigen. **a** Domain organisation of the *Akkermansia muciniphila* GH110 enzyme candidates for the B antigen conversion and the previously reported counterpart FragA from *Bacteroides fragilis* (Genbank: AM039447). **b** The amino acid sequence identity (%) matrix between the enzymes in a. **c** Domain organisation of the *A. muciniphila* GH27 and the previously reported counterparts: the α -galactosidase from *Coffea arabica* (Genbank: AAA33022) and the α -N-acetylgalactosaminidase from *Acremonium* sp. (Genbank: BAB08149), which displayed activities on the B and A antigens, respectively. **d** The amino acid sequence identity (%) matrix between the enzymes in c. **e** Domain organisation of the *A. muciniphila* GH36 candidates, and previously characterized GH36 α -galactosidases from *Thermotoga maritima* (Genbank: CAH68692) and *Lactobacillus acidophilus* (Genbank: AAO21867), and the recently discovered galactosaminidase *FpGalNase* (Genbank: ANU41754), which was able to convert group A RBCs in combination with a deacetylase. **f** Amino acid sequence identity (%) matrices of the enzymes in e. Domain annotations are based on the CAZy, DBcan, and Interpro databases.



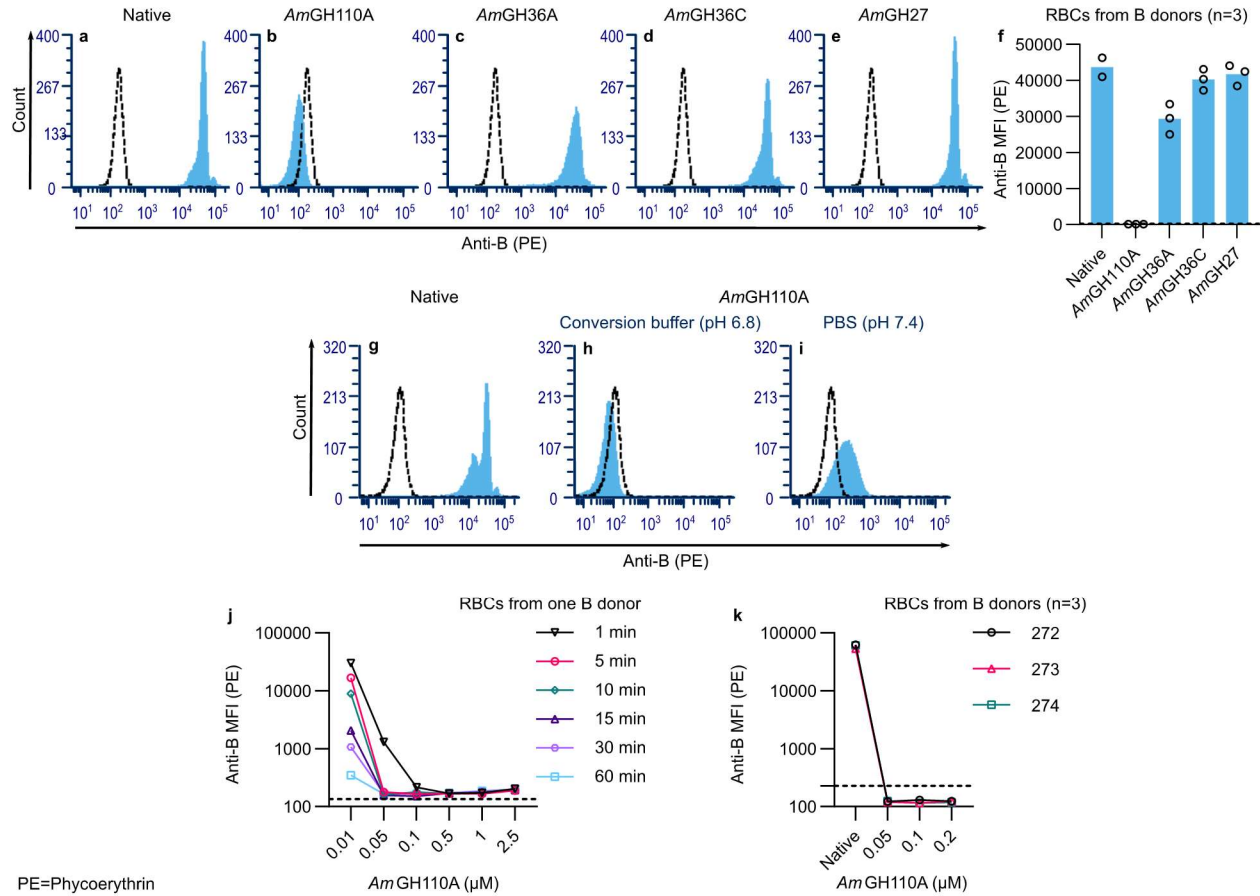
Supplementary Fig. 3: Activity profiles of *A. muciniphila* enzymes against the B type 2 hexaose. **a-f**, Activity of the enzyme candidates toward the B antigen hexaose (type 2 chain, see Fig. 1a) analysed by thin layer chromatography. The assay was performed by monitoring the time-course hydrolysis of 2mM B type 2 hexaose substrate at 37°C and 1 μM enzyme. The TLC data are from two independent experiments, which gave similar results. **g** The HPAEC-PAD chromatograms show galactose and B type 2 hexaose standards at 0.5 mM, and the release of galactose from the B type 2 hexaose (2 mM total) at different times after incubation with 20nM *AmGH110A* at 37°C. **h** The initial rate of *AmGH110A* was measured from the time course release of galactose using a galactose standard curve generated from HPAEC-PAD analyses based on the area under the galactose peak. The HPAEC-PAD data are the means and the error bars are standard deviations of 3 independent experiments. The data are provided in the source data file.



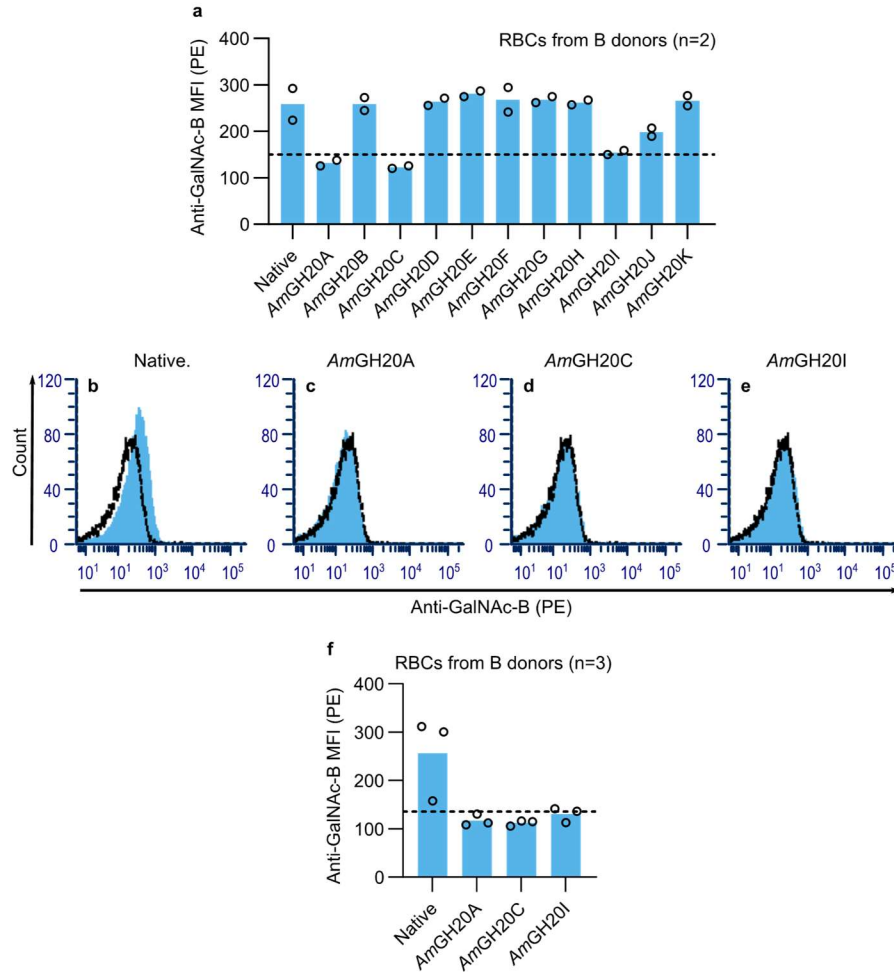
Supplementary Fig. 4: Modular organisation and sequence divergence of enzymes against the extended GalNAc-B antigen. **a** Domain organisation of *A. muciniphila* GH20 enzyme candidates for the GalNAc-B antigen conversion. The five enzymes *AmGH20A*, *AmGH20B*, *AmGH20F*, *AmGH20G*, *AmGH20J* contain one or more unannotated domains at the C-terminus, and one of the enzymes (*AmGH20J*) harbours an additional unannotated sequence patch at the N-terminus. **b** Amino acid sequence identity (%) matrix between the GH20 enzymes based on alignment of catalytic domains. Domain annotations are based on the CAZy, dbCAN, and Interpro databases (see Methods).



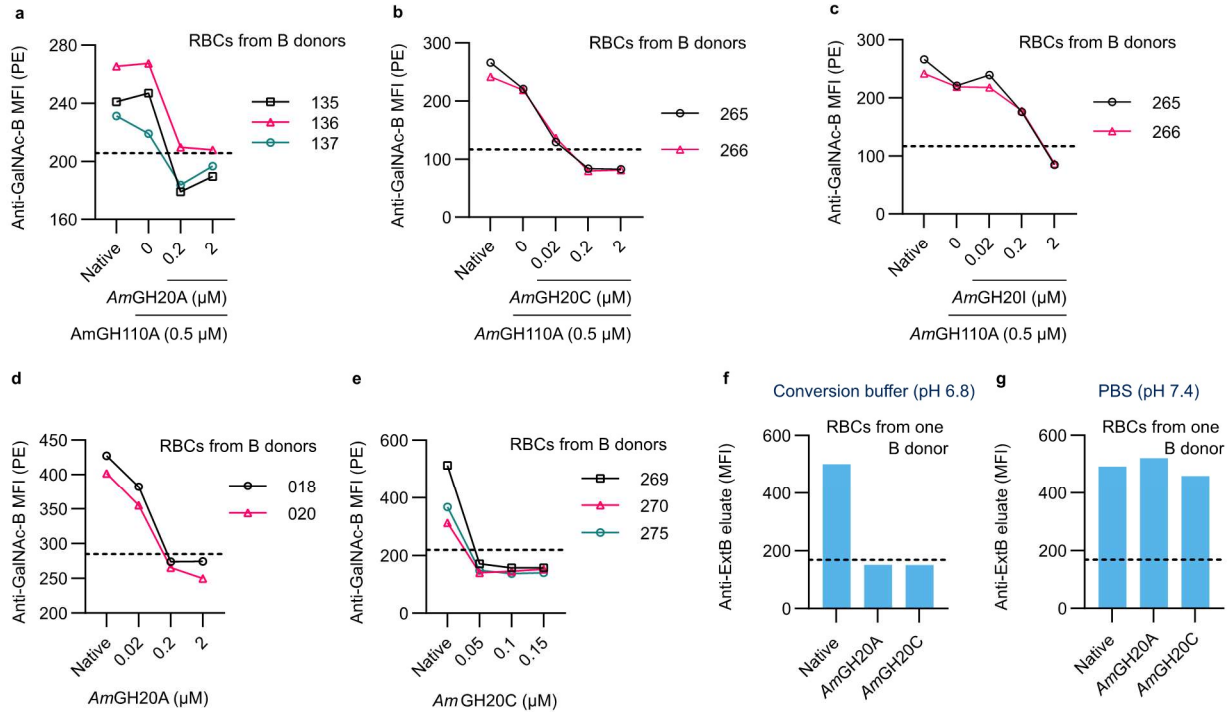
Supplementary Fig. 5: Phylogenetic tree of GH20 sequences from *A. muciniphila*. The phylogenetic tree was generated from blast results of all eleven GH20 enzymes from *A. muciniphila* and previously characterized GH20 β-N-acetyl-6-sulfo-glucosaminidase, endo-β-N-acetylglucosaminidase, and lacto-N-triose activities. The blast against the non-redundant database was performed using default setting, set to retrieve 500 maximum target sequences, which were filtered for 90-100 query coverage and 30-95% sequence identity. Sequences from all blast results were combined and further filtered for redundancy using CD-HIT with a 95% sequence identity cut-off, and the tree was generated from the remaining 1758 sequences using the neighbour-joining algorithm on the MAFFT server with bootstraps performed with 1000 iterations. Previously characterized enzymes are indicated with enzyme name, GenBank accession and substrates, where enzyme reactivity has been demonstrated. The different clusters are denoted with Roman numerals.



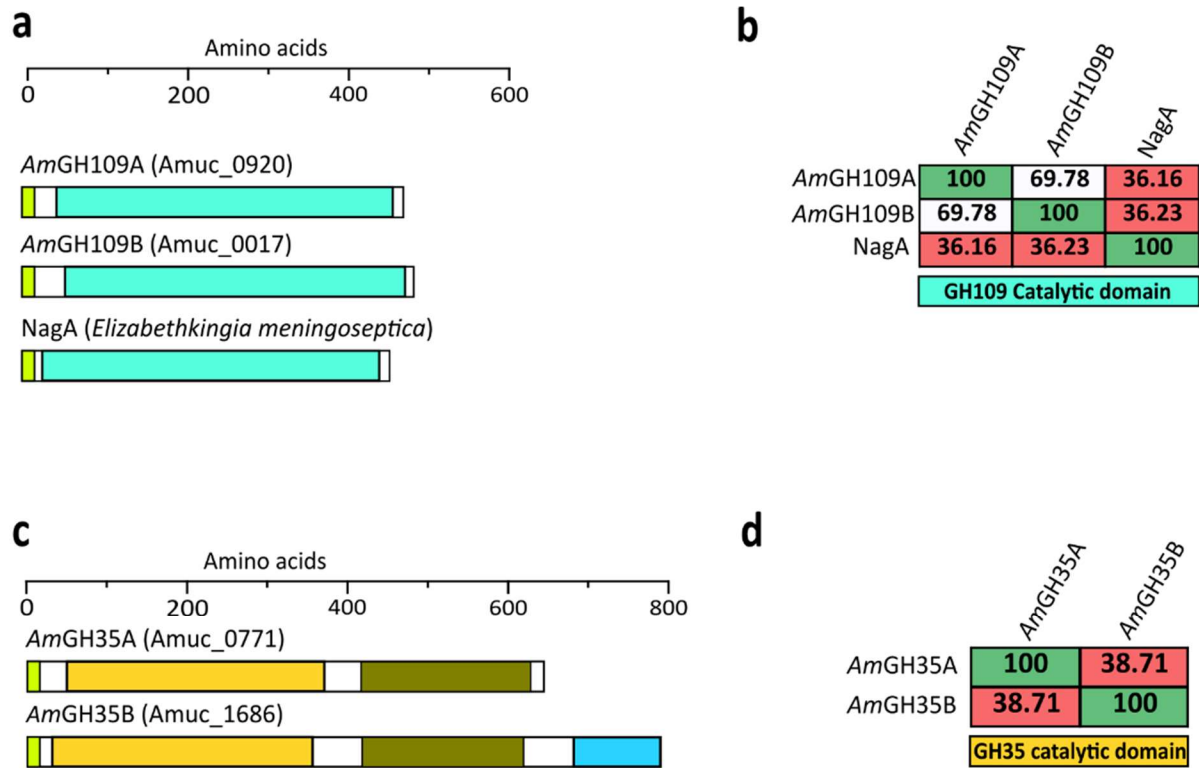
Supplementary Fig. 6: Activity of enzyme candidates towards B antigens on blood group B RBCs. Representative flow cytometry histograms of anti-B (clone 9621A8)/anti-RAM κ -Phycoerythrin (PE)-stained blood group B RBCs (blue filled) and native group O RBCs (dotted black line) as staining negative controls. **a** Native B RBCs and **b, c, d, and e** RBCs post treatment with $1\mu\text{M}$ of *AmGH110A*, *AmGH36A*, *AmGH36C* or *AmGH27*, respectively for 30 min at room temperature in conversion buffer (200mM glycine, 3 mM NaCl, pH 6.8). **f** Overview of the data in a-e. The bars are means of the median fluorescence intensity (MFI) of the PE-fluorescence signal of RBC aliquots from three donors ($n=3$). The hollow circles are the MFI from each donor. **g** Native group B RBCs. **h** and **i**, RBCs from the same donor as in **g**, after treatment with $0.12\mu\text{M}$ *AmGH110A* for 60 min at room temperature in conversion buffer and PBS (phosphate buffered saline), respectively. **j** Removal of B-antigens monitored by changes of MFI at different concentrations of *AmGH110A* over time and with RBCs from a single group B donor. **k** Removal of B-antigens monitored by changes of MFI at different concentrations of *AmGH110A* in 30min reactions with RBCs from three donors ($n=3$). The MFI on the y-axes in **j** and **k** is shown in logarithmic scales and the dotted line in each graph mark the MFI levels of the negative controls (group O RBCs). All the reactions have been performed with 38% haematocrit (volumetric ratio between RBCs and the source blood).



Supplementary Fig. 7: Activity of enzyme candidates towards the extended GalNAc-B antigen on group B RBCs. a Eleven GH20 candidates were evaluated in treatments of GalNAc-B carrying B RBCs from two donors (n=2), all at 1 μ M concentration for 30 min and 38% haematocrit in conversion buffer and at room temperature. The conversion was monitored by changes in mean median fluorescence intensity (MFI) of enzyme treated and anti-GalNAc-B-stained RBCs relative to native RBCs. **b** Representative flow cytometry histograms of anti-GalNAc-B-stained blood group B RBCs (blue filled) and group O RBCs (dotted black line) as negative staining controls. **c-e** The RBCs from the same donors in **b** after treatment with *AmGH20A*, *AmGH20C* and *AmGH20I*, respectively. **f** Overview data from three group B donors (n=3), including the example shown in panels **c-e**. The dotted line in panel **a** and **f** graphs shows the MFI level of group O RBC, the negative staining control. The data are shown as the mean MFI (blue bars) of three donors (n=3) (n=3) shown as open circles.



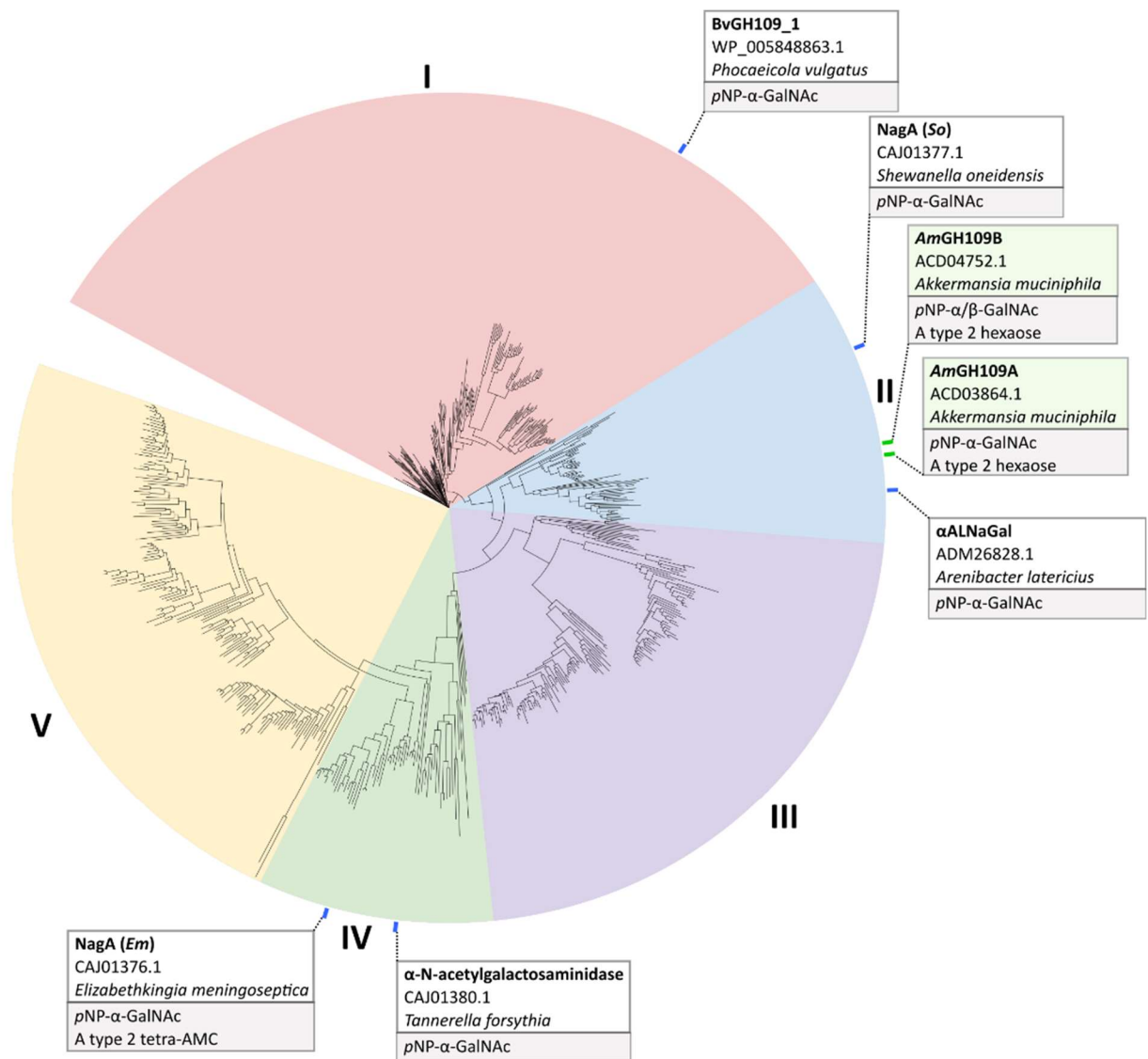
Supplementary Fig. 8: Effect of enzyme concentration and reaction buffer on removal of the GalNAc-B antigen on group B RBCs. **a-c** The conversion of the GalNAc-B antigen as a function of concentration of AmGH20A, AmGH20C and AmGH20I, respectively in one pot 30min reactions with the B-antigen-removing AmGH110A. **d,e** The effect of AmGH20A and AmGH20C concentration on the conversion of the GalNAc B antigen, respectively. **f,g** The removal of the GalNAc-B antigen on group B RBCs in conversion buffer and PBS, respectively, using AmGH20A (0.2μM) and AmGH20C (0.1μM). The graphs in show the mean MFI and the dotted lines mark MFI level of the negative staining control, O RBCs. Each scatter plot represents data from a single anonymous donor, denoted with the number of this donor.



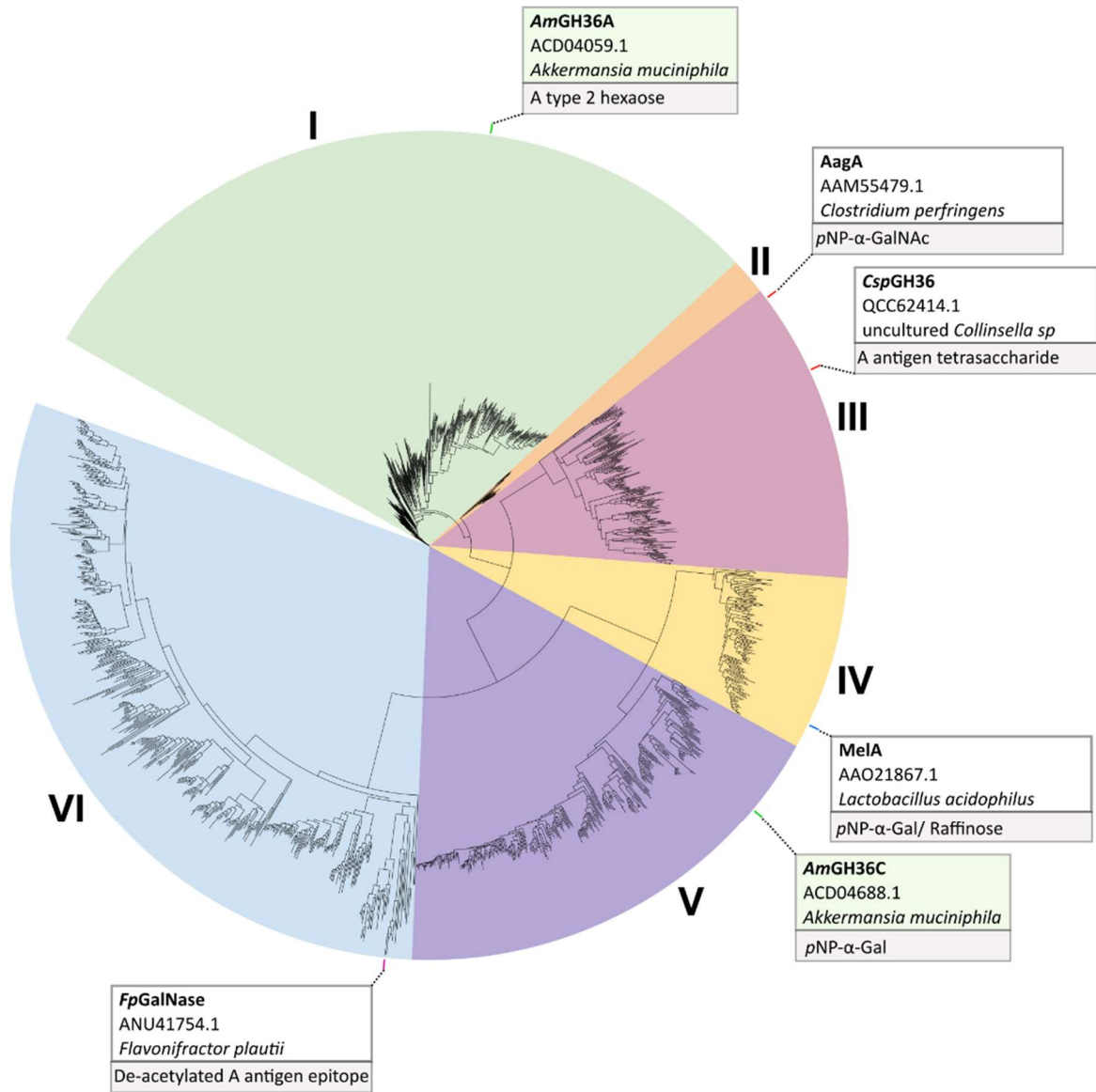
Legend

■ Signal peptide
 ■ GH109 catalytic domain
 ■ GH35 catalytic domain
 ■ Galactose binding-like domain
 ■ CBM32

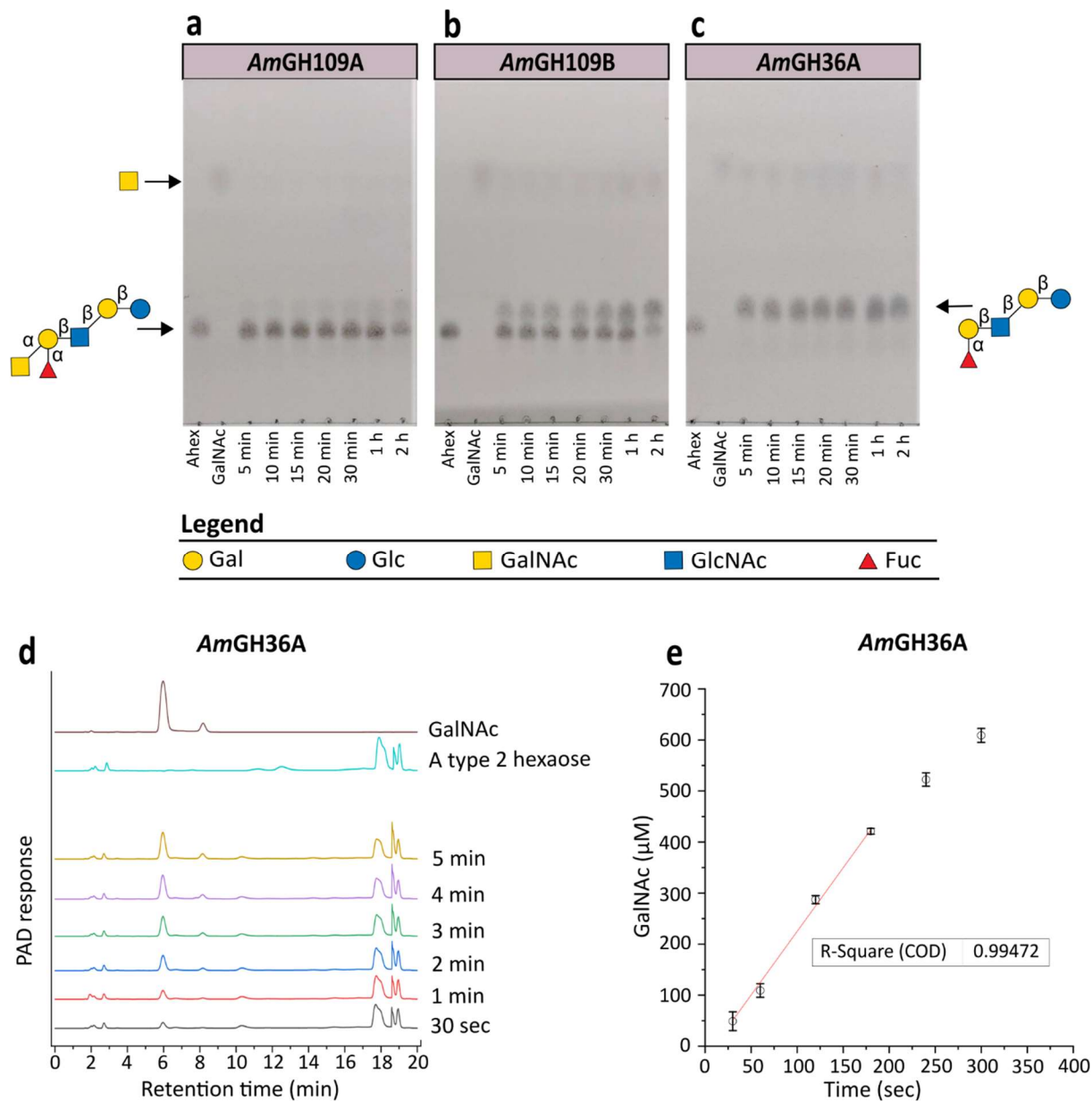
Supplementary Fig. 9: Modular organization and sequence divergence of enzyme candidates against A and extended Gal-A antigens. **a** Organisation of GH109 α -N-acetylgalactosaminidases from *A. muciniphila* GH109 and the previously characterized NagA from *Elizabethkingia meningoseptica*, which is active on the A antigen. **b** Amino acid sequence identity (%) matrix between the enzymes in a. **c** Domain organization of *A. muciniphila* GH35 enzyme candidates for the conversion of the extended Gal-A antigen. **d** Amino acid sequence identity (%) between the catalytic domain of the two GH35 candidates. Domain annotations are based on the CAZy, dbCAN, and Interpro databases (see Methods).



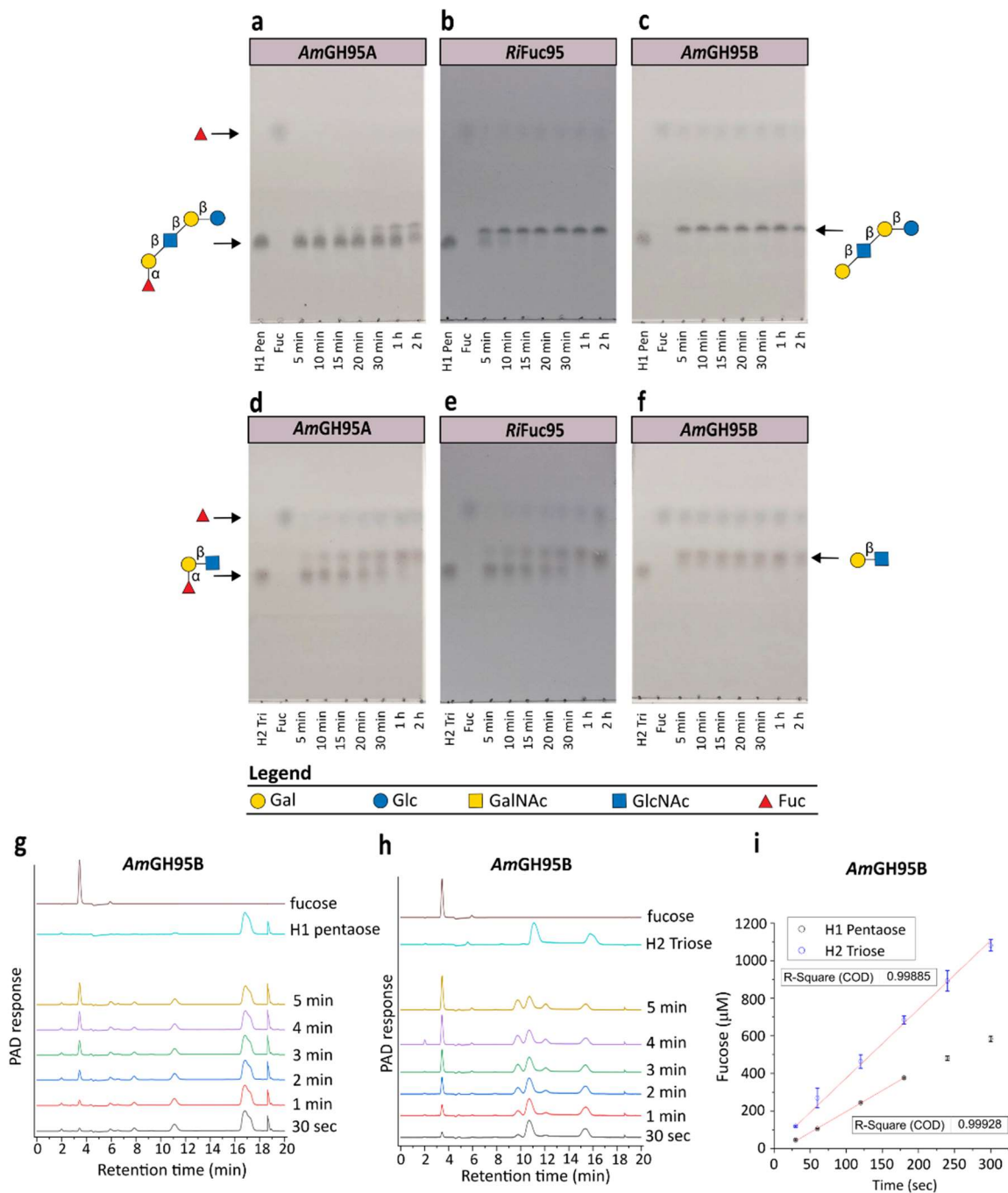
Supplementary Fig. 10: Phylogenetic tree of GH109 enzymes from the CAZy database. A phylogenetic tree generated from 2118 sequences of GH109, retrieved from the CAZy database (April 2023). The phylogenetic analysis is based on the multisequence alignment of GH109 catalytic domains from the CAZy database based on the catalytic domain of *AmGH109B* (residues 27-481) and filtered using CD-HIT at a 95% identity cut-off and MaxAlign with default settings. This resulted in the alignment of 646 sequences that were used to compute the phylogenetic tree using the neighbour-joining algorithm on the MAFFT server with bootstraps performed with 1000 iterations. The different clusters are denoted with Roman numerals. The α -N-acetylgalactosaminidase *AmGH109B* that catalyses the removal of the A-antigen, populates cluster II.



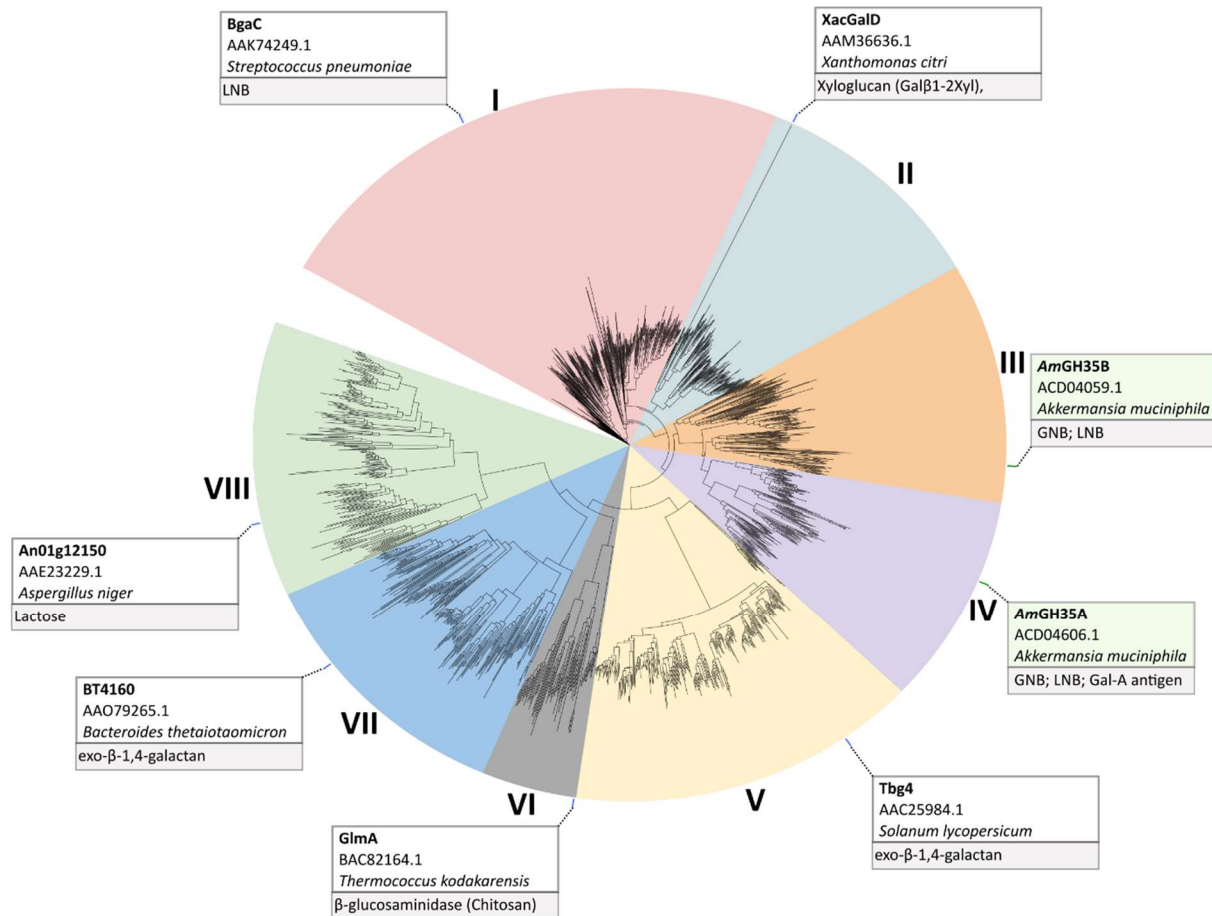
Supplementary Fig. 11: Phylogenetic tree of GH36 enzymes. The phylogenetic analysis is based on the multi-sequences that are retrieved by a blast search using the catalytic domains of AmGH36A, AmGH36C, AagA (*Clostridium perfringens*), MeIA (*Lactobacillus acidophilus*), and FpGalNase. The catalytic domain of AmGH36A (residues 385-749) was used as a query to retrieve the regions corresponding to the catalytic domains of orthologues in the blast search against the non-redundant database, which were then filtered for redundancy using CD-HIT with a 95% cut-off. This search resulted in the alignment of 1726 sequences that were used to compute the phylogenetic tree using the neighbour-joining algorithm on the MAFFT server with bootstraps performed with 1000 iterations. The different clusters are denoted with Roman numerals. The α -N-acetylgalactosaminidases populate clusters I-III, whereas the remaining specificities populate a different clade with clusters IV-VII.



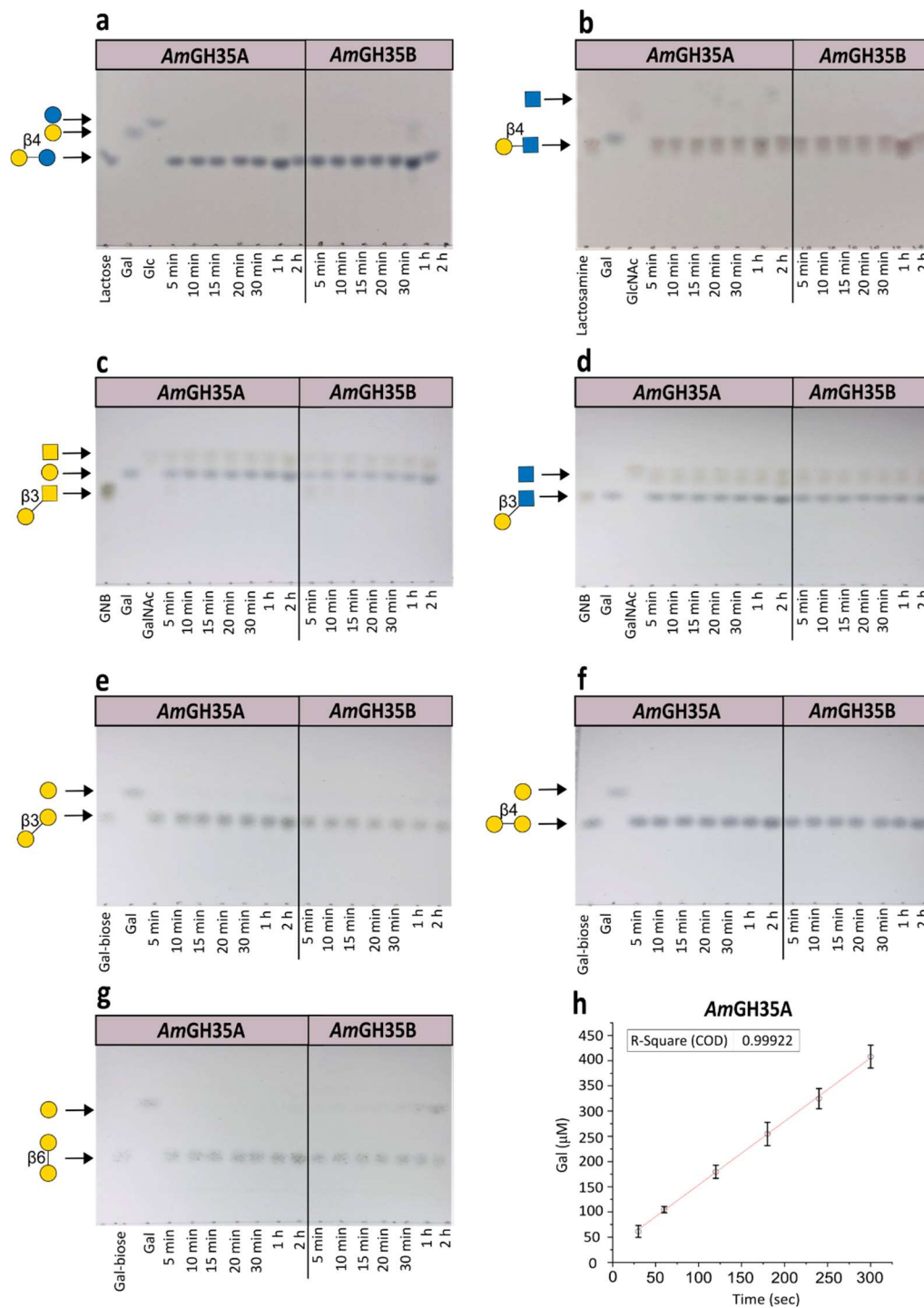
Supplementary Fig. 12: Activity of *A. muciniphila* enzymes against the A type 2 hexaose. **a-c** Time course of hydrolysis of the terminal nonreducing GalNAc from the A type 2 hexaose by *AmGH109A*, *AmGH109B* and *AmGH36A* respectively. The reactions were performed using 1 μM of each enzyme and 2 mM of each B antigen oligosaccharide in 20mM sodium phosphate, 150mM NaCl, 0.005% (v/v) Triton X-100, pH7.0 at 37°C. The TLC data are from two independent experiments showing similar results. **d** The HPAEC-PAD chromatograms showing GalNAc and A type 2 hexaose standards 1mM, and the release of GalNAc from the A type 2 hexaose at different incubation times with 20nM *AmGH36A* at 37°C. **e** The initial rate of *AmGH36A* against the A type 2 hexaose from the time course release of GalNAc in the linear range between 30-180 s using the area under the peak of a GalNAc standard curve at 6min elution time in HPAEC-PAD chromatograms. The data are means and standard deviations of three independent experiments (n=3).



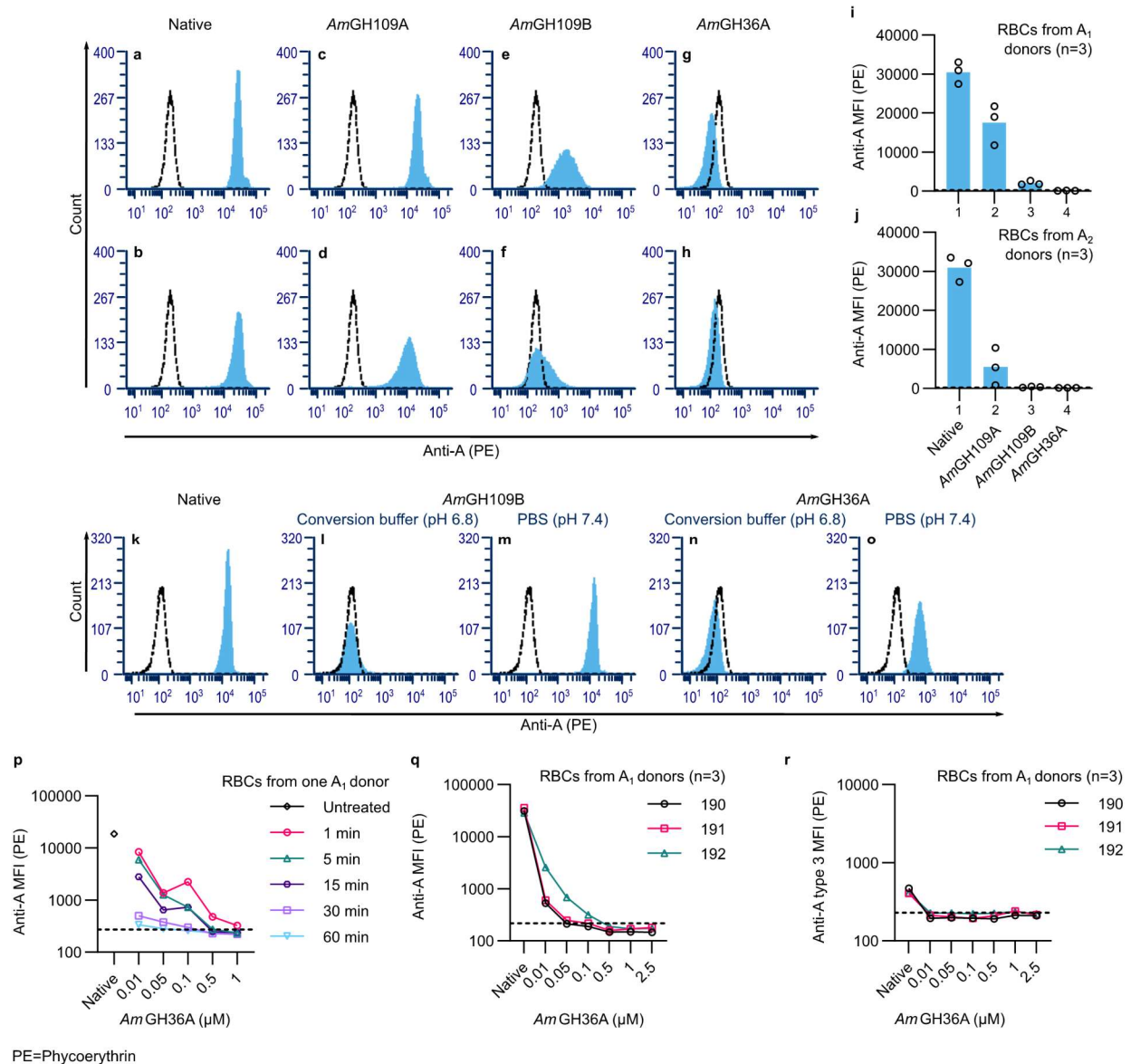
Supplementary Fig. 13: Activity of α -1,2-fucosidases against the H type 1 pentaose and H type 2 triose. **a-c** Time course defucosylation of the H type 1 pentaose with *AmGH95A*, *RiFuc95* and *AmGH95B*, respectively. **d-f** Reactions of enzymes in a-c, against the H type 2 triose. All reactions were performed using $1\mu\text{M}$ of each enzyme and 2mM of each H antigen oligosaccharide in 20mM sodium phosphate, 150mM NaCl, 0.005% (v/v) Triton X-100, pH 7.0 at 37°C . The TLC analysis was performed in two independent experiments, which showed similar results. **g-h** The HPAEC-PAD chromatograms showing fucose, H type 1 pentaose and H type 2 triose standards at 1mM , and the time course release of fucose from the H antigens after incubation with 20nM *AmGH95B* at 37°C . **i** The initial rate of *AmGH95B* was measured from the fucose time-course release using a standard curve generated from the area under the peak at 3.5 min elution time in the HPAEC-PAD chromatograms. The data are means and standard deviations of three independent experiments ($n=3$).



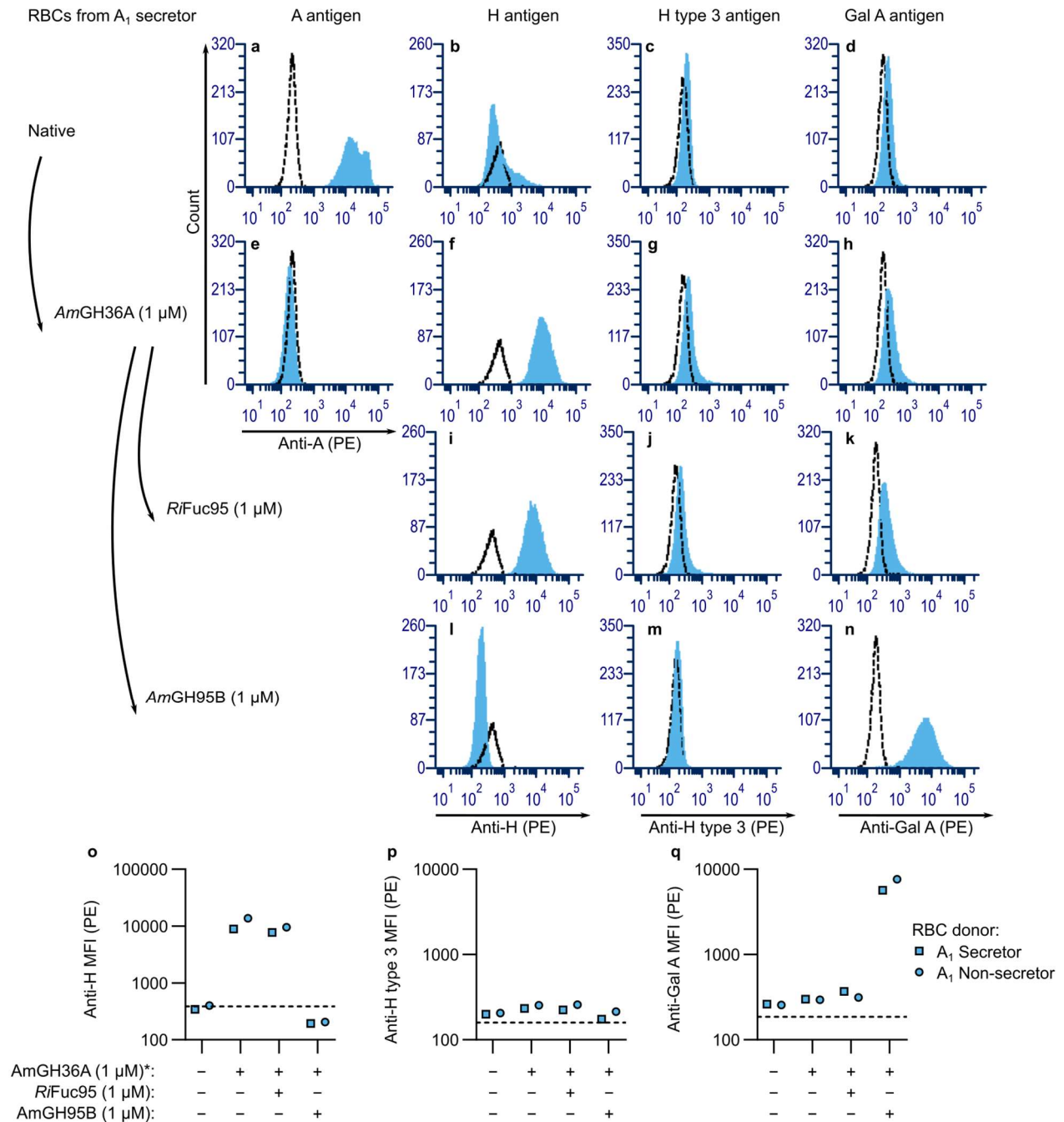
Supplementary Fig. 14: Phylogenetic tree of GH35 enzymes from the CAZy database. The phylogenetic analysis is based on 7902 sequences that were retrieved from the CAZy database (April 2023). The sequences were all trimmed to include only catalytic domains as retrieved using the *AmGH35A* catalytic domain (residues 26-379) as a query. The sequences were filtered for redundancy using CD-HIT with a 95% sequence identity cut-off, and further reduced using MaxAlign at default settings. This resulted in the alignment of 1763 sequences that were used to compute the phylogenetic tree using the MAFFT neighbour-joining algorithm with bootstraps performed with 1000 iterations. The different clusters are denoted with Roman numerals. Biochemically characterised enzymes and the substrates they have been shown to be active on are indicated. The β -1,3-galactosidase *AmGH35A* that is active on the extended Gal-A antigen populates cluster IV.



Supplementary Fig. 15: Activity profiles of the *A. muciniphila* GH35 enzymes towards galactobioses. The time course hydrolysis of **a** Lactose, **b** Lactosamine, **c** Galacto-*N*-biose (GNB), **d** Lacto-*N*-biose (LNB), **e** β -1,3-Galactobiose, **f** β -1,4-Galactobiose, and **g** β -1,6-Galactobiose, by *AmGH35A* and *AmGH35B* (1 μ M of each enzyme, 20mM sodium phosphate, 150mM NaCl, 0.005% (v/v) Triton X-100, pH7.0, 37°C). The TLC data are representative sets from two independent experiments, which yielded similar outcomes. **h** The initial rate of *AmGH35A* was determined from the time course release of galactose from GNB based on a standard curve generated of galactose (0-500 μ M) using the Megazyme L-arabinose/D-galactose kit. The shown data points are means and standard deviations of three independent experiments (n=3).

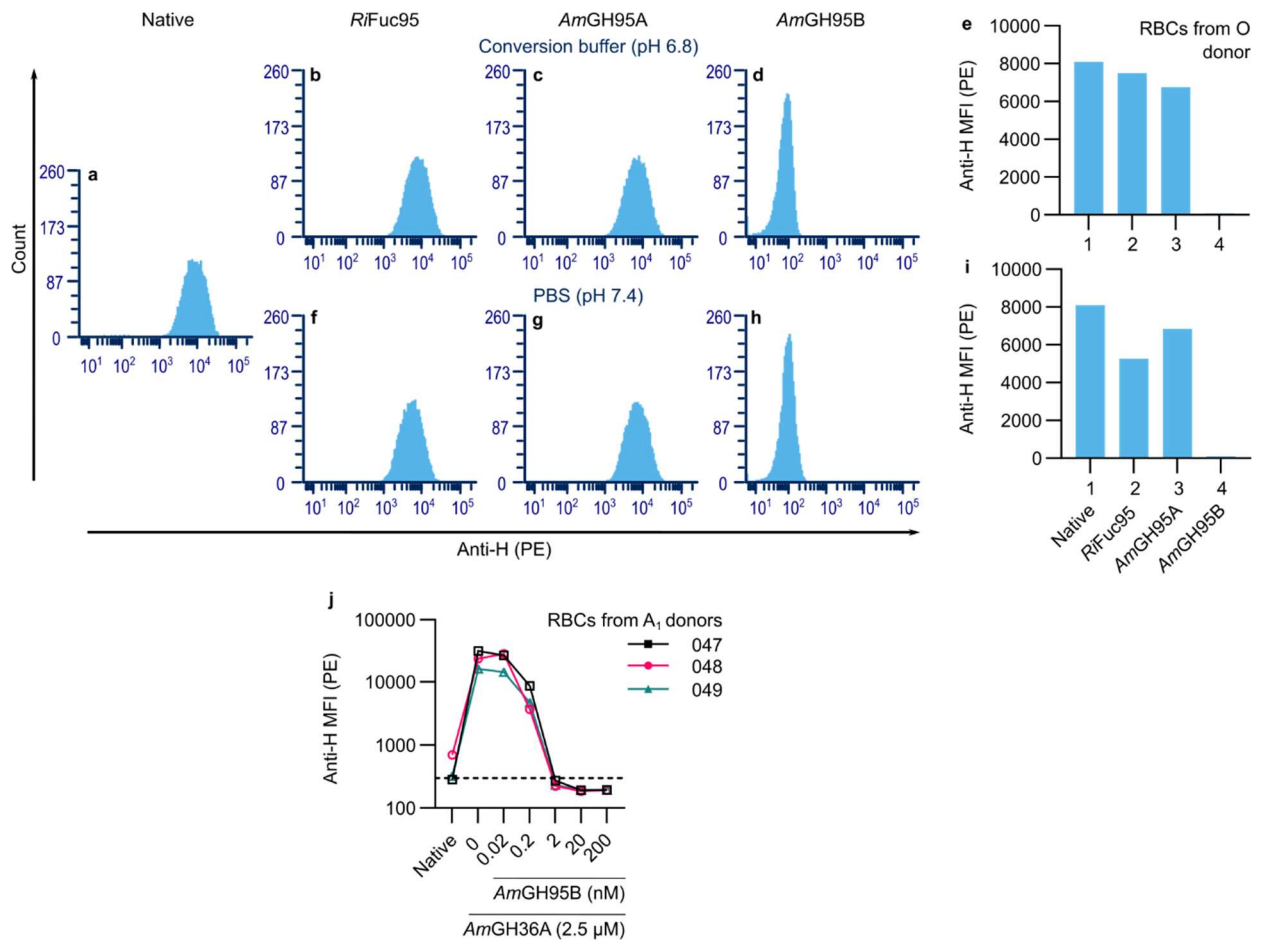


Supplementary Fig. 16: Activity of *A. muciniphila* enzymes towards A antigens on group A RBCs. **a,b** Representative histograms (blue) of immuno-A-stained RBCs from A₁ and A₂ donors, respectively, or the same after a 30min incubation at room temperature with 1 μM of **c,d** *AmGH109A*; **e,f** *AmGH109B*, or **g,h** *AmGH36A*, all compared to negative staining control O RBCs (dotted line histograms). **i,j** An overview of the data from three donors (n=3) of A₁ and A₂ phenotype, respectively, based on the median fluorescence intensity (MFI). The bars are the MFI means from three donors, represented as open circles. **k** Histograms of A-stained RBCs (blue), and negative staining control (dotted line histogram), **l-o** RBCs from the same donors in **k**, post 60 min treatment with *AmGH109B* (8μM) or *AmGH36A*(1.1μM) in PBS or conversion buffer, respectively. **p** Levels of A antigens of A₁ RBCs after different incubation times and concentrations of *AmGH36A*. **q** Levels of A antigens on A₁ RBCs from three more donors after incubation with different *AmGH36A* concentrations for 30min. **r** Levels of the extended A type 3 antigen on the same RBCs as in **q**. In graphs **p-q** the y-axis has logarithmic scales starting at 100. The dotted horizontal line in each graph is the MFI level of O RBCs as negative control. The A-antigen levels were assayed by flow cytometry after staining with anti-A monoclonal antibodies (ES-15) and for anti-A type 3 the monoclonal antibody TH1. For both, a secondary antibody rat anti-mouse-κ conjugated to Phycoerythrin (PE) was used.

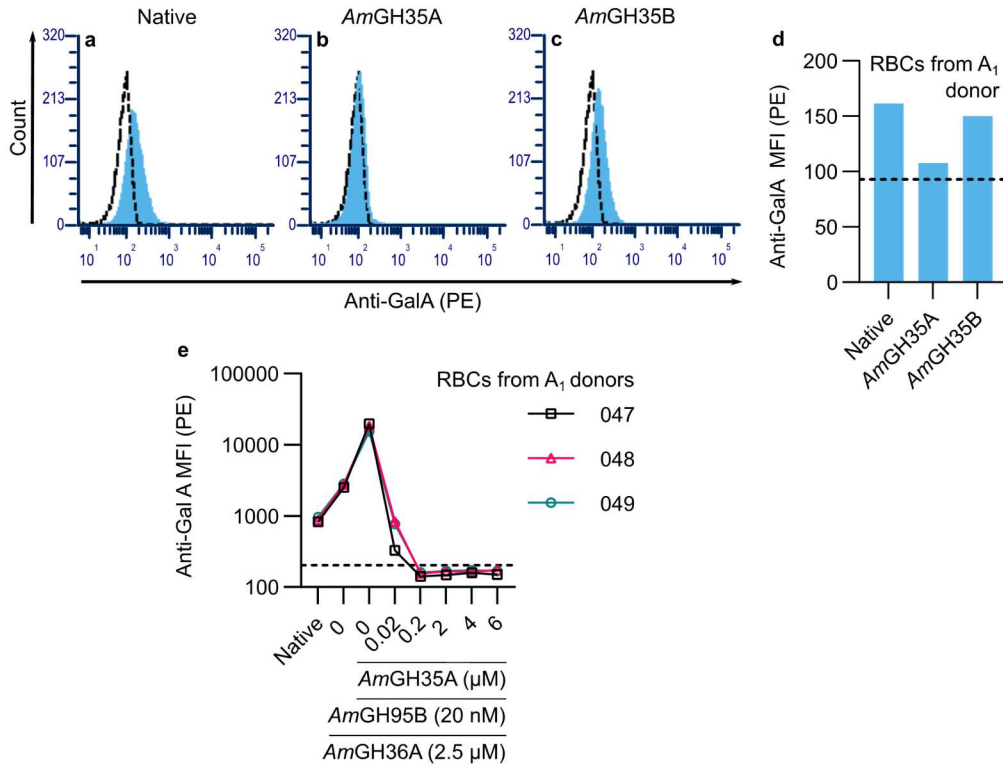


*Pretreatment

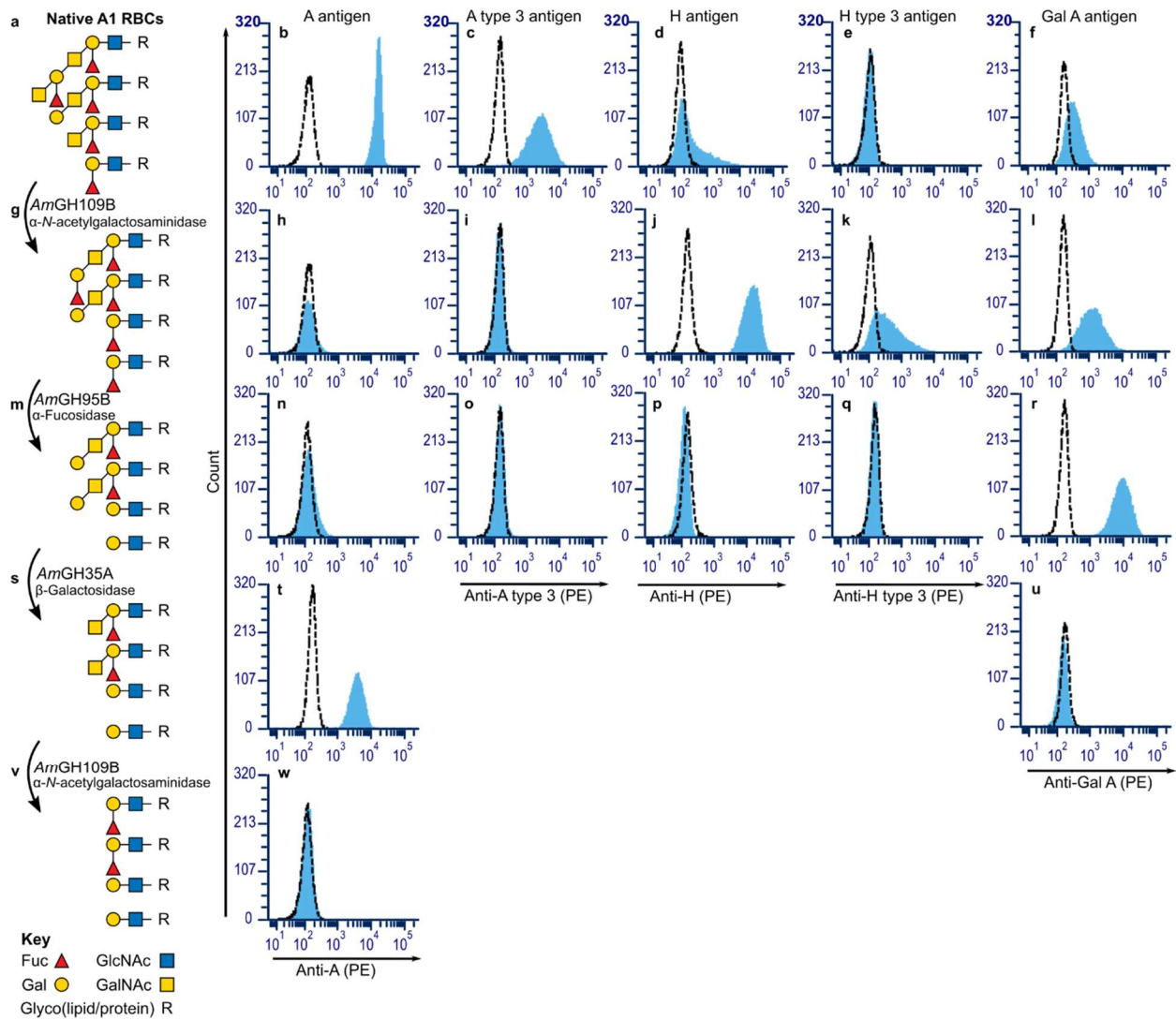
Supplementary Fig. 17: Activity of the α -1,2-fucosidases RiFuc95 and AmGH95B towards H and H type 3 antigens on group A RBCs. a-d Representative blue histograms depict levels of A, H, H type 3, and Gal-A antigens, respectively, on untreated A₁ RBCs. Dotted line histograms are of O or Bombay RBCs as negative controls. e-h The RBCs of the same donor as in a-d after treatment with AmGH36A (in conversion buffer, 30 minutes at room temperature). i-k The cells in f-h after treatment of RiFuc95. l-n The cells in f-h after treatment of AmGH95B. o-q Comparison of the α -1,2-fucosidase on A₁ RBCs between a secretor and nonsecretor donor phenotype background.



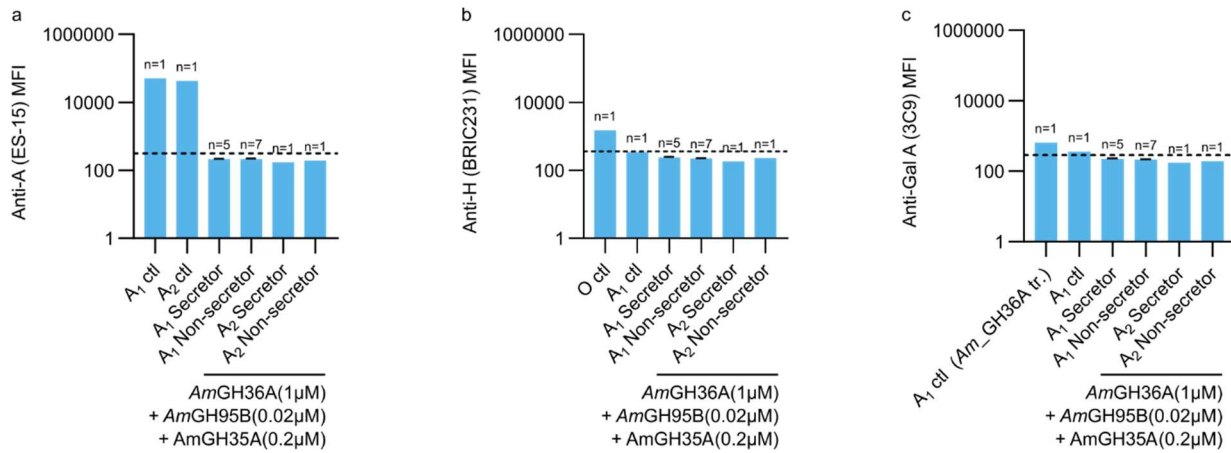
Supplementary Fig. 18: Dependence of α -1,2-fucosidase activity towards H antigens on RBCs on buffer and enzyme concentration. **a** Representative histograms of native group O RBCs stained with anti-H (BRIC231), and RAMk-Phycoerythrin (PE). **b-d** are RBCs from the same donor after treatment with of *Ri*GH95, *Am*GH95A and *Am*GH95B (each at 1.2 μ M) in conversion buffer during 60min incubations at room temperature, respectively. **e** Overview of the data in a-d showing the median fluorescence intensity (MFI). **f, g** and **h** The same as b, c, and d, but the reactions are performed in PBS. **i** Overview graph of the MFI in a, f-h. **j** The activity of *Am*GH95B after 60min incubations towards H antigens at different concentration in the presence of *Am*GH36A.



Supplementary Fig. 19: Evaluation of GH35 β -galactosidases towards the extended Gal-A antigen. **a** Representative histogram (blue) of the Gal-A antigen of A₁ RBCs from a single donor, compared with negative control O RBCs (dotted black line histogram), stained with anti-T/Gal-A (clone 3C9) followed by RAMk-Phycoerythrin (PE). **b** and **c** RBCs from the same donor after treatment with *AmGH35A* (5.8 μ M) and *AmGH35B* (7.2 μ M), respectively, during 60min incubations in conversion buffer at room temperature. **d** Overview of the data b and c showing the median fluorescence intensity (MFI). The dashed black line represents the MFI level noted in the negative group O control RBCs. **e** The activity of *AmGH35A* at different concentrations in the presence of *AmGH36A* and *AmGH95B* after incubation for 30min. For comparison, RBCs only with *AmGH36A* and the one-pot treatment with *AmGH36A* and *AmGH95B* are included. Note the y-axis have logarithmic scales starting at 100.



Supplementary Fig. 20: Sequential treatment of A₁ RBCs with AmGH109B, AmGH95B and AmGH35A. The left column (panels a, g, m, s, and v) depicts the schematic glycan structures of type 2 antigens on the RBC surfaces, and which are monitored in the representative histograms of A₁ RBCs that follow in each row. **b-f** Histograms show native RBCs or post-treatment with **h-l**, AmGH109B, **n-r**, AmGH95B, **t-u** AmGH35A and **w** AmGH109B second treatment. RBCs were immunostained, targeting the antigen of interest after each treatment. The blue histograms show native and enzyme-treated A₁ RBCs while dotted line display negative controls, which are group O RBCs for anti-A, Anti-A type 3, Anti-H type 3 and Anti-Gal-A, and Bombay-RBCs for anti-H.



Supplementary Fig. 21: One-pot treatment of A₁ RBCs with AmGH36A, AmGH95B and AmGH35A. RBCs from five A₁ secretors, seven A₁ nonsecretors, one A₂ secretor, and one A₂ nonsecretor. **a-c** The levels of A, H antigens or Gal-A antigens after one-pot treatments with AmGH36A, AmGH95B and AmGH35A. The analysis shows no detectable levels of all the three analysed antigens after the enzymatic treatment. The bars in the graphs are the mean MFI and the error bars are the standard errors of the mean (SEM). The A₁ cells that were initially treated with AmGH36A to expose the Gal-A antigen were included as a control (ctl) in panel c. The dashed black lines represent the MFI levels noted in the negative control RBCs for each staining (group O for a and c; Bombay for b).

		Native	<i>AmGH110A</i> (0.5 μ M)										<i>AmGH110A</i> (0.5 μ M) + <i>AmGH20C</i> (0.4 μ M)										<i>AmGH110A</i> (0.5 μ M) + <i>AmGH20A</i> (0.4 μ M)												
		1	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10			
B secretor RBCs	a	3+	1+	-	-	-	-	1+	-	-	-	-	1+	-	-	-	(-)	1+	-	-	-*	-	(-)	-	-	-	-	2+	-	-	-	-	-		
	b	4+	-	1+	(-)	-	-	-	-	-	-	-	-	1+	-	-	-	-	-	-	-	-	-	-	1+	-	-	-	-	-	-	-	-	-	
	c	4+	-	1+	-	-	2+	-	-	1+	(1+)	-	-	(1+)	-	-	1+	-	-	(1+)	-	-	-	-	-	-	(-)	-	-	(1+)	-	-	-	-	
	d	4+	-	-	-	(-)	-	-	-	-	-	1+	-	-	-	-	-	-	-	-	-	(-)	(-)	-	-	-	-	-	-	-	-	-	-	-	
	e	4+	-	1+	-	-	(-)	-	-	-	-	2+	-	(-)	1+	-	-	-	-	-	-	-	1+	-	-	-	(-)	-	(-)	-	-	-	-	-	
	f	3+	-	-	-	1+	-	-	-	2+	2+	-	-	-	-	(-)	(1+)	-	-	2+	-	-	-	-	-	-	(1+)	-	-	(1+)	-	-	-	-	
	g	4+	-	-	(-)	-	1+	-	-	-	-	-	-	-	-	(1+)	(-)	-*	-	-	-	-	-	-	-	-	(-)	-	-	-	-	-	-	-	
	h	3+	2+*	2+	-	(-)	-	-	-	-	-	1+	-	1+	1+	-	-	-	-	-	-	-	(1+)	-	(1+)	1+	-	-	-	-	-	-	1+	-	
	i	3+	-	-	-	-	-	-	(-)	-	(1+)	1+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(1+)	-	-	
	j	4+	-	-	1+	-	-	-	-	-	-	1+	-	-	(1+)	-	-	-	-	-	-	-	(1+)	-	-	-	-	-	-	-	-	-	(-)	-	-
B non-secretor RBCs	a	4+	1+	-	-	-	-	1+	-	-	-	-	1+	-	-	(-)	(-)	1+	-	-	-	-	(-)	-	-	-	1+	-	-	-	-	-	-	-	
	b	4+	-	1+	(-)	-	-	-	-	(-)	-	-	-	1+	-	-	(-)	-	(-)	-	-	-	-	-	1+	-	(-)	-	-	-	-	-	-	-	-
	c	4+	-	(+)	-	-	2+	-	-	2+	-	-	-	(-)	-	1+	-	-	1+	-	-	-	-	-	-	-	-	-	-	(-)	-	-	-	-	-
	d	4+	-	-	-	-	-	-	-	-	1+	-	-	(-)	-	-	-	-	-	-	-	(-)	-	-	-	-	-	-	-	-	-	-	-	-	-
	e	4+	-	1+	-	-	-	-	-	-	1+	-	-	(1+)	(-)	-	-	-	(-)	-	(-)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	f	4+	-	-	-	1+	-	-	-	2+	1+	-	-	(-)	(-)	-	-	-	1+	(-)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	g	4+	-	-	1+	-	1+	-	-	-	-	-	-	-	1+	(+)	-	(-)	(-)	-	-	-	-	-	-	(-)	-	(-)	-	-	-	-	-	-	-
	h	3+	2+*	2+	-	-	-	-	-	-	-	1+	-	1+	1+	-*	-	-	-	-	-	-	1+	-	-	-	-	-	-	-	-	(1+)	-	-	-
	i	3+	-	-	-	-	-	-	-	-	-	2+	-	(-)	(-)	-	-	-	-	-	-	(-)	2+	-	-	-	-	-	-	-	(-)	2+	-	-	-
	j	4+	-	-	1+	-	-	-	-	-	-	1+	-	-	(-)	(-)	-	-	-	-	-	(1+)	-	-	-	-	-	-	-	-	-	(-)	-	-	-

* = Fibrinogen observed in crossmatch

Supplementary Fig. 22: Crossmatch reactivity of O plasmas towards enzymatically converted B RBCs. The crossmatch data are generated for a single treatment with *AmGH110A* or a one pot treatment of RBCs with *AmGH110A* (0.5 μ M) and either *AmGH20A* or the *AmGH20C* (both at 0.4 μ M). The enzymatic treatments were performed for 30min in conversion buffer. The colour-scale (red-pink-white) corresponds to the strength of crossmatch reactivity against O-plasma samples (n=100).

		AB plasma donors (n=11)										A plasma donors (n=12)												
		A ₁ B					A ₂ B					A ₁						A ₂						
		213	214	215	216	217	218	220	219	221	223	224	199	201	202	203	204	205	206	207	208	209	200	210
B secretor RBCs	Native	-	(-)	-	-	-	-	-	(-)	-	-	-	3+	2+	1+	3+	4+	3+	3+	3+	3+	3+	4+	3+
	<i>AmGH110A</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1+	-
	<i>AmGH110A + AmGH20A</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(-)	-	-	-	-	-	(-)	-
	<i>AmGH110A + AmGH20C</i>	-	(-)	-	-	-	-	-	(-)	-	-	-	(-)	(-)	(-)	(-)	-	(-)	-	-	-	-	(-)	-
B non-secretor RBCs	Native	-	-	-	-	-	-	-	-	-	-	-	3+	2+	1+	4+	4+	3+	3+	3+	3+	3+	3+	3+
	<i>AmGH110A</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(-)	-
	<i>AmGH110A + AmGH20A</i>	-	(-)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	<i>AmGH110A + AmGH20C</i>	-	(-)	-	(-)	-	-	-	-	-	-	-	-	-	-	-	-	(-)	-	-	-	-	(-)	-

Supplementary Fig. 23: Crossmatch reactions of AB and A plasmas towards enzymatically converted B RBCs. The crossmatches of AB plasmas (n=11) and B plasmas (n=12) towards B RBCs treated with only *AmGH110A* (0.5µM) or a combination of *AmGH110A* (0.5µM) and either *AmGH20A* or *AmGH20C* (both at 0.4µM). The colour-scale (red-pink-white) corresponds to crossmatch reactivity levels.

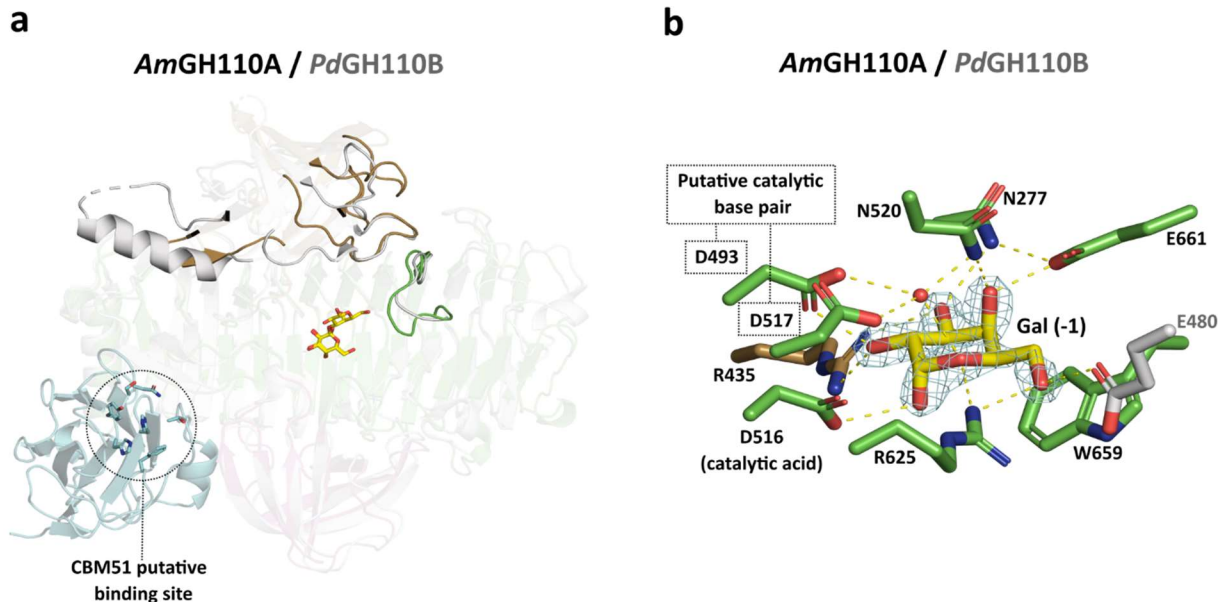
	Plasma donor														
	O			A ₂ B		A ₁ B			A ₁		A ₂		B		
	P136	P137	P138	P139	P140	P141	P142	P143	P144	P145	P146	P147	P148	P149	
<i>AmGH36A</i>	4+	4+	4+	-	-	-	-	-							
<i>AmGH36A -> AmGH95B</i>	4+	2+	(+)	-	-	-	-	-							
<i>AmGH36A -> AmGH95B -> AmGH35A</i>	3+	-	2+	2+	(+)	-	-	-							
<i>AmGH36A -> AmGH95B -> AmGH35A -> AmGH36A</i>	Cells not available														
<i>AmGH109A</i>	2+	-	-	-	-	-	-	-	-	-	-	-	-	2+	
<i>AmGH109A -> AmGH95B</i>	4+	2+	(+)	-	-	-	-	-	(+)						
<i>AmGH109A -> AmGH95B -> AmGH109A</i>	3+	(-)	2+	1+	-	-	-	-							
<i>AmGH109A -> AmGH95B -> AmGH35A -> AmGH109A</i>	Cells not available														
	3+	-	(+)	-	-	-	-	-	-	-	-	-	-	2+	
<i>AmGH95B</i>	-	-	-	-	-	-	-	-							
<i>AmGH35A</i>	-	-	-	-	-*	-	-	-							
<i>AmGH95B -> AmGH35A</i>	-	-	(+)	-	(+)	-	-	-							
<i>AmGH95B</i>	4+	4+	4+	-	-	-	-	-							
<i>AmGH35A</i>	4+	4+	4+	-	-	-	-	-							
<i>AmGH95B -> AmGH35A</i>	4+	4+	4+	-	-	-	-	-							

Supplementary Fig. 24: Crossmatch reactivity of donor plasmas of different types towards enzymatically converted A₁ RBCs. The crossmatches were performed with A₁ RBCs cells or O RBCs (green font denotes the enzymes treatments of group O RBCs) after treatment with single treatment with *AmGH36A* (1μM) or sequential treatments with *AmGH36A* (1μM), *AmGH95B* (0.02μM) and *AmGH35A* (0.2μM) as denoted by the arrows (->). All incubations were performed for 30min in conversion buffer. The colour-scale (red-pink-white) corresponds to the strength of the crossmatch reactivity.

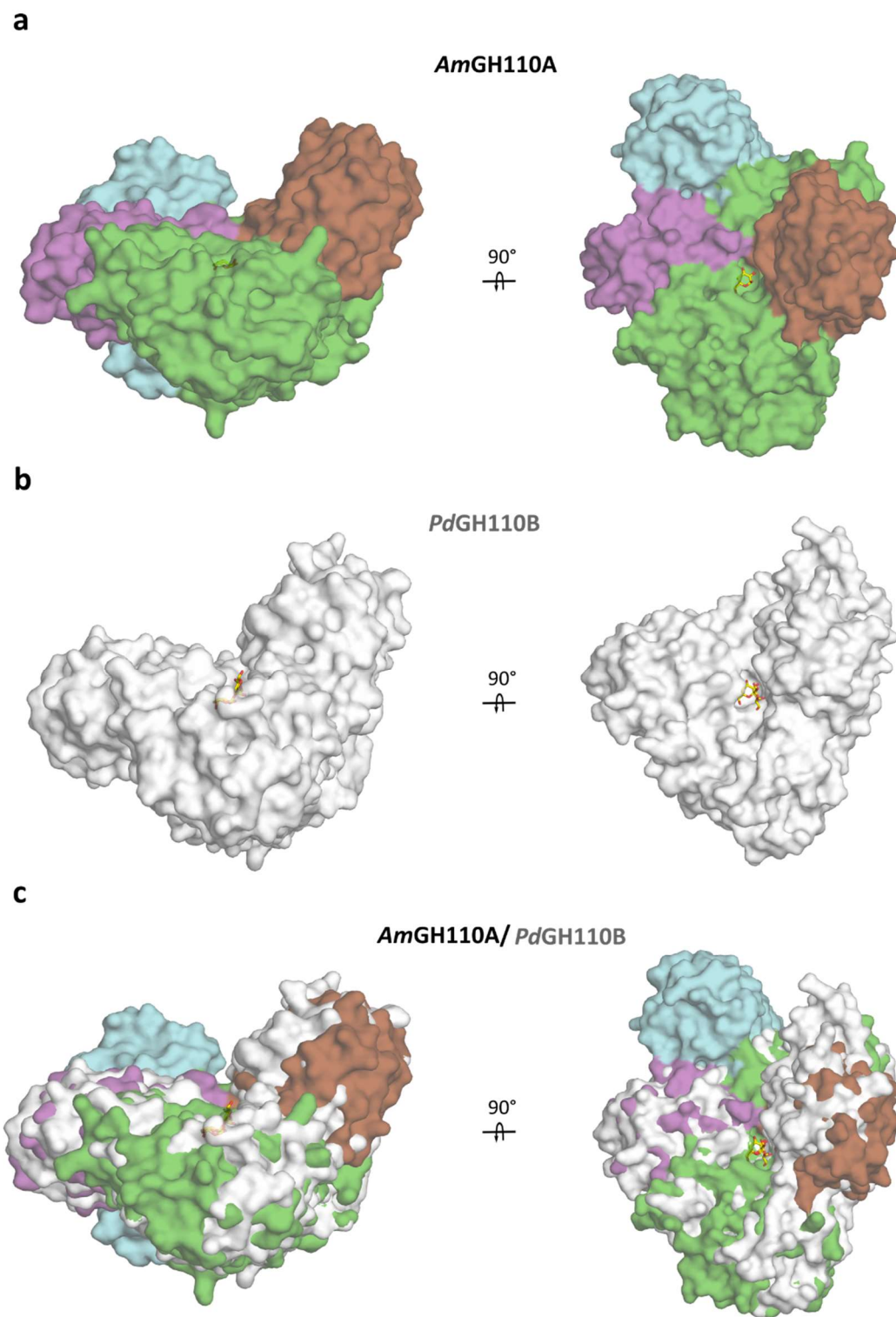
		Native										<i>AmGH36A</i> (1 μ M)										<i>AmGH36A</i> (1 μ M) + <i>AmGH95B</i> (20 nM) + <i>AmGH35A</i> (0.2 μ M)												
		1	1	2	3	4	5	6	7	8	9	10	1	1	2	3	4	5	6	7	8	9	10	1	1	2	3	4	5	6	7	8	9	10
A ₁ secretor RBCs	a	3+	-	-	3+	2+	-	3+	-	3+	2+	2+	-	2+	(-)	(-)	1+	2+	-	3+	1+*	2+	-	2+	(-)	(-)	1+	2+	-	3+	1+*	2+		
	b	3+	2+	3+	3+	-	3+	-	3+	2+	3+	(1+)	-	3+	1+	-	3+	(1+)	3+	-	3+	-	3+	-	-	3+	1+	-	3+	(1+)	3+	-	3+	-
	c	4+	3+	2+	3+	1+	3+	-	3+	1+	3+	1+	2+	1+	2+	-	2+	-	3+	1+	3+	-	3+	-	2+	1+	2+	-	2+	-	3+	1+	3+	-
	d	4+	1+	3+	2+	3+	3+	-	2+	3+	2+	3+	-	-	2+	(1+)	(1+)	-	(-)	-	1+	-	1+	-	-	-	2+	(1+)	(1+)	-	(-)	-	1+	-
	e	4+	3+	3+	3+	3+	-	3+	2+	3+	3+	2+	3+	3+	(-)	1+	-	-	-	-	(1+)	2+	3+	3+	(-)	1+	-	-	-	-	(1+)	2+		
	f	4+	2+	3+	1+	3+	1+	2+	1+	3+	3+	3+	-	(1+)	-	(1+)	(-)	2+	m	2+	3+	2+	-	(1+)	-	(1+)	(-)	2+	m	2+	3+	2+		
	g	4+	1+	-	-	-	2+	1+	1+m	1+	3+	2+	1+	-	-	-	2+	-	m	(1+)	-	1+	1+	-	-	-	2+	-	m	(1+)	-	1+		
	h	4+	3+	2+	3+	(1+)	2+	-	1+	3+	3+	-	2+	-	3+	-	-	-	1+	1+	3+	-/m	2+	-	3+	-	-	-	1+	1+	3+	-/m		
	i	3+	-	2+	2+	3+	-	3+	3+	3+	1+	(1+)	-	-	1+	-	-	3+	-	4+	-	-	-	-	1+	-	-	3+	-	4+	-	-		
	j	4+	-	(1+)	3+	2+	1+	2+	3+	3+	3+	2+	-	-	2+	-	-	-	-	3+	3+	(-)	-	-	2+	-	-	-	-	3+	3+	(-)		
A ₁ non-secretor RBCs	a	3+	-	2+	3+	2+	-	3+	-	3+	2+	2+	-	3+	(1+)	-	1+	2+	-	3+	1+*	2+	-	3+	(1+)	-	1+	2+	-	3+	1+*	2+		
	b	3+	2+	3+	2+	-	3+	-	3+	2+	3+	(1+)*	-	3+	1+	-	3+	-	3+	-	3+	-	3+	-*	-	3+	1+	-	3+	-	3+	-	3+	-*
	c	4+	3+	1+	3+	1+	3+	-	3+	2+	3+	1+*	-	1+	1+	-	2+	-	3+	2+	3+	-*	2+	-*	-	1+	1+	-	2+	-	3+	2+	3+	-*
	d	4+	2+	3+	1+	3+	3+	-	3+	2+	2+	3+	-	-	(1+)	(1+)	(-)	-	-	-*	2+	-	-	-	(1+)	(1+)	(-)	-	-	-*	2+	-		
	e	4+	3+	3+	3+	3+	-	3+	2+	3+	3+	2+	3+	3+	(-)	1+	-	-	-	-	-	1+	3+	3+	(-)	1+	-	-	-	-	-	1+		
	f	4+	2+	3+	1+	3+	2+	3+	2+*	3+	3+	3+	-	(1+)	-	-	-	3+	(1+) m	2+*	3+	1+	-	(1+)	-	-	-	3+	(1+) m	2+*	3+	1+		
	g	4+	1+	-	-	-	2+	2+	m	-*	3+	2+	1+	-	-	-	2+	-	m	-*	-	1+*	1+	-	-	-	2+	-	m	-*	-	1+*		
	h	4+	3+	2+	3+	-	2+	-	1+*	3+	3+	m	2+	-	3+	-	-	1+*	2+*	2+	2+	-*	2+	-	3+	-	-	1+*	2+*	2+	2+	-*		
	i	3+	-	1+	2+	3+	-	3+	3+	4+	1+	-	-	-	1+	-	-	3+	-*	3+	-	-	-	-	1+	-	-	3+	-*	3+	-	-		
	j	4+	-	2+	3+	2+	-	3+	3+	3+	3+	2+	-	-	1+	-	-	-	-*	3+	2+	-	-	-	1+	-	-	-	-*	3+	2+	-		

* = Fibrinogen observed in the cross match
m = mixed field

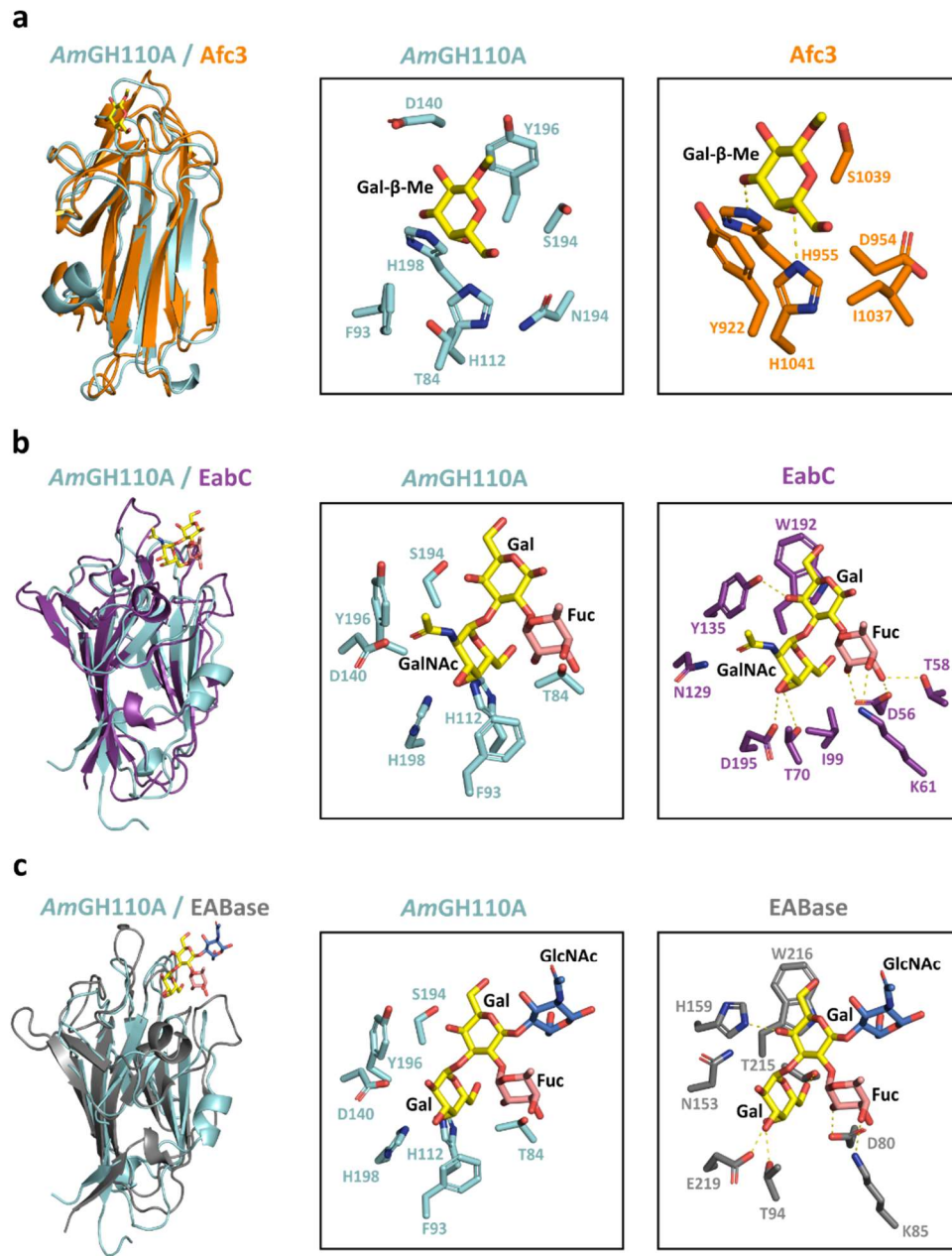
Supplementary Fig. 25: Crossmatch reactivity of O plasmas towards enzymatically converted A₁ RBCs. The crossmatches were performed before and after treatment with *AmGH36A* (1 μ M) or a one-pot treatment with *AmGH36A* (1 μ M), *AmGH95B* (0.02 μ M) and *AmGH35A* (0.2 μ M). The colour-scale (red-pink-white) corresponds to the strength of the crossmatch reactivity towards the O-plasma samples (n=100).



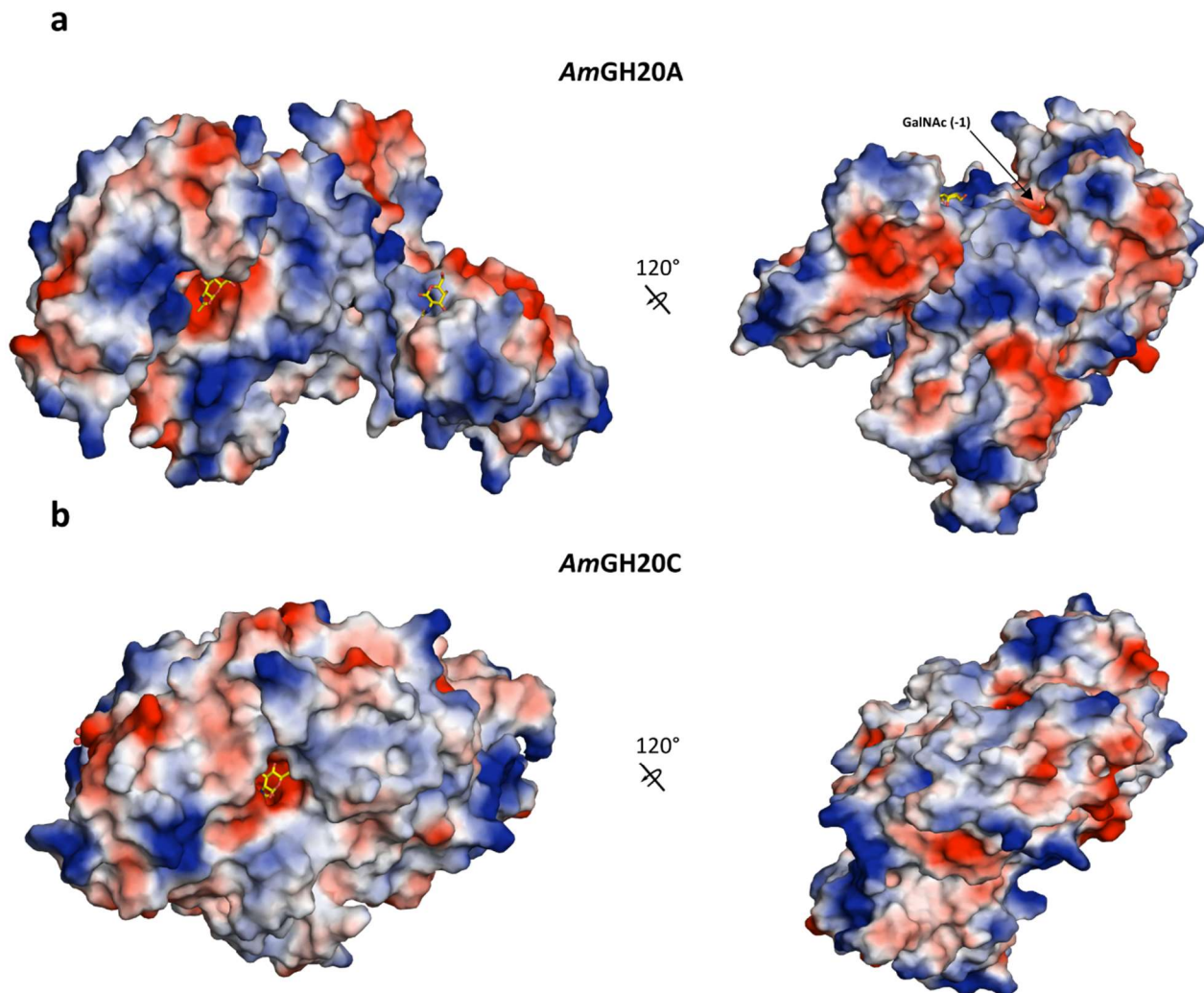
Supplementary Fig. 26: Structural comparison between *AmGH110A* and *PdGH110B*. **a** Structural superimposition of *PdGH110B* (light grey) in complex with α -galactobiose (PDB: 7JWF), onto *AmGH110A*. Divergent loops in the β 1 domain and the catalytic domain, which effectuate differences in the active site architecture between the two enzymes are highlighted, along with the additional N-terminal CBM51 domain of *AmGH110A*, which is lacking in *PdGH110B*. The putative binding site of the CBM51 is on the same side of the enzyme as the active site. **b** The $2F_o - F_c$ electron density (teal mesh) of the galactose (Gal) bound at the subsite -1 of *AmGH110A*. Direct polar interactions between active site residues and the galactose are shown as yellow dotted lines. All the residues that define the -1 subsite are conserved between *AmGH110A* and *PdGH110B*, apart from E480 in *PdGH110B* that forms a potential hydrogen bond to the C6-OH of the galactose unit.



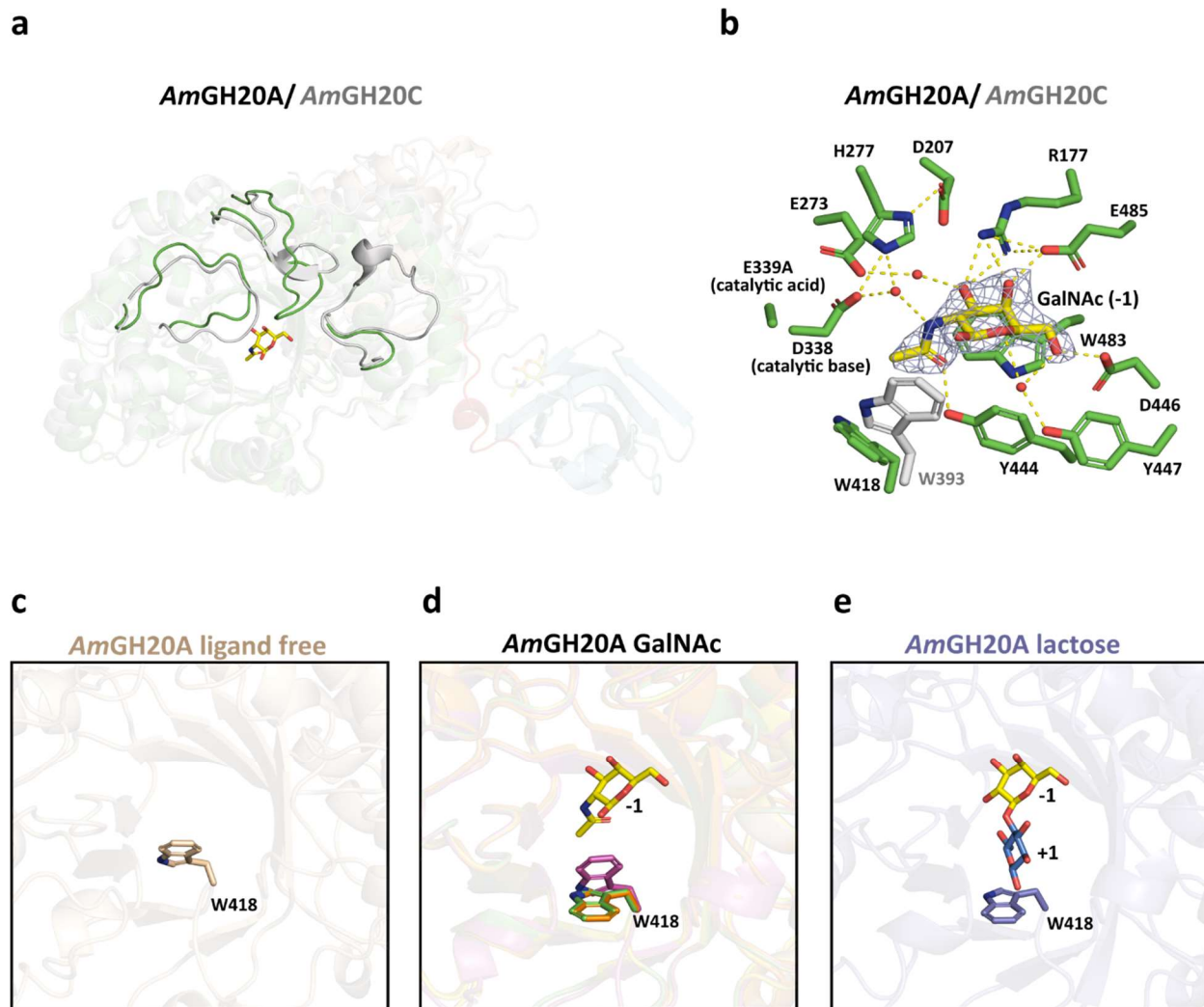
Supplementary Fig. 27: Surface topology of *AmGH110A* and *PdGH110B*. **a** Top and side views of a surface representation of *AmGH110*. The active site pocket is flanked by a flat surface, and the surface of the β 1-domain at one side of the active forms a 120° angle with the horizontal plane of the active site surface. **b** Same as a, but showing *PdGH110B*, which has a bulkier β 1-domain, which results in an almost orthogonal wall at the side of the active site of the enzyme. **c** Overlay of the enzymes in a and b. Showing topological differences between the two enzymes.



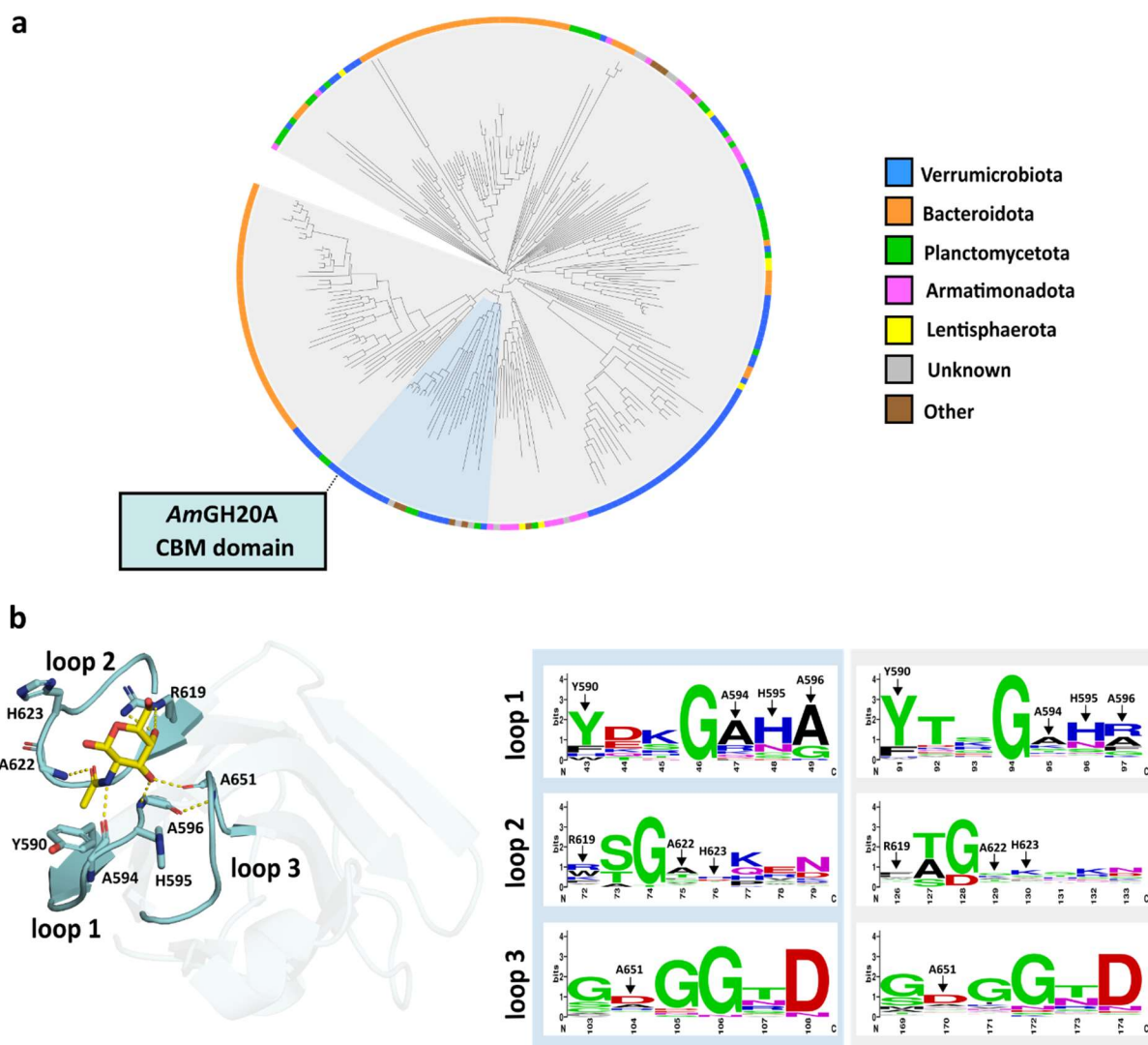
Supplementary Fig. 28: Comparison of the *AmGH110A* CBM51 domain to structural orthologues. a Superimposition of the CBM51 from the *Clostridium perfringens* α -L-fucosidase (*Afc3*) (orange) with a bound Methyl- β -D-galactose (Gal- β -Me) (PDB: 2VMG) onto the CBM51 of *AmGH110A* (cyan), **b,c** The putative binding sites of the CBM51 domains of *AmGH110A* and *Afc3* CBM51, respectively. The residues shown are within 4 Å from the Gal- β -Me bound in *Afc3*. The two histidine residues that form polar interaction with the ligand in *Afc3* are conserved in *AmGH110A*. Two tyrosine residues are at opposite locations of the bound Gal unit, highlighting the divergence of the two domains. **d** Superimposition of the *Clostridium perfringens* A- and B- antigen cleaving endo- β -galactosidase (*EabC*) CBM51 (purple) with a bound A antigen trisaccharide (PDB: 2VNG) onto *AmGH110A*, and the residues of the binding sites. **e,f** Comparison of the ligand binding sites for the CBM51 domains shown in **d**. Residues within 4 Å from the ligand in the binding sites are shown in sticks. **g** Superimposition of the *Streptococcus pneumoniae* A- and B- antigen cleaving endo- β -galactosidase (*EABase*) CBM51 (dark grey) with a bound B antigen tetrasaccharide (PDB: 2VNG) onto *AmGH110A*. **h,i** Comparison of the ligand binding sites for the CBM51 domains shown in **g**. Residues within 4 Å from the ligand in the binding sites are shown in sticks.



Supplementary Fig. 29: Surface electrostatics of *AmGH20A* and *AmGH20C*. **a** Top and side views of the surface electrostatic potential of *AmGH20A*. **b** Same as **a** for *AmGH20C*. The surface electrostatic potential is generated using the APBS plugin in Pymol. Positive potential is shown in blue, negative potential is in red, and apolar surface in grey. The active site of *AmGH20A* is flanked by positive electrostatic potential to a larger extent than *AmGH20C*.



Supplementary Fig. 30: Comparison of *AmGH20A* in free and ligand bound forms and with *AmGH20C*. **a** Superimposition of *AmGH20A* in complex with *N*-acetylgalactosamine (yellow) onto *AmGH20C* (grey) (PDB: 7CBO). Loop differences around the active site are shown. **b** The recognition of the GalNAc bound at subsite -1 of *AmGH20A*. The $2F_o - F_c$ electron density map (blue mesh) for the GalNAc units is shown, Polar interactions are shown as dotted yellow lines and water molecules as red spheres. The residues that define the -1 subsite are identical in *AmGH20A* and *AmGH20C*, however W418 adopts a closed conformation in the GlcNAc-bound *AmGH20C*, as compared to the *AmGH20A* in complex with GalNAc. **c** The W418 residue in the active site of the ligand-free *AmGH20A* adopts a closed conformation. **d** Two of the molecules in the asymmetric unit of *AmGH20A* in complex with GalNAc show an open conformation for W418, while the third shows a closed conformation of this residue. **e** The two lactose-bound molecules in the asymmetric unit of *AmGH20A* show an open conformation of W418, while this residue adopts a closed conformation in the ligand-free third molecule (chain A). The closed conformation of W418 would make a steric clash with the glucosyl units (blue) of the lactose, but it would not hinder the binding of GalNAc or GlcNAc, which indicates a potential conformational selection, where the open conformation of W418 is stabilised by substrate binding at the +1 subsite. In the absence of substrate both the open and closed conformations are feasible.

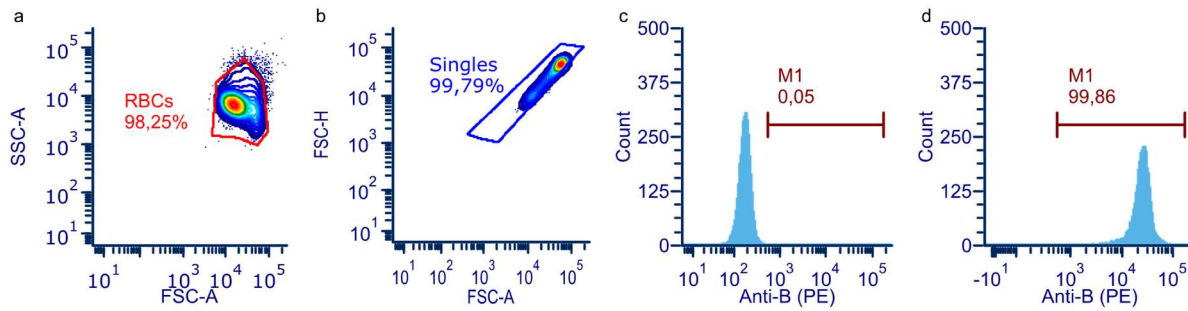


Supplementary Fig. 31: Phylogenetic analysis of the *AmGH20A* CBM. **a** Phylogenetic tree generated from a search with the C-terminal CBM of *AmGH20A* (aa 555–665) as a query against the non-redundant database using Delta-Blast (Domain enhanced lookup time accelerated blast) algorithm against sequences with 20–95 % sequence identity and 90–100% sequence coverage. The search retrieved 549 sequences, mostly derived from other putative HexNAcases, but also a few putative M14 peptidases and putative Gfo/Idh/MocA family oxidoreductases. The sequences were initially aligned using MAFFT, and then trimmed to only include the segment corresponding the CBM domain of *AmGH20A*. The redundancy of the sequences was then reduced using CD-HIT with a 95% identity cut-off, and MaxAlign at default settings. A phylogenetic tree was generated from the remaining 253 sequences and annotated based on phylum affiliation. The cluster shaded with blue harbours the *AmGH20A* CBM. **b** The *N*-acetylgalactosamine binding residues that are presented by three loops are shown and polar interactions are denoted with yellow dotted lines. Sequence logos of the loops forming the GalNAc binding site, showing the conservation in the cluster harbouring the *AmGH20A* CBM, as compared to the conservation in all sequences in the tree, respectively. The corresponding positions for residues shown in the cartoon are marked with black arrows.

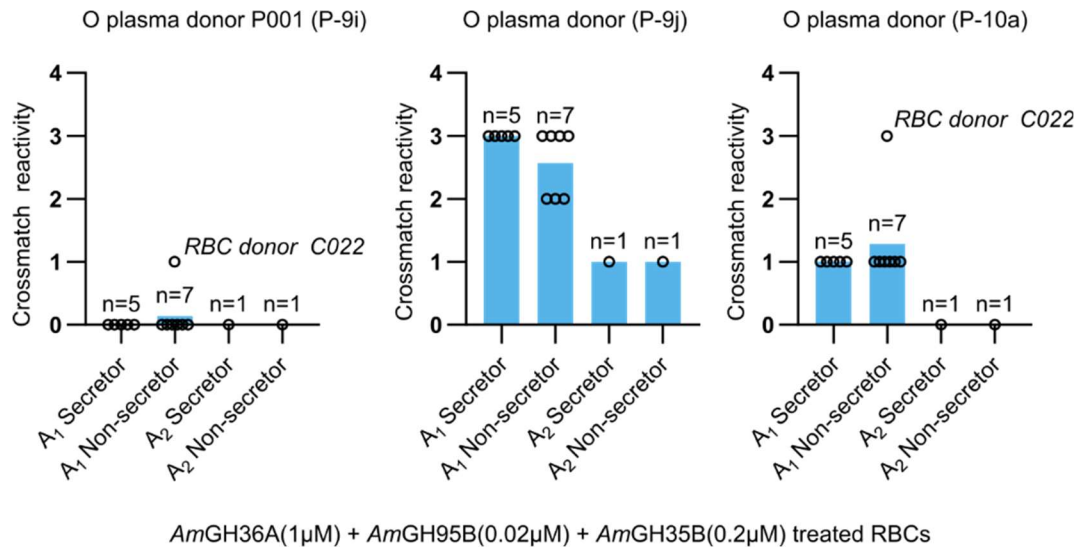
		AB plasma donors (n=11)										B plasma donors (n=12)												
		A ₁ B					A ₂ B																	
		213	214	215	216	217	218	220	219	221	223	224	225	226	227	228	229	230	231	232	233	234	235	236
A ₁ secretor RBCs	Native	-	-	-	-	-	-	-	-	-	-	4+	3+	3+	3+	2+	3+	4+	3+	3+	4+	4+	4+	
	AmGH36A	-	-	-*	-	-	-	2+	-	-	-	-	1+	(1+)	1+	-	-	-	3+	3+	-	-	1+	3+
	AmGH36A + AmGH95B + AmGH35A	-	-	-*	-	-	-	-	-	-	-	1+	(1+)	-	-	-	-	-	-	(1+)	-	-	(1+)	-
A ₁ non- secretor RBCs	Native	-	-	-	-	-	-	-	-	-	-	4+	3+	3+	3+	2+	3+	4+	4+	2+	3+	4+	3+	
	AmGH36A	-	-	-*	-	-	-	-	-	-	-	3+	2+	2+	-	-	-	3+	3+	-	-	2+	3+	
	AmGH36A + AmGH95B + AmGH35A	-	-	-*	-	-	-	-	-	-	-	1+	(1+)	-	-	-	-	-	-	1+*	-	-	1+	-

* = Fibrinogen observed in crossmatch

Supplementary Fig. 32: Crossmatch reactivity of AB and B plasmas towards ECO-A RBCs. The crossmatches of AB plasmas (n=11) and B plasmas (n=12) toward A₁ RBCs treated with AmGH36A alone or a one-pot treatment of AmGH36A (1μM), AmGH95B (0.02μM) and AmGH35A (0.2μM) for 30min in conversion buffer. The crossmatches were performed before and after treatment with AmGH36A (1μM) alone or a one-pot combination of AmGH36A (1μM), AmGH95B (0.02μM) and AmGH35A (0.2μM). The colour-scale (red-pink-white) corresponds to the strength of the crossmatch reactivity.



Supplementary Fig. 33. Representative gating strategy for flow cytometry data. Example of the gating strategy used in flow cytometry experiments shown in figure 3 and supplementary figures 13-14, and 16-20. For each sample, 20 000 events were recorded in total on a BD FACS Canto II and analysed in FCS express v6 by gating for the RBC population. **a** Followed by gating on the singlets. **b** The PE-signal of the singlets population was then used to compare the sample of interest to relevant negative control, either by displaying the histograms, or comparing MFI and percentages of positive. The latter according to marker set by negative and positive control as exemplified in section **c** and **d**, respectively, to gate the positive population. Note, in table “ExtB” only the MFI was compared.



Supplementary Fig. 34: Crossmatch reactions of group O plasmas towards group A RBCs from secretors and nonsecretors after enzymatic conversion. The crossmatches were performed before and after one-pot conversion with AmGH36A (1 μ M), AmGH95B (0.02 μ M) and AmGH35A (0.2 μ M) for 30min in conversion buffer. Each graph shows the reactivity of one plasma (numbering according to Supplementary Fig. 25) with each of the fourteen RBC samples (open rings) of known secretor status, and the bars represent mean reactivities.

Supplementary Tables

Supplementary Table 1: Cloning primers of *A. muciniphila* genes encoding enzymes for the conversion of ABO blood group antigens and their extensions.

Enzyme product	Locus tag	Accession (GenBank)	Strand	Sequence (5'→3')
AmGH110A	Amuc_0480	ACD04318	Sense	AGGAGATATA CCATGTCCGGAACCATAGCCGTA AAA
			Antisense	GGTGGTGGT GCTCGA GTCCACCAAAGCGGACCT
AmGH110B	Amuc_1463	ACD05285	Sense	AGGAGATATA CCATGGCGGATTATCCGGAACGGAC
			Antisense	GGTGGTGGT GCTCGA GCCGGATATCCACCGCCTTTTC
AmGH27	Amuc_1187	ACD05012	Sense	AGGAGATATA CCATGGCAGAGTTCCTAATCCCTATC
			Antisense	GGTGGTGGT GCTCGA GCTTCTTTTTGCTGAACCGGAAC
AmGH36A	Amuc_0216	ACD04059	Sense	AGGAGATATA CCATGACAGCCAGCGCGGAACAATC
			Antisense	GGTGGTGGT GCTCGA TTTGACCTTTCCCTTCGTAC
AmGH36B	Amuc_0517	ACD04355	Sense	AGGAGATATA CCATGGCCCAGCCCCGGACAACCGCC
			Antisense	GGTGGTGGT GCTCGA GTTCGCTGTTACAGTGATGTCCAGG
AmGH36C	Amuc_0855	ACD04688	Sense	AGGAGATATA CCATGATTCGGATTCCGTT
			Antisense	GGTGGTGGT GCTCGA GTGAAGCCGGAGCATCAGG
AmGH20A	Amuc_0369	ACD04208	Sense	AGGAGATATA CCATGCAAGAACAAATCATCCCGAAAC
			Antisense	GGTGGTGGT GCTCGA GTGGCGTTTCAAATACACTGTC
AmGH20B	Amuc_0397	ACD04236	Sense	AGGAGATATA CCATGCTTTTGTTCAGAGTGTTCAGCATT
			Antisense	GGTGGTGGT GCTCGA GGCGGGCATCCGGACGGGGGTC
AmGH20C	Amuc_0868	ACD04701	Sense	AGGAGATATA CCATGGCGCCCCATACCATTCCCT
			Antisense	GGTGGTGGT GCTCGA GACGGCGTTCCCGCGTAATG
AmGH20D	Amuc_1032	ACD04858	Sense	AGGAGATATA CCATGGGAACCAATCCCTACAACATTATTC
			Antisense	GGTGGTGGT GCTCGA GTTTTGTCGCGTTTTGCT
AmGH20E	Amuc_1669	ACD05487	Sense	AGGAGATATA CCATGCAGGATCAGATAATACCCCGGC
			Antisense	GGTGGTGGT GCTCGA GCTCCTTTTACCGTGCC
AmGH20F	Amuc_1815	ACD05633	Sense	AGGAGATATA CCATGCAGGATACGCCCCGTCTTT
			Antisense	GGTGGTGGT GCTCGA GTTCCGGACGTGTGGAAAT
AmGH20G	Amuc_1924	ACD05737	Sense	AGGAGATATA CCATGTGGACATCCCCGCACCCT
			Antisense	GGTGGTGGT GCTCGA GTGGCACC GGAGAGACGT
AmGH20H	Amuc_2018	ACD05828	Sense	AGGAGATATA CCATGGCACGTCCCTTACCCATTCTC
			Antisense	GGTGGTGGT GCTCGA GCTTGTGTTTTCCAGCGG
AmGH20I	Amuc_2019	ACD05829	Sense	AGGAGATATA CCATGGCCGAGCGGACAAATACA
			Antisense	GGTGGTGGT GCTCGA GTTTGAAAGCGGTTCTGTTTTCT
AmGH20J	Amuc_2136	ACD05945	Sense	AGGAGATATA CCATGCAGGACGCCAAGCAGATTG
			Antisense	GGTGGTGGT GCTCGA GCAACGGCTGTACGTT CATATG
AmGH20K	Amuc_2148	ACD05956	Sense	AGGAGATATA CCATGGAAATGCTGCCGGCCGAG
			Antisense	GGTGGTGGT GCTCGA GGTCTTCTTCATCCATGGGGC
AmGH35A	Amuc_0771	ACD04606	Sense	AGGAGATATA CCATGGGCCCTGAGGCTGTCCAG
			Antisense	GGTGGTGGT GCTCGA GCCACTGGTTGTCCAGTTTTTG
AmGH35B	Amuc_1686	ACD05504	Sense	AGGAGATATA CCATGGCTGCTCCCATGCCTTG
			Antisense	GGTGGTGGT GCTCGA GCTTGGCAGGCTTGAACG

The bold bases in the primers are the homologous flanks for the recombination into the target plasmid.

Supplementary Table 2: Predicted properties for enzyme candidates screened for the conversion of B and extended GalNAc-B antigens.

Target	Enzyme	Signal peptide ^a	pI	Mw (kDa)	ϵ_{280} M ⁻¹ cm ⁻¹	Yield ^d (mg/L)
B	<i>AmGH110A</i>	57(63) ^b	8.87	81.6	1.00	15
	<i>AmGH110B</i>	21	8.64	63.8	0.95	30
	<i>AmGH27</i>	21	8.58	55.8	2.02	70
	<i>AmGH36A</i>	24(33) ^b	8.95	91	1.84	15
	<i>AmGH36B</i>	23/24	8.89	72	1.71	30
	<i>AmGH36C</i>	No SP	6.09	80.8	1.80	10
GalNAc-B	<i>AmGH20A</i>	19	8.92	72.8	1.38	30
	<i>AmGH20B</i>	No SP	9.2	82.3	2.04	5
	<i>AmGH20C</i>	28	7.82	61.8	1.50	35
	<i>AmGH20D</i>	17	8.9	59	1.51	25
	<i>AmGH20E</i>	24	9.08	61.4	1.51	25
	<i>AmGH20F</i>	23	8.79	83.3	1.73	5
	<i>AmGH20G</i>	23	8.95	75.1	1.87	15
	<i>AmGH20H</i>	21	8.56	56	1.77	10
	<i>AmGH20I</i>	31	9	57.6	1.72	30
	<i>AmGH20J</i>	19	9.04	83.7	1.30	20
	<i>AmGH20K</i>	25	5.89	61.4	1.68	25

^aThe signal peptide (number of amino acids) as predicted by SignalP (V.6.0) and deepTMHMM (V.1.0.20).

^bThe starting amino acid number in the recombinant enzyme is indicated in parenthesis. The absence of parenthesis means that the construct starts at the first amino acid following the signal peptide. The computed properties in the table are based on the amino acid sequence of recombinant enzymes that comprise the mature peptides lacking signal peptides fused to a hexa-histidine C-terminal tags. The properties were computed using ExPASy ProtParam¹⁸, with the extinction coefficients assuming using all Cys residues are reduced.








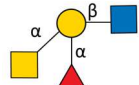
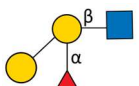
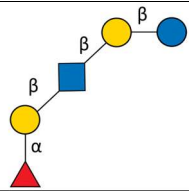
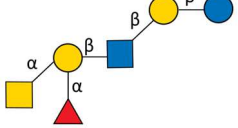
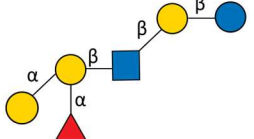
^cThe yields indicated are mg enzyme after two purification steps (IMAC followed by gelfiltration) per L non-optimised shake-flask culture.

Supplementary Table 3: Normalised reaction rates of the screened enzymes candidates against Aryl substrate analogues.

Enzyme	Normalised rate (V_0/E) (s^{-1})	
	pNP-α-Gal	pNP-α-GalNAc
<i>AmGH110A</i>	1.58 \pm 0.09	N.D.
<i>AmGH110B</i>	(5.23 \pm 0.02) $\times 10^{-4}$	N.D.
<i>AmGH27</i>	0.55 \pm 0.01	N.D.
<i>AmGH36A</i>	N.D.	45.1 \pm 0.34
<i>AmGH36B</i>	N.D.	N.D.
<i>AmGH36C</i>	28.1 \pm 0.86	N.D.
	pNP-β-GalNAc	pNP-β-GlcNAc
<i>AmGH20A</i>	60.8 \pm 5.7	109.7 \pm 11.7
<i>AmGH20B</i>	0.31 \pm 0.20	0.16 \pm 0.08
<i>AmGH20C</i>	8.3 \pm 2.9	31.3 \pm 11.1
<i>AmGH20D</i>	0.06 \pm 0.03	0.44 \pm 0.32
<i>AmGH20E</i>	2.1 \pm 0.6	4.6 \pm 0.8
<i>AmGH20F</i>	0.29 \pm 0.17	0.05 \pm 0.03
<i>AmGH20G</i>	8.0 \pm 5.6	0.05 \pm 0.03
<i>AmGH20H</i>	0.17 \pm 0.09	2.3 \pm 1.2
<i>AmGH20I</i>	4.8 \pm 1.9	18.16 \pm 6.53
<i>AmGH20J</i>	0.3 \pm 0.2	45.2 \pm 33.4
<i>AmGH20K</i>	(0.9 \pm 0.1) $\cdot 10^{-3}$	(0.25 \pm 0.14) $\cdot 10^{-1}$
	pNP-β-Gal	pNP-β-GalNAc
<i>AmGH35A</i>	(8.6 \pm 0.04) $\cdot 10^{-1}$	N.D.
<i>AmGH35B</i>	21.9 \pm 3.9	N.D.

Activity data for B antigen candidates and Gal-A candidates (green and red, respectively) are from n=3 independent experiments. The data for GalNAc-B candidates (blue) are from n=2 independent experiments.

Supplementary Table 4: Overview of oligosaccharide used for enzyme activity analyses.

Compound	Structure ^a	Purity	CAS	Supplier
Lactose (Lac)		Not stated	63-42-3	Sigma
Lactosamine		98%	32181-59-2	Sigma
Lacto- <i>N</i> -Biose (LNB)		90%	50787-09-2	Carbosynth
Galacto- <i>N</i> -Biose (GNB)		94%	20972-29-6	Carbosynth
β -1,3-Galactobiose		95%	5188-48-7	Megazyme
α -1,3-Galactobiose		95%	13168-24-6	Carbosynth
H type 2 Triose		90%	GLY031-2-90%*	Elicityl
A type 2 Tetraose		90%	GLY035-2-90%*	Elicityl
B Type 2 Tetraose		90%	GLY038-2-90%*	Elicityl
H Type 1 Pentaose		90%	7578-25-8	Elicityl
A Type 2 Hexaose		90%	GLY037-2-90%*	Elicityl
B Type 2 Hexaose		90%	GLY040-2-90%*	Elicityl

^aSubstrate structures are represented according to the Symbol Nomenclature for Glycans (SNFG)

(<https://www.ncbi.nlm.nih.gov/glycans/snfg.html>)

*Catalogue number for the mentioned supplier is provided if a CAS number is unavailable.

Supplementary Table 5: Normalised reaction rates of ABO antigen active enzyme candidates against blood antigen oligosaccharides.

Enzyme	Substrate¹	Normalised rate (V_0/E) (s^{-1})	Specific Activity ($U\ mg^{-1}$)
<i>AmGH110A</i>	B type 2 hexaose	122 ± 3.2	89.6 ± 2.3
<i>AmGH36A</i>	A type 2 hexaose	127 ± 4.1	83.9 ± 2.7
<i>AmGH95B</i>	H type 1 pentasoe	101 ± 3.5	69.6 ± 2.4
	H type 2 triose	176 ± 4.8	121.4 ± 3.3
<i>AmGH35A</i>	GNB	12.6 ± 1.1	11.4 ± 1.0

¹Oligosaccharide substrate structures, purity and suppliers are listed in Supplementary Table 4. Activity data are from three independent experiments (n=3).

Supplementary Table 6: Predicted properties of enzyme candidates for canonical and extended A antigen conversion.

Target epitope	Enzyme	Signal peptide	pI	Mw (kDa) ²	ϵ_{280} M ⁻¹ cm ⁻¹	Yield ^a (mg L ⁻¹)
A/A type 3	<i>AmGH109A</i>	29	6.78	48.8	1.44	20
	<i>AmGH109B</i>	31	8.46	51.6	1.78	100
Gal-A	<i>AmGH35A</i>	24	9.06	66.2	1.95	35
	<i>AmGH35B</i>	17	8.70	86.5	1.64	20
H type 3	<i>AmGH95A</i>	23	8.90	84.7	1.44	5
	<i>AmGH95B</i>	13	8.94	87.1	2.02	30

Size of the signal peptide regions as predicted by SignalP (V.6.0) and deepTMHMM (V.1.0.20). The first amino acid of the final recombinant enzymes starts at the next amino acid following the signal peptides. The computed properties in the table are based on the amino acid sequence of recombinant enzymes that comprise the mature peptides lacking signal peptides fused to a hexa-histidine C-terminal tags, with the extinction coefficient assuming that all cysteine residues are reduced.

^aThe yields indicated are mg enzyme after two purification steps (IMAC followed by gelfiltration) per L non-optimised shake-flask culture.

Table 7. Data collection and refinement statistics for *AmGH110A*.

PDB accession	8PVS
Resolution range (Å)	48.99 - 1.68 (1.74 - 1.68)
Space group	C 2 2 21
Unit cell (Å, °)	118.527 134.035 215.346, 90 90 90
Total reflections	2694319 (264474)
Unique reflections	193688 (19152)
Multiplicity	13.9 (13.8)
Completeness (%)	99.89 (99.29)
Mean I/sigma(I)	10.01 (0.44)
Wilson B-factor	28.64
R-merge	0.168 (>1)
R-meas	0.174 (>1)
R-pim	0.046 (>1)
CC1/2	0.999 (0.346)
CC*	1 (0.717)
Reflections used in refinement	193545 (19080)
Reflections used for R-free	9475 (948)
R-work	0.184 (0.385)
R-free	0.210 (0.389)
CC(work)	0.970 (0.628)
CC(free)	0.956 (0.603)
Number of non-hydrogen atoms	12482
macromolecules	11384
ligands	61
solvent	1037
Protein residues	1462
RMS(bonds, Å)	0.009
RMS(angles, °)	0.94
Ramachandran favored (%)	95.34
Ramachandran allowed (%)	4.66
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.81
Clashscore	1.32
Average B-factor (Å ²)	34.88
macromolecules	34.26
ligands	35.59
solvent	41.61
Number of TLS groups	11

Statistics for the highest-resolution shell are shown in parentheses.

Supplementary Table 8: AmGH110A top structural orthologues*.

Name	Z-score	RMSD (Å)	Aligned Res	Total Res	Identity%
<i>Pseudoalteromonas distincta</i> PdGH110B (PDB: 7JWF-A)	55.0	1.8	554	584	29
Poly (β-D-Mannuronate) C5 epimerase 4 (PDB: 2PYH-B)	30.8	2.3	325	374	19
Poly (β-D-Mannuronate) C5 epimerase 6 (PDB: 5LW3-A)	30.5	2.3	325	381	20
α-1,3-glucanase (PDB: 5ZRU-A)	30.1	2.4	349	576	16
Pectate Lyase (PDB: 5OLP-A)	28.0	2.9	325	432	16
α-1,3-glucanase (PDB: 6K0V-B)	27.9	2.5	330	562	15
Dettilom tailspike protein gb208 (PDB: 6F7K-D)	27.2	5.8	323	547	12
Bifunctional tail protein (PDB: 4URR-E)	27.0	6.1	328	508	10
Tailspike protein of bacteriophage Sf6 (PDB: 2VBE-A)	27.0	6.1	327	510	10
Cell wall surface anchor family protein (PDB: 3ZPP-A)	26.9	2.6	337	421	17

*Dali search performed on 09-10-2023.

Supplementary Table 9: AmGH110A-CBM51 top structural orthologues*

Name	Z-score	RMSD (Å)	Aligned Res	Total Res	Identity%
ZmpB-CBM51 (PDB: 7JRL)	21.4	1.8	139	457	26
Afc3-CBM51 (PDB: 2VMG)	21.1	1.8	137	145	23
SnGH98-CBM51 (PDB: 2YGL)	18.6	2.5	145	351	21
CpGH98-CBM51 (PDB: 2VNG)	18.3	2.1	141	171	18
Lys-gingipain W83 (PDB: 4ITC)	9.9	2.9	119	168	12
Lysine specific cysteine protease (PDB: 3M1H)	9.9	3.1	124	170	11
Reelin (PDB: 6A48)	9.8	3.4	125	651	6
Endo- β -1,4-mannanase (PDB: 1WKY)	9.0	2.9	116	446	4
Arabinogalactan endo-1,4- β -galactosidase (PDB: 2XON)	9.0	3.5	119	145	8
GDSL-like protein (PDB: 6HH9)	8.9	3.2	117	363	9

*Dali search performed on 09-10-2023

Supplementary Table 10: Isoelectric point (pI) of characterized GH110 enzymes that are active against the B antigen.

Enzyme	Source	Genbank accession	pI
<i>AmGH110A</i>	<i>Akkermansia muciniphila</i>	ACD04318.1	8.87
FragA	<i>Bacteroides fragilis</i>	CAH09922.1	6.99
Gal110B	<i>Bacteroides fragilis</i>	CAJ33351.1	6.53
Gal110A	<i>Bacteroides thetaiotaomicron</i>	AAO78266.1	6.36
Gal110B	<i>Bacteroides thetaiotaomicron</i>	AAO79356.1	5.76
<i>AgaBb</i>	<i>Bifidobacterium bifidum</i>	BAM76380.1	5.26
Gal110A	<i>Streptomyces avermitilis</i>	CAJ33349.1	6.17

Supplementary Table 11: Data collection and refinement statistics for *AmGH20A*.

	GalNAc	Ligand-free	Lactose
PDB accession	8PXT	8PXU	8PXV
Resolution range (Å)	62.32 -2.25 (2.33 -2.25)	49.4 -1.99 (2.061- 1.99)	49.25 -2.5 (2.589- 2.5)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell (Å, °)	92.23 169.06 171.97, 90 90 90	96.7066 170.738 172.368, 90 90 90	96.26 170.42 171.94, 90 90 90
Total reflections	1448660	2711269 (268863)	447171 (44546)
Unique reflections	127867 (12605)	194009 (19084)	93815 (9509)
Multiplicity	11.3 (11.2)	14.0 (14.1)	4.8 (4.7)
Completeness (%)	99.96 (99.99)	99.30 (98.64)	95.34 (97.89)
Mean I/sigma(I)	12.86 (1.72)	9.65 (0.56)	10.36 (1.34)
Wilson B-factor	47.04	46.33	45.99
R-merge	0.123 (>1)	0.148 (>1)	0.146 (1)
R-meas	0.129 (>1)	0.153 (1)	0.163 (1)
R-pim	0.038 (0.502)	0.041 (0.841)	0.069 (0.658)
CC1/2	0.998 (0.64)	0.995 (0.428)	0.995 (0.457)
CC*	1 (0.884)	0.999 (0.774)	0.999 (0.792)
Reflections used in refinement	127843 (12605)	193888 (19059)	93805 (9509)
Reflections used for R-free	6403 (626)	9704 (975)	4657 (474)
R-work	0.180 (0.307)	0.214 (0.371)	0.193 (0.316)
R-free	0.210 (0.346)	0.238 (0.368)	0.225 (0.338)
CC(work)	0.963 (0.844)	0.952 (0.603)	0.960 (0.700)
CC(free)	0.953 (0.748)	0.945 (0.588)	0.939 (0.679)
Number of non-hydrogen atoms	15667	15945	15590
macromolecules	15036	15161	15081
ligands	215	20	53
solvent	521	764	456
Protein residues	1905	1923	1912
RMS(bonds, Å)	0.005	0.007	0.003
RMS(angles, °)	0.64	0.78	0.67
Ramachandran favored (%)	97.20	97.07	97.69
Ramachandran allowed (%)	2.75	2.93	2.31
Ramachandran outliers (%)	0.05	0.00	0.00
Rotamer outliers (%)	0.32	0.44	0.32
Clashscore	2.13	2.79	1.59
Average B-factor (Å ²)	57.39	57.23	51.17
macromolecules	57.01	57.42	51.34
ligands	61.43	75.91	49.53
solvent	49.75	52.96	45.79
Number of TLS groups	23	18	19

Statistics for the highest-resolution shell are shown in parentheses.

Supplementary Table 12: AmGH20A top structural orthologues*

Name	Z-score	RMSD (Å)	Aligned Res	Total Res	Identity%
<i>AmGH20C</i> (PDB: 7CBN-A)	52.8	2.0	478	526	38
<i>Bacteroides thetaiotaomicron</i> β - <i>N</i> -acetylglucosaminidase (PDB: 6Q63-A)	52.1	2.8	513	751	34
<i>Bacteroides thetaiotaomicron</i> β -6-sulfo- <i>N</i> - acetylglucosaminidase (PDB: 7DVA-A)	51.4	1.9	498	523	34
β - <i>N</i> -acetylhexosaminidase (PDB: 7DUP-A)	51.1	1.8	499	524	34
β - <i>N</i> -acetylhexosaminidase (PDB: 3RCN-A)	50.1	1.8	480	525	35
β - <i>N</i> -acetylhexosaminidase (PDB: 6JEB-A)	48.7	2.0	439	469	31
β - <i>N</i> -acetylhexosaminidase NAG2 (PDB: 6K35-A)	48.6	1.8	469	640	33
β - <i>N</i> -acetylhexosaminidase NAG2 (PDB: 6EZX-A)	47.1	2.1	469	639	33
β - <i>N</i> -acetylhexosaminidase (PDB: 4C7F-B)	45.6	1.9	464	494	33
β - <i>N</i> -acetylhexosaminidase (PDB: 1M03-A)	45.3	1.9	461	499	34

*Dali search performed on 09-10-2023.

Supplementary Table 13: AmGH20A CBM domain top structural orthologues*

Name	Z-score	RMSD (Å)	Aligned Res	Total Res	Identity%
Neur_chan_LBD Domain-containing protein (PDB: 6V4S-B)	7.9	2.8	90	585	8
Tripeptidyl-peptidase 2 (PDB: 3LXU-X)	7.6	2.9	100	1217	8
Putative α -L-fucosidase (PDB: 3GZA-A)	7.4	2.7	87	438	6
Phage-like Element PBSX protein XEPA (PDB: 6IA5-C)	7.3	2.8	90	278	11
Echovirus 18 empty particle (PDB: 6HBJ-B)	7.3	3.4	96	237	9
p24delta1 GOLD domain (PDB: 5AZX-A)	7.3	3.5	88	100	8
Lytic cassette protein XepA (PDB: 6I56-C)	7.2	2.8	90	278	11
Turnip yellow mosaic virus (PDB: 1AUY-A)	7.1	3.4	102	163	10
Turnip yellow mosaic virus empty capsid (PDB: 1W39-B)	7.1	3.5	103	189	10
Fab fragment protein (PDB: 1QGC)	7.1	3.3	96	218	15

*Dali search performed on 09-10-2023.

Supplementary Table 14: Comparison of conditions for enzymatic conversion of ABO antigens on RBCs

Target antigen	Enzyme	Dosage ¹ ($\mu\text{g mL}^{-1}$)/nM	Temperature (° C)	Time (min)	Heamatocrit (%)	Buffer system
B	<i>AmGH110B</i>	4/5	RT	30	38	Conversion buffer ⁶
	<i>FragA</i> ²	10/150	26	60	30	Conversion buffer ⁶
GalNAc-B	<i>AmGH20A</i>	14.6/200	RT	30	38	Conversion buffer ⁶
A	<i>AmGH36A</i>	45.5/500	RT	30	38	Conversion buffer ⁶
	<i>NagA</i> ²	300/6200	26	60	30	Conversion buffer ⁶
	<i>NagA</i> ³	0.3/6	26	180	20	250 mM glycine pH 6.8
	<i>FpGalNAcDeAc</i> ⁴	0.5/ 90*	37	60	10	PBS (pH7.4)
	<i>FpGalNase</i> ⁴	0.5/4*	37	60	10	PBS (pH7.4)
	<i>FpGalNAcDeAc</i> ⁴	5/ 900	37	60	10	PBS (pH7.4)
	<i>FpGalNase</i> ⁴	5/43	37	60	10	PBS (pH7.4)
H types 1-4	<i>AmGH95B</i>	0.17/2	RT	30	38	Conversion buffer ⁶
	<i>FucOB</i> ⁵ (<i>AmGH95B</i>)	5/57	37	30	4	PBS (pH 7.4)
Gal-A	<i>AmGH35A</i>	13.2/200	RT	30	38	Conversion buffer ⁶

Data from previously published RBC-active enzymes are highlighted in orange.

* Concentrations depicted in the presence of 40 kDA dextran (300 mg/ml) as an additive

RT = room temperature

¹Dosages in the lowest concentration that result in complete conversion at the given conditions

²Liu, Q. P. *et al.* (2007). *Nat. Biotechnol.* **25**, 454–464

³Gao, H. W. *et al.* (2016). *Blood Transfus.* **14**, 168–174

⁴Rahfeld, P. *et al.* (2019). *Nat. Microbiol.* **4**, 1475–1485

⁵Anso, I. *et al.* (2023). *Nat. Commun.* **14**, 1765

⁶ 200mM glycine, 3mM NaCl, pH6.8

Supplementary Table 15: Overview of enzyme assay conditions for the determination of normalized rates.

Enzyme	Substrate	[S] (mM)	[E] (μM)	Time (min)^a	Volume (μL)
<i>AmGH20A</i>	<i>p</i> NP- β -GlcNAc	2	0.01	10	100
	<i>p</i> NP- β -GalNAc	2	0.01	10	100
<i>AmGH35A</i>	<i>p</i> NP- β -Gal	2	0.1	10	100
	Galacto- <i>N</i> -Biose	1	0.02	5	77
<i>AmGH36A</i>	<i>p</i> NP- α -GalNAc	2	0.02	10	100
	A Type 2 Hexaose	2	0.02	5	77
<i>AmGH95A</i>	H Type 1 Pentaose	2	0.02	5	77
	H Type 2 Triose	2	0.02	5	77
<i>AmGH109B</i>	<i>p</i> NP- α -GalNAc	2	0.5	30	100
	A Type 2 Hexaose	2	1	120	77
<i>AmGH110A</i>	<i>p</i> NP- α -Gal	2	1	30	100
	B Type 2 Hexaose	2	0.02	5	77

^aAll reactions were performed at 37°C in a 2mM phosphate, 150 mM NaCl, 0.005% triton X-100, pH7 assay buffer.

Supplementary Table 16: Overview of enzyme crystallization conditions.

Enzyme	Ligand (co-crystallized)	Screening Kits	Optimised conditions	Cryo- protectant	Beam- line
<i>AmGH110A</i> (35 mg ml ⁻¹) PDB: 8PVS	Galactose (20 mM)	LMB	7% PEG350MME 7% PEG20000 0.1 M Tris 1.4 M NaHCOO pH 7.5	20% Ethylene Glycol	BioMAX MAX IV
<i>AmGH20A</i> (15 mg ml ⁻¹) PDB: 8PXV	Lactose (10 mM)	LMB	11% PEG8000, 0.5 M NaCl, 0.1 M C ₂ H ₆ AsNaO ₂ , pH 6.2	20% (w/v) glycerol	BioMAX MAX IV
<i>AmGH20A_E339A</i> (ligand free) (20 mg ml ⁻¹) PDB: 8PXU	B-antigen type 2 tetrasaccharide	No screens	11% PEG8000, 0.5 M NaCl, 0.2 M C ₂ H ₆ AsNaO ₂ , pH 6.2	Soak with B- antigen type 2 tetrasaccharid e	BioMAX MAX IV
<i>AmGH20A_E339A</i> (GalNAc) (20 mg ml ⁻¹) PDB: 8PXT	GalNAc (20 mM)	No screens	11% PEG8000, 0.5 M NaCl, 0.3 M C ₂ H ₆ AsNaO ₂ , pH 6.2	Soak with GalNAc & α- galactobiose	Petra III P13

Chapter 3 – Structural analysis of *Akkermansia muciniphila* enzymes active on RBCs

The surfaces of RBCs presents dense negative potential, attributed to a variety of sialyl-glycoconjugates¹⁶⁷. This chapter will focus on structural comparisons of *AmGH110A*, *AmGH20A*, and a novel structure of *AmGH109B*, to their characterized orthologues, or to orthologues we have verified to be inactive on RBCs. Both orthologues that have been previously crystallized, and AlphaFold models are analysed, to elucidating structural divergence between enzymes that efficiently convert RBC antigens, to counterparts that exhibit lower, or no, activity on RBCs.

3.1. Methods

3.1.1. Crystallization and structure determination of *AmGH109B*

Recombinant *AmGH109B* was produced and purified as described in Chapter 2, and then buffer exchanged into a minimal buffer (2 mM MES, 50 mM NaCl) and concentrated to 28 mg mL⁻¹. Initial crystal screening was performed by sitting-drop vapor diffusion using the JCSG crystal screening kit (Molecular Dimensions, Sheffield S1 4DP, UK), in a 96 well plate (MD3-11, Molecular Dimensions). Drops were set up in a 1:1 ratio of enzyme solution to reservoir solution by the mosquito[®] Xtal3 handling robot (SPT Labtech, Melbourne, Australia) to a final volume of 200 nL. The most promising crystal conditions were chosen for optimization by incomplete factorial screens using a Gilson GX-281 liquid handler. Final crystal conditions that resulted in the structure of *AmGH109B* were 16% w/v PEG3350, 0.14 M Ca(OAc)₂, pH 8. Crystals were cryo-protected in 20% glycerol and flash frozen in liquid nitrogen. X-Ray diffraction data was collected remotely at 100K using the BioMAX beamline at the MAX IV synchrotron radiation facility (Lund, Sweden).

X-Ray diffraction data were initially processed and integrated with xia2²⁶⁸ using the 3dii pipeline, with XDS (version: January 31, 2020). The phase was solved by molecular replacement using the Phaser version, included in Phenix package (v.1.19.2-4158). The structures of *AmGH109A* (PDB: 6T2B) was used to solve the structure of *AmGH109B*. The structure was refined using phenix.refine²⁶⁹, then manual rebuilding and addition of solvent content was performed in COOT²⁷⁰. The models were validated using Molprobit, and the data statistics and refinement details are listed in Table 2.

3.1.2. AlphaFold model predictions

Amino acid sequences encoding target glycoside hydrolases were obtained from the UniProt database ([UniProt](#)). AlphaFold2 structural model predictions were performed using Colab v1.5.2 with default settings^{271,272}. The highest ranked models were used for analysis in PyMOL Molecular Graphics System, version 2.4.1. (Schrodinger, LLC). Sequences encoding signal peptides were removed. Models used for analysis include *AmGH110B* (Uniprot ID: B2UL12) and *AmGH20E* (Uniprot ID: B2UM43).

3.2. *Akkermansia muciniphila* encodes two distant GH110 enzymes

A. muciniphila encodes two GH110 enzymes, whose founding member is an α -galactosidases able to cleave the B antigen on RBCs¹⁶². In contrast to *AmGH110A*, *AmGH110B* was inactive against the B antigen. The amino acid sequence identity between the catalytic module of *AmGH110A* and *AmGH110B* is 28.36 (%), populating two distant clusters of the phylogenetic tree (Figure 15). *AmGH110A* belongs to cluster II, in which all sequences are taxonomically related and stemming from the Verrumicrobiota phylum. By contrast, *AmGH110B* populates cluster VIII, also dominated by sequences belonging to Verrumicrobiota (91 %), but containing a few sequences from Bacteriodota (9 %). The segregation of *Akkermansia* sequences into distant clusters is suggestive of different specificities. Previously, the phylogenetic tree of GH110A was divided into two sub-families, A and B, defined by their specificity²⁵⁷. Sub-family A is specific for the fucosylated B antigen, whereas members of sub-family B were active on both linear α -gal oligosaccharide substrates and fucosylated B antigens. Interestingly, *PdGH110B* (sub-family B), which belongs to a cluster of mainly marine bacteria, lacks activity against the B antigen²⁷³. While *AmGH110A* falls into sub-family B, *AmGH110B* may populate a hitherto undescribed sub-family.

A comparison of the crystal structure of *AmGH110A* to a model of *AmGH110B* reveals differences at the surface of the active site. *AmGH110B* consists of the canonical catalytic β -helix domain flanked by two smaller β -barrel domains, albeit without the N-terminal CBM51 present in *AmGH110A* (Figure 16). The active site surface of *AmGH110B* is flanked by positive potential apart from a negative cavity adjacent to the putative -1 subsite. This cavity is not present in the structure of *AmGH110A* owing to the elongation of a surface loop (15 amino acids, Figure 17), which is significantly shorter in *AmGH110B* (8 amino acids). The elongation of this loop is a conserved feature in cluster II of the phylogenetic tree, while shortened loop is conserved entirely in cluster VIII (Figure 15).

Superimposition of the two enzymes reveal that the -1 subsites is fully conserved (not shown). However, differences in the putative +1 subsite may explain the inability of *AmGH110B* to act on the B antigen oligosaccharide. Notably, *AmGH110A* has an aromatic amino acid, Tyrosine (Y433), located on a loop connecting the end of the B1 sheet domain to the catalytic β -helix fold domain. This tyrosine is flanking a pocket adjacent to the C2-OH of the Gal in the putative +1 subsite, presenting a possible aromatic stacking interaction for the fucose in the putative +1' subsite (Figure 17). This tyrosine is strictly conserved (100 %) in cluster II (Figure 15), while the neighboring cluster I, harboring a B-antigen converting GH110 from *Bacteroides fragilis*, Gal110B²⁵⁷, has a W or Y in the corresponding position (W: 65 %; Y: 28.2 %). Additionally, an arginine (R544) flanks the potential fucose binding site. R544 is an invariant and unique residue in cluster II (Figure 15), and may play a role in coordination of the saccharide ring stacked onto Y433. Another invariant arginine (R623), which is conserved in clusters I – IV (sub-family B), has been shown to be involved in the binding of Gal in the +1 subsite²⁷³. Interestingly, both aforementioned arginines in *AmGH110A* (R544 and R623), are replaced by 95 % conserved prolines in cluster VIII including *AmGH110B* (P394 and P473, respectively). The open cavity in the direction of the C6 of the -1 subsite Gal is flanked by a glutamic acid (E503) and a methionine (M497) residue. *AmGH110B* lacks the aromatic platform on loop 2, instead containing a threonine in the corresponding position (T283). Additionally, *AmGH110B* has an arginine (R281), neighboring T283 and pointing towards the putative +1 subsite (Figure 18). A tyrosine (Y231) is present in close proximity to R281, coming from the B1 domain. Compared to *AmGH110A*, the potential saccharide binding cavity is offset as a result of the divergent amino acid placement, which may indicate different stereochemical substrate preferences.

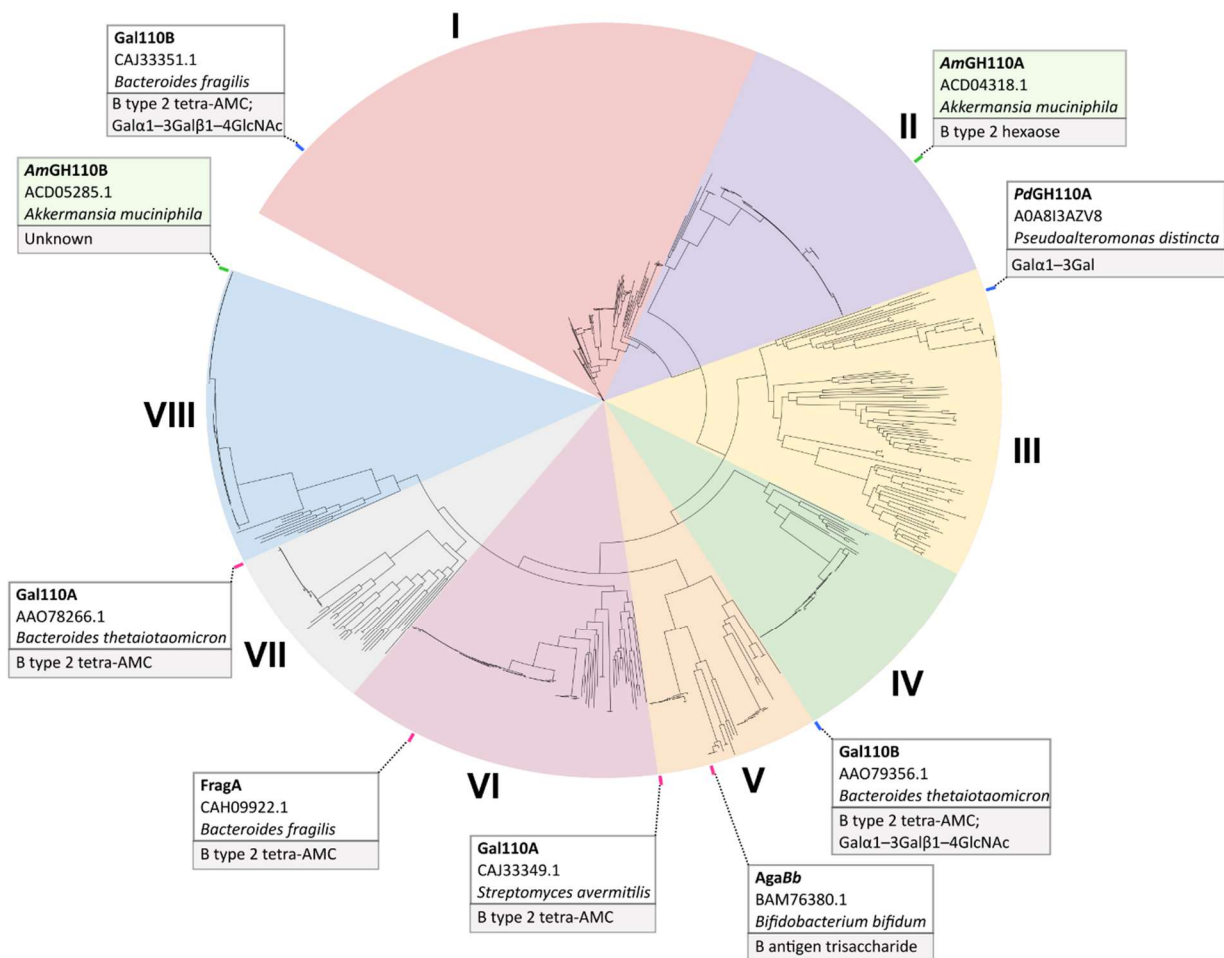


Figure 15. Phylogenetic tree of GH110 sequences from the CAZy database. Previously characterized enzymes that harbor group B antigen activity or λ -carrageenan activity are indicated with enzyme name, Genbank accession and substrates, where enzyme reactivity has been demonstrated. The different clusters are denoted with Roman numerals. A total of 966 annotated GH110 sequences, retrieved from CAZy, were initially aligned using MAFFT V.7 and then trimmed to include only the catalytic domains corresponding to residues 213–795 of *AmGH110A*. The sequences were filtered using MaxAlign with the default settings, and a final list of 748 sequences were subsequently realigned, and a phylogenetic tree was then made using the neighbor-joining algorithm on the MAFFT server with bootstraps performed with 1000 iterations. The figure is from Chapter 2.

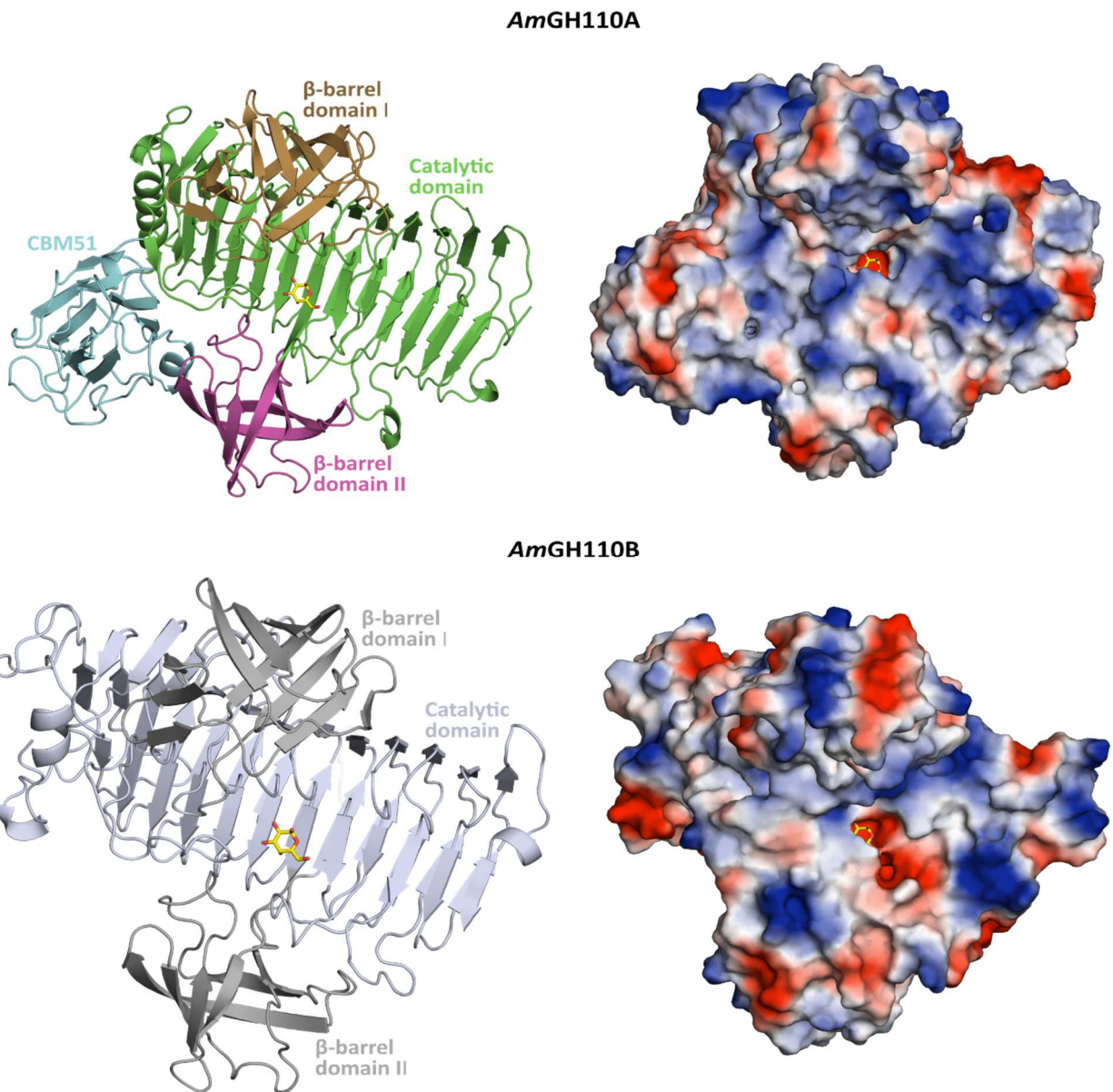


Figure 16. Structural overview of *AmGH110A* and *AmGH110B*. Overall structure of *AmGH110A* in complex with α -galactose (yellow) and an AlphaFold model of *AmGH110B*, showing domain organization. The model of *AmGH110B* is superimposed onto *AmGH110A* to show the putative α -galactose in the active site pocket. Surface electrostatic potential is generated using the APBS electrostatics plug-in in PyMOL, with positive potential in blue and negative potential in red.

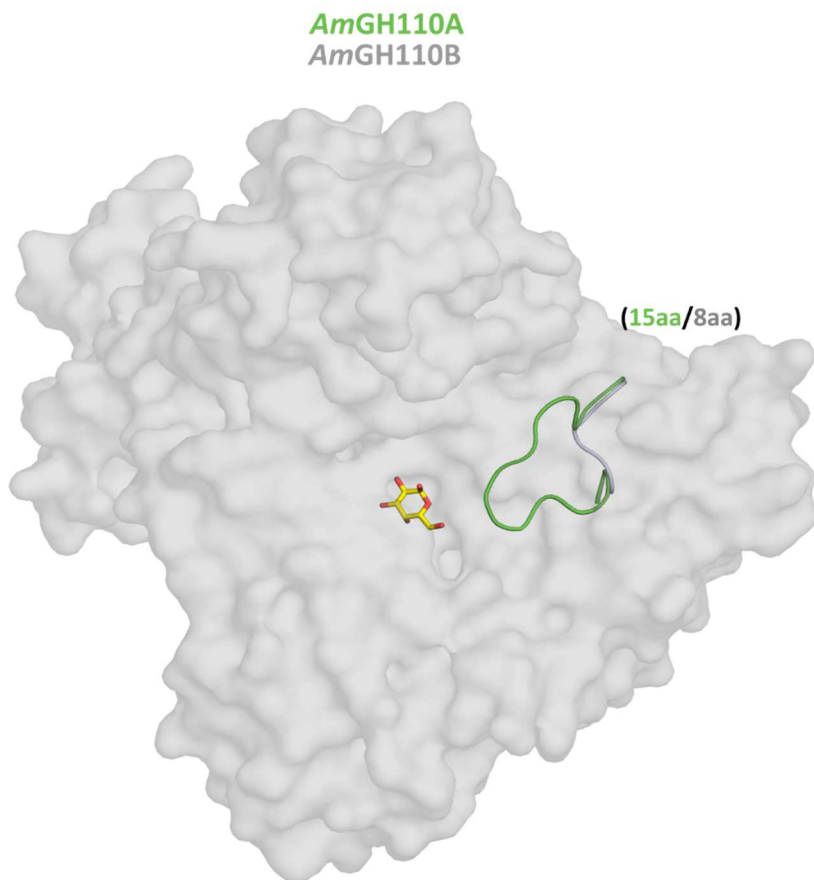
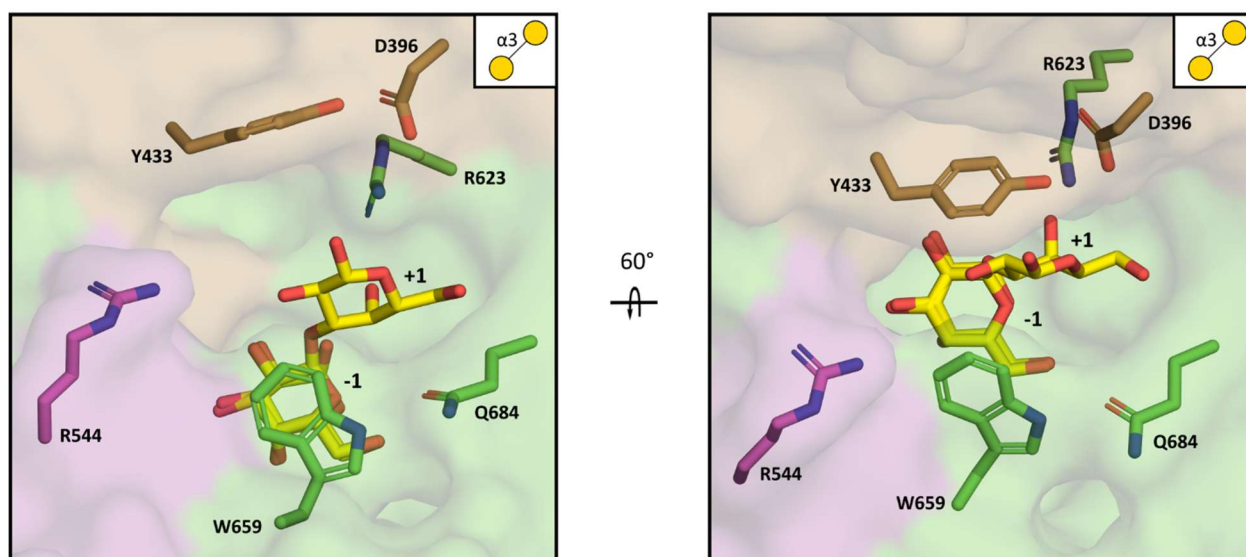


Figure 17. Loop divergence between *AmGH110A* and *AmGH110B*. Structural superimposition of *AmGH110B* (light grey) onto *AmGH110A* in complex with α -galactose. The extended loop of *AmGH110A* is shown in green. The number of amino acid residues of each loop is shown in green for *AmGH110A* and grey for *AmGH110B*. Semi-transparent surface view of *AmGH110B* is shown to highlight topological differences caused by the loop shortening.

AmGH110A



AmGH110B

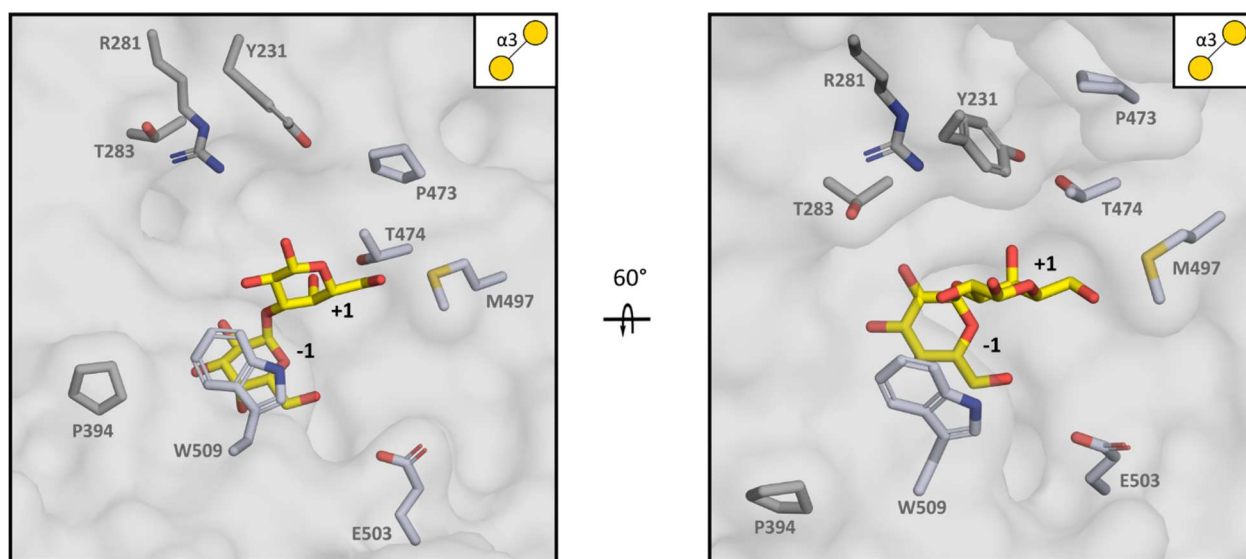


Figure 18. Comparison of the putative +1 subsite of *AmGH110A* and *AmGH110B*. Semi-transparent surface of *AmGH110A* and *AmGH110B* around the -1 and putative +1 subsite. A structural superimposition of *PdGH110B* in complex with α -galactobiose onto *AmGH110A* and *AmGH110B* is used to highlight the putative +1 subsite. Images on the right show a 60° rotation, giving a top-down view of the putative +1 subsite.

3.3. Active site loop divergence of *A. muciniphila* GH20 enzymes

The GH20 family is the most well represented family of CAZymes in the genome of *A. muciniphila* with 11 encoded genes, likely attributed to the abundance of HexNAcs in mucin¹⁰⁵. In our phylogenetic analysis, the *A. muciniphila* enzymes populate 5 hitherto undescribed clusters (Figure 19). Out of the 11 recombinant GH20 enzymes, only 3 were able to convert the extended GalNAc-B (ExtB) antigen on RBCs: *AmGH20A*, *AmGH20C*, and *AmGH20I*. *AmGH20A* and *AmGH20C* share 41.82 % amino acid sequence identity in the catalytic domains, while *AmGH20A* and *AmGH20I* share 33.33 % identity. These enzymes populate two distinct clusters, with *AmGH20A* belonging to cluster II, while *AmGH20C* and *AmGH20I* sharing cluster VII (Figure 19).

A structural comparison of *AmGH20A* to the available structure of *AmGH20C*²⁷⁴ (PDB: 7CBO) (Chapter 2) revealed divergent loop organization around the active site. Crystal structures of *AmGH20J*^{159,275} (PDB: 6JQF and 7EZT) and *AmGH20H*²⁷⁶ (PDB: 6JEA), both of which are inactive on ExtB, have been solved previously. *AmGH20J* consists of the canonical hexosaminidase domain and (β/α)₈ catalytic domain, along with a smaller uncharacterized N-terminal, and a larger C-terminal β -sandwich domain (Figure 20a). In accordance with our previous observations, the active site of the enzyme is flanked by positive potential. *AmGH20H* consists of the canonical hexosaminidase and (β/α)₈ catalytic domains, and contains a cleft of negative potential adjacent to the *N*-acetyl of the bound GlcNAc (Figure 20b). Structural superimposition of both enzymes onto *AmGH20A* reveals divergent loop organization around the surface of the active sites (denoted as loop 1, 2, or 3) (Figure 21a), resulting in differences in the -1 and putative +1 subsites. Notably, *AmGH20J* shows unique organization of loop 1, which folds into a short α -helix as opposed to the elongated loops seen in *AmGH20A* and *AmGH20H*. On this loop is a highly conserved Isoleucine, I275, (99.5 % in cluster IX) positioned over the C4 of the bound GalNAc in *AmGH20A* (Figure 21c). The distance between the C4-OH and I275 of *AmGH20J* is 2.9 Å, which may explain the significant preference of *AmGH20J* against *p*NP- β -GlcNAc over *p*NP- β -GalNAc. It may be less favorable for *AmGH20J* to accommodate a GalNAc in the -1 subsite, with the axial C4-OH position pointing towards I275, as opposed to GlcNAc having a C4-OH in plane with the pyranose ring.

AmGH20A has a tryptophan (W235) closing in over the active site (4 Å between the C6 of the tryptophan to the C6 of the GalNAc) from loop 2. This tryptophan is highly conserved (94.5 %) in cluster 2 and may play a role in binding the saccharide ring at the putative +1 subsite. Notably, the loop in this position (loop 2) of *AmGH20J* is coming from the C-terminal domain. *AmGH20H* similarly has a tryptophan flanking over the active site, in this case coming from the elongated loop 3. Likewise, *AmGH20J* has a tyrosine (Y497) in this position on loop 3. In both *AmGH20A* and *AmGH20H*, an aspartic acid (D446, *AmGH20A* numbering) from loop 3 interacts with the C6-OH of the bound substrate. In *AmGH20J*, however, the corresponding residue is the hydrophobic valine (V492). These differences suggest that the substrate preference of the GH20 enzymes is dictated by the divergent loops around the active site of the enzyme.

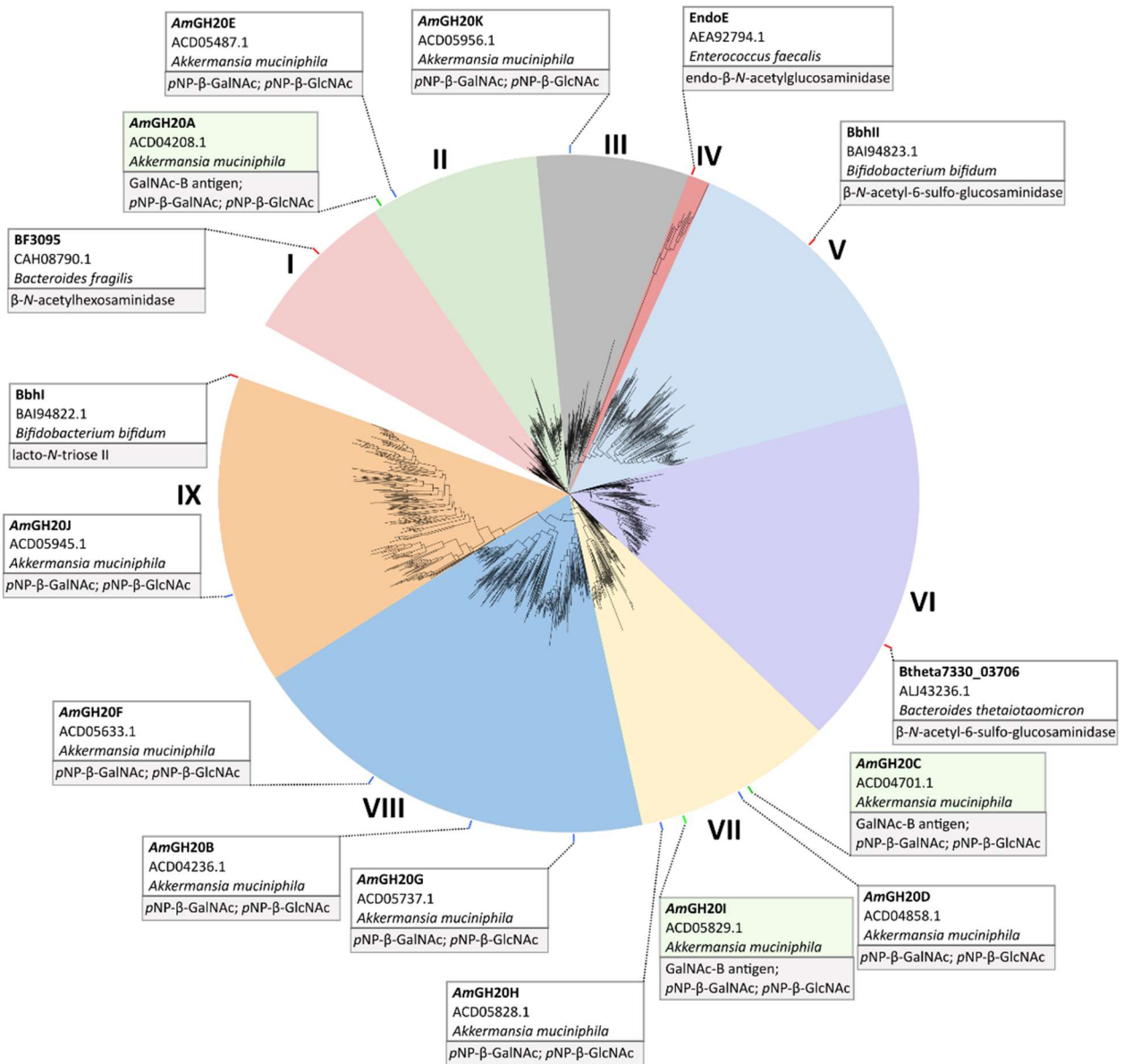


Figure 19: Phylogenetic tree of GH20 sequences from *Akkermansia muciniphila*. The phylogenetic tree was generated from blast results of all 11 GH20 enzymes from *A. muciniphila* and the previously characterized GH20 enzymes with β-N-acetyl-6-sulfo-glucosaminidase, endo-β-N-acetylglucosaminidase, and lacto-N-triose activity. The blast against the non-redundant database was performed using default setting, set to retrieve 500 maximum target sequences, which were filtered for 90–100 query coverage and 30–95% sequence identity. Sequences from all blast results were combined and further filtered for redundancy using CD-HIT with a 95% sequence identity cut-off, and the tree was generated from the remaining 1758 sequences using the neighbor-joining algorithm on the MAFFT server with bootstraps performed with 1000 iterations. Previously characterized enzymes are indicated with enzyme name, GenBank accession and substrates, where enzyme reactivity has been demonstrated. The clusters are denoted with Roman numerals. This figure is from Chapter 2.

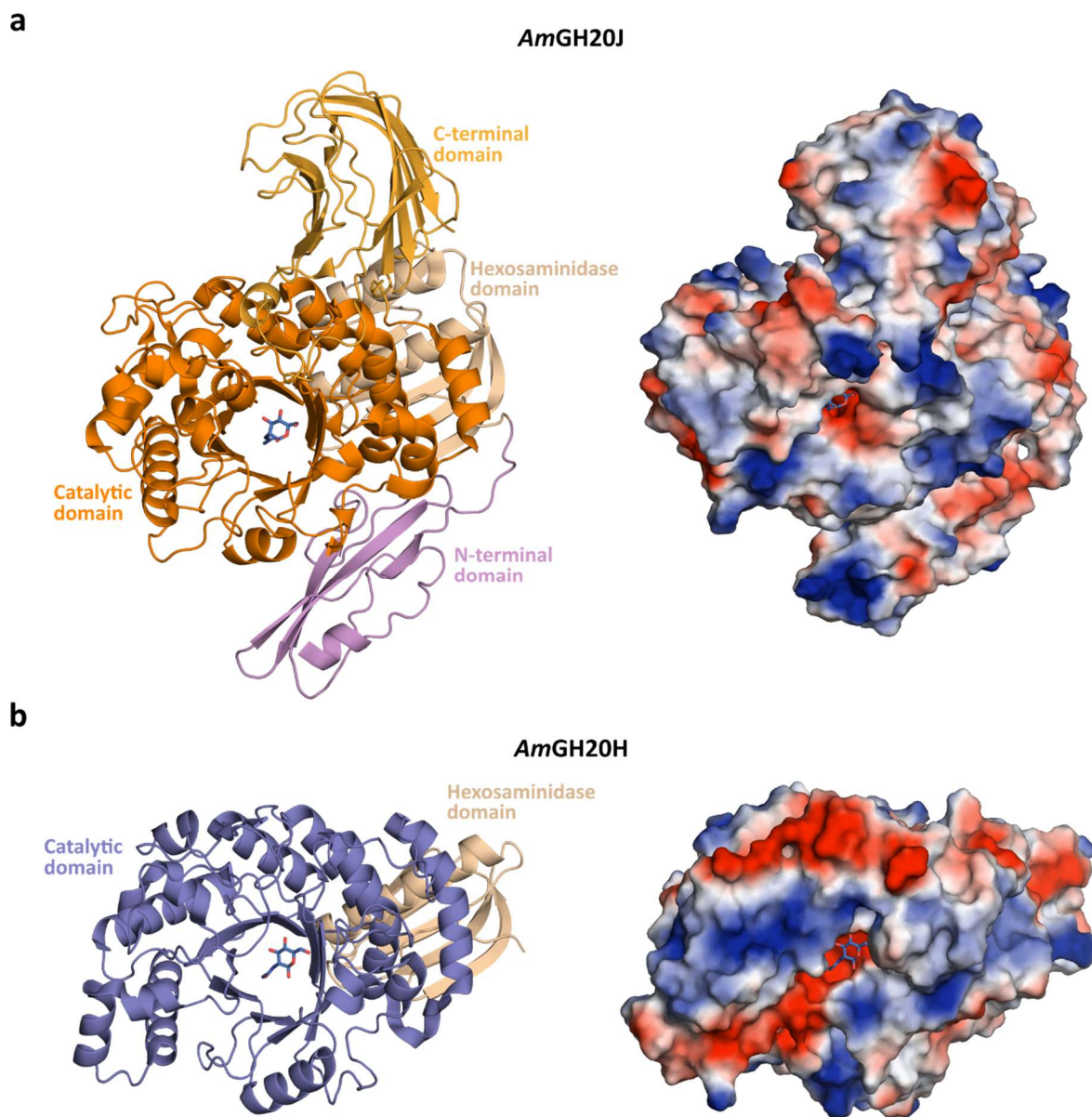


Figure 20. Structural overview of *AmGH20J* and *AmGH20H*. a. Domain overview and surface electrostatic potential of *AmGH20J* (PDB: 6JQF) in complex with β -GlcNAc b. Domain overview and surface electrostatic potential of *AmGH20H* (PDB: 6JEA) in complex with β -GlcNAc. Surface electrostatic potential is generated using the APBS electrostatics plug-in in PyMOL, with positive potential in blue and negative potential in red.

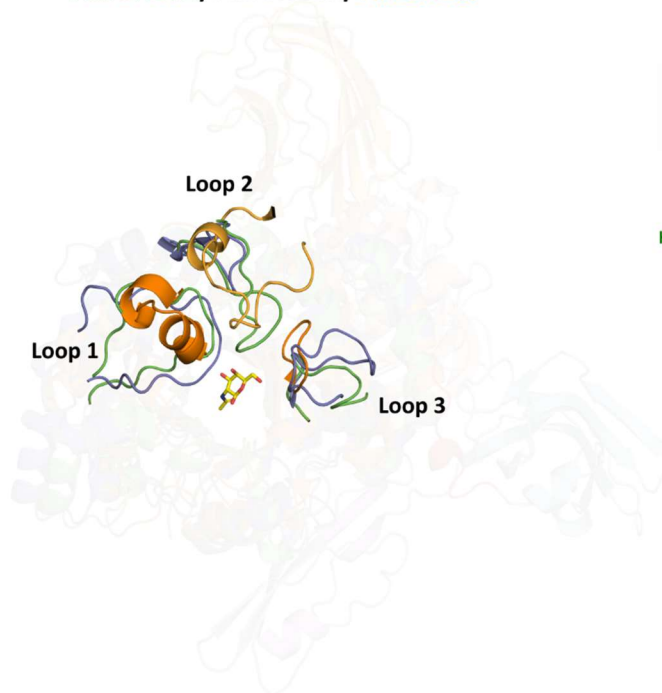
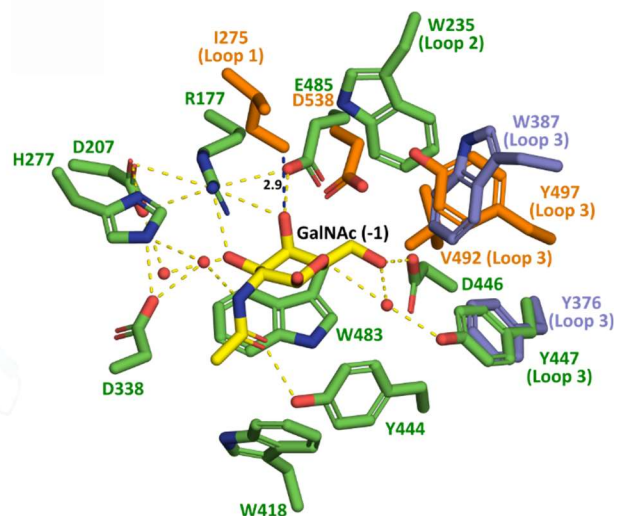
a*AmGH20A/AmGH20H/AmGH20J***b***AmGH20A/AmGH20H/AmGH20J*

Figure 21. Structural comparison between *AmGH20A*, *AmGH20H*, and *AmGH20J*. Superimposition of *AmGH20H* and *AmGH20J* onto *AmGH20A* in complex with β -N-acetylgalactosamine (yellow). Loop differences around the active site are shown. **b.** The recognition of the β -GalNAc bound at subsite -1 of *AmGH20A*. Polar interactions are shown as dotted yellow lines and water molecules as red spheres. The residues that define the -1 subsite are mostly identical, however, residues from *AmGH20H* and *AmGH20J* are shown in the case of a difference.

3.4. *AmGH20E* shares characteristics with a 6S-GlcNAcase

Interestingly, *AmGH20E* is also a member of cluster II and shares the highest amino acid sequence identity to *AmGH20A* (60.1 % between catalytic domains), but lack activity on the GalNAc-B antigen on RBCs. A structural comparison reveals interesting deviations between these two enzymes (Figure 22). The length and organization of the loops around the active site appear similar, in contrast to the remaining *A. muciniphila* GH20 enzymes (Figure 22a). The surface of *AmGH20E* also appears to contain more negative potential around the active site compared to *AmGH20A* (Figure 22b). Strikingly, a subtle difference in the residues adjacent to the C6 of the -1 bound HexNAc (loop 3) differentiates *AmGH20E* from other members of cluster II. At this site, *AmGH20A* has an aspartic acid (D446) and a tyrosine (Y447) coordinating the C6-OH position of the bound HexNAc (Figure 21b), which are 88.1 % and 87.2 % conserved in cluster II, respectively (Figure 19). In contrast, *AmGH20E* has an asparagine (N451) and a lysine (K419) in the corresponding position (Figure 8c-d), which shares similarity with the sulfate-coordinating residues of the β -*N*-acetyl-6-sulfo-glucosaminidase from *Bacteroides thetaiotaomicron* (*Bt0459*)²⁷⁷. A superimposition of *Bt0459* onto *AmGH20A* reveals that the sulfate would be blocked by the position of the tyrosine (Figure 22c). In contrast, a superimposition onto *AmGH20E* reveals the possible accommodation of a sulfate group on this position (Figure 22d). This provides a possible explanation for the difference in the ability to cleave the ExtB antigen may arise from the selectivity of *AmGH20E* to 6S-HexNAcs, although this substrate preference remains to be validated.

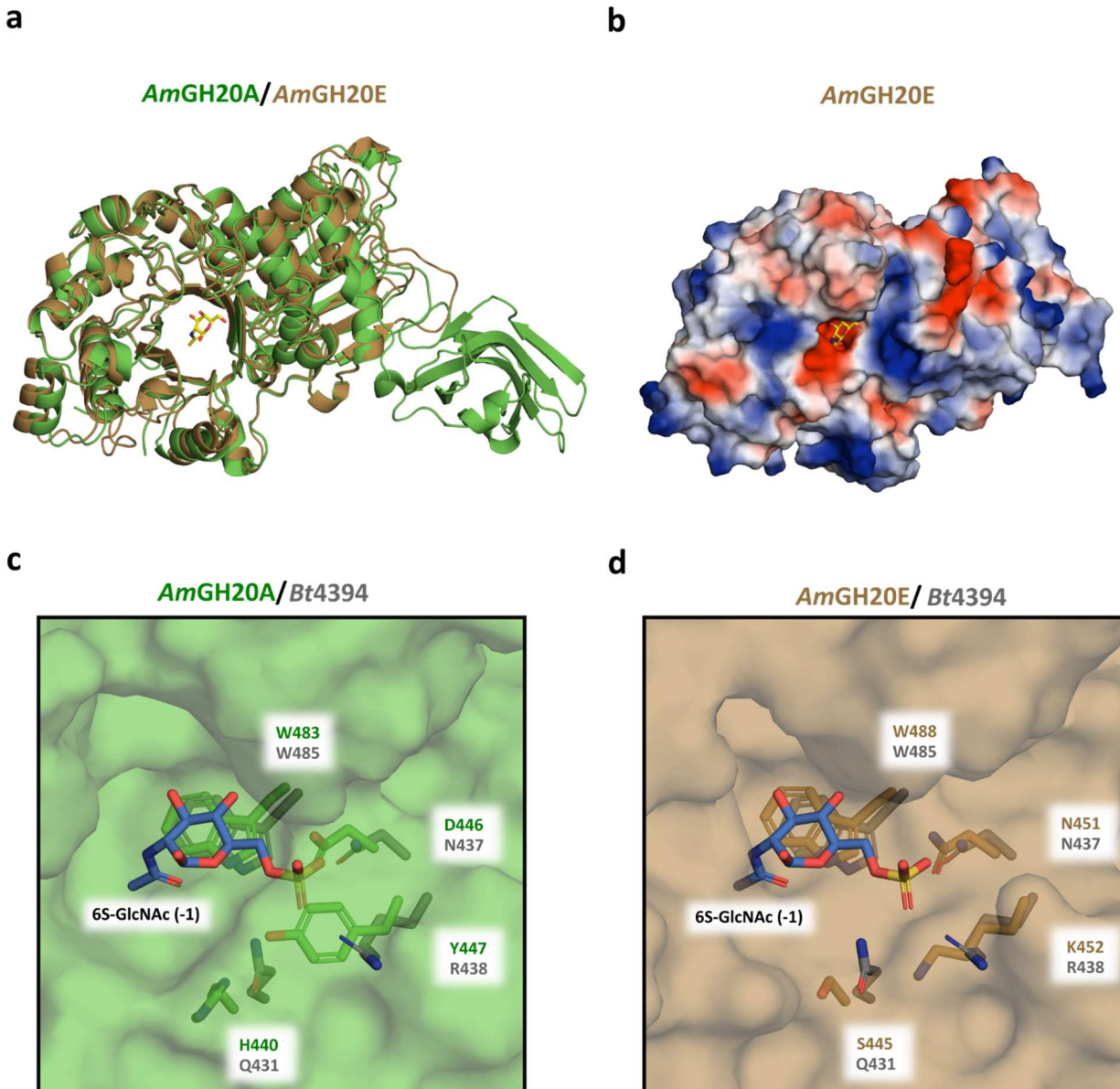


Figure 22. Structural comparison of *AmGH20E* to *AmGH20A*. **a.** Superimposition of *AmGH20E* onto *AmGH20A* in complex with β -GalNAc. **b.** Electrostatic surface potential of *AmGH20E*, using the superimposition in **a** to show the putative binding pocket. **c.** Superimposition of *Bt4394* in complex with 6-S-GlcNAc (PDB: 7DVA) onto *AmGH20A*, showing residues around the sulfate. Semi-transparent surface of *AmGH20A* shows that the sulfate clashes with the tyrosine (Y447). **d.** Superimposition of *Bt4394* in complex with 6-S-GlcNAc (PDB: 7DVA) onto *AmGH20E*, showing residues around the sulfate. Semi-transparent surface of *AmGH20E* shows that the sulfate can potentially be accommodated, coordinated by N451 and K452.

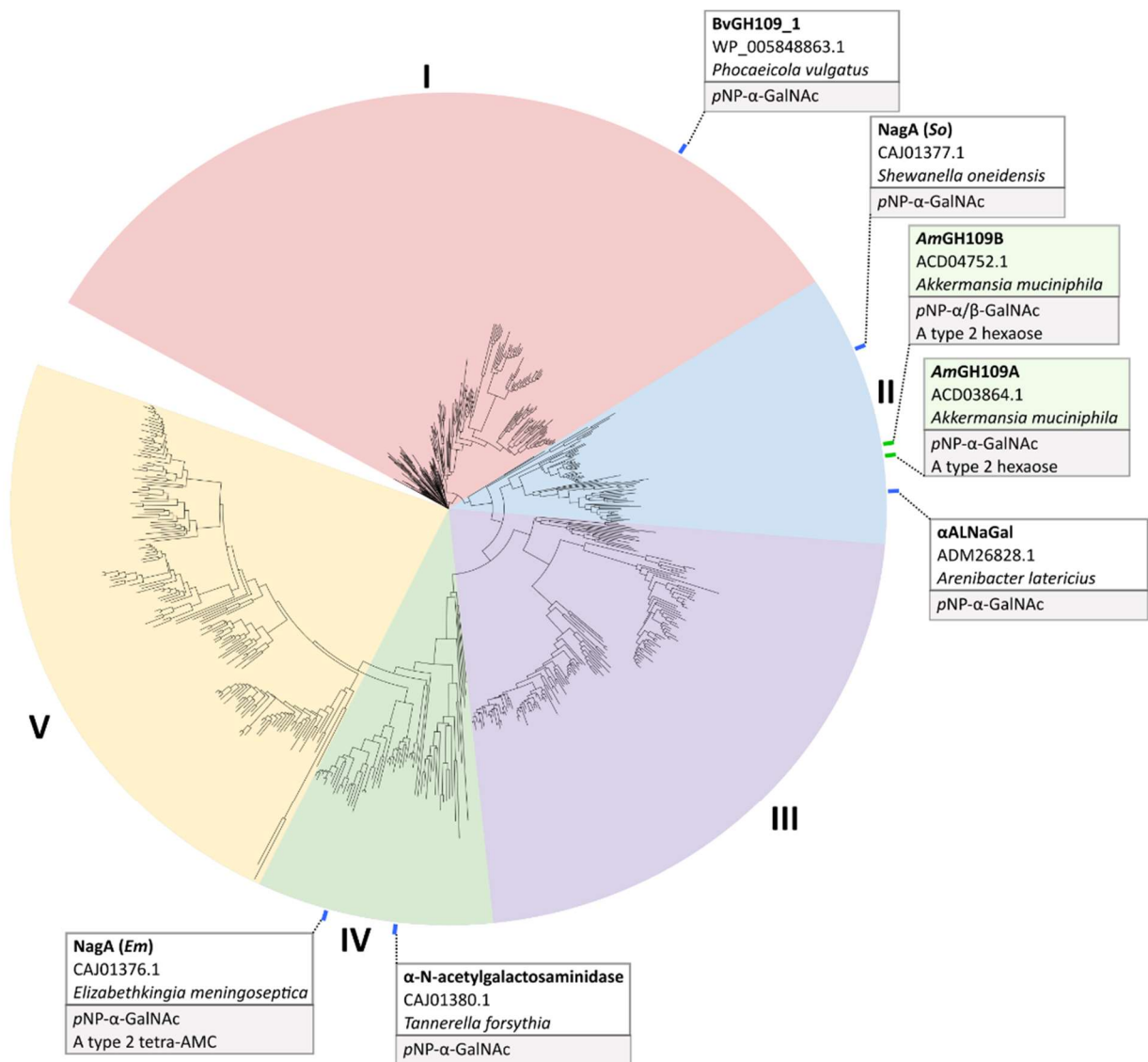
3.5. *AmGH109B* shortened loops confer a more open active site compared to the previously described blood converting GH109 enzyme

The GH109 family was founded based on the A antigen converting member, NagA, from *Elizabethkingia meningosepticum*¹⁶². The two GH109 enzymes encoded by *A. muciniphila*, *AmGH109A* and *AmGH109B*, were previously demonstrated to adopt both α -retaining and β -inverting activity²⁵⁶. Both enzymes were previously tested using α/β -pNP-GalNAc and natural α/β -GalNAc disaccharides, but not on A antigen substrates. *AmGH109B* showed higher activity towards the A antigen compared to NagA¹⁶². *AmGH109A*, however, was far less active against the same substrate, despite sharing 69.8 % amino acid sequence identity and populating the same phylogenetic cluster as *AmGH109B* (Figure 23). NagA falls into a separate cluster, sharing only 36 % amino acid sequence identity to both *AmGH109A* and *AmGH109B*.

We solved the structure of *AmGH109B* in complex with α -GalNAc, which crystallized in the P 1 21 1 space group, with 2 molecules in the asymmetric unit (Table 5). *AmGH109B* has the classical GH109 domain organization, consisting of a single catalytic domain with an N-terminal Rossman fold and a following (α/β)-fold (Figure 24a). Structural comparison to NagA (PDB: 2IXB), reveals that the residues of the -1 subsite are conserved, apart from a threonine (T248) in *AmGH109B* adjacent to the C6-OH of α -GalNAc, in place of glutamic acid (E209) in NagA (Figure 24b). In the structure of NagA, E209 interacts with a neighboring arginine (R213), orienting R231 to make a polar interaction with the C4-OH of the bound GalNAc. In *AmGH109B*, the corresponding arginine (R252) instead makes an interaction with another glutamic acid residue, E341, pointing away from the -1 subsite. Consequently, R252 is not within range to interact with the C4-OH of the GalNAc. Importantly, the shortening of two loops in *AmGH109B* on opposite sides of the active site pocket as compared to NagA, results in a more spacious active site around the catalytic site (Figure 24c-d). In contrast, the active site pocket of NagA appears more obscured. Furthermore, the catalytic site of *AmGH109B* is flanked by a higher distribution of positive potential (Figure 24e), compared to NagA (Figure 24f). This suggests that the decreased catalytic efficiency on RBCs may be due to lower cell surface affinity and a more obscure active site pocket.

The reduced loop sizes of *AmGH109B* was also observed previously for *AmGH109A* (PDB: 6T2B)²⁵⁶. Despite that, there is a significant difference in their ability to cleave the A antigen on RBCs is suggestive of divergent substrate selectivity. Superimposition of *AmGH109A* onto *AmGH109B* shows no noticeable topological differences (Figure 25a). Of note, *AmGH109B* has a higher computed *pI* value (8.46) than *AmGH109A* (6.78). However, the electrostatic surface potential of *AmGH109A* has a similar high density of positive potential, apart from a small patch of negative potential, resulting from an aspartic acid (D236) in place of the corresponding glutamine (Q243) in *AmGH109B*, oriented towards the active site pocket (Figure 25b-c). Additionally, *AmGH109A* has glycine (G77) in place of a glutamic acid (E84) on *AmGH109B*, located below the C6-OH of the bound GalNAc, which consequently exposes the NAD⁺ binding site compared to the more closed state in *AmGH109B* (Figure 24e). Finally, like NagA, *AmGH109A* also contains a glutamic acid (E240) at subsite -1, in place of the aforementioned threonine (T248) of *AmGH109B*. An intriguing observation is that despite the overlap of secondary structure and the catalytic amino acids in the superimposition (Figure 25a,c), the bound α -GalNAc and parts of the NAD⁺ molecule do not align well. This may be an artifact of the glutamic acid residues present in both NagA (E209) and *AmGH109A* (E240), in place of the threonine (T248) in *AmGH109B* (Figure 25d). In respective native structures, both glutamic acid residues and the threonine of *AmGH109B*, make polar interactions with the bound NAD⁺, but not to the bound α -GalNAc. The NAD⁺ position in *AmGH109B* is shifted to allow a polar contact to the threonine,

while the glutamic acid residues would be too close (2.3 Å) for the NAD⁺ to adopt this position in the structure of *AmGH109A* and NagA. This difference does not explain the inability of *AmGH109A* to act on the A antigen, however, since NagA has been shown to act on the A antigen despite sharing this feature, and the key determinant of the specificity remains elusive.



Supplementary Fig. 23: Phylogenetic tree of GH109 enzymes from the CAZy database. A phylogenetic tree generated from 2118 sequences of GH109, retrieved from the CAZy database (April 2023). The phylogenetic analysis is based on the multisequence alignment of GH109 catalytic domains from the CAZy database based on the catalytic domain of *AmGH109B* (residues 27-481) and filtered using CD-HIT at a 95% identity cut-off and MaxAlign with default settings. This resulted in the alignment of 646 sequences that were used to compute the phylogenetic tree using the neighbour-joining algorithm on the MAFFT server with bootstraps performed with 1000 iterations. The different clusters are denoted with Roman numerals. The α-N-acetylgalactosaminidase *AmGH109B* that catalyses the removal of the A-antigen, populates cluster II. This figure is from Chapter 2.

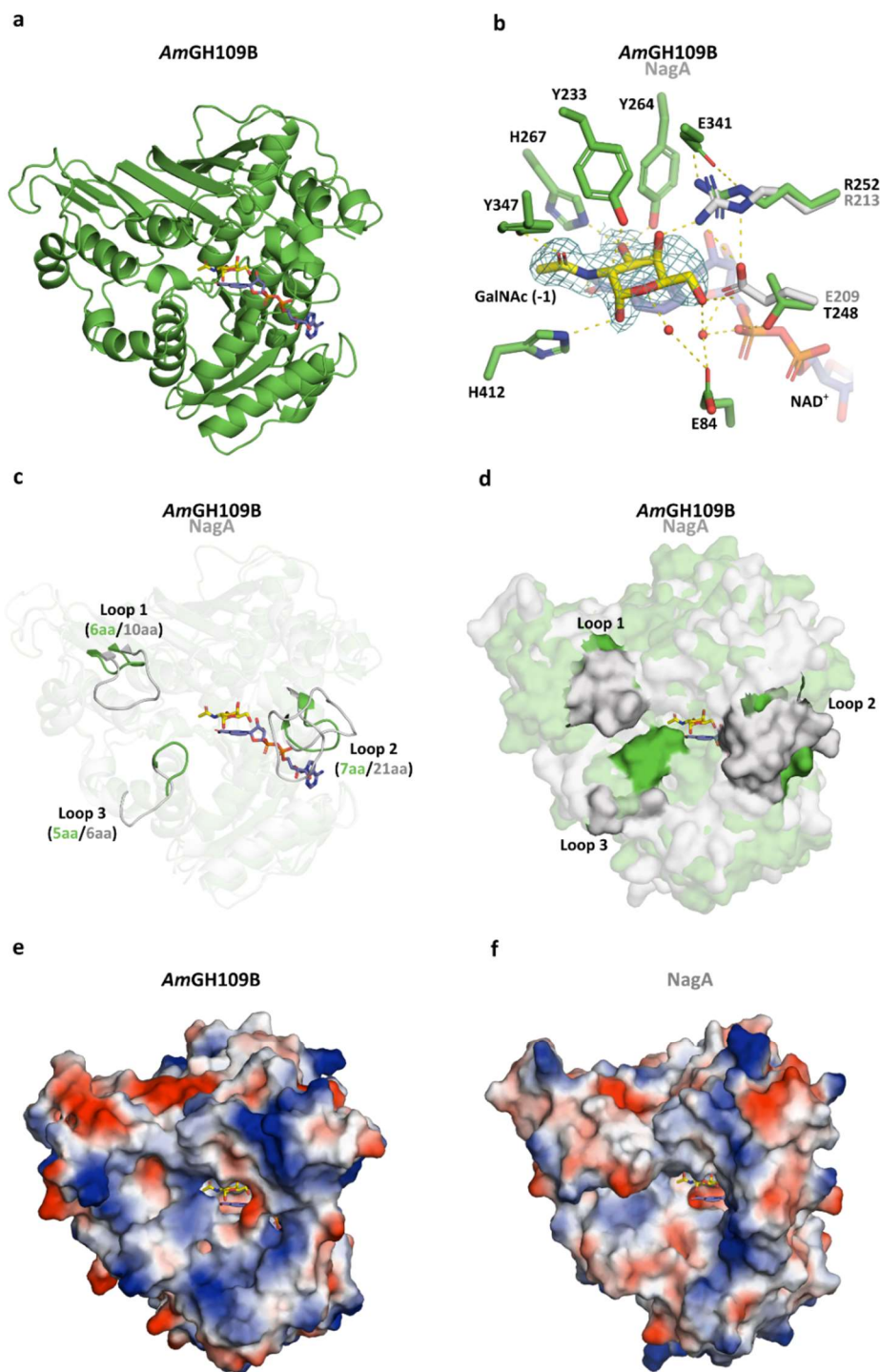


Figure 24. Structural comparison between *AmGH109B* and *NagA*. a. Structure of *AmGH109B* in complex with GalNAc. b. Structural imposition of *NagA* onto *AmGH109B*, zoomed in on the subsite -1 showing the recognition of the GalNAc. The $2F_o - F_c$ electron density map (blue mesh) for the GalNAc unit is shown. Polar interactions are shown as dotted yellow lines and water molecules as red spheres. The residues that define the -1 subsite are identical in *AmGH109B* and *NagA*, apart from a glutamic acid (E209) of *NagA* instead of a threonine (T248) in *AmGH109B*. c. Loop differences around the active site are shown. d. A surface representation showcasing the effects of the loop differences. e-f. Electrostatic surface potential generated using the APBS electrostatics plug-in in PyMOL, with positive potential in blue and negative potential in red.

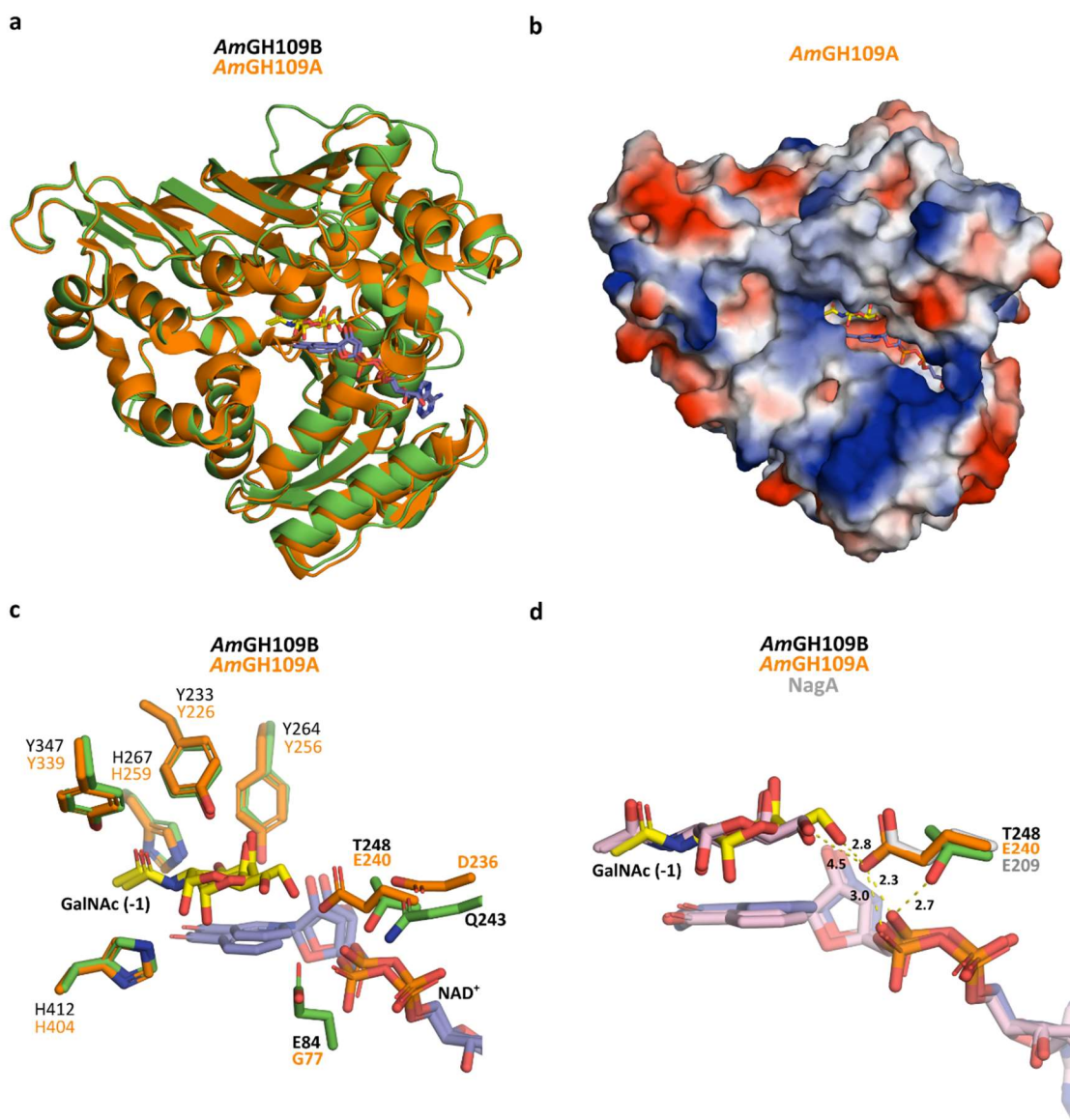


Figure 25. Structural comparison between *AmGH109A* and *AmGH109B*. **a.** Superimposition of *AmGH109A* (PDB: 6T2B) onto *AmGH109B*. **b.** Electrostatic surface potential of *AmGH109A* generated using the APBS electrostatics plug-in in PyMOL, with positive potential in blue and negative potential in red. **c.** Amino acids around the bound GalNAc, with differences highlighted. **d.** Substate and NAD⁺ position of *AmGH109B* varies compared to previous structures. Distances are indicated by dotted yellow lines. For clarity, the GalNAc in the subsite -1 and NAD⁺ molecules of NagA and *AmGH109A* are shown in pink, while the corresponding structure in *AmGH109B* is shown in yellow and blue, respectively.

Table 5: Data collection and refinement statistics for *AmGH109B*

PDB accession		Not submitted
Resolution range (Å)	43.42 - 1.79	(1.854 - 1.79)
Space group		P 1 21 1
Unit cell (Å, °)	70.712 55.013 107.662 90 90.377 90	
Total reflections		538957 (52408)
Unique reflections		77775 (6740)
Multiplicity		6.9 (6.9)
Completeness (%)		97.64 (86.85)
Mean I/sigma(I)		9.01 (0.59)
Wilson B-factor		30.80
R-merge		0.120 (>1)
R-meas		0.130 (>1)
R-pim		0.049 (>1)
CC1/2		0.999 (0.552)
CC*		1 (0.843)
Reflections used in refinement		76463 (6728)
Reflections used for R-free		3730 (286)
R-work		0.191 (0.436)
R-free		0.248 (0.445)
CC(work)		0.977 (0.787)
CC(free)		0.962 (0.806)
Number of non-hydrogen atoms		7560
macromolecules		7103
ligands		184
solvent		346
Protein residues		900
RMS(bonds, Å)		0.012
RMS(angles, °)		1.12
Ramachandran favored (%)		96.32
Ramachandran allowed (%)		3.68
Ramachandran outliers (%)		0.00
Rotamer outliers (%)		0.40
Clashscore		3.42
Average B-factor (Å ²)		50.02
macromolecules		50.51
ligands		45.15
solvent		41.52
Number of TLS groups		13

Statistics for the highest-resolution shell are shown in parenthesis

3.6. Conclusion

The structural analysis highlights distinct hallmarks shared by *A. muciniphila* enzymes. Active site surface topology follows the model of appearing flat and open, as opposed to having the active site in a deeper clefts/ pockets. Additionally, positive surface potential flanks the active site of the enzymes, reflecting the high *pI* values. These adaptations may confer increased affinity to the highly varied and negatively charged mucin *O*-glycans, by improving Coulombic attraction and increasing active site steric accessibility to target motifs. *A. muciniphila* enzymes within the same GH family potentially harbor divergent specificities. This is underscored both by phylogenetic analysis, wherein the enzymes populate different clusters, and based on the difference in activity we observe on RBCs. Even enzymes sharing relatively high amino acid sequence identity, e.g. the GH109 enzymes and *AmGH20A/E*, appear to have different substrate selectivities. Strikingly, structural comparisons suggest that the selectivity of glycan substrates may be dictated by divergence in the surface loops of the active site. Differences in loop organization and amino acid residues may play a role in substrate coordination in the putative +*n* subsites. Crystal complexes with the A, B, or extended antigen structures, would help to elucidate determinants of the substrate specificity, which in turn could motivate rational engineering to improve substrate specificity.

Chapter 4 – Discussion

Blood transfusions are an indispensable part of the health care system, and mismatching of donor blood to recipients has led to otherwise avoidable deaths²⁷⁸. Employing enzymes to catalyze conversion of the immunogenic A/B blood group antigens into the non-immunogenic H antigen would help mitigate transfusion risks, and alleviate shortages of rare blood types. Initial research into Enzyme Converted O (ECO) blood led to the discovery of eukaryotic enzymes able to cleave the A and B antigens, respectively, albeit at undesirable conditions for clinical implementation (high enzyme load, low pH optimum)^{227,229,235}. Nonetheless, clinical trials of enzymatically converted B cells commenced, showing a general increase in transfusion safety, although residual, unexplained crossmatch reactions between ECO Red Blood Cells (RBCs) and recipient plasma persisted^{231–233}. At the time, the leading hypothesis was that low catalytic efficiency of the enzymes resulted in incomplete antigen conversion, and residual B antigens left on the surfaces of the ECO RBCs were the cause of the crossmatch reactions. Later, discoveries of bacterial exoglycosidases with improved rates of conversion failed to solve the compatibility issue, as evident from a study by Gao *et al.* reporting 55% and 18% of ECO-A and ECO-B cells, respectively, exhibited crossmatches when mixed with group O plasma samples¹⁵. These results suggest additional underlying issues besides enzyme efficacy are at play. Notably, the clinical significance of four extensions of the A and B antigens remained unexplored. To address this gap, we report the discovery of six enzymes from the Human Gut Microbiota (HGM) member, *A. muciniphila*, collectively able to convert all known extensions, alongside canonical A and B antigens. Using these enzymes, we provide evidence that enzymatic removal of the extended antigens significantly improves compatibility of ECO RBCs with recipient plasma samples. Finally, crystal structures of three of the enzymes revealed characteristic hallmarks of enzymatic activity against RBCs, which may help guide future discovery, or rational engineering of blood-active enzymes.

4.1. Enzyme source and discovery strategy

In the search for novel ABO antigen converting enzymes, we employed a targeted approach by looking into the atypical mucolytic bacterium, *A. muciniphila*, exclusively utilizing mucin as a carbon and nitrogen source¹³⁵. Mucin O-glycans present the A, B, and Lewis antigens in addition to an abundance of sialylated and sulfated glycans^{105,279}. Therefore, we hypothesized that the evolutionary adaptation of *A. muciniphila* enzymes towards conjugated mucin O-glycans^{151,158}, may extend to enzymatic activity against antigens on RBCs due to the overlap of glycan structures and negative charge. Prior to this project, the most efficient enzymatic conversion of the A and B antigens was performed using enzymes from *Flavonifractor plautii*¹⁶ and *Bacteroides fragilis*¹⁶², respectively. *B. fragilis*, a member of Bacteroidota, has been shown to utilize and grow on mucin, in addition to plant polysaccharides^{280–282}. *F. plautii* is a gram positive member of Bacillota (order Eubacteriales²⁸³), and is a known butyrate producer^{284,285}. Although the association of *F. plautii* to mucin has not yet been reported, other members of Eubacteriales (e.g. *Roseburia inulinivorans* and *Ruminococcus Gnavus*) are known to share the outer mucus layer together with *A. muciniphila*, and cross-feed on mucin^{286–288}. In accordance with this, a search of the annotated genome of *F. plautii* strain YL31 revealed the presence of at least 3 putative carbohydrate ABC transporters ([Flavonifractor plautii strain YL31 chromosome, complete genome - Nucleotide - NCBI \(nih.gov\)](#)). The modular organization of the A antigen converting enzymes from *F. plautii*, *FpGalNAcDeAc* and *FpGalNAse*¹⁶, consist of a signal peptide and one-or-more ancillary Carbohydrate Binding Modules (CBMs) in addition to the catalytic domain. This general domain organization resembles a signature of primary degraders, including

members of the *Roseburia-Eubacterium* group²⁸⁷, in which extracellular cell-attached enzymes with ancillary CBMs aid in anchoring cells onto target glycans. In this case, the enzymes potentially plays a role in adherence to mucin, and the degradation of the terminal A antigen epitope.

Although the notion of mining the HGM for enzymes is not novel, our approach differs from those employed previously. In 2019, Rahfeld *et al.* searched for HGM enzymes for A and B antigen conversion using a metagenomics library screening approach¹⁶. The emergence of high-throughput metagenome screening has made it possible to screen thousands of enzyme genes within a relatively compact timeframe²⁸⁹. This approach is powerful, throwing a large net for discovery, potentially catching enzymes from a plethora of bacteria that have yet to be cultured²⁹⁰. Despite this advantage, native environmental sequences may express poorly in the *Escherichia coli* libraries. It is estimated that *E. coli* only expresses about 40 % of heterologous proteins from randomly cloned metagenomics libraries²⁹¹. The transcriptional machinery of *E. coli* may not recognize heterologous regulatory elements (e.g. promoters) in foreign DNA²⁹². Expression can additionally be hampered by codon bias (e.g. AUG preferred to other start codons), and GC content²⁹³. It has been estimated that only about 7% of high GC content DNA from metagenome libraries is expressed in *E. coli*²⁹⁴. Inefficient translation of environmental proteins containing foreign signal peptides or transmembrane domain regions may also result in protein miss-folding and aggregation^{295,296}. Poor expression combined with potential low activity against inexpensive, initial reporter substrates, may leave several good enzyme candidates undiscovered. For example, Rahfeld *et al.* reported no hits from *A. muciniphila*, despite its abundance in the HGM⁹⁷.

Our targeted approach limits the amount of enzymes we can reasonable screen, but allows for optimization of the expression system and a more thorough screening process of individual enzyme. Most importantly, however, this approach allows us to choose suitable enzymes from particular sources of interest, in this case *A. muciniphila*. An important evolutionary adaptation of *A. muciniphila*, which primed our interest, is the ability to use mucin as both a carbon and nitrogen source. This was indicative of a potential to accumulate and uptake mucin oligomer as revealed recently by Davey *et al.*¹³⁹ and by our recent work showing extracellular decapping of sialic acid from T and Tn antigens¹⁵¹, which allows for specific cleavage of fragments by glycopeptidases¹⁵²⁻¹⁵⁵. Thus, secreted and periplasmic enzymes alike, are likely adapted to target conjugated glycans, which we reasoned could be an advantage for enzymes to act on RBC surface glycoconjugates that share a similar sense and negatively charged glycosylation with mucin. This system of mucin capture on the cell surface, cleavage to oligomers and periplasmic uptake appears to be tailored by *A. muciniphila*, but remains unknown for any other mucolytic bacteria. For instance, Actinomycetota and Bacillota typically employ ABC transporters to import oligosaccharides and monosaccharides released by extracellular enzymes^{131,286,288}. Although the details of the mucin glycan uptake system in *Bacteroides* spp. remains unknown, the study by Davey *et al.* observed no intracellular accumulation of mucin fragments in *Bacteroides thetaiotaomicron*¹³⁹. Ultimately, our approach, which integrated a functional and ecological analyses of the strategy of mucin utilization systems, has proved fruitful, and may be useful to guide further mining of human glycan-utilizing bacteria.

4.2. Conversion of ABO blood group antigens on RBCs

Here, we report the first enzymatic conversion of all known extensions of the A and B antigens, in addition to novel enzymes for the conversion of the canonical A and B antigens. In the context of clinical implementation, conversion conditions are crucial to optimize. Low enzyme concentrations reduce the

economic burden associated with enzyme production. Furthermore, shorter reaction time and lower temperature is preferred to alleviate unnecessary stress on the cells during enzymatic conversion, as well as lowering potential side reactions of the enzymes. In this project, we aimed at full conversion of RBC antigens within 30 minutes at room temperature. We employed a low-ionic strength “conversion buffer” consisting of 200 mM glycine, 3 mM NaCl, pH 6.8 and the higher ionic strength buffer, PBS. Notably, our enzymes systematically performed better using the conversion buffer, as opposed to PBS. Most likely, this is attributed to high pI of the enzymes, particularly the density of positive electrostatic potential around the surface of the active site. RBCs are dense in sialyl-glycoconjugates, and thus carry a shell of negative charge around their surface. Coulombic attraction is likely a key factor contributing to the high affinity of the enzymes to RBCs. Increasing the ionic strength of the buffer will result in an electrostatic screening effect, in which the ions in the buffer dampen the electric field of the enzymes and the RBCs, reducing coulombic attraction^{297–299}. This allows us in turn to use PBS washing to remove our enzymes from the RBCs post conversion, which is a critical step in regards to potential clinical implementation. Additionally, we used the hitherto highest density of RBCs during conversion reactions, using 38% v/v RBCs (38% hematocrit). Under these conditions, we could obtain complete conversion of all known antigen structures using our enzymes at concentrations of ≤ 500 nM (Chapter 2, Supplementary table 14), underscoring the high efficiency of the enzymes.

Direct comparisons to enzymes of previous studies is not strait forward, due to the variations in reaction conditions. However, overall comparison suggest that our enzymes exhibit improved conversion rates, using conditions that are more favorable in a clinical perspective. Prior to this project, the best enzyme for B antigen conversion, a GH110 from *Bacteroides fragilis* (FragA)¹⁶², required 150 nM to convert group B RBCs at 30% hematocrit within 60 minutes, at 26 °C, in conversion buffer. Under our conditions (explained above), AmGH110A obtained complete conversion of the B antigen on RBCs at 50 nM concentrations. This increase in efficiency may stem from increased affinity to RBCs due to favorable electrostatics and the appended CBM51.

Complete conversion of the A₁ and A type 3, antigens on 3 individual A₁ donors was obtained using a concentration of 500 nM of AmGH36A, under the conditions outlined above. The hitherto best candidates, FpGalNAcDeAc and FpGalNAse, were reported to completely convert group A RBCs at 10% hematocrit within 60 minutes using 0.5 $\mu\text{g mL}^{-1}$ of both enzymes (90 and 4 nM, respectively), at 37 °C in PBS. Notably, these concentrations worked only in the presence of a molecular crowder (40 kDA dextran) at 300 mg mL^{-1} . In the absence of dextran, 10 times higher enzyme concentration (900 and 40 nM, respectively) was required for complete conversion. Both of these enzymes appear to be more acidic ($pI= 4.14$ and 4.27, respectively), which suggests possible Coulombic repulsion may occur. This is in line with the use of PBS as a buffer system, as electrostatic screening has been shown to reduce the strength of electrostatic interactions and thereby allow for more rapid reaction rates than at lower ionic strengths³⁰⁰, in contrast to the effect observed on our basic enzymes. Notably, the high apparent K_m on A type 2 oligosaccharide¹⁶, may suggest low affinity of the *F. plautii* enzymes to their substrate. It should be noted, that an excess amount of FpGalNAcDeAc may be warranted in order to alleviate the risk of acquired B syndrome, resulting from incomplete conversion. The intermediate of the two-step reaction is GalN, and has recently been shown to result in crossmatches with anti-B antibodies³⁰¹. Thus, in the absence of a molecular crowder, AmGH36A appears to be catalytically superior, but more importantly a potentially safer option.

Strikingly, AmGH95B showed an unprecedented conversion rate, being able to completely convert all H antigen structures on RBCs after 60 minutes at the low concentration of 2 nM. This rate differs from that

reported earlier this year by Anso *et al.*, who used the same enzyme (referred to as FucOB) for conversion of universal O blood into the rare Bombay type¹⁷. In their study, complete conversion of RBCs at 4% hematocrit was reported using at least 57 nM AmGH95B, for 30 minutes at 37 °C in PBS buffer. As opposed to our other enzymes, AmGH95B obtained complete conversion of the H antigen in both conversion buffer and PBS during our initial screening at 1.2 μM concentrations, although further optimization of conversion in PBS was not performed in order to keep a uniform approach for all our enzymes. Based on our work and the work by Anso *et al.*, AmGH95B works efficiently in either buffer system. However, the marginal difference in hematocrit suggests that the conversion buffer we chose may be more efficient for RBC conversion. This is further underscored by the higher incubation temperature (37 °C) used by Anso *et al.*, as the activity rate should in theory be increased at higher temperatures³⁰².

An important advantage of our work is the use of the highest RBC concentrations to date, at 38 % hematocrit, followed by Liu *et al.* in 2007 using 30% hematocrit. Full conversion using higher hematocrit further underscores the efficacy of our enzymes compared to previous enzymes. The most recent studies by Rahfeld *et al.* and Anso *et al.* report conversion using 10% and 4% hematocrit, respectively. Altogether, data presented in this thesis show complete conversion of all respective antigens on RBCs within 30 minutes. Compared to previous work, we present more favorable conditions in regards to clinical implementation.

4.3. Compatibility of ECO RBCs with recipient plasma

Until now, data showing immunogenic compatibility of ECO RBCs is scarce. Of note, early phase I and II clinical trials of ECO-B RBCs, reported residual unexplained crossmatches^{231–233}. A more recent study by Gao *et al.* in 2016 reported hemagglutination experiments using ECO RBCs treated with the GH109 and GH110 enzymes discovered by Liu *et al.* in 2007¹⁶². Strikingly, 55% of ECO-A cells and 18% of ECO-B cells showed various levels of hemagglutination, despite apparent complete removal of the A and B antigens¹⁵. A possible explanation for the observed crossmatches could be the presence of extended A^{7–9} and B²⁰ structures, which have not been targeted for degradation by enzymes prior to this work. To test this theory, we performed large-scale hemagglutination experiments of ECO RBCs after converting all canonical and known extended structures, compared to the conversion of only the A and B antigens. Our data shows a significant decrease in crossmatches when both canonical and extended antigen structures were removed, compared to removal of only canonical A and B antigens, providing premier evidence that extended antigens are immunogenic. Despite the decreased reactivity, remaining crossmatches are still present, and need to be addressed to achieve complete compatibility. Particularly, 52% of group O plasma samples exhibited crossmatches with ECO-A RBCs, whereas only 9% exhibited crossmatches with ECO-B RBCs. This reflects the more complex and diverse glycan phenotype of the A group.

In order to perform an extensive evaluation of ECO compatibility, we used plasma samples from 100 group O, 12 group A, 12 group B, and 11 group AB donors. Among group A and group AB plasmas, 10 and 7 samples were from A₁ donors, respectively, and the rest from A₂ donors. In addition, RBCs from both secretors and non-secretors of group A₁, A₂, and B donors, were used for the hemagglutination experiments. Using this setup allowed us to judge whether occurring crossmatch reactions correlated to differences between backbone chain types, as secretors express the FUT2 fucosyltransferase responsible for the synthesis of the H type 1 antigen. Although not statistically significant, we observed a general trend showing better compatibility of non-secretor RBCs. A possible explanation is the exposure of underlying

non-ABO antigens, from the cleavage of the type 1 chains. Potential underlying antigens, uncovered as a result of enzymatic activity are termed “cryptantigens”. In particular, *AmGH95B* exhibits a broad selectivity, targeting H type 1, 2, and 3 antigens. Cleavage of the H type 1 antigen would uncover the type 1 precursor consisting of Gal β -1,3-GlcNAc- epitope. This is a potential target for *AmGH35A*, which exhibits activity against LNB, and would result in the exposure of the underlying GlcNAc. To test whether this was a cause for reaction, we treated group O RBCs from secretors with one or both of these enzymes, and mixed with group O or group AB plasma samples. Although no crossmatches were observed after single enzyme treatment using either *AmGH95B* or *AmGH35A* alone, low level crossmatches were observed after treatment with both enzymes for one group O and A₁B plasma sample. This indicates that the exposed Gal under the H antigen (by *AmGH95B*) appears to be safe, but exposed GlcNAc on the type 1 chain is a potential cryptantigen for some recipient. The low levels of these reactions correlates with the reported low levels of H type 1 as opposed to H type 2 antigens. Furthermore, these results provide one plausible explanation for the subtle decrease in crossmatch reactivity between non-secretors and secretors.

Undiscovered extensions of ABO blood group antigens not targeted by our enzymatic treatment, may also exist. Notably, during sequential treatment of group A RBCs with our enzymes, initial treatment with *AmGH36A* resulted in a slight increase in Gal-A antigen levels. Currently, there is no reported structure consisting of a non-fucosylated GalNAc α -1,3-Gal β -1,3-Gal epitope, within the A phenotype. Our data raise questions on the existence of additional unknown extension, which has eluded characterization, potentially due to lower abundance.

Taken together, we provide evidence for the antigenicity of the extended A and B antigen structures, and show a significant improvement in compatibility by targeting these extensions for enzymatic removal. Additionally, the variety in crossmatch reactivity levels among group O plasma samples mixed with ECO-A RBCs, is evidence for inter-human variations in antibody specificity, reactivity, or titer, underscoring the importance of exhaustive crossmatch analysis. Until now, the paradigm was that removal of the A and B antigens would be sufficient for generating universal O type blood, and low enzyme efficacy was the bottleneck to overcome. According to data presented here, the ABO blood group is much more complicated than previously assumed, and removal of the A and B antigens alone will not be sufficient for generating safe, universal donor blood. This work sets the stage for a new paradigm, in which conversion of the extended antigen structures is a necessary step in the route towards truly safe and universal donor blood.

4.4. Structural signatures of enzymes active on RBCs

We have determined the structures of three efficient, blood-converting enzymes: *AmGH110A*, *AmGH20A*, and *AmGH109B*. For the purposes of A/B antigen conversion, enzymes with more basic properties are preferred to perform the conversion at a neutral pH^{6,162}. In accordance, all enzymes in the present work, have high theoretical isoelectric potential ($pI \approx 9$) and exhibit the hitherto best performance on RBCs according to the conditions highlighted in the previous section. Structural observations further reveal that the active site surface of these enzymes are generally flanked by dense positive potential. This appears to be an evolutionary hallmark of these enzymes to target human O-glycans on mucins, which are decorated in negatively charged sialylated and sulfated glycoconjugates. Comparisons to orthologues reveals a general flat surface topology resulting from decreased loop size or reorganization of loops. Increased

accessibility of the active site pocket is an advantage for activity on conjugated human glycans, counteracting potential steric hindrance from neighboring glycan structures. Furthermore, the presence of ancillary carbohydrate binding modules (CBMs) on *AmGH110A* and *AmGH20A* may also increase affinity to specific glyco-conjugate motifs presented on RBC surfaces. CBMs are known to be particularly involved in increasing the affinity of enzymes to insoluble substrates⁴², or to facilitate anchoring of enzymes and bacteria to surfaces (*e.g.* mucosal anchoring)¹⁹. The A converting enzyme pair, *FpGalNAcDeAc* and *FpGalNAse* include ancillary CBM domains as well, consistent with the potential importance of CBMs to cell-surface glycoconjugates. It should be noted, however, that CBMs are not a pre-requisite for high activity, as evident from the high conversion rates of *AmGH95B* and *AmGH109B*. In some cases, increased affinity comes at the expense of catalytic turnover, due to a decreased rate of dissociation. According to the Sabatier principle³⁰³, there is a characteristic optimum of affinity and catalytic turnover, optimized during evolution.

In this thesis, we report the first structure of a blood-active enzyme of GH110 (*AmGH110A*). Prior to this project, the only available structure belongs to a marine GH110, *PdGH110B*, which is inactive on the B antigen²⁷³. Although efforts to soak in a B antigen substrate were unsuccessful, structural comparisons between *AmGH110A* and *PdGH110B*, in complex with α -galactobiose, revealed differences in the putative +1 and +1' subsites. Based on this, we propose likely candidates for the recognition of the fucosyl unit present in the B antigen. A highly conserved tyrosine (Y433) and arginine (R544) flank a spacious cavity in the direction of the C2-OH of the galactose moiety in the putative subsite +1 of *AmGH110A*. In *PdGH110B*, the presence of a lysine (K207) blocks this space. Additionally, *AmGH110A* is the first report of a CBM51-GH110 combination, only present in 10% of GH110 sequences from CAZy. The CBM51 module appears to be intimately associated with the catalytic GH110 domain, with the putative binding site aligned with the active site surface. To date, characterized CBM51s have displayed two modes of ligand binding: terminal Galactose and N-acetylglucosamine binding, or blood group antigen binding^{64,304}. *AmGH110A*-CBM51 shows higher structural similarity with the former type, sharing most of the binding site residues with the orthologue, *ZmpB*-CBM51, in complex with galactose³⁰⁴. One notable difference is the substitution of a phenylalanine (F93) over a valine, which would block the putative *N*-acetyl binding pocket, suggesting galactosyl may be preferred over HexNAc. Comparisons of *AmGH110A*-CBM51 to the A/B antigen binding CBM51 modules from GH98 endo-galactosidases enzymes, *EABase* and *Endo-ABase*⁶⁴, reveals significant binding site differences. In particular, the GH98-CBMs contain elongated loops with residues contributing to recognition of the A/B-antigen trisaccharides. Binding assays using monosaccharide and B antigen substrates may help elucidate affinity of *AmGH110A*-CBM51 towards various potential ligands.

A phylogenetic analysis of GH110 sequences revealed that both GH110 enzymes encoded by *A. muciniphila*, *AmGH110A* and *AmGH110B*, populate hitherto undescribed phylogenetic clusters. The low amino acid identity shared between the catalytic domain of these enzymes, as well as structural differences around the active site, are suggestive of yet undescribed specificities in this enzyme family.

We also report the structure of the ExtB converting enzyme, *AmGH20A*, in complex with β -GalNAc. Interestingly, a GalNAc unit was bound in the unannotated C-terminal domain, in addition to the active site pocket. This prompted us to assign this domain as a novel CBM, based on low structural similarity and high amino acid sequence divergence from known structures. A phylogenetic tree reveals sequence homologues are present in a variety of phyla, but mostly from Verrumicrobiota (34 %) and Bacteroidota (35 %). Interestingly, different sequences from the same taxonomic group populated distinct clusters, which may reflect the presence of different binding specificities. Correspondingly, although the binding

residues of *AmGH20A*-CBM appear conserved, there is a degree of divergence in the residues present on particularly two of the loops (loops 1 and 2) carrying binding residues (Chapter 2, Supplementary figure 31). Additionally, sequence alignment shows that loop 2 is elongated in some sequences, which also suggest more elaborate substrate recognition. This falls in accordance with other human glycan-binding CBM families, such as CBM32 and CBM51, whose binding site residues are not conserved throughout the family⁵⁶, however, any additional specificities remain to be determined. With regard to the appended modules to members of this new CBM family, putative HexNAcases were dominant. This potentially reflects a general role of the CBM family in conferring affinity towards HexNAc-containing glycan structures.

Structural insights gained during this project reveal common characteristics of enzymes with high conversion rates on RBCs. Flat surface topology around the active site, flanked by positive surface potential appear as important adaptations to access conjugated human *O*-glycans, and appended CBMs may also increase affinity towards RBC surfaces. These findings expand our understanding of enzyme-RBC interactions, unraveling structural hallmarks that may be useful for further mining of enzymes active on RBCs or other conjugated glycans.

Conclusions and perspectives

The work presented in this thesis expands the concept of enzymatically converting A and B blood antigens to H antigen, which defines the universal O group. We recombinantly produced 24 enzymes: 23 from the mucolytic specialist, *Akkermansia muciniphila*, and 1 from the mucin-associated *Roseburia inulinivorans*. Ultimately, enzymes with activity against all known A- and B-core antigens were discovered and characterized in more detail. We formulated one-pot enzymatic cocktails for A and B donor cells, respectively, and obtained seemingly complete conversion using the lowest enzyme concentrations, shortest incubation times, and highest RBC density to date. Importantly, we show premier evidence that targeting extended A/B antigens significantly reduces the crossmatch level and severity of both enzymatically converted A- , and B- cells, which has previously remained uncharted. However, the prevalence of crossmatch reactions occurring after enzymatic digestion, showcases the complexity and diversity of donor RBC glycan phenotypes and recipient antibody titers.

Structural analyses revealed enzymatic signatures that correlate with high conversion efficacy. We report the structure of the first blood-active GH110 α -galactosidase, which has an appended CBM51 domain, making it also the first modular enzyme from this family. The structure of a GH20 β -*N*-acetylhexosaminidase with a GalNAc-binding putative CBM from a previously unknown family was also determined. Finally, we report the structure of GH109 α -*N*-acetylgalactosaminidase exhibiting the hitherto most efficient A antigen conversion in the family. Compared to characterized orthologues, all these enzymes harbor shortened and reorganized loops around their active sites, resulting in relatively shallow catalytic sites, which are flanked by a relatively flat and positively charged area. These features appear to be an important adaptation that confers both steric and electrostatic compatibility to RBC glycoconjugates. This importance of electrostatics is further supported by the strong dependence of enzyme activity on ionic strength. Using a higher ionic strength buffer (PBS) impairs enzymatic conversion with our enzymes, compared to the low ionic strength conversion buffer. A plausible explanation, is that electrostatic screening reduces the strength of Coulombic attraction between enzymes and RBC surfaces. Our analysis sheds light on the molecular adaptations of enzymes from *A. muciniphila* to efficiently target human glycoconjugates, and highlight this organism as a source for efficient enzymes for human glycan conversion.

The success of our enzyme discovery strategy motivates further mining of the human-glycan utilizing ecological niche, to expand our repertoire of enzymes. Additionally, crystal structures of enzymes in complex with the different ABO blood group antigens would expose the molecular determinants of substrate selectivity, which may motivate rational engineering to fine-tune enzyme specificities. The occurrence of crossmatch reactions between enzymatically converted RBCs to group O recipient plasma warrants further analysis of the RBC glycan landscape and the potential promiscuity of our enzymes. Finally, it would be interesting to investigate the role of ancillary CBM domains in the catalytic efficacy of enzymes on human glycoconjugates. In summary, we shift the current text book paradigm that the A and B antigens are solely responsible for incompatibility between donor and recipients, and provide compelling evidence for the antigenicity of the extended antigens structures, laying ground for a new generation of enzymatically converted O blood.

References

1. World Health Organization. WHO model list of essential medicines - 22nd list, 2021. *Tech. Doc.* 2021 (2021).
2. Daniels, G. The molecular definition of red cell antigens. *ISBT Sci. Ser.* **5**, 300–302 (2010).
3. K, L. Ueber Agglutinationserscheinungen normalen menschlichen Blutes. *Wiener Klin. Wochenschrift* **14**, 1132–1134 (1901).
4. Landsteiner, K. Individual Differences in Human Blood. *Science*. **73**, 403–409 (1931).
5. Daniels, G. & Reid, M. E. Blood groups: The past 50 years. *Transfusion* **50**, 281–289 (2010).
6. Storry, J. R. & Olsson, M. L. The ABO blood group system revisited: A review and update. *Immunohematology* **25**, 48–59 (2009).
7. Clausen, H., Lavery, S. B., Nudelman, E., Tsuchiya, S. & Hakomori, S. Repetitive A epitope (type 3 chain A) defined by blood group A1-specific monoclonal antibody TH-1: Chemical basis of qualitative A1 and A2 distinction. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 1199–1203 (1985).
8. Clausen, H., Holmes, E. & Hakomori, S. I. Novel blood group H glycolipid antigens exclusively expressed in blood group A and AB erythrocytes (type 3 chain H). II. Differential conversion of different H substrates by A1 and A2 enzymes, and type 3 chain H expression in relation to secretor status. *J. Biol. Chem.* **261**, 1388–1392 (1986).
9. Clausen, H., Lavery, S. B., Kannagi, R. & Hakomori, S. I. Novel blood group H glycolipid antigens exclusively expressed in blood group A and AB erythrocytes (type 3 chain H). I. Isolation and chemical characterization. *J. Biol. Chem.* **261**, 1380–1387 (1986).
10. Ricci Hagman, J. *et al.* β 1,3GalNAc-T1-dependent extension of the human blood group B antigen results in a novel ABO-related glycolipid structure on erythrocytes. *Vox Sang.* **114**, 53 (2019).
11. Lomas-Francis, C. Clinical significance of antibodies to antigens in the International Society of Blood Transfusion collections, 700 series of low-incidence antigens, and 901 series of high-incidence antigens. *Immunohematology* **34**, 39–45 (2018).
12. Poole, J. & Daniels, G. Blood Group Antibodies and Their Significance in Transfusion Medicine. *Transfus. Med. Rev.* **21**, 58–71 (2007).
13. Fujisawa, K., Furukawa, K., Akabane, J. Studies on A and B blood group substances of human erythrocytes by blood group-specific enzymes. *Kita Kanto Igatu*, **13**, 19–39 (1963).
14. Harpaz, N., Flowers, H. M. & Sharon, N. Studies on B-antigenic sites of human erythrocytes by use of coffee bean α -galactosidase. *Arch. Biochem. Biophys.* **170**, 676–683 (1975).
15. Gao, H. W. *et al.* Evaluation of group A1B erythrocytes converted to type as group O: Studies of markers of function and compatibility. *Blood Transfus.* **14**, 168–174 (2016).
16. Rahfeld, P. *et al.* An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat. Microbiol.* **4**, 1475–1485 (2019).
17. Anso, I. *et al.* Turning universal O into rare Bombay type blood. *Nat. Commun.* **14**, 1765 (2023).
18. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut

- microbiome. *Front. Genet.* **5**, (2015).
19. Raba, G. & Luis, A. S. Mucin utilization by gut microbiota: recent advances on characterization of key enzymes. *Essays Biochem.* **67**, 345–353 (2023).
 20. Clausen, H. & Olsson, M. L. Towards universally acceptable blood. *Nat. Microbiol.* **4**, 1426–1427 (2019).
 21. Varki, A. & Kornfeld, S. Historical Background and Overview. *Essentials of Glycobiology* (eds. Varki, A. et al.) (CSH Laboratory Press, 2022).
 22. Seeberger, P. H. Monosaccharide Diversity. *Essentials of Glycobiology* (eds. Varki, A. et al.) (CSH Laboratory Press, 2022).
 23. McNaught, A. D. Nomenclature of carbohydrates. *Adv. Carbohydr. Chem. Biochem.* **52**, 47–177 (1997).
 24. Lebrilla, C. B., Liu, J., Widmalm, G. & Prestegard, J. H. Oligosaccharides and Polysaccharides. *Essentials of Glycobiology* (eds. Varki, A. et al.) (CSH Laboratory Press, 2022).
 25. Johansson, M. E. V., Sjövall, H. & Hansson, G. C. The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 352–361 (2013).
 26. Lillehoj, E. P., Kato, K., Lu, W. & Kim, K. C. *Cellular and Molecular Biology of Airway Mucins. International Review of Cell and Molecular Biology* vol. 303 (Elsevier, 2013).
 27. Bouvet, V. & Ben, R. N. Antifreeze glycoproteins: Structure, conformation, and biological applications. *Cell Biochem. Biophys.* **39**, 133–144 (2003).
 28. Duman, J. G. Animal ice-binding (antifreeze) proteins and glycolipids: An overview with emphasis on physiological function. *J. Exp. Biol.* **218**, 1846–1855 (2015).
 29. Vasudevan, D., Takeuchi, H., Johar, S. S., Majerus, E. & Haltiwanger, R. S. Peters plus syndrome mutations disrupt a noncanonical ER quality-control mechanism. *Curr. Biol.* **25**, 286–295 (2015).
 30. Xu, C. & Ng, D. T. W. O-mannosylation: The other glycan player of ER quality control. *Semin. Cell Dev. Biol.* **41**, 129 (2015).
 31. Williams, C. *et al.* Assessing the role of surface glycans of extracellular vesicles on cellular uptake. *Sci. Rep.* **9**, 1–14 (2019).
 32. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).
 33. Drula, E. *et al.* The carbohydrate-active enzyme database: Functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).
 34. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).
 35. Bissaro, B., Monsan, P., Fauré, R. & O’Donohue, M. J. Glycosynthesis in a waterworld: New insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* **467**, 17–35 (2015).
 36. Bourne, Y. & Henrissat, B. Glycoside hydrolases and glycosyltransferases: Families and functional modules. *Curr. Opin. Struct. Biol.* **11**, 593–600 (2001).

37. Biely, P. Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnol. Adv.* **30**, 1575–1588 (2012).
38. Lombard, V. *et al.* A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J.* **432**, 437–444 (2010).
39. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. & Henrissat, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**, 1–14 (2013).
40. Henrissat, B. *et al.* 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 (1995).
41. Khosla, C. & Harbury, P. B. Modular enzymes. *Nature* **409**, 247–252 (2001).
42. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781 (2004).
43. Koshland, D. E. Stereochemistry and the Mechanism of Enzymatic Reactions. *Biol. Rev.* **28**, 416–436 (1953).
44. Puchart, V. Glycoside phosphorylases: Structure, catalytic properties and biotechnological potential. *Biotechnol. Adv.* **33**, 261–276 (2015).
45. Bojarová, P. & Křen, V. Glycosidases: a key to tailored carbohydrates. *Trends Biotechnol.* **27**, 199–209 (2009).
46. Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases [1]. *Biochem. J.* **321**, 557–559 (1997).
47. Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R. & Pettersson, G. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. Separation of functional domains. *FEBS Lett.* **204**, 223–227 (1986).
48. Tormo, J. *et al.* Crystal structure of a bacterial family-III cellulose-binding domain: A general mechanism for attachment to cellulose. *EMBO J.* **15**, 5739–5751 (1996).
49. Lehtiö, J. *et al.* The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 484–489 (2003).
50. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J. & Williamson, M. P. The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *J. Biol. Chem.* **275**, 41137–41142 (2000).
51. Pell, G. *et al.* Importance of hydrophobic and polar residues in ligand binding in the family 15 carbohydrate-binding module from *Cellvibrio japonicus* Xyn10C. *Biochemistry* **42**, 9316–9323 (2003).
52. Boraston, A. B. *et al.* Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and ‘galactose-binding’ domains. *J. Mol. Biol.* **327**, 659–669 (2003).
53. Henshaw, J. L. *et al.* The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. *J. Biol. Chem.* **279**, 21552–21559 (2004).

54. Pires, V. M. R. *et al.* The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5A in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. *J. Biol. Chem.* **279**, 21560–21568 (2004).
55. Baroroh, U. *et al.* The Importance of Surface-Binding Site towards Starch-Adsorptivity Level in α - Amylase: A Review on Structural Point of View. *Enzyme Res.* **2017**, (2017).
56. Ficko-Blean, E. & Boraston, A. B. Insights into the recognition of the human glycome by microbial carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* **22**, 570–577 (2012).
57. Ficko-Blean, E. *et al.* Carbohydrate recognition by an architecturally complex α -n-acetylglucosaminidase from *Clostridium perfringens*. *PLoS One* **7**, (2012).
58. Ficko-Blean, E. & Boraston, A. B. N-Acetylglucosamine Recognition by a Family 32 Carbohydrate-Binding Module from *Clostridium perfringens* NagH. *J. Mol. Biol.* **390**, 208–220 (2009).
59. Ficko-Blean, E. & Boraston, A. B. Structural analysis of a bacterial exo-d-N-acetylglucosaminidase in complex with an unusual disaccharide found in class III mucin. *Glycobiology* **22**, 590–595 (2012).
60. Abbott, D. W., Eirín-López, J. M. & Boraston, A. B. Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules. *Mol. Biol. Evol.* **25**, 155–167 (2008).
61. Moustafa, I. *et al.* Sialic acid recognition by *Vibrio cholerae* neuraminidase. *J. Biol. Chem.* **279**, 40819–40826 (2004).
62. Xu, G. *et al.* Crystal Structure of the NanB Sialidase from *Streptococcus pneumoniae*. *J. Mol. Biol.* **384**, 436–449 (2008).
63. Boraston, A. B., Wang, D. & Burke, R. D. Blood group antigen recognition by a *Streptococcus pneumoniae* virulence factor. *J. Biol. Chem.* **281**, 35263–35271 (2006).
64. Gregg, K. J., Finn, R., Abbott, D. W. & Boraston, A. B. Divergent modes of glycan recognition by a new family of carbohydrate-binding modules. *J. Biol. Chem.* **283**, 12604–12613 (2008).
65. Anderson, K. M. *et al.* A clostridial endo- β -galactosidase that cleaves both blood group A and B glycotopes: The first member of a new glycoside hydrolase family, GH98. *J. Biol. Chem.* **280**, 7720–7728 (2005).
66. Singh, A. K. *et al.* Unravelling the Multiple Functions of the Architecturally Intricate *Streptococcus pneumoniae* β -galactosidase, BgaA. *PLoS Pathog.* **10**, (2014).
67. Thursby, E. & Juge, N. Introduction to the human gut microbiota. *Biochem. J.* **474**, 1823–1836 (2017).
68. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **14**, 1–14 (2016).
69. Neish, A. S. Microbes in Gastrointestinal Health and Disease. *Gastroenterology* **136**, 65–80 (2009).
70. Cho, I. & Blaser, M. J. The human microbiome: At the interface of health and disease. *Nat. Rev. Genet.* **13**, 260–270 (2012).
71. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141

- (2014).
72. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2015).
 73. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
 74. Briggs, J. A., Grondin, J. M. & Brumer, H. Communal living: glycan utilization by the human gut microbiota. *Environ. Microbiol.* **23**, 15–35 (2021).
 75. Jakobsson, H. E. *et al.* Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by Caesarean section. *Gut* **63**, 559–566 (2014).
 76. Schmidt, T. S. B., Raes, J. & Bork, P. The Human Gut Microbiome: From Association to Modulation. *Cell* **172**, 1198–1215 (2018).
 77. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4554–4561 (2011).
 78. Biedermann, L. *et al.* Smoking Cessation Induces Profound Changes in the Composition of the Intestinal Microbiota in Humans. *PLoS One* **8**, (2013).
 79. Dubinkina, V. B. *et al.* Links of gut microbiota composition with alcohol dependence syndrome and alcoholic liver disease. *Microbiome* **5**, 1–14 (2017).
 80. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
 81. Leeming, E. R., Johnson, A. J., Spector, T. D. & Roy, C. I. L. Effect of diet on the gut microbiota: Rethinking intervention duration. *Nutrients* **11**, 1–28 (2019).
 82. Glowacki, R. W. P. & Martens, E. C. If you eat it or secrete it, they will grow: The expanding list of nutrients utilized by human gut bacteria. *J. Bacteriol.* **203**, (2020).
 83. Cockburn, D. W. & Koropatkin, N. M. Polysaccharide Degradation by the Intestinal Microbiota and Its Influence on Human Health and Disease. *J. Mol. Biol.* **428**, 3230–3252 (2016).
 84. Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* **12**, 661–672 (2014).
 85. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **7**, 189–200 (2016).
 86. Corrêa-Oliveira, R., Fachi, J. L., Vieira, A., Sato, F. T. & Vinolo, M. A. R. Regulation of immune cell function by short-chain fatty acids. *Clin. Transl. Immunol.* **5**, 1–8 (2016).
 87. Statovci, D., Aguilera, M., MacSharry, J. & Melgar, S. The impact of western diet and nutrients on the microbiota and immune response at mucosal interfaces. *Front. Immunol.* **8**, (2017).
 88. García-Montero, C. *et al.* Nutritional components in western diet versus mediterranean diet at the gut microbiota-immune system interplay. implications for health and disease. *Nutrients* **13**, 1–53 (2021).

89. Greenwood-Van Meerveld, B., Johnson, A. C., & Grundy, D. *Gastrointestinal Physiology and Function. Gastrointestinal Pharmacology* (Springer International Publishing, 2017). doi:10.1007/164.
90. Lamont, R. J., Koo, H. & Hajishengallis, G. The oral microbiota: dynamic communities and host interactions. *Nat. Rev. Microbiol.* **16**, 745–759 (2018).
91. Baker, J. L., Mark Welch, J. L., Kauffman, K. M., McLean, J. S. & He, X. The oral microbiome: diversity, biogeography and human health. *Nat. Rev. Microbiol.* (2023) doi:10.1038/s41579-023-00963-6.
92. May, M. & Abrams, J. A. Emerging Insights into the Esophageal Microbiome. *Curr. Treat. Options Gastroenterol.* **16**, 72–85 (2018).
93. Malfertheiner, P. *et al.* Helicobacter pylori infection. *Nat. Rev. Dis. Prim.* **9**, (2023).
94. Müller, M., Canfora, E. E. & Blaak, E. E. Gastrointestinal transit time, glucose homeostasis and metabolic health: Modulation by dietary fibers. *Nutrients* **10**, (2018).
95. Procházková, N. *et al.* Advancing human gut microbiota research by considering gut transit time. *Gut* **72**, 180–191 (2023).
96. Pereira, F. C. & Berry, D. Microbial nutrient niches in the gut. *Environ. Microbiol.* **19**, 1366–1378 (2017).
97. Derrien, M., Collado, M. C., Ben-Amor, K., Salminen, S. & De Vos, W. M. The mucin degrader *Akkermansia muciniphila* is an abundant resident of the human intestinal tract. *Appl. Environ. Microbiol.* **74**, 1646–1648 (2008).
98. Tropini, C., Earle, K. A., Huang, K. C. & Sonnenburg, J. L. The Gut Microbiome: Connecting Spatial Organization to Function. *Cell Host Microbe* **21**, 433–442 (2017).
99. Hansson, G. C. Role of mucus layers in gut infection and inflammation. *Curr. Opin. Microbiol.* **15**, 57–62 (2012).
100. Corfield, A. P. Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 236–252 (2015).
101. Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339–1353.e21 (2016).
102. Larsson, J. M. H. *et al.* Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. *Inflamm. Bowel Dis.* **17**, 2299–2307 (2011).
103. Johansson, M. E. V. *et al.* Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut* **63**, 281–291 (2014).
104. Bansil, R. & Turner, B. S. Mucin structure, aggregation, physiological functions and biomedical applications. *Curr. Opin. Colloid Interface Sci.* **11**, 164–170 (2006).
105. Luis, A. S. & Hansson, G. C. Intestinal mucus and their glycans: A habitat for thriving microbiota. *Cell Host Microbe* **31**, 1087–1100 (2023).
106. Audie, J. . *et al.* Human Mucin Genes in Respiratory, Digestive, and Reproductive Tracts Ascertained. **41**, 1479–1481 (1993).

107. Ermund, A., Schütte, A., Johansson, M. E. V., Gustafsson, J. K. & Hansson, G. C. Studies of mucus in mouse stomach, small intestine, and colon. I. Gastrointestinal mucus layers have different properties depending on location as well as over the Peyer's patches. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **305**, 341–347 (2013).
108. Lueschow, S. R. & McElroy, S. J. The Paneth Cell: The Curator and Defender of the Immature Small Intestine. *Front. Immunol.* **11**, 1–12 (2020).
109. Schütte, A. *et al.* Microbial-induced meprin β cleavage in MUC2 mucin and a functional CFTR channel are required to release anchored small intestinal mucus. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12396–12401 (2014).
110. Birchenough, G. M. H., Johansson, M. E. V., Gustafsson, J. K., Bergström, J. H. & Hansson, G. C. New developments in goblet cell mucus secretion and function. *Mucosal Immunol.* **8**, 712–719 (2015).
111. Paone, P. & Cani, P. D. Mucus barrier, mucins and gut microbiota: The expected slimy partners? *Gut* **69**, 2232–2243 (2020).
112. Hansson, G. C. Mucins and the Microbiome. *Annu. Rev. Biochem.* **89**, 769–793 (2020).
113. Bennett, E. P. *et al.* Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22**, 736–756 (2012).
114. González-Ramírez, A. M. *et al.* Structural basis for the synthesis of the core 1 structure by C1GalT1. *Nat. Commun.* **13**, (2022).
115. Martens, E. C., Neumann, M. & Desai, M. S. Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier. *Nat. Rev. Microbiol.* **16**, 457–470 (2018).
116. Robbe, C., Capon, C., Coddeville, B. & Michalski, J. C. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J.* **384**, 307–316 (2004).
117. Iwai, T. *et al.* Molecular cloning and characterization of a novel UDP-GlcNAc: GalNAc-peptide β 1,3-N-acetylglucosaminyltransferase (β 3Gn-T6), an enzyme synthesizing the core 3 structure of O-glycans. *J. Biol. Chem.* **277**, 12802–12809 (2002).
118. Robbe-Masselot, C. *et al.* Expression of a core 3 disialyl-Lex hexasaccharide in human colorectal cancers: A potential marker of malignant transformation in colon. *J. Proteome Res.* **8**, 702–711 (2009).
119. Xia, L. Core 3-derived O-glycans are essential for intestinal mucus barrier function. *Methods in Enzymology* vol. 479 (Elsevier Inc., 2010).
120. Marcos, N. T. *et al.* Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated Sialyl-Tn antigen. *Cancer Res.* **64**, 7050–7057 (2004).
121. Thomsson, K. A., Karlsson, H. & Hansson, G. C. Sequencing of sulfated oligosaccharides from mucins by liquid chromatography and electrospray ionization tandem mass spectrometry. *Anal. Chem.* **72**, 4543–4549 (2000).
122. Thomsson, K. A., Bäckström, M., Holmén Larsson, J. M., Hansson, G. C. & Karlsson, H. Enhanced detection of sialylated and sulfated glycans with negative ion mode nanoliquid chromatography/mass spectrometry at high pH. *Anal. Chem.* **82**, 1470–1477 (2010).

123. Yao, Y. *et al.* Mucus sialylation determines intestinal host-commensal homeostasis. *Cell* **185**, 1172-1188 (2022).
124. Rouquier, S. *et al.* Molecular cloning of a human genomic region containing the H blood group $\alpha(1,2)$ fucosyltransferase gene and two H locus-related DNA restriction fragments. Isolation of a candidate for the human Secretor blood group locus. *J. Biol. Chem.* **270**, 4632–4639 (1995).
125. Rossez, Y. *et al.* Almost all human gastric mucin O-glycans harbor blood group A, B or H antigens and are potential binding sites for *Helicobacter pylori*. *Glycobiology* **22**, 1193–1206 (2012).
126. Yamamoto, F. I., Clausen, H., White, T., Marken, J. & Hakomori, S. I. Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**, 229–233 (1990).
127. Henry, S., Oriol, R. & Samuelsson, B. Lewis Histo-Blood Group System and Associated Secretory Phenotypes. *Vox Sang.* **69**, 166–182 (1995).
128. Harduin-Lepers, A. *et al.* The human sialyltransferase family. *Biochimie* **83**, 727–737 (2001).
129. Kaoutari, A. El, Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11**, 497–504 (2013).
130. Leth, M. L. *et al.* Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* **3**, 570–580 (2018).
131. Ejby, M. *et al.* Two binding proteins of the ABC transporter that confers growth of *Bifidobacterium animalis* subsp. *lactis* ATCC27673 on β -mannan possess distinct manno-oligosaccharide-binding profiles. *Mol. Microbiol.* **112**, 114–130 (2019).
132. Theilmann, M. C., Fredslund, F., Svensson, B., Leggio, L. Lo & Hachem, M. A. Substrate preference of an ABC importer corresponds to selective growth on -(1,6)-galactosides in *Bifidobacterium animalis* subsp. *lactis*. *J. Biol. Chem.* **294**, 11701–11711 (2019).
133. Berlemont, R. & Martiny, A. C. Glycoside Hydrolases across Environmental Microbial Communities. *PLoS Comput. Biol.* **12**, 1–16 (2016).
134. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* **54**, 1469–1476 (2004).
135. Tramontano, M. *et al.* Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nat. Microbiol.* **3**, 514–522 (2018).
136. Collado, M. C., Derrien, M., Isolauri, E., De Vos, W. M. & Salminen, S. Intestinal integrity and *Akkermansia muciniphila*, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl. Environ. Microbiol.* **73**, 7767–7770 (2007).
137. Kostopoulos, I. *et al.* *Akkermansia muciniphila* uses human milk oligosaccharides to thrive in the early life conditions in vitro. *Sci. Rep.* **10**, 1–17 (2020).
138. Ojima, M. N. *et al.* Diversification of a Fucosyltransferase Transporter within the Genus *Bifidobacterium*. *Appl. Environ. Microbiol.* **88**, (2022).
139. Davey, L. E. *et al.* A genetic system for *Akkermansia muciniphila* reveals a role for mucin foraging

- in gut colonization and host sterol biosynthesis gene expression. *Nat. Microbiol.* **8**, 1450–1467 (2023).
140. Cani, P. D. & de Vos, W. M. Next-generation beneficial microbes: The case of *Akkermansia muciniphila*. *Front. Microbiol.* **8**, 1–8 (2017).
 141. Cani, P. D., Depommier, C., Derrien, M., Everard, A. & de Vos, W. M. *Akkermansia muciniphila*: paradigm for next-generation beneficial microorganisms. *Nat. Rev. Gastroenterol. Hepatol.* **19**, 625–637 (2022).
 142. Everard, A. *et al.* Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9066–9071 (2013).
 143. Ottman, N. *et al.* Genome-Scale Model and Omics Analysis of Metabolic Capacities of *Akkermansia muciniphila* Reveal a Preferential Mucin_Degrading Lifestyle. *Appl. Environ. Microbiol.* **83**, (2017).
 144. Fang, J. *et al.* Slimy partners: the mucus barrier and gut microbiome in ulcerative colitis. *Exp. Mol. Med.* **53**, 772–787 (2021).
 145. Yoon, H. S. *et al.* *Akkermansia muciniphila* secretes a glucagon-like peptide-1-inducing protein that improves glucose homeostasis and ameliorates metabolic disease in mice. *Nat. Microbiol.* **6**, 563–573 (2021).
 146. Plovier, H. *et al.* A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat. Med.* **23**, 107–113 (2017).
 147. Wang, J. *et al.* The outer membrane protein Amuc_1100 of *Akkermansia muciniphila* promotes intestinal 5-HT biosynthesis and extracellular availability through TLR2 signalling. *Food Funct.* **12**, 3597–3610 (2021).
 148. Ou, Z. *et al.* Protective effects of *Akkermansia muciniphila* on cognitive deficits and amyloid pathology in a mouse model of Alzheimer's disease. *Nutr. Diabetes* **10**, (2020).
 149. Bush, A. I. & Tanzi, R. E. The galvanization of β -amyloid in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7317–7319 (2002).
 150. Zhang, T., Ji, X., Lu, G. & Zhang, F. The potential of *Akkermansia muciniphila* in inflammatory bowel disease. *Appl. Microbiol. Biotechnol.* **105**, 5785–5794 (2021).
 151. Shuoker, B. *et al.* Sialidases and fucosidases of *Akkermansia muciniphila* are crucial for growth on mucin and nutrient sharing with mucus-associated gut bacteria. *Nat. Commun.* **14**, 1833 (2023).
 152. Trastoy, B., Naegeli, A., Anso, I., Sjögren, J. & Guerin, M. E. Structural basis of mammalian mucin processing by the human gut O-glycopeptidase OgpA from *Akkermansia muciniphila*. *Nat. Commun.* **11**, 1–14 (2020).
 153. Medley, B. J. *et al.* A previously uncharacterized O-glycopeptidase from *Akkermansia muciniphila* requires the Tn-antigen for cleavage of the peptide bond. *J. Biol. Chem.* **298**, 102439 (2022).
 154. Shon, D. J. *et al.* An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 21299–21307 (2020).
 155. Taleb, V. *et al.* Structural and mechanistic insights into the cleavage of clustered O-glycan

- patches-containing glycoproteins by mucinases of the human gut. *Nat. Commun.* **13**, 1–15 (2022).
156. Davey, L. *et al.* Mucin foraging enables *Akkermansia muciniphila* to compete against other microbes in the gut and to modulate host sterol biosynthesis. *Res. Sq.* (2022).
 157. Crouch, L. I. *et al.* Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown. *Nat. Commun.* **11**, (2020).
 158. Kosciow, K. & Deppenmeier, U. Characterization of three novel β -galactosidases from *Akkermansia muciniphila* involved in mucin degradation. *Int. J. Biol. Macromol.* **149**, 331–340 (2020).
 159. Li, C. C. *et al.* Nucleotide binding as an allosteric regulatory mechanism for *Akkermansia muciniphila* β -N-acetylhexosaminidase Am2136. *Gut Microbes* **14**, (2022).
 160. Katoh, T. *et al.* A bacterial sulfoglycosidase highlights mucin O-glycan breakdown in the gut ecosystem. *Nat. Chem. Biol.* **19**, 778–789 (2023).
 161. Luis, A. S. *et al.* A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* **598**, 332–337 (2021).
 162. Liu, Q. P. *et al.* Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **25**, 454–464 (2007).
 163. Shimada, Y. *et al.* α -N-Acetylglucosaminidase from *Bifidobacterium bifidum* specifically hydrolyzes α -linked N-acetylglucosamine at nonreducing terminus of O-glycan on gastric mucin. *Appl. Microbiol. Biotechnol.* **99**, 3941–3948 (2015).
 164. Koutsioulis, D., Landry, D. & Guthrie, E. P. Novel endo- α - N -acetylgalactosaminidases with broader substrate specificity. **18**, 799–805 (2008).
 165. Kiyohara, M. *et al.* α - N -Acetylgalactosaminidase from Infant-associated *Bifidobacteria* Belonging to Novel Glycoside Hydrolase Family 129 Is Implicated in Alternative Mucin Degradation. *J. Biol. Chem.* **287**, 693–700 (2012).
 166. Klinken, S. P. Cells in focus Red blood cells. *Int. J. Biochem. Cell Biol.* **34**, 1513–1518 (2002).
 167. Bua, R. O. *et al.* N-glycomics of human erythrocytes. *Int. J. Mol. Sci.* **22**, 1–14 (2021).
 168. Schroeder, H. W. & Cavacini, L. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, S41–S52 (2010).
 169. Crout, D. H. G. & Vic, G. ChemInform Abstract: Glycosidases and Glycosyl Transferases in Glycoside and Oligosaccharide Synthesis. *ChemInform* **29**, 98-111 (2010).
 170. ISBT. Table of Blood Group Systems. *Isbt* **6**, 9–11 (2017).
 171. Jajosky, R. P. *et al.* ABO blood group antigens and differential glycan expression: Perspective on the evolution of common human enzyme deficiencies. *iScience.* **26**, 1-31 (2023).
 172. Hakomori, S. itiroh, Stellner, K. & Watanabe, K. Four antigenic variants of blood group A glycolipid: Examples of highly complex, branched chain glycolipid of animal cell membrane. *Biochem. Biophys. Res. Commun.* **49**, 1061–1068 (1972).
 173. Furukawa, J. I. *et al.* Comprehensive Glycomic Approach Reveals Novel Low-Molecular-Weight

- Blood Group-Specific Glycans in Serum and Cerebrospinal Fluid. *J. Proteome Res.* **20**, 2812–2822 (2021).
174. Cserti, C. M. & Dzik, W. H. The ABO blood group system and Plasmodium falciparum malaria. *Blood* **110**, 2250–2258 (2007).
 175. Goel, R. *et al.* ABO blood group and COVID-19: a review on behalf of the ISBT COVID-19 Working Group. *Vox Sang.* **116**, 849–861 (2021).
 176. Boukhari, R. *et al.* ABO Blood Group Incompatibility Protects Against SARS-CoV-2 Transmission. *Front. Microbiol.* **12**, (2022).
 177. Oriol, R. Genetic control of the fucosylation of ABH precursor chains. Evidence for new epistatic interactions in different cells and tissue. *J. Immunogenetics.* **17**, 235–245 (1990).
 178. Saliva, T. & Lymphocytes, C. S. F. H Blood Group System. **2**, 489–498 (2012).
 179. Ravn, V. & Dabelsteen, E. Tissue distribution of histo-blood group antigens. *Apms* **108**, 1–28 (2000).
 180. Chester, M. A. & Olsson, M. L. The ABO blood group gene: A locus of considerable genetic diversity. *Transfus. Med. Rev.* **15**, 177–200 (2001).
 181. Kelly, R. J., Rouquier, S., Giorgi, D., Lennon, G. G. & Lowe, J. B. Sequence and expression of a candidate for the human Secretor blood group $\alpha(1,2)$ fucosyltransferase gene (FUT2). Homozygosity for an enzyme- inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J. Biol. Chem.* **270**, 4640–4649 (1995).
 182. Bhatia, H. M. The ‘ Bombay ’ (Oh) Blood Group. **53**, 152–153 (1987).
 183. Kelly, R. J. *et al.* Molecular basis for H blood group deficiency in Bombay (O(h)) and para- Bombay individuals. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5843–5847 (1994).
 184. Levine, P., Robinson, E., Celano, M., Briggs, O. & Falkenburg, L. Gene interaction resulting in suppression of blood group substance B. *Blood* **10**, 1100–1108 (1955).
 185. Schenkel-Brunner, H. Studies on Blood-Groups A₁ and A₂. *Eur. J. Biochem.* **122**, 511–514 (2005).
 186. Svensson, L., Rydberg, L., De Mattos, L. C. & Henry, S. M. Blood group A1 and A2 revisited: An immunochemical analysis. *Vox Sang.* **96**, 56–61 (2009).
 187. Hakomori, S. I. Antigen structure and genetic basis of histo-blood groups A, B and O: Their changes associated with human cancer. *Biochim. Biophys. Acta - Gen. Subj.* **1473**, 247–266 (1999).
 188. Cameron, H. S., Szczepaniak, D. & Weston, B. W. Expression of human chromosome 19p $\alpha(1,3)$ -fucosyltransferase genes in normal tissues: Alternative splicing, polyadenylation, and isoforms. *J. Biol. Chem.* **270**, 20112–20122 (1995).
 189. Combs, M. R. Lewis blood group system review. *Immunohematology* **25**, 112–118 (2009).
 190. Nishihara, S. *et al.* Molecular genetic analysis of the human Lewis histo-blood group system. *J. Biol. Chem.* **269**, 29271–29278 (1994).
 191. Cass, D. M. . M. and L. E. . Glycosphingolipids with Lewis Blood Group Activity : Uptake by Human

- Erythrocytes. *Am. Assoc. Adv. Sci.* **164**, 553–555 (2023).
192. Fukuda, M., Fukuda, M. N. & Hakomori, S. Developmental change and genetic defect in the carbohydrate structure of band 3 glycoprotein of human erythrocyte membrane. *J. Biol. Chem.* **254**, 3700–3703 (1979).
 193. Magnet, A. D. & Fukuda, M. Expression of the large I antigen forming P-I,6-A⁺-acetylglucosaminyltransferase in various tissues of adult mice. *Glycobiology* **7**, 285–295 (1997).
 194. Cheong, S. S. *et al.* Pleiotropic effect of a novel mutation in GCNT2 causing congenital cataract and a rare adult i blood group phenotype. *Hum. Genome Var.* **4**, 1–5 (2017).
 195. Stenfelt, L. *et al.* Missense mutations in the C-terminal portion of the B4GALNT2-encoded glycosyltransferase underlying the Sd(a⁻) phenotype. *Biochem. Biophys. Reports* **19**, 100659 (2019).
 196. Groux-Degroote, S., Vicogne, D., Cogez, V., Schulz, C. & Harduin-Lepers, A. B4GALNT2 Controls Sda and SLex Antigen Biosynthesis in Healthy and Cancer Human Colon. *ChemBioChem* **22**, 3381–3390 (2021).
 197. Kaczmarek, R., Buczkowska, A., Mikołajewicz, K., Krotkiewski, H. & Czerwinski, M. P1PK, GLOB, and FORS Blood Group Systems and GLOB Collection: Biochemical and clinical aspects. Do we understand it all yet? *Transfus. Med. Rev.* **28**, 126–136 (2014).
 198. Stenfelt, L., Hellberg, A., Westman, J. S. & Olsson, M. L. The P1PK blood group system: Revisited and resolved. *Immunohematology* **36**, 99–103 (2020).
 199. Å. Hellberg, J.S. Westman, and M. L. O. An update on the Glob blood group system. *Immunohematology* **36**, 58–59 (2020).
 200. Hellberg, Å., Westman, S., Thuresson, B. & Olsson, M. L. P1PK: The blood group system that changed its name and expanded. *Immunohematology* **29**, 25–33 (2013).
 201. Thuresson, B., Westman, J. S. & Olsson, M. L. Identification of a novel A4GALT exon reveals the genetic basis of the P1/P2 histo-blood groups. *Blood* **117**, 678–687 (2011).
 202. Westman, J. S. *et al.* Identification of the molecular and genetic basis of PX2, a glycosphingolipid blood group antigen lacking on globoside-deficient erythrocytes. *J. Biol. Chem.* **290**, 18505–18518 (2015).
 203. Kaczmarek, R. *et al.* Human Gb3/CD77 synthase reveals specificity toward two or four different acceptors depending on amino acid at position 211, creating Pk, P1 and NOR blood group antigens. *Biochem. Biophys. Res. Commun.* **470**, 168–174 (2016).
 204. Kaczmarek, R. *et al.* Single nucleotide polymorphisms in A4GALT spur extra products of the human Gb3/CD77 synthase and underlie the P1PK blood group system. *PLoS One* **13**, (2018).
 205. Svensson, L. *et al.* Forssman expression on human erythrocytes: Biochemical and genetic evidence of a new histo-blood group system. *Blood* **121**, 1459–1468 (2013).
 206. Yamamoto, M., Tarasco, M. C., Cid, E., Kobayashi, H. & Yamamoto, F. ABO blood group A transferase and its codon 69 substitution enzymes synthesize FORS1 antigen of FORS blood group system. *Sci. Rep.* **9**, 1–11 (2019).

207. Hult, A. K. & Olsson, M. L. The FORS awakens: Review of a blood group system reborn. *Immunohematology* **33**, 64–72 (2017).
208. Black, C. A. A brief history of the discovery of the immunoglobulins and the origin of the modern immunoglobulin nomenclature. *Immunol. Cell Biol.* **75**, 65–68 (1997).
209. Keyt, B. A., Baliga, R., Sinclair, A. M., Carroll, S. F. & Peterson, M. S. Structure, function, and therapeutic use of IgM antibodies. *Antibodies* **9**, 1–35 (2020).
210. Vidarsson, G., Dekkers, G. & Rispens, T. IgG subclasses and allotypes: From structure to effector functions. *Front. Immunol.* **5**, 1–17 (2014).
211. Papadea, C. & Check, I. J. Human immunoglobulin g and immunoglobulin g subclasses: Biochemical, genetic, and clinical aspects. *Crit. Rev. Clin. Lab. Sci.* **27**, 27–58 (1989).
212. Woof, J. M. & Mestecky, J. Mucosal Immunoglobulins. *Mucosal Immunol. Fourth Ed.* **1–2**, 287–324 (2015).
213. Rogentine, G. N., Rowe, D. S., Bradley, J., Waldmann, T. A. & Fahey, J. L. Metabolism of human immunoglobulin D (IgD). *J. Clin. Invest.* **45**, 1467–1478 (1966).
214. Edholm, E. S., Bengten, E. & Wilson, M. Insights into the function of IgD. *Dev. Comp. Immunol.* **35**, 1309–1316 (2011).
215. Platts-Mills, T. A. E. The role of immunoglobulin E in allergy and asthma. *Am. J. Respir. Crit. Care Med.* **164**, (2001).
216. Winter, W. E., Hardt, N. S. & Fuhrman, S. Immunoglobulin E: Importance in parasitic infections and hypersensitivity responses. *Arch. Pathol. Lab. Med.* **124**, 1382–1385 (2000).
217. Ukita, M. *et al.* IgG Subclasses of Anti-A and Anti-B Antibodies Bound to the Cord Red Cells in ABO Incompatible Pregnancies. *Vox Sang.* **56**, 181–186 (1989).
218. Branch, D. R. Anti-A and anti-B: what are they and where do they come from? *Transfusion* **55**, S74–S79 (2015).
219. Bullock, T. *et al.* Bombay phenotype (O h) and high-titer anti-H in pregnancy: two case reports and a review of the literature. *Transfusion* **58**, 2766–2772 (2018).
220. Harris, S. B., Josephson, C. D., Kost, C. B. & Hillyer, C. D. Nonfatal intravascular hemolysis in a pediatric patient after transfusion of a platelet unit with high-titer anti-A. *Transfusion* **47**, 1412–1417 (2007).
221. Balbuena-Merle, R., West, F. B., Tormey, C. A. & Hendrickson, J. E. Fatal acute hemolytic transfusion reaction due to anti-B from a platelet apheresis unit stored in platelet additive solution. *Transfusion* **59**, 1911–1915 (2019).
222. Subramaniyan, R. Serological characteristics of Lewis antibodies and their clinical significance – A case series. *Hematol. Transfus. Cell Ther.* **45**, 159–164 (2023).
223. Combs, M. R. An update on the Lewis blood group system. *Immunohematology* **35**, 65–66 (2020).
224. Chaplin, H., Hunter, V. L., Malecek, A. C., Kilzer, P. & Rosche, M. E. Clinically significant allo-anti-I in an I-negative patient with massive hemorrhage. *Transfusion* **26**, 57–61 (1986).

225. Hult, A. K. & Olsson, M. L. May the FORS be with you: A system sequel. *Immunohematology* **36**, 14–18 (2020).
226. Levy, G. N. & Aminoff, D. Purification and properties of α -N-acetylgalactosaminidase from *Clostridium perfringens*. *J. Biol. Chem.* **255**, 11737–11742 (1980).
227. Goldstein, J., Siviglia, G., Hurst, R., Lenny, L. & Reich, L. Group B Erythrocytes Enzymatically Converted to Group O Survive Normally in A, B, and O Individuals. *Science*. **215**, 168–170 (1982).
228. Olsson, M. L. *et al.* Universal red blood cells - Enzymatic conversion of blood group A and B antigens. *Transfus. Clin. Biol.* **11**, 33–39 (2004).
229. Goldstein, J. Preparation of transfusable red cells by enzymatic conversion. *The Red Cell: Sixth Ann Arbor Conference*. 139–157 (1984).
230. Zhu, A., Monahan, C., Wang, Z. K. & Goldstein, J. Expression, purification, and characterization of recombinant α -N-acetylgalactosaminidase produced in the yeast *Pichia pastoris*. *Protein Expr. Purif.* **8**, 456–462 (1996).
231. Lenny, L. L., Hurst, R., Goldstein, J., Benjamin, L. J. & Jones, R. L. Single-unit transfusions of RBC enzymatically converted from group B to group O to A and O normal volunteers. *Blood* **77**, 1383–1388 (1991).
232. Lenny, L. L., Hurst, R. & Galbraith, R. A. Transfusions to group O subjects of 2 units of red cells enzymatically converted from group B to group O. *Transfusion* **34**, 209–214 (1994).
233. Kruskall, M. S. *et al.* Transfusion to blood group A and O patients of group B RBCs that have been enzymatically converted to group O. *Transfusion* **40**, 1290–1298 (2000).
234. Kasowaki, S., *et al.* Isolation and Characterization of a Blood Group A Substance-degrading α -N-Acetylgalactosaminidase from *Acremonium* sp. *Agric. Biol. Chem.* **53**, 111–120 (1989).
235. Izumi, K., Yamamoto, K., Tochikura, T. & Hirabayashi, Y. Serological study using α -N-acetylgalactosaminidase from *Acremonium* sp. *BBA - Gen. Subj.* **1116**, 72–74 (1992).
236. Shoji, T., *et al.* Molecular cloning of a full-length cDNA for human α -N-acetylgalactosaminidase (α -galactosidase B). *Biochem. Biophys. Res. Commun.* **163**, 1–2 (1989).
237. Kulik, N. *et al.* The α -galactosidase type A gene *aglA* from *Aspergillus niger* encodes a fully functional α -N-acetylgalactosaminidase. *Glycobiology* **20**, 1410–1419 (2010).
238. Hatada, Y. *et al.* Hyper-production of an isomalto-dextranase of an *Arthrobacter* sp. by a proteases-deficient *Bacillus subtilis*: Sequencing, properties, and crystallization of the recombinant enzyme. *Appl. Microbiol. Biotechnol.* **65**, 583–592 (2004).
239. Ichinose, H. *et al.* A β -L-arabinopyranosidase from *Streptomyces avermitilis* is a novel member of glycoside hydrolase family 27. *J. Biol. Chem.* **284**, 25097–25105 (2009).
240. Lansky, S. *et al.* Structure-specificity relationships in Abp, a GH27 β -L-arabinopyranosidase from *Geobacillus stearothermophilus* T6. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 2994–3012 (2014).
241. McKee, L. S. & Brumer, H. Growth of *Chitinophaga pinensis* on plant cell wall glycans and characterisation of a glycoside hydrolase family 27 β -L-arabinopyranosidase implicated in

- arabinogalactan utilisation. *PLoS One* **10**, 1–22 (2015).
242. BRUMER III, H., SIMS, P. F. G. & SINNOTT, M. L. Lignocellulose degradation by *Phanerochaete chrysosporium*: purification and characterization of the main α -galactosidase. *Biochem. J.* **339**, 43 (1999).
 243. Vocadlo, D. J., Davies, G. J., Laine, R. & Withers, S. G. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**, 835–838 (2001).
 244. Garman, S. C., Hannick, L., Zhu, A. & Garboczi, D. N. The 1.9 Å structure of α -N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. *Acta Crystallogr. Sect. A Found. Crystallogr.* **58**, c271–c271 (2002).
 245. Fujimoto, Z., Kaneko, S., Momma, M., Kobayashi, H. & Mizuno, H. Crystal structure of rice α -galactosidase complexed with D-galactose. *J. Biol. Chem.* **278**, 20313–20318 (2003).
 246. Garman, S. C. & Garboczi, D. N. The molecular defect leading to fabry disease: Structure of human α -galactosidase. *J. Mol. Biol.* **337**, 319–335 (2004).
 247. Guce, A. I. *et al.* Catalytic mechanism of human α -galactosidase. *J. Biol. Chem.* **285**, 3625–3632 (2010).
 248. Bakunina, I. Y. *et al.* α -galactosidase of the marine bacterium *Pseudoalteromonas* sp. KMM 701. *Biochem.* **63**, 1209–1215 (1998).
 249. Bakunina, I. Y. *et al.* α -N-acetylgalactosaminidase from marine bacterium *Arenibacter latericius* KMM 426T removing blood type specificity of A-erythrocytes. *Biochem.* **67**, 689–695 (2002).
 250. Hsieh, H., Mitra, M., Wells, D. C. & Smith, D. Purification and Characterization of α -N-Acetylgalactosaminidase from *Clostridium perfringens*. *Life.* **50**, 91–97 (2000).
 251. Hsin-Yeh, H., Chapman, L. F., Calcutt, M. J., Mitra, M. & Smith, D. S. Recombinant *Clostridium perfringens* α -N-acetylgalactosaminidase blood group A 2 degrading activity. *Artif. Cells. Blood Substit. Immobil. Biotechnol.* **33**, 187–199 (2005).
 252. Calcutt, M. J., Hsieh, H. Y., Chapman, L. F. & Smith, D. S. Identification, molecular cloning and expression of an α -N-acetylgalactosaminidase gene from *Clostridium perfringens*. *FEMS Microbiol. Lett.* **214**, 77–80 (2002).
 253. Falk, P., Hoskins, L. C., Lindstedt, R., Svanborg, C. & Larson, G. Deantigenation of human erythrocytes by bacterial glycosidases-Evidence for the noninvolvement of medium-sized glycosphingolipids in the Dolichos biflorus lectin hemagglutination. *Arch. Biochem. Biophys.* **290**, 312–319 (1991).
 254. Hoskins, L. C., Boulding, E. T. & Larson, G. Purification and characterization of blood group A-degrading isoforms of α -N-acetylgalactosaminidase from *Ruminococcus torques* strain IX-70. *J. Biol. Chem.* **272**, 7932–7939 (1997).
 255. Hoskins, L. C. & Boulding, E. T. Changes in immunologic properties of group A RBCs during treatment with an A-degrading exo- α -N-acetylgalactosaminidase. *Transfusion* **41**, 908–916 (2001).
 256. Teze, D. *et al.* The Catalytic Acid-Base in GH109 Resides in a Conserved GGHGG Loop and Allows for Comparable α -Retaining and β -Inverting Activity in an N-Acetylgalactosaminidase from

- Akkermansia muciniphila. *ACS Catal.* **10**, 3809–3819 (2020).
257. Liu, Q. P. *et al.* Identification of a GH110 subfamily of α 1,3-galactosidases: Novel enzymes for removal of the α 3Gal xenotransplantation antigen. *J. Biol. Chem.* **283**, 8545–8554 (2008).
258. Kwan, D. H. *et al.* Toward efficient enzymes for the generation of universal blood through structure-guided directed evolution. *J. Am. Chem. Soc.* **137**, 5695–5705 (2015).
259. Higgins, M. A. *et al.* Differential recognition and hydrolysis of host carbohydrate antigens by *Streptococcus pneumoniae* family 98 glycoside hydrolases. *J. Biol. Chem.* **284**, 26161–26173 (2009).
260. Ellis, R. J. Macromolecular crowding: Obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597–604 (2001).
261. Gerbal, A., Maslet, C. & Salmon, C. Immunological Aspects of the Acquired B Antigen. *Vox Sang.* **28**, 398–403 (1975).
262. Lesley, S. A. *et al.* Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11664–11669 (2002).
263. Comfort, D. A. *et al.* Biochemical analysis of *Thermotoga maritima* GH36 α -galactosidase (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. *Biochemistry* **46**, 3319–3330 (2007).
264. Katayama, T. *et al.* Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2- α -L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95). *J. Bacteriol.* **186**, 4885–4893 (2004).
265. Rogowski, A. *et al.* Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, (2015).
266. Bode, L. Human milk oligosaccharides: Every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).
267. Nagae, M. *et al.* Structural basis of the catalytic reaction mechanism of novel 1,2- α -L-fucosidase from *Bifidobacterium bifidum*. *J. Biol. Chem.* **282**, 18497–18509 (2007).
268. Gildea, R. J. & Winter, G. Determination of Patterson group symmetry from sparse multi-crystal data sets in the presence of an indexing ambiguity. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 405–410 (2018).
269. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).
270. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
271. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
272. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

273. McGuire, B. E. *et al.* The structure of a family 110 glycoside hydrolase provides insight into the hydrolysis of α -1,3-galactosidic linkages in l-carrageenan and blood group antigens. *J. Biol. Chem.* **295**, 18426–18435 (2020).
274. Xu, W., Yang, W., Wang, Y., Wang, M. & Zhang, M. Structural and biochemical analyses of β -N-acetylhexosaminidase Am0868 from *Akkermansia muciniphila* involved in mucin degradation. *Biochem. Biophys. Res. Commun.* **529**, 876–881 (2020).
275. Chen, X., Li, M., Wang, Y., Tang, R. & Zhang, M. Biochemical characteristics and crystallographic evidence for substrate-assisted catalysis of a β -N-acetylhexosaminidase in *Akkermansia muciniphila*. *Biochem. Biophys. Res. Commun.* **517**, 29–35 (2019).
276. Chen, X. *et al.* Crystallographic evidence for substrate-assisted catalysis of β -N-acetylhexosaminidase from *Akkermansia muciniphila*. *Biochem. Biophys. Res. Commun.* **511**, 833–839 (2019).
277. Zhang, Z. *et al.* Mechanistic and Structural Insights into the Specificity and Biological Functions of Bacterial Sulfoglycosidases. *ACS Catal.* **13**, 824–836 (2023).
278. Storch, E. K., Rogerson, B. & Eder, A. F. Trend in ABO-incompatible RBC transfusion-related fatalities reported to the FDA, 2000–2019. *Transfusion* **60**, 2867–2875 (2020).
279. Aoki, T. A comprehensive review of our current understanding of red blood cell (RBC) glycoproteins. *Membranes (Basel)*. **7**, 1–19 (2017).
280. Macfarlane, G. T. & Gibson, G. R. Formation of glycoprotein degrading enzymes by *Bacteroides fragilis*. *FEMS Microbiol. Lett.* **77**, 289–294 (1991).
281. Pudlo, N. A. *et al.* Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *MBio* **6**, (2015).
282. Glover, J. S., Ticer, T. D. & Engevik, M. A. Characterizing the mucin-degrading capacity of the human gut microbiota. *Sci. Rep.* **12**, 1–14 (2022).
283. Rajilić-Stojanović, M. & de Vos, W. M. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol. Rev.* **38**, 996–1047 (2014).
284. Carlier, J. P., Bedora-Faure, M., K'ouas, G., Alauzet, C. & Mory, F. Proposal to unify *Clostridium orbiscindens* Winter *et al.* 1991 and *Eubacterium plautii* (Séguin 1928) Hofstad and Aasjord 1982, with description of *Flavonifractor plautii* gen. nov., comb. nov., and reassignment of *Bacteroides capillosus* to *Pseudoflavonifrac*. *Int. J. Syst. Evol. Microbiol.* **60**, 585–590 (2010).
285. Shetty, S. A. *et al.* Inter-species Metabolic Interactions in an In-vitro Minimal Human Gut Microbiome of Core Bacteria. *npj Biofilms Microbiomes* **8**, (2022).
286. Pichler, M. J. *et al.* Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. *Nat. Commun.* **11**, (2020).
287. Leth, M. L., Pichler, M. J. & Hachem, M. A. Butyrate-producing colonic clostridia: picky glycan utilization specialists. *Essays Biochem.* **67**, 415–428 (2023).
288. Bell, A. *et al.* Elucidation of a sialic acid metabolism pathway in mucus-foraging *Ruminococcus gnavus* unravels mechanisms of bacterial adaptation to the gut. *Nat. Microbiol.* **4**, 2393–2404 (2019).

289. Handelsman, J. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiol. Mol. Biol. Rev.* **69**, 195–195 (2005).
290. Amann, R. I. *et al.* Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
291. Gabor, E. M., Alkema, W. B. L. & Janssen, D. B. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* **6**, 879–886 (2004).
292. Van Der Helm, E., Genee, H. J. & Sommer, M. O. A. The evolving interface between synthetic biology and functional metagenomics. *Nat. Chem. Biol.* **14**, 752–759 (2018).
293. Uchiyama, T. & Miyazaki, K. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr. Opin. Biotechnol.* **20**, 616–622 (2009).
294. Robinson, S. L., Piel, J., Sunagawa, S. A roadmap for metagenomic enzyme discovery. *Nat. Prod. Rep.* **38**, 1994–2023 (2021).
295. Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A. & Ghasemi, Y. A comprehensive review of signal peptides: Structure, roles, and applications. *Eur. J. Cell Biol.* **97**, 422–441 (2018).
296. Mathieu, K. *et al.* Functionality of membrane proteins overexpressed and purified from *E. coli* is highly dependent upon the strain. *Sci. Rep.* **9**, 1–15 (2019).
297. Schwarzl, S. M., Fischer, S. & Smith, J. C. Solvent Electrostatic Screening in Protein Simulations. *Bioinforma. Genome Regul. Struct.* 179–184 (2004) doi:10.1007/978-1-4419-7152-4_19.
298. Lee, K. K., Fitch, C. A., Lecomte, J. T. J. & García-Moreno, E. B. Electrostatic effects in highly charged proteins: Salt sensitivity of pKa values of histidines in staphylococcal nuclease. *Biochemistry* **41**, 5656–5667 (2002).
299. Gebbie, M. A., Dobbs, H. A., Valtiner, M. & Israelachvili, J. N. Long-range electrostatic screening in ionic liquids. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7432–7437 (2015).
300. Kowalski, A. *et al.* Effective screening of Coulomb repulsions in water accelerates reactions of like-charged compounds by orders of magnitude. *Nat. Commun.* **13**, 1–9 (2022).
301. Hagman, J. R. *et al.* Use of a recombinant deacetylase to convert A₁ red blood cells to the acquired B phenotype for quality control purposes. *Blood Transfus.* 1–5 (2023) doi:10.2450/BloodTransfus.584.
302. Daniel, R. M. & Danson, M. J. Temperature and the catalytic activity of enzymes: A fresh understanding. *FEBS Lett.* **587**, 2738–2743 (2013).
303. Kari, J., Schaller, K., Molina, G. A., Borch, K. & Westh, P. The Sabatier principle as a tool for discovery and engineering of industrial enzymes. *Curr. Opin. Biotechnol.* **78**, 102843 (2022).
304. Pluvinae, B. *et al.* Architecturally complex O-glycopeptidases are customized for mucin recognition and hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

Appendix – Poster Abstracts

Poster #1

This poster was presented at the **EMBL symposium**, DTU, in 2021.

Structural Comparison of Glycoside Hydrolase 109 (GH109) Enzymes: Variance in Active site Topology

Mathias Jensen¹, Bashar Shouker¹, Jens Preben Morth¹, Maher Abou Hachem¹

1. Department of Biotechnology and Biomedicine, Technical University of Denmark, DK-2800, Kongens Lyngby.

Enzymes of the glycoside hydrolase family 109 (or GH109) are characterized by their ability to cleave the terminal non-reducing α -galactosamine in a range of *O*-glycoconjugates. For example, this family of enzymes was used successfully to cleave the GalNAc in the A blood antigen. The mechanism of this family of hydrolases proceeds via an oxidative intermediate that is generated by a tightly bound NAD-cofactor. Recently, we have identified the elusive catalytic base in this family that resides in a highly conserved loop motif comprising GGHG. Very recently we determined the structure of a microbial GH109 enzyme, which allowed us to perform structural comparison of all three structurally characterized enzymes in this family. These differences in active site topology and electrostatics will be presented and discussed.

Poster #2 and oral presentations

This poster was first presented at the **Carbohydrate Bioengineering Meeting 14**, 2022, in Norway, including an oral flash presentation. Updated versions of the poster, containing new and updated content, was subsequently presented at the **Enzyme Future Symposium**, DTU, in 2022, and **Eurocarb 21**, 2023, in France, and at the **Gordon Research Conference: Carbohydrate-Active Enzymes for Glycan Conversions**, 2023, in the USA. Additionally, a 5 minute oral presentation was given at Eurocarb21.

For simplicity, only the latest version of the abstract is included below.

Structural and functional diversity of the *N*-acetyl hexosaminidases from *Akkermansia muciniphila*

Mathias Jensen¹, Bashar Shouker^{1,2}, Jens Preben Morth¹, Maher Abou Hachem¹

¹ Section of Protein Chemistry & Enzyme Technology, Department of Biotechnology & Biomedicine, DTU, 2800 Kongens Lyngby, Denmark.

² Division of Biotechnology, Department of Chemistry, Lund University, 22100 Lund, Sweden.
E-mail: matjen@dtu.dk

Akkermansia muciniphila is an emerging probiotic of the human Gut Microbiota (HGM), which colonizes the outer layer of gastrointestinal mucus [1]. *A. muciniphila* specializes in the degradation of mucin, a heavily *O*-glycosylated glycoprotein that form the major component of mucus [2].

The most represented family of glycoside hydrolases (GHs) in the genome of *A. muciniphila* is family 20, comprising 11 diverse enzymes. Here, we screen the specificity of the aglycone (-1) subsites of the aforementioned enzymes, using *para*-nitrophenyl β -D-*N*-acetyl glucosamine/galactosamine (*p*NP- β -GlcNAc/*p*NP- β -GalNAc) to shed light on their functional diversity. The GH20 enzymes displayed mostly dual HexNAc activity, but some displayed a strong preference for one of these two HexNAc components of mucin.

We also report the structure of a previously uncharacterized GH20 from *A. muciniphila* (*AmGH20A*). Comparison of the new structure to previously characterized counterparts from the same bacterium, revealed commonalities and differences, most notably, the presence of a C-terminal uncharacterized and structurally unique CBM-like domain with a putative GalNAc binding site, and unique loop conformations around the active site. We will present the latest data from these analyses, which promote our understanding of the exoglycosidases machinery that *A. muciniphila* deploys to harness mucin *O*-glycans for growth.

[1] Bae, M., Cassilly, C. D., Liu, X., Park, S. M., Tusi, B. K., Chen, X., Kwon, J., Filipčík, P., Bolze, A. S., Liu, Z., Vlamakis, H., Graham, D. B., Buhrlage, S. J., Xavier, R. J., and Clardy, J. *Nature* (2022). Vol 608, 168-173

[2] Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. *Front. Genet.* (2015). Vol 6, Article 81 .