nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-019-0469-7

An enzymatic pathway in the human gut microbiome that converts A to universal O type blood

Peter Rahfeld[1,2], Lyann Sim[1,2,8], Haisle Moon[3,4,8], Iren Constantinescu[3,4], Connor Morgan-Lang[5,6], Steven J. Hallam[5], Jayachandran N. Kizhakkedathu[3,4] and Stephen G. Withers[1,2,7]*

Access to efficient enzymes that can convert A and B type red blood cells to 'universal' donor O would greatly increase the supply of blood for transfusions. Here we report the functional metagenomic screening of the human gut microbiome for enzymes that can remove the cognate A and B type sugar antigens. Among the genes encoded in our library of 19,500 expressed fosmids bearing gut bacterial DNA, we identify an enzyme pair from the obligate anaerobe *Flavonifractor plautii* that work in concert to efficiently convert the A antigen to the H antigen of O type blood, via a galactosamine intermediate. The X-ray structure of the *N*-acetylgalactosamine deacetylase reveals the active site and mechanism of the founding member of an esterase family. The galactosaminidase expands activities within the CAZy family GH36. Their ability to completely convert A to O of the same rhesus type at very low enzyme concentrations in whole blood will simplify their incorporation into blood transfusion practice, broadening blood supply.

Correct matching of blood types is a central requirement of transfusion since plasma of individuals of blood group A contains antibodies against the B antigen and vice versa; thus, incompatible transfusions can result in activation of complement and red blood cell (RBC) lysis, which can be lethal[1]. These cell surface antigens are carbohydrate structures terminating in α-1,3-linked-*N*-acetylgalactosamine (GalNAc) or galactose (Gal) for A and B type blood, respectively (Fig. 1). O type RBCs, on the other hand, contain neither of these terminal sugars, and thus can be transfused universally to patients of the same rhesus type[2]. Accordingly, a good supply of group O RBCs is needed in blood banks to deal with emergency situations where the patient's blood type is unknown or unclear, or when supply is limited.

The concept of enzymatic removal of the GalNAc or Gal structures from A or B RBCs as a means of converting A or B RBCs to O was first proposed and demonstrated by Goldstein et al.[3]. Using an α-galactosidase from green coffee bean, B type RBCs were converted to O and a subsequent successful transfusion was performed[4]. However, gram quantities of enzyme were needed, rendering the approach impractical. Conversion of type A to O is yet more challenging, largely because type A antigen is present in many subtypes that differ in their internal linkages[5] (Supplementary Fig. 1). A major advance towards practical conversions, including of type A, was made by screening of a library of bacteria for both A and B conversion activities, using tetrasaccharide substrates. Two families of glycosidases were found that show high antigen-cleavage activity at neutral pH values: the CAZy GH109 α-*N*-acetylgalactosaminidases and the GH110 α-galactosidases[6]. Both enzymes converted their corresponding RBCs with complete removal of the respective antigens. However, substantial amounts of enzyme and very specific buffer conditions were still needed for conversion, especially for type A (60 mg enzyme per unit of blood). Along with the observation of some low-level reactivity of converted B cells in cross-matching studies, this potentially limited practical application[4]. Complete removal of both the A and B antigens with a single enzyme was explored using an endo-galactosidase (EABase) that cleaves the terminal trisaccharides, but was limited by its narrow subtype specificity[7]. Directed evolution approaches were employed to broaden specificity, but the challenges of creating a fully universal enzyme and the formation of a GlcNAc-terminated non-H type antigen caused this approach to be abandoned[8]. Identification of better enzymes would allow this field to move forward.

Screening of metagenomic libraries offers an efficient approach for the discovery of interesting enzyme activities[9] encoded within the multitude of microorganisms that have not yet been cultured or even identified[10]. Such approaches require thoughtful selection of the metagenomics source to optimize chances of successful enzyme discovery. We reasoned that the human gut microbiome was a good source since the A and B antigen structures are present within the mucins that line the gut wall[11]. These mucins serve as a barrier to the invasion of gut bacteria, but also serve as points of attachment and a source of nutrition for members of the microbiome. Consequently, some of these bacteria should express glycosidases that cleave the A and B antigens, which in turn could be useful for conversion to O type blood.

[1]Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada. [2]Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada. [3]Department of Pathology and Laboratory Medicine, Life Sciences Centre, University of British Columbia, Vancouver, British Columbia, Canada. [4]Centre for Blood Research, Life Sciences Centre, University of British Columbia, Vancouver, British Columbia, Canada. [5]Department of Microbiology and Immunology, Life Sciences Centre, University of British Columbia, Vancouver, British Columbia, Canada. [6]Graduate Program in Bioinformatics, Genome Sciences Centre, University of British Columbia, Vancouver, British Columbia, Canada. [7]Department of Biochemistry, Life Sciences Centre, University of British Columbia, Vancouver, British Columbia, Canada. [8]These authors contributed equally: Lyann Sim, Haisle Moon. *e-mail: withers@chem.ubc.ca

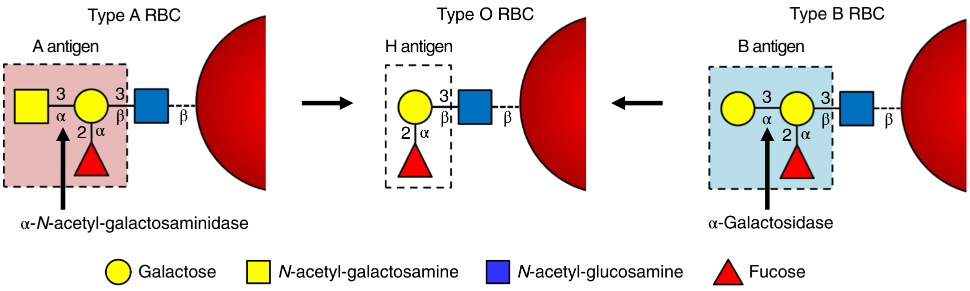

Fig. 1 | Basic A, B and H antigens on RBCs. Antigens are presented as type 1 linkages; other possible linkages are presented in Supplementary Fig. 1. Black arrows mark the cleavage points for the α-*N*-acetyl-galactosaminidase or α-galactosidase during the conversion to the H antigen. Sugars are shown using the CFG notation[29].

Metagenomic library construction and screening

We constructed a metagenomic library that contains large (35–65 kb) fragments of DNA extracted from faecal samples provided by a male donor of AB^+ blood type. Such a library contains multiple genes per bacterium, increasing the probability of expression of at least some of those genes and allowing expression of small 'pathways' of multiple genes. Our library comprised ~19,500 clones in 51 × 384-well plates, potentially around 800,000 genes; thus, initial screening of such a library with expensive A antigen substrates was impractical. Rather, we first screened with simple, sensitive fluorogenic substrates—the methylumbelliferyl (MU) α-glycosides of galactose and *N*-acetyl-galactosamine (Gal-α-MU and GalNAc-α-MU; Supplementary Fig. 2). This initial screen, with a mixture of the 2 substrates, yielded a subset of 226 hits. These were re-screened against each individual substrate, identifying 44 with α-GalNAcase and 166 with α-galactosidase activity. We then performed a second round of screening on these hits using the A and B antigen tetrasaccharide glycoside substrates shown in Supplementary Fig. 2, through a coupled enzyme assay[8] (Supplementary Fig. 3), along with a no-substrate control: only if the initial Gal or GalNAc is cleaved can the coupling enzymes act and release MU. Eleven of these hits contained A-antigen-cleaving activity, one of which also cleaved B antigen, while six produced fluorescence in the absence of substrate, thus encoding pathways that generate unrelated fluorescent products.

Sequencing and initial analysis of hits

The eleven fosmids were sequenced on an Illumina MiSeq and open reading frames (ORFs) therein that are present in the CAZy database (http://www.cazy.org/)[12] were identified using Metapathways software[13]. Owing to the considerable depth of human microbiome sequencing data now available, the organisms from which all hit fosmids were derived could be identified. Their sequences (Fig. 2) can be grouped into five clusters since eight of the eleven derived from overlapping fragments of the genomes of just two *Bacteroides* sp. The only CAZy protein encoded by a gene common to all fosmids in cluster B is a GH109 enzyme (*Bacteroides vulgatus*); cluster A also contains a GH109 (*B. vulgatus*), while a GH109 is the only CAZy protein encoded in the other *Bacteroides*-derived fosmid (*Bacteroides stercoris*). Fosmid N08, from the obligate anaerobe *Flavonifractor plautii*[14], contains three ORFs found within CAZy: an apparent carbohydrate-binding module CBM32, and two potential glycoside hydrolases—GH36 and a GH4. Finally, fosmid K05 from a *Collinsella* sp., probably *Collinsella tanakaei*, contains no obvious CAZy-related ORFs. Here the generation of a sub-library of fosmid K05 allowed the identification of the ORF with A-antigen-cleaving activity, later identified as a GH36 (Supplementary Fig. 4 and Supplementary Table 1). The genes encoding those proteins were all cloned, expressed in *Escherichia coli* and the protein products purified and characterized.

Analysis of the GH109 enzymes

The GH109 family was founded on the basis of its A-antigen-cleaving activity and shown to employ an unusual NAD^+-dependent mechanism first uncovered in enzymes from GH4[6,15]. Kinetic parameters were determined for the three GH109 enzymes identified here, *Bs*GH109, *Bv*GH109_1 and *Bv*GH109_2 (Supplementary Table 2), as well as the GH109 from *Elizabethkingia meningosepticum* (*Em*GH109), which has the highest published specific activity for A antigen cleavage[6] (Supplementary Table 3). Promisingly, similar catalytic efficiencies were seen for all four enzymes with each of the three A subtype substrates tested. Their ability to remove A antigen from A^+ RBCs was then tested in the presence of 40 kDa dextran as a macromolecular crowder to concentrate enzyme at the cell surface[16], using approved Micro Typing System (MTS) cards. These are spin columns containing antibodies against A or B antigen, and are scored from 4 to 0 according to their level of retardation of RBCs, and thus their antigen content. In the absence of crowder, none of the enzymes was effective, while in its presence only *Em*GH109 worked (Supplementary Table 4). Low ionic strength conditions have also been shown to boost the activity of *Em*GH109 on cells[6], but the enzyme is not effective in whole blood without additives.

Analysis of GH36 enzyme from *Collinsella* sp. fosmid K05

The GH36 protein within the fosmid K05 (named *Csp*GH36) cleaved both GalNAc-α-MU and the A antigen tetrasaccharide (Supplementary Fig. 4). This is consistent with membership of the GH36 family, which contains primarily α-galactosidases and α-*N*-acetylgalactosaminidases[17]. Phylogenetic analysis aligned its sequence within cluster 4 of the GH36 subfamilies[18] (Supplementary Fig. 5), a group that contains, in close proximity, a GH36 from *Clostridium perfringens* that is also known to cleave A antigen structures[19]. However, *Csp*GH36 showed limited ability to remove A antigens from RBCs, scoring only a 3 on MTS cards, even when used with a macromolecular crowder (Supplementary Table 1).

Analysis of fosmid N08 from *Flavonifractor plautii*

Since the enzymes we had found offered no advantages over the previously discovered *Em*GH109 enzyme[6], our attention turned to the CAZy-related enzymes encoded in the N08 fosmid from *F. plautii*, especially since they cleave both A and B antigens. Surprisingly, when we tested the individual purified proteins against the small-molecule A and B tetrasaccharide substrates, the only cleavage observed was of the B antigen substrate (α-Gal) by *Fp*GH36. We therefore tested pairwise combinations of these enzymes and found that the combination of *Fp*CBM32 and *Fp*GH36 rapidly cleaved the A antigen tetrasaccharide to H antigen trisaccharide (Fig. 3a and Supplementary Table 5). Thin-layer chromatography (TLC) analysis of reaction mixtures with the individual enzymes revealed that *Fp*CBM32 catalysed the conversion of A antigen to a more polar

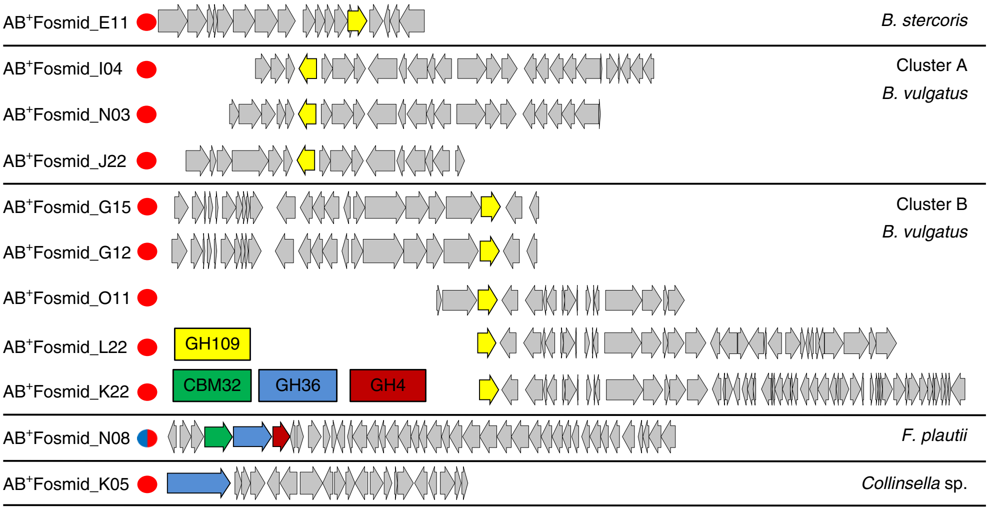

Fig. 2 | Overview of ORFs on the sequenced fosmids. Overview of the ORFs within the single sequenced fosmids. A red dot signals activity on A antigen, and a red and blue dot signals activity on A and B antigen substrates. The arrows present the ORF direction of each gene. Fosmids with overlapping sequences are aligned in clusters. The colours of the arrows represent the family classification within the CAZy database. Red, GH4; blue, GH36; yellow, GH109; green, CBM32.

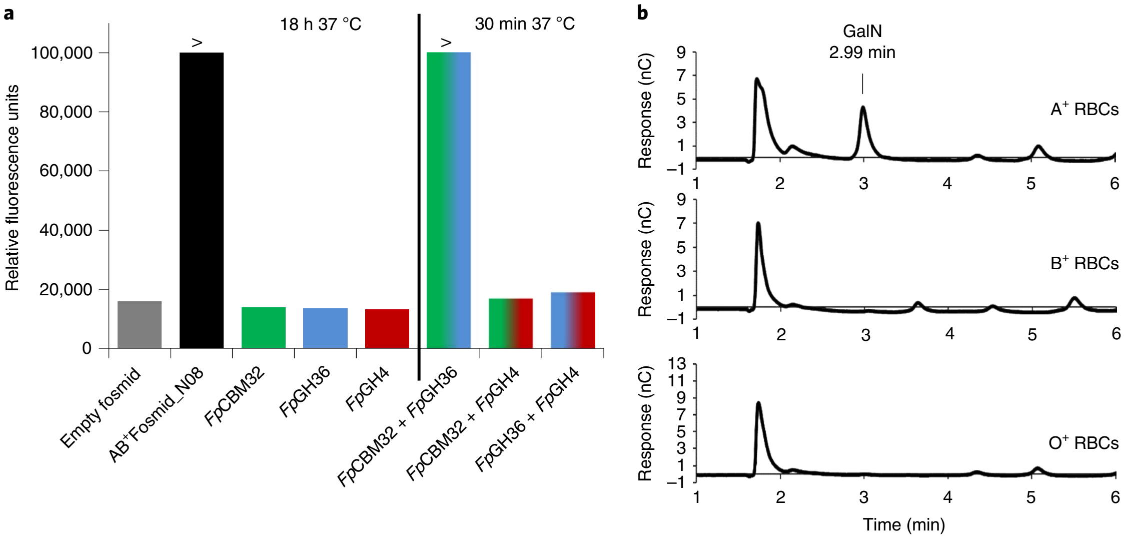

Fig. 3 | Combinatorial activity of *Fp*CBM32 and *Fp*GH36 on A antigen substrate. a, Coupled enzyme assay for A antigen type1$_{\mathrm{tetra}}$-MU cleavage using the individual enzymes and pairwise combinations of *Fp*CBM32 (green), *Fp*GH36 (blue) and *Fp*GH4 (red). Full activity was seen only if *Fp*CBM32 and *Fp*GH36 were combined. An empty (pCC1FOS$^{\mathrm{TM}}$) fosmid was used as a negative and the active AB$^+$Fosmid_N08 as a positive control. The black line separates the data obtained after different incubation times, left 18 h and right 30 min. Mixed colour bars represent the different combinations of enzymes. The '>' symbol indicates that reaction had reached the maximum detectable fluorescence signal. **b**. A$^+$, B$^+$ and O$^+$ RBCs were treated with a combination of *Fp*GalNAcDeAc and *Fp*GalNase. The HPAE-PAD assay was used to identify the cleavage product, by separating the monosaccharides. In this case, only the A$^+$ RBCs released galactosamine (GalN, 2.99 min). Sugars were identified on the basis of standards, as presented in Supplementary Fig. 6. Both experiments were performed twice, independently, with similar results.

but still ultraviolet-active product, while subsequent addition of *Fp*GH36 released galactosamine, along with H antigen trisaccharide. Mass spectrometry analysis of reaction mixtures demonstrated that *Fp*CBM32 is an A antigen deacetylase, hence the decrease of 42 in *m*/*z* and the more polar product, while *Fp*GH36 is a galactosaminidase (Fig. 4). This was further confirmed by high-performance anion exchange chromatography with pulsed amperometric detection (HPAE-PAD) analysis of the reaction (Supplementary Fig. 6), which showed that treatment of A antigen with both enzymes released galactosamine, while the individual enzymes did not. Similar results were obtained with gastric mucin substrates, for which this enzyme presumably evolved. These two enzymes are henceforth referred to as *Fp*GalNAc deacetylase (*Fp*GalNAcDeAc) and *Fp*Galactosaminidase (*Fp*GalNase) (Supplementary Table 2).

While this pathway for degradation of the A antigen was previously uncharacterized, fascinatingly it had been suggested over 50 years ago as an explanation for the so-called 'acquired' B phenomenon wherein patients with A type sepsis infected with *Clostridium tertium* underwent an apparent change in blood type to B type[20], as also did forensic samples of human tissue that had been submerged

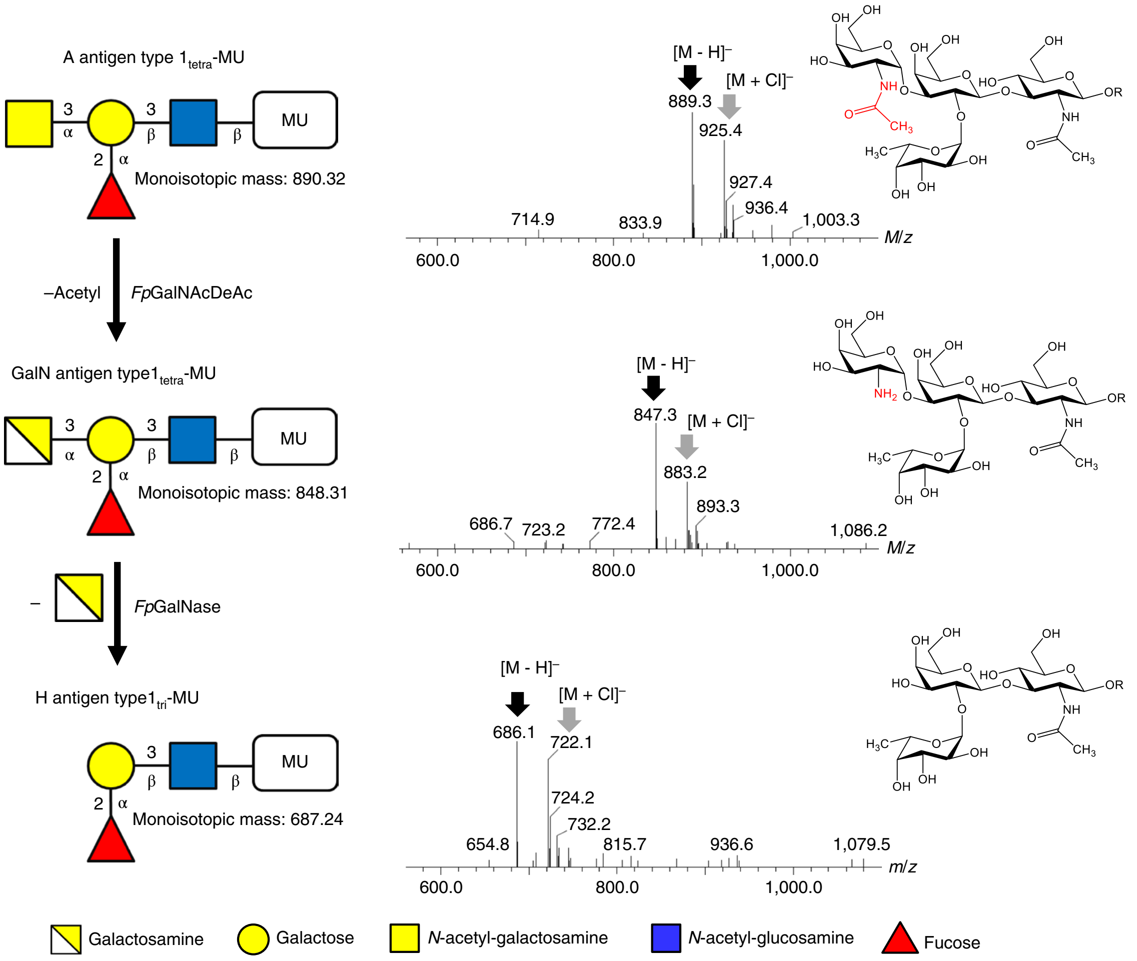

Fig. 4 | Deacetylation pathway for A antigen cleavage. The A antigen type1$_{tetra}$-MU is deacetylated by *Fp*GalNAcDeAc, followed by cleavage of galactosamine by *Fp*GalNase, yielding H antigen. Mass spectra show the mass loss of 42 on deacetylation by *Fp*GalNAcDeAc and 161 mass loss following cleavage of the galactosamine linkage. The black arrow indicates the deprotonated species and the grey arrow indicates the chloride adduct of the product. Sugars are presented as chemical structures (red-labelled part of the chemical structure is the functional group being converted by *Fp*GalNAcDeAc) and symbols using the CFG notation[29].

in the river Thames[21]. This presumably arose because the polyclonal anti-B antibodies used in typing were unable to distinguish between terminal Gal and GalN. Indeed, bacterial isolates from patients with sepsis were purified and protein fractions with proposed deacetylase and galactosaminidase activity were found but not characterized[22,23]. Our uncovering of these two enzymes therefore further explains this interesting phenomenon while also providing a useful entrée into controlled RBC type conversion.

Investigation of the third enzyme in the fosmid, the GH4, showed that, while it hydrolyses Gal-α-*p*-nitrophenol (*p*NP), GalN-α-*p*NP and GlcN-α-*p*NP, it does not cleave any A-antigen-based substrates (Supplementary Table 6), and thus does not seem to play a direct role in conversion of A antigen.

Characterization of *Fp*GalNAc deacetylase

Closer bioinformatic analysis of this protein with Phyre[2,24] indicated a ~308-amino-acid domain at the amino terminus and an ~145-amino-acid CBM32 near the carboxy terminus, with linker regions between. Truncation analysis (Supplementary Fig. 7) confirmed this basic structure since all constructs containing the intact deacetylase domain were indeed catalytically active (Supplementary Table 7). This protein is therefore classified as the founding member of a carbohydrate esterase family with specificity for A antigens.

Acetamidosugar deacetylases have all proved to be metalloenzymes, requiring divalent metal ions[25]. Consonant with this, treatment with 100 uM EDTA largely obliterated the enzyme activity, while addition of Mn^{2+}, Co^{2+}, Ni^{2+} or Zn^{2+} increased it. Other inhibitors of (non-metallo) amidases had no effect (Supplementary Table 8). The enzyme has a broad pH profile with an optimum around pH 8 (Supplementary Fig. 8) and a narrow substrate specificity (Supplementary Table 6), restricted to the different A subtypes and shorter versions thereof. However, within those subtypes, it is not very discriminatory, there being only an approximately twofold difference in specific activity across the range (Supplementary Table 7). Such a pH dependence and specificity profile is ideal for RBC conversion since all subtypes of A will be deacetylated, but nothing else.

The specificity of the CBM portion of the protein was explored using the glycan array of the Consortium for Functional Glycomics (CFG). As seen in Supplementary Fig. 9 and Supplementary Dataset, Tab 1, the preferred targets were glycans with repeating *N*-acetyl lactosamine (LacNAc) structures, as also seen for the founding member of the CBM32 family[26]. However, unlike that CBM, ours shows no high-affinity binding to blood antigen structures. Repeating LacNAc structures are a common component of cell surfaces[27], being a universal component of complex and hybrid *N*-glycans, as well as some *O*-glycans and glycolipids. Here, they

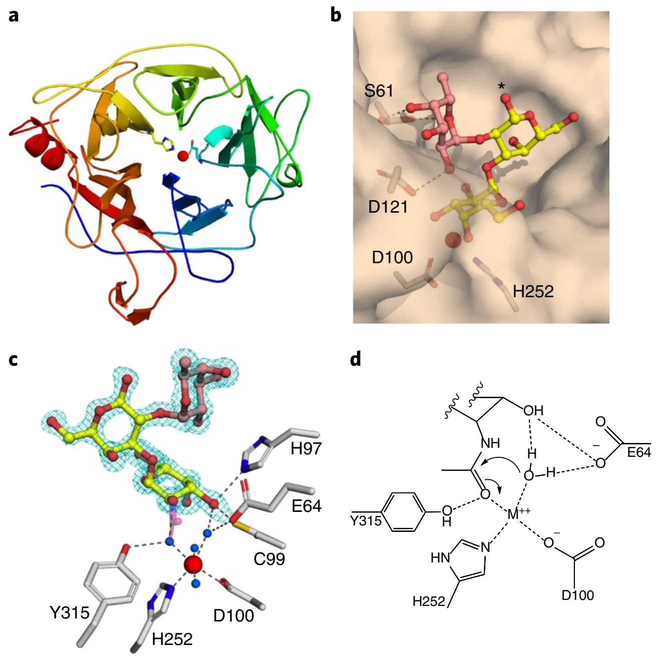

Fig. 5 | Crystal structure of FpGalNAc deacetylase. **a**, A schematic representation of the five-fold beta-propeller structure of *Fp*GalNAc DeAc_D1ext coloured from N (blue) to C (red) termini. The His 252 and Asp 100 residues that coordinate the divalent metal ion (red sphere) are shown in yellow and blue sticks, respectively. **b**, Surface representation of the active-site pocket with bound B antigen trisaccharide. Galactosyl residues are shown in yellow sticks and the fucosyl residue in pink. The asterisk denotes the location of the C1-OH group of the reducing-end galactosyl group to which a GlcNAc residue is normally attached in the natural B antigen ligand. **c**, Active-site residues involved in interactions with the non-reducing end galactosyl group. Polar interactions in **b** and **c** are shown as dashed grey lines and water molecules are shown as blue spheres. The *N*-acetyl group of the GalNAc group in a modelled A antigen trisaccharide is shown by a semi-transparent magenta stick. An electron density ($2F_o - F_c$) map of the B antigen trisaccharide ligand, contoured at 1.0σ, is shown in cyan mesh. **d**, Proposed deacetylation mechanism; first step.

presumably serve as the anchor point for attachment of the deacetylase domain, bringing its catalytic domain close to the A antigen without competing for its own substrate. In support of this model, removal of the domain resulted in a decreased activity on RBCs (Supplementary Table 9), with no effect on rates of soluble substrate cleavage (Supplementary Table 7).

Crystallographic analysis of *Fp*GalNAc deacetylase

Solution of the *Fp*GalNAcDeAc_D1ext crystal structure revealed a catalytic domain that adopts a five-fold beta propeller structure with an active site harbouring a divalent metal ion coordinated by Asp 100 and His 252 (Fig. 5 and Supplementary Table 10). Co-crystallization of the enzyme with B antigen trisaccharide as a close analogue of the reaction product unveiled its binding mode. At the base of the active-site pocket, the non-reducing end galactosyl moiety, which is the distinguishing group between A and B antigen, makes hydrogen-bonding interactions with His 97, Glu 64 and two of the metal coordinated waters (Fig. 5c). Modelling of the *N*-acetyl group of the A trisaccharide onto this structure (Fig. 5c) allowed us to make rational mutations of the nearby amino acids that may be involved in substrate deacetylation. Glu 64 proved to be critical for activity since its alanine and leucine mutants were inactive (Supplementary Table 11), suggesting a direct role, probably in activation of the nucleophilic water molecule within a mechanism such as that shown in Fig. 4d. Likewise, mutation of the residues that coordinate the divalent metal, Asp 100, and His 252 to asparagine and phenylalanine, respectively, resulted in ~5,000-fold rate decreases, consistent with their role in the mechanism. The rest of the ligand is largely surface-exposed with the C1-OH group of the reducing end galactosyl moiety pointing into the solvent (Fig. 5b), thereby explaining how the linkage of the substrate to the cell surface is accommodated by the enzyme. Most importantly, the polar interactions seen between the fucosyl group and the Ser 61 and Asp 121 side chains (Fig. 5b) explain the high observed specificity of *Fp*GalNAcDeAc towards core fucose-containing substrates. This specificity is crucial for use of this enzyme in selective A antigen removal (Supplementary Table 6). The structure of this complex also shows that it is highly unlikely that internal GalNAc residues of the repeat type 3 trisaccharide structures would be cleaved, just the terminal GalNAc.

Characterization of *Fp*GalNase

Phylogenetic analysis of the sequence places *Fp*GalNase in a subgroup (5) of the GH36 family (Supplementary Fig. 5)[18]. The 390-amino-acid catalytic domain is located in the centre of this large (1,079 amino acid) protein, with a potential carbohydrate-binding domain at the C terminus (Supplementary Fig. 7). Removal of this C-terminal domain had no effect on kinetic parameters of the enzyme with soluble substrates (Supplementary Table 7) but led to reduced efficiency in cleavage of deacetylated A$^+$ RBCs (Supplementary Fig. 10). The enzyme is specific for galactosamine-containing sugars and will not cleave GalNAc residues in any context tested. However, it has a fairly broad specificity for cleavage of de-*N*-acetylated galactosaminides ranging from the simple aryl glycoside GalN-α-*p*NP to tetrasaccharide antigens (Supplementary Table 6). Indeed (Supplementary Table 7), k_{cat}/K_M values for the three A subtypes tested were all similar to each other and to those of the deacetylase. Values of k_{cat}/K_M for cleavage of B antigen were over 2,000 times lower than for the corresponding GalN antigen but nonetheless were sufficient to yield a positive hit on the original screen. This specificity for de-acetylated alpha galacto-configured substrates, coupled with its pH optimum of ~6.5–7.0, suits it well for use in blood type conversion in conjunction with the deacetylase (Supplementary Fig. 11 and Supplementary Table 12).

Cleavage of A antigen from RBCs

To assess whether our enzymes can be used for RBC conversion, type A$^+$, B$^+$ and O$^+$ RBCs were incubated with *Fp*GalNAcDeAc and *Fp*GalNase, individually and as a mixture, and the released sugars were analysed on a HPAE-PAD ion chromatogram. As we show in Supplementary Fig. 12, neither of the enzymes used individually released any sugar products. However, when the mixture of the two was employed, galactosamine was clearly released from type A$^+$ RBCs but no sugar was released from B$^+$ or O$^+$ (Fig. 3b), consistent with their high specificity towards only the A antigen. This is very important as it shows that GalNAc is not released from the RBC surface in any other context.

We then moved onto testing for antigen removal from RBCs using the industry-standard MTS cards. Treatment with *Fp*GalNase alone did not remove A or B antigenicity at the concentration employed (Supplementary Table 13), consistent with its inactivity on GalNAc substrates, and its low activity on Gal. However, incubation with *Fp*GalNAcDeAc alone removed A antigenicity, due to conversion of the acetamide to an amine thereby compromising the binding of the anti-A antibody employed. Amounts of *Fp*GalNAcDeAc down to 3 μg ml^{-1} were sufficient without assistance from dextran, while inclusion of 300 mg ml^{-1} dextran reduced the required loading to 0.5 μg ml^{-1} (Supplementary Table 13). By comparison, the best previous enzyme (Supplementary Fig. 13), *Em*GH109, was inactive on RBCs in the absence of dextran unless low-salt buffers were

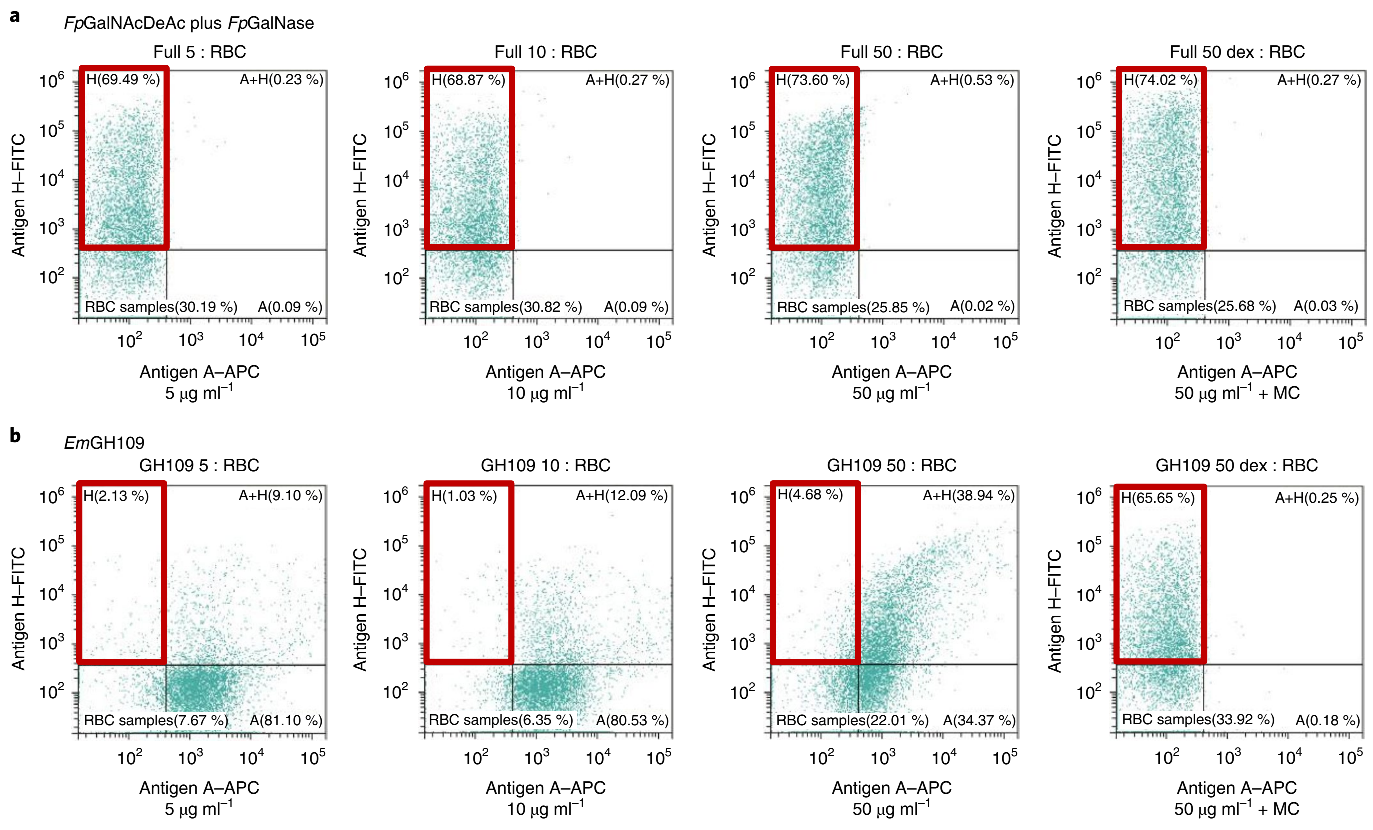

Fig. 6 | Validation of enzymatic blood type conversion. a,**b**, FACS analysis of A$^+$ RBCs treated with different concentrations of *Fp*GalNAcDeAc plus *Fp*GalNase (**a**) or *Em*GH109 (**b**) for 1 h at 37 °C with and without macromolecular crowder (MC). The enzymatically modified RBCs were treated with anti-H antibody (plus secondary FITC-labelled antibody) and APC-labelled anti-A antibody to determine the presence of the respective antigens. The region containing H antigens is marked with a red box. Untreated controls are shown in Supplementary Fig. 14. The experiment was repeated with two independent A$^+$ RBC donors, showing similar results.

employed, while in the presence of dextran the minimum effective concentration was 15 μg ml^{-1}, a 30-fold higher loading.

To assess whether conversion of A type RBCs by our enzyme pair was complete, we probed for the formation of the H antigens that should be present on enzyme-converted O type RBCs, both by fluorescence-activated cell sorting (FACS) analysis and by measurement of agglutination times. Anti-A antibodies and anti-H antibodies that do not cross-react with GalN-containing RBCs were used in these studies (Supplementary Fig. 14). As shown in Fig. 6, use of just 5 μg ml^{-1} of our enzyme pair, even in the absence of macromolecular crowder, resulted in apparently complete conversion of A antigens to H antigens, based on the analytical methods used. Likewise, anti-H antibody agglutination data showed the successful conversion of up to 26 different A$^+$ RBC donors (Supplementary Table 14) with no detectable residual A antigen. Importantly, these enzymes worked at least as well in blood as in a buffer system: such an ability to function in blood has not been shown by any enzymes previously tested. To fully confirm this, the conversion of a whole unit of blood was successfully validated (Supplementary Table 15). Further, we showed that the enzymes used in the conversion were removed by the simple centrifugation-based washing procedures used during normal RBC processing (Supplementary Fig. 15). Clearly, more work is needed to confirm removal of all traces of A antigen that might be recognized by human polyclonal anti-A antibody, but perhaps not by the anti-A antibodies that are commonly used in typing. These could arise from poorly accessible substrates. Studies with papain-treated RBCs to better expose buried antigens[28] and cross-matching studies to further probe conversion are underway.

Our combination of a deacetylase and a galactosaminidase converts A$^+$ RBCs to O type 'universal donor' RBCs via a unique mechanism, using low enzyme loadings. The very high activity and specificity of these enzymes in both buffer solutions and whole blood make these very promising candidates for cost-efficient implementation into the already existing automated routines of blood collection, processing and storage, with major implications for the flexibility of our blood supply and possible applications in organ transplantation.

Methods

Chemicals and commercial enzymes used in this study were purchased from Sigma-Aldrich unless otherwise stated. Monosaccharide methylumbelliferyl glycosides were a gift from H. Chen and the A antigen type1_{penta}-MU was a gift from D. Kwan[30].

Human faeces metagenomic library screening. For the generation of the human metagenomics fosmid library, human fresh faecal samples were collected from a healthy male with blood group AB$^+$. Informed consent was obtained from the voluntary participant before his donation, based on the guidelines of the Clinical Research Ethics Board of the University of British Columbia. The direct DNA extraction and fosmid library creation were performed according to the procedure described in Armstrong et al.[31]. A brief protocol follows.

The human metagenomics fosmid library was prepared from human fresh faecal samples that were collected from a healthy Asian male of blood group AB$^+$. The participant did not eat any food containing prebiotics or probiotics, nor did he receive any antibiotics 30 days before sample collection. Direct DNA extraction was performed using a modified chemical lysis procedure[31] adapted from a previously published procedure[31,32]. A 10 ml aliquot of extraction buffer (100 mM Tris/HCl pH 7.0, 100 mM NaH_2PO_4 pH 7.0, 100 mM EDTA pH 8.0, 1.5 M NaCl; before use add 1% (v/v) CTAB, 0.5 mg ml^{-1} lysozyme, 0.2 mg ml^{-1} Achromopeptidase) was added to 5 mg of fresh sample, and incubated for 10 min on ice in a 50 ml conical tube, with inversion once per minute, followed by five cycles of freezing in liquid nitrogen for 5 min and thawing at 42 °C in a water bath. Afterwards, 1 ml of extraction buffer II (5% SDS (wt/v), 4 M guanidinium thiocyanate, 10 mM Tris/HCl pH 7.0, 1 mM EDTA pH 8.0; before use add 0.5 mg ml^{-1} lysozyme, 0.2 mg ml^{-1} Achromopeptidase, 5 mg ml^{-1} proteinase K, 10 µl ml^{-1} 2-mercaptoethanol) was added and the sample was incubated for 2 h in a 65 °C water bath with shaking at 100 r.p.m. The sediments were precipitated by centrifugation (4,000 *g*, 4 °C, 10 min) and the supernatant was transferred to a new 50 ml conical tube. The remaining pellet was resuspended and extracted with a mixture of 5 ml extraction buffer I and 0.5 ml extraction buffer II. After sedimentation, the two supernatants were combined. To isolate the environmental DNA, 20 ml of fresh ultrapure phenol/chloroform/isoamyl alcohol (24:24:1) was added and the suspension was mixed for 10 min at 10 °C with shaking at 100 r.p.m. The two phases were separated by centrifugation (13,200 *g*, 10 °C, 10 min) and the supernatant was carefully removed without disturbing the organic phase. Precipitation of the DNA present in the supernatant was performed by addition of 0.1 volume of 3 M sodium acetate, pH 5.2 and 0.7 volumes of ice cold 100% 2-propanol. After incubation at 4 °C for 30 min, the solution was centrifuged (13,200 *g*, 4 °C, 20 min) and the supernatant was discarded. The DNA pellet obtained was washed with ice-cold 70% ethanol, incubated at 4 °C for 30 min, centrifuged (13,200 *g*, 4 °C, 20 min) and then the supernatant was discarded. The DNA was resuspended in TE buffer (10 mM Tris/HCl, 1 mM EDTA pH 8) at 4 °C overnight. The phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation steps were repeated on the following day and the final pellet was resuspended in 10 mM Tris, pH 8. The yield was approximately 800 µg of DNA from a 14 g sample. End-repair (CopyControl Fosmid Library Production Kit (Epicentre)) was performed with 20 µg of DNA through incubation at room temperature for 1 h and inactivated at 70 °C for 10 min. Size selection was performed using pulse field gel electrophoresis apparatus on a 1% agarose gel using ultrapure low-melting-point agarose (Invitrogen) and 0.5× TAE with the following settings: included angle of 120°, initial switch time of 0.10 s, final switch time of 21.79 s, with a linear ramping factor and voltage gradient of 6.0 V. After the run, the DNA was stained in the gel with SYBR Gold (ThermoFisher) for 1 h in the dark. With a clean razor blade, the DNA was cut out in a range of 36 to 65 kbp. The isolated gel was melted at 70 °C for 30 min; then, after being sure that all of the gel had melted, the sample temperature was dropped to 45 °C for 10 min. GELase (12 µl) was added to the molten agarose along with the appropriate amount of GELase buffer and the reaction was incubated at 45 °C for 3 h with occasional mixing, after which it was transferred to 70 °C and incubated at 10 min to inactivate the GELase. Amicon and Microcon filters (Millipore) were used to replace the melted agarose with water and to concentrate the DNA to 14 µl. Ligation, phage packaging and transduction of the size-selected DNA were performed using the commercially available CopyControl Fosmid Library Production Kit (Epicentre). A 20 µg sample of isolated DNA yielded approximately 40,000 clones, 20,000 of which were arrayed in 51 × 384-well plates filled with LB medium (12.5 µg ml^{-1} chloramphenicol, 50 µg ml^{-1} kanamycin, 2% maltose, 10 mM $MgSO_4$) and 15% glycerol using a Qpix2 colony picker (Genetix). After overnight incubation at 37 °C, the plates were stored at −70 °C.

For the screening, 51 × 384-well AB$^+$Blood Fosmid library plates were thawed at room temperature and replicated into 384-well plates containing 50 µl screening LB medium (12.5 µg ml^{-1} chloramphenicol, 25 µg ml kanamycin, 100 µg ml^{-1} arabinose, 0.2% (v/v) maltose, 10 mM $MgSO_4$). Plates were incubated at 37 °C for 18 h in a sealed container containing a reservoir of water to prevent excessive evaporation. Onto the grown screening plates, 45 µl of the reaction mixture (100 mM NaH_2PO_4, pH 7.4, 2% (v/v) Triton X-100, 100 µM GalNAc-α-MU, 100 µM Gal-α-MU) was added using the QFill instrument (Genetix). The plates were then incubated at 37 °C (in a sealed container) for 24 h, and the fluorescence (excitation: 365 nm, emission: 435 nm, sweep-mode, gain 80) of each plate was measured at hours 1, 2, 4, 8 and 24 via a Synergy H1 plate reader (BioTek). For all wells, a *Z*-score was calculated, which is given by the formula: *Z*-score = (fluorescence-median value) / standard deviation (ref. [33]).

All positive hits above a certain threshold (Supplementary Fig. 16), were re-arrayed in a new 384-well plate, designated the 'simple substrate hit' plate and stored at −70 °C. Two screening plates were replicated from the 'simple substrate hit' plate and re-screened for either GalNAc-α-MU or Gal-α-MU activity to verify and deconvolute the previously detected activity.

To determine which of the hits can cleave A or B antigen structures, their activity on 50 µM A antigen type1_{tetra}-MU or 50 µM B antigen type1_{tetra}-MU (Supplementary Fig 2) was determined using a coupled enzyme assay. A version of this coupled assay (Supplementary Fig 3) was described previously by Kwan et al.[30]. Our assay was modified to also detect cleavage of the type 1A antigen, by use of BgaC[34] instead of BgaA[35] as the coupling enzyme. Potential α-*N*-acetylgalactosaminidases or α-galactosidases would cleave the terminal sugar, releasing the H antigen typeI_{tri}-MU. Subsequently, an α-fucosidase (AfcA[36]), a β-galactosidase (BgaC[34]) and a β-hexosaminidase (SpHex[37]) will cleave the residual sugars in exo-fashion, until 4-methylumbelliferyl alcohol is released; detectable as an increase of the fluorescence. To achieve this, 50 µg ml^{-1} of each enzyme was added to the reaction mixture. All positive hits above a certain threshold were re-screened in triplicate and a host cell strain containing a vector lacking any insert was used as a negative control. All verified hits were stored separately at −70 °C in LB medium (12.5 µg ml^{-1} chloramphenicol, 25 µg ml^{-1} kanamycin, 15%(v/v) glycerol, 0.2%(v/v) maltose, 10 mM $MgSO_4$).

Fosmid hit sequencing. To isolate the fosmid DNA for sequencing, the positive-hit fosmid glycerol stocks were used to inoculate 5 ml of TB medium (12.5 µg ml^{-1} chloramphenicol, 25 µg ml^{-1} kanamycin, 100 µg ml^{-1} arabinose, 0.2%(v/v) maltose, 10 mM $MgSO_4$), and incubated overnight at 37 °C and 220 r.p.m. Fosmid isolation was performed using the GeneJet plasmid miniprep kit (Thermo Fisher). The isolated fosmids were purified from contaminating linear *E. coli* DNA using Plasmid-Safe ATP-Dependent DNase (Epicentre), followed by another round of purification with a GeneJet PCR purification kit (Thermo Fisher). Concentration was calculated with a Quant-iT dsDNA HS Assay Kit (Invitrogen) on a Qbit fluorimeter (ThermoFisher). Expected DNA size was validated with a 1% agarose gel. For full fosmid sequencing, 2 ng of each fosmid was sent to the UBC Sequencing Centre (Vancouver, British Columbia, Canada). Each fosmid was individually barcoded and sequenced using an Illumina MiSeq system.

All Illumina MiSeq raw sequence data were trimmed and assembled using a python script available on GitHub at https://github.com/hallamlab/FabFos. Briefly, Trimmomatic was used to remove adapters and low-quality sequences from the reads[38]. These reads were screened for vector and host sequences using BWA[39] and then filtered using samtools and a bam2fastq script to remove contaminants. These high-quality and purified reads were assembled by MEGAHIT with *k*-mer values ranging between 71 and 241, increasing by increments of 10 (ref. [40]). Since these libraries often had in excess of 20,000 times coverage, to prevent the accumulation of sequencing errors interfering with proper sequence assembly, the minimum *k*-mer multiplicity was set to 1% of the estimated coverage of a fosmid. Outside the python script, assemblies that yielded more than one contig were then scaffolded using minimus2[41]. Parameterized commands can be found in both documentation on the GitHub page and in the python script itself.

Fosmid ORF prediction and hit validation. Fosmid ORFs were identified using the metagenomic version of Prodigal[42] and compared to the CAZy database using BLASTP as part of the MetaPathways v2.5 software package[13]. MetaPathways parameters: length >60, BLAST score >20, blast score ratio >0.4, $E_{Value} < 1 \times 10^{-6}$.

All predicted ORFs with annotations to members of a GH or CBM family (with known or suspected α-galactosidase and/or α-*N*-acetylgalactosaminidase activities) were cloned into pET16b plasmid using the Golden Gate cloning strategy[43]; the primer sequences are noted in Supplementary Table 16. The proteins were expressed in BL21(DE3), cultured in 10 ml ZY5052 auto induction medium[44] for 20 h at 37 °C and 220 r.p.m. Cells were collected by centrifugation (4,000 *g*, 4 °C, 10 min) and resuspended in 1 ml lysis buffer (100 mM NaH_2PO_4, pH 7.4, 2%(v/v) Triton-X 100, 1× Protease Inhibitor EDTA-free (Pierce)). A coupled assay[30] was performed with 50 µl crude cell lysate from the candidates mixed with 50 µl assay buffer (100 mM NaH_2PO_4, pH 7.4, 50 µg ml^{-1} SpHex, 50 µg ml^{-1} AfcA, 50 µg ml^{-1} BgaC, 100 µM A antigen type 1_{tetra}-MU or 100 µM B antigen type 1_{tetra}-MU) and incubated at 37 °C. All reactions were performed as triplicates in a black 96-well plate. Fluorescence (365/435 nm) was monitored continuously for 4 h using a Synergy H1 plate reader (BioTek). Assays from crude extracts showing cleavage activity for A or B antigen were repeated, this time without the coupled enzymes, and the reaction product was isolated via an HF Bond Elut C18 column and analysed with liquid chromatography–mass spectrometry and/or TLC. TLC was performed using TLC Silica Gel 60 F254 TLC plates (EMD Millipore).

Hit validation assay for AB$^+$Fosmid_N8. To identify the enzymes responsible for the A- and B-antigen-cleavage activity, purified *Fp*CBM, *Fp*GH36 and *Fp*GH4 were tested in 100 mM NaH_2PO_4, pH 7.4 at 37 °C. Reaction was performed in 100 µl volumes containing 100 µM A or B antigen type1_{tetra}-MU, 0.1 mg ml^{-1} SpHex, AfcA, 0.1 mg ml^{-1} BgaC. Incubation time was either 18 h or 30 min, depending on the progress of the reaction. The fluorescence signal (365/435 nm) resulting from MU release by hydrolysis was monitored using a Synergy H1 plate reader (BioTek) and the results are presented in relative fluorescence units. The AB$^+$Fosmid_N8 was used as a positive and the empty (pCC1FOS) fosmid as a negative control.

Antigen type syntheses. The syntheses of the A and B antigen types $1/2/4_{tetra}$-MU were performed with a slightly modified version of the protocol (Supplementary Fig. 17) described in Kwan et al.[8], as follows.

Two-step H antigen type $1/2/4_{tri}$-MU synthesis. All three syntheses were performed on scales of 20 mg GalNAc-α-MU/GlcNAc-α-MU in 10 ml 50 mM Tris/HCl, 200 mM NaCl, pH 7.4, 10 mM $MnCl_2$, 50 U alkaline phosphatase, 1.5 equivalent UDP-Gal, 1.2 equivalent GDP-Fuc (scaled on LacNAc-MU product). Depending on the desired product, the appropriate glycosyltransferases were added at a concentration of 100 µg ml^{-1}: for type I CgtB, S42 and Te2FT; for type II, HP0826 and WbgL; for type IV, LgtD and Te2FT (Supplementary Table 17). The reaction was performed at 37 °C and the progress was monitored via TLC (mobile phase, ethyl acetate/methanol/water with a ratio of 6:2:1), with detection via ultraviolet (360 nm) after 10% H_2SO_4 dip and heating. When the reaction had halted, the mixture was applied to an HF Bond Elut C18 column, washed with several column volumes of 5% methanol, and product was eluted with 25% methanol. The solvent was then removed in vacuo.

A antigen type 1/2/4$_{tetra}$-MU synthesis. The final synthesis step was performed using 10 mg H antigen type 1/2/4$_{tri}$-MU in 5 ml 50 mM Tris/HCl, 200 mM NaCl, pH 7.4, 10 mM $MnCl_2$, 25 U alkaline phosphatase, 1.5 equivalent UDP-GalNAc and 100 ug ml^{-1} BgtA (Supplementary Table 17) at 37 °C. Progress was followed via TLC, and once complete the reaction mixture was applied to an HF Bond Elut C18 column, washed with several column volumes of 5% methanol, and product was eluted with 25% methanol. The solvent was then removed in vacuo and the final product, dissolved in water, was further purified on a 1.5 × 46 cm HW-40F size-exclusion column and then freeze-dried.

B antigen type 1/2/4$_{tetra}$-MU synthesis. The final synthesis step was performed using 10 mg H antigen type 1/2/4$_{tri}$-MU in 5 ml 50 mM Tris/HCl, 200 mM NaCl, pH 7.4, 25 U alkaline phosphatase, 1.5 equivalent UDP-Gal and 100 μg ml^{-1} BoGT6a (Supplementary Table 17) at 37 °C. Reaction progress was followed via TLC, and once complete the reaction mixture was applied to an HF Bond Elut C18 column, washed with several column volumes of 5% methanol, and product was eluted with 25% methanol. The solvent was then removed in vacuo and the final product was dissolved in water and further purified on a 1.5 × 46 cm HW-40F size exclusion column, then freeze-dried.

GalN antigen type 1$_{penta}$-MU synthesis. A antigen type 1$_{penta}$-MU (10 mg) was incubated with 1 μg ml^{-1} *Fp*GalNAcDeAc in 5 ml 100 mM NaH_2PO_4 at 37 °C for 30 min and then reaction was stopped through addition of 1 mM EDTA. The complete conversion of the substrate was confirmed via TLC and the reaction mixture was applied to an HF Bond Elut C18 column, washed with several column volumes of 2% methanol, and then the product was eluted with 10% methanol. The solvent was then removed in vacuo.

Protein purification. All proteins and their truncations were cloned via Golden Gate cloning[43] or PIPE cloning[45] into pET16b or pET28a vectors. The primer sequences are noted in Supplementary Table 16.

The production of proteins for extended characterization was performed in BL21(DE3) cells, cultured in 200 ml ZY5052 auto induction medium[44] for 20 h at 37 °C and 220 r.p.m. inoculated with 100 μl of an overnight LB culture. Cells were collected by centrifugation (4,000 *g*, 4 °C, 10 min) and resuspended in 10 ml lysis buffer (50 mM Tris/HCl, 150 mM NaCl, 1%(v/v) glycerol, 40 mM imidazole, pH 7.4, 2 mM dithiothreitol (DTT), 1× Protease Inhibitor EDTA-free (Pierce), 2 U Benzonase (Novagen), 0.3 mg ml^{-1} lysozyme, 10 mM $MgCl_2$), followed by sonification (3 min pulse time; 5 s pulse, 10 s pause, 35% amplitude) on ice. After removal of cell debris by centrifugation (14,000 *g*, 4 °C, 30 min), supernatant was collected and loaded on a nickel affinity chromatography column (5 ml HisTrap HP column (GE)) using a peristaltic pump. The elution was performed and monitored on an AKTApurifier system (GE) with a 10–75% gradient of 50 mM Tris/HCl, 400 mM imidazole, pH 7.4, 2 mM DTT. The fractions containing the protein were identified via SDS–polyacrylamide gel electrophoresis (PAGE) and then pooled. Buffer exchange into 50 mM Tris/HCl, 150 mM NaCl, pH 7.4, 2 mM DTT and concentration were performed in Amicon Ultra-15 centrifugal filter units (MWCO 10 kDa; Millipore).

*Fp*GalNAcDeAc, *Fp*GalNase and their truncations were subjected to a second round of purification. Amicon Ultra-15 centrifugal filter units (MWCO 10 kDa; Millipore) were used to exchange the buffers before loading the proteins on a hydrophobic interaction chromatography column (10 ml Phenyl Sepharose High Performance column (Pharmacia Biotech)). Loading, washing and elution (gradient 0–100%) of the column was handled through an AKTApurifier system (GE), utilizing the following buffer conditions: *Fp*GalNAcDeAc; binding 1× PBS, 800 mM NH_4HPO_4, pH 7.4 and elution 1× PBS, pH 7.4 and *Fp*GalNase; binding 25 mM Tris/HCl, 1 M NaCl, pH 7.4 and elution 25 mM Tris/HCl pH 7.4. The fractions containing the protein were identified via SDS–PAGE and then pooled. Buffer exchange into 50 mM Tris/HCl, 150 mM NaCl, pH 7.4 and concentration was performed using Amicon Ultra-15 centrifugal filter units (MWCO 10 kDa; Millipore).

Protein characterization. *Optimum pH value.* The general pH range for activity of *Fp*GalNAcDeAc and *Fp*GalNase for A antigen type1$_{penta}$-MU and GalN antigen type1$_{penta}$-MU, respectively, was determined by monitoring product formation at varying pH values by TLC analysis. Reaction was performed on 100 μl scales at 37 °C with 50 μM substrate and 1 μg ml^{-1} enzyme in the appropriate buffer system. Buffers for the pH range 4 to 6 were based on a 50 mM citric acid/sodium citrate buffer, buffers for the pH range 6–8 were based on a 50 mM sodium phosphate buffer and buffers for the pH range 8–10 were based on a 50 mM glycine/sodium hydroxide buffer.

To determine the optimal pH value of the galactosaminidase, *Fp*GalNase (5 μg ml^{-1}) was incubated in 100 μl 50 mM sodium phosphate buffer containing 200 μM GalN-α-pNP at a series of pH values (5.8–8.0). The absorption (at 405 nm) resulting from *p*NP release was monitored by a Synergy H1 plate reader (BioTek) for 1 h at 37 °C and results were corrected for the extinction coefficient at each pH value.

To determine the pH optimum for the deacetylase, *Fp*GalNAcDeAc (5 μg ml^{-1}) and 50 μM A antigen type1$_{penta}$-MU were pre-incubated for 10 min at 37 °C in 25 mM sodium phosphate buffer at a series of pH values (5.8–10.0). The reaction was quenched with 100 mM sodium phosphate buffer pH 7.5, 100 μM EDTA, 5 μg ml^{-1} *Fp*GalNase, 50 μg ml^{-1} SpHex, 50 μg ml^{-1} AfcA and 50 μg ml^{-1} BgaC, final volume 100 μl. The fluorescence signal (365/435 nm) resulting from MU release by hydrolysis was monitored by a Synergy H1 plate reader (BioTek) for 30 min at 37 °C.

FpGalNAcDeAc inhibition. Different potential inhibitors of *Fp*GalNAcDeAc were tested in 96-well plate format using the coupled assay. Reaction was performed on a 100 μl scale at 37 °C with 50 μM A antigen type1$_{penta}$-MU and 5 μg ml^{-1} *Fp*GalNAcDeAc in 100 mM NaH_2PO_4 pH 7.4 with 10 μg ml^{-1} *Fp*GalNase, 50 μg ml^{-1} SpHex, 50 μg ml^{-1} AfcA, 50 μg ml^{-1} BgaC. EDTA (1, 10, 100 μM), marimastat (1, 10, 100, 1,000 μM), dimethylsulfoxide (2%, 4%) and Protease Inhibitor Cocktail EDTA-free (Pierce) (1×, 2× and 4×) were tested as inhibitors. Fluorescence (365/435 nm) was monitored continuously for 1 h using a Synergy H1 plate reader (BioTek). Those compounds or mixtures showing strong inhibitory effects were retested without the coupled enzymes and product formation was analysed via TLC, to ensure that the inhibition observed was not of the coupling enzymes.

*FpGalNAcDeAc divalent metal influence. Fp*GalNAcDeAc was incubated with 2 mM EDTA for 10 min at room temperature in 50 mM Tris/HCl pH 7.4; then the protein was washed once with 100 μM EDTA, twice with 100 nM EDTA, three times without EDTA, all in 50 mM Tris/HCl pH 7.4, using Amicon Ultra-0.5 ml centrifugal filter units (MWCO 10 kDa; Millipore). The buffers used were pre-treated with Chelex 100 (Bio-Rad). A 1 μg ml^{-1} sample of EDTA-treated *Fp*GalNAcDeAc was incubated with 1, 5 and 10 μM concentrations of each of $MgCl_2$, $MnCl_2$, $CaCl_2$, $CoCl_2$, $NiCl_2$, $CuCl_2$, $FeCl_2$ or $ZnCl_2$ individually for 10 min at 37 °C. After incubation, 10 μg ml^{-1} *Fp*GalNase, 50 μg ml^{-1} SpHex, 50 μg ml^{-1} AfcA, 50 μg ml^{-1} BgaC and 50 μM A antigen type1$_{penta}$-MU were added. The fluorescence (365/435 nm) was monitored continuously for 1 h at 37 °C using a Synergy H1 plate reader (BioTek). As controls, no metal, 2 mM EDTA and no *Fp*GalNAcDeAc plus 50 μM GalN antigen type1$_{penta}$-MU as substrate were used.

*Circular dichroism spectroscopy and secondary-structure prediction. Fp*GalNAcDeAc (0.17 mg ml^{-1}) and *Fp*GalNase (0.29 mg ml^{-1}) were dialysed into 5 mM NaH_2PO_4. Circular dichroism spectroscopic data were acquired in a Jasco J-815 CD spectrophotometer with the following parameters: 190–250 nm measuring range, speed 50 nm min^{-1}, pitch 0.5, scans 10, room temperature. The obtained data were interpreted with the web tool BeStSel (http://bestsel.elte.hu/index.php).

HPAE-PAD assay. The analysis of the enzymatic release of galactosamine was carried out on an HPAE-PAD (Dionex) high-performance liquid chromatography system. Cleavage activity of the different proteins was tested on the following substrates: 7.5 μg μl^{-1} mucin from porcine stomach type II in 100 mM NaH_2PO_4 pH 7.4; 5 mM A antigen type 1$_{penta}$-MU in 100 mM NaH_2PO_4 pH 7.4 and RBCs (50% haematocrit) from A$^+$, B$^+$ and O type donors in 1× PBS pH 7.4. Samples containing 10 μg ml^{-1} enzyme were incubated for 2 h at 37 °C and then stored at −80 °C for further analysis. Small aliquots of the reaction (10 μl) were diluted in H_2O (100 μl) and submitted to analysis on the HPAE-PAD instrument. Separation was performed on a CarboPAC PA200 (150 mm) column with a guard column, and detection was achieved using a disposable gold on polytetrafluoroethylene electrode and a four-potential waveform. The separation conditions were as follows: 100 mM sodium hydroxide and a sodium acetate gradient from 70 to 300 mM over the first 10 min of the separation. The eluent was held at the final gradient conditions for 1 min and then returned to the starting conditions over the next minute. The flow rate was 1.0 ml min^{-1} and an injection was made every 27 min. A standard of the free sugars GalNAc, Gal and GalN (10 μM) was also applied to HPAE-PAD to determine the peak elution time for reference.

Kinetic assays. All kinetic assays utilizing 4-methylumbelliferone as the leaving group were performed through measurement of fluorescence. To avoid measurement errors based on the inner filter effect[46], standard curves were used to validate the linear range of the fluorophore.

FpGalNAcDeacetylase. Michaelis–Menten parameters were determined for A antigen type 1$_{penta}$-MU in 100 mM NaH_2PO_4, pH 7.4 at 37 °C using the coupled assays described previously[30]. The assay was modified to allow detection of cleavage of the type 1 (and later type 4) antigen, by use of BgaC[34] instead of BgaA[35] as the β-galactosidase. In addition, since A antigen type 1$_{penta}$-MU contains an additional galactose, the concentration of BgaC was increased to 0.2 mg ml^{-1} to compensate for its need to cleave both the Gal-β-1,3-β-GlcNAc-β-1,3-Gal-β-MU and Gal-β-MU. Further, *Fp*Galactosaminidase was included to allow the cleavage of the galactosamine-containing intermediate. Reaction set-up in 100 μl was 3 nM *Fp*GalNAcDeacetylase (4.52 nM *Fp*GalNAcDeAc_D1ext, 3.55 nM *Fp*GalNAc-DeAc_D1 + 2) and 0.01 mg ml^{-1} *Fp*Galactosaminidase, 0.1 mg ml^{-1} SpHex, AfcA, 0.2 mg ml^{-1} BgaC and varying concentrations of substrate (5 μM–2.5 mM). The reactions were run as a series of four with controls (no *Fp*GalNAcDeacetylase) as duplicates. The fluorescence signal (365/435 nm) resulting from MU release by hydrolysis was monitored on a Synergy H1 plate reader (BioTek) and converted to units of concentration using MU standard concentration curves determined under identical reaction conditions. Initial rates (μM s^{-1}) were determined and plotted in Grafit 7.0 to determine the kinetic parameters.

k_{cat}/K_M parameters were determined for A antigen type 1/2/4_{tetra}-MU at pH 7.4 at 37 °C. Reactions (total volume of 100 μl) were performed in black 96-plate wells and as coupled assays in 100 mM NaH_2PO_4 (pH 7.4) with 12 nM *Fp*GalNAcDeacetylase, 0.1 mg ml^{-1} SpHex, BgaC (BgaA for type 2), AfcA, at varying concentrations of substrate (25 μM, 20 μM, 15 μM, 10 μM, 7.5 μM, 5 μM). The reactions were run as a series of four with controls (no *Fp*GalNAcDeacetylase) as duplicates. The fluorescence signal (365/435 nm) resulting from MU release by hydrolysis was monitored on a Synergy H1 plate reader (BioTek) and converted to units of concentration using MU standard concentration curves determined under identical reaction conditions. Initial rates (μM s^{-1}) were determined and plotted in Grafit 7.0 to determine the k_{cat}/K_M ($s^{-1}mM^{-1}$) parameters.

FpGalNase. Michaelis–Menten parameters were determined for GalN antigen type1_{penta}-MU and A antigen type1_{penta}-MU in 100 mM NaH_2PO_4, pH 7.4 at 37 °C. Reaction was performed in 100 μl volumes containing 3.4 nM *Fp*GalNase (5.31 nM *Fp*GalNase_truncA), 0.1 mg ml^{-1} SpHex, AfcA, 0.2 mg ml^{-1} BgaC and varying concentrations of substrate (5 μM–2 mM). The reactions were run as a series of four with controls (no *Fp*GalNase) as duplicates. The fluorescence signal (365/435 nm) resulting from MU release by hydrolysis was monitored using a Synergy H1 plate reader (BioTek) and results were converted to units of concentration using MU standard concentration curves determined under identical reaction conditions. Initial rates (μM s^{-1}) were determined and plotted in Grafit 7.0 to determine the kinetic parameters.

Values of k_{cat}/K_M were determined for GalN antigen type1/2/4_{tetra}-MU and B antigen type1_{tetra}-MU at pH 7.4 and 37 °C. Reactions (total volume of 100 μl) were performed in black 96-well plates as coupled assays in 100 mM NaH_2PO_4 (pH 7.4) with 8.63 nM *Fp*GalNase, 0.1 mg ml^{-1} each of SpHex, BgaC (BgaA for Type 2) and AfcA, and varying concentrations of substrate (25 μM, 20 μM, 15 μM, 10 μM, 7.5 μM, 5 μM). The reactions were run as a series of four with controls (no *Fp*GalNase) as duplicates. The fluorescence signal (365/435 nm) resulting from MU release was monitored on a Synergy H1 plate reader (BioTek) and results were converted to units of concentration using MU standard concentration curves determined under identical reaction conditions. Initial rates (μM s^{-1}) were determined and plotted in Grafit 7.0 to determine the k_{cat}/K_M ($s^{-1}mM^{-1}$) parameters.

Michaelis–Menten parameters were determined for GalN-α-pNP in clear 96-plates at 37 °C using 863.2 nM *Fp*GalNase (in 100 mM NaH_2PO_4, pH 7.4) or 369.9 nM *Fp*GH4 (in 50 mM Tris/HCl, pH 7.4, 100 μM NAD^+, 1 mM $MnCl_2$) and concentrations of substrate from 10 μM to 5 mM in a volume of 100 μl. The reactions were run as a series of three with two controls (no enzyme). The absorption (at 405 nm) resulting from *p*NP release was monitored on a Synergy H1 plate reader (BioTek) and rates were converted to units of concentration using *p*-nitrophenol standard concentration curves determined under identical reaction conditions. Initial rates (μM s^{-1}) were determined and plotted in Grafit 7.0 to determine the kinetic parameters.

Kinetic analysis of GH109 with different A type substrates. Values of k_{cat}/K_M for A antigen type1/2/4_{tetra}-MU were determined at pH 7.4 and 37 °C. Reactions (total volume of 100 μl) were performed in black 96-well platess and performed as coupled assays in 100 mM NaH_2PO_4, pH 7.4 with 86.02 nM *Bv*GH109_1, 100.49 nM *Em*GH109, 80.52 nM *Bv*GH109_2, 87.4 nM *Bs*GH109 and 5 μM NAD^+, along with 0.1 mg ml^{-1} each of SpHex, BgaC (BgaA for type 2), AfcA, and varying concentrations of substrate (25 μM, 20 μM, 15 μM, 10 μM, 7.5 μM, 5 μM). The reactions were run as a series of four with controls (no α-*N*-acetylgalactosaminidase) as duplicates. The fluorescence signal (365/435 nm) resulting from MU release was monitored on a Synergy H1 plate reader (BioTek) and rates were converted to units of concentration using MU standard concentration curves determined under identical reaction conditions. Initial rates (μM s^{-1}) were determined and plotted in Grafit 7.0 to determine the k_{cat}/K_M ($s^{-1}mM^{-1}$) values.

Limited proteolysis. To investigate whether smaller, stable subdomains of *Fp*GalNase could be formed, a limited proteolysis was performed. *Fp*GalNase was treated with thermolysin (10:1 protein/protease mass ratio) at various temperatures (20 °C, 37 °C, 42 °C, 50 °C, and 65 °C) for 1.5 h. Samples were then run on an SDS–PAGE gel and a stable fragment was identified with a mass of ~70 kDa (down from the initial 118 kDa) with nearly complete digestion to this form achieved at the 50 °C incubation temperature. This fragment was sent to the UBC proteomics core facility for peptide identification and was determined to be a C-terminal-truncated version of the full-length protein with a cleavage site between amino acids 690 and 700 (Supplementary Fig. 18).

Crystallography. Before crystallization, *Fp*GalNAcDeAc_D1ext was digested with thrombin (Novagen) at a concentration of 1 mg ml^{-1} overnight using the manufacturer's suggested protocol. Protein was then purified by HisTrap FF column and the flowthrough was collected, buffer-exchanged into 10 mM Tris pH 8.0 + 75 mM NaCl, and concentrated to 12 mg ml^{-1}.

*Crystallization. Fp*GalNAcDeAc_D1ext (12 mg ml^{-1}) was crystallized by use of the hanging-drop diffusion method using a reservoir solution composed of 0.2 M $CaCl_2$, 0.1 M MES pH 6, 18% PEG 4000 and 20 mM $MnCl_2$ at a 1:1 protein/reservoir ratio. A quick bromide soak was used to derivatize crystals for phasing and was prepared by transferring the crystal to a solution of 1 M NaBr, 25% glycerol, 18% PEG4000, 20 mM $CaCl_2$ and 0.1 M MES pH 6 for 30 s and flash-frozen in liquid nitrogen. Crystal complexes with blood group B antigen trisaccharide (B_tri) were prepared as stable analogues of the natural substrate, which would otherwise by hydrolysed, by preincubating protein (12 mg ml^{-1}) with 10 mM B_tri for 2 h before setting up drops under the same conditions as above, but omitting $MnCl_2$. Crystals were cryoprotected with reservoir solution supplemented with 25% glycerol.

Data collection, phasing and structure determination. Datasets for the Br derivative and B_tri complex of the *Fp*GalNAcDeAc_D1ext crystals were collected at the Canadian Light Source beamline 08ID-1 at wavelengths of 0.919 Å and 0.979 Å, respectively. Data were integrated using XDS[47] and scaled with Aimless[48]. Single-wavelength anomalous dispersion phasing and automated structure solution was performed using CRANK2[49] in the CCP4i2 program suite[50]. Structure was checked and refined using alternating cycles of Coot[51] and Refmac[52]. The B_tri structure complex was solved by difference Fourier analysis and the ligand was manually built in Coot as were the water and metal ions. Difference density maps confirmed the presence of Mn^{2+} in the apo structure and Ca^{2+} in the liganded structure. Models were validated by Coot and Molprobity[53]. Data collection and refinement statistics are summarized in Supplementary Table 10. Atomic coordinates and structure factors of the apo and B_tri complex have been deposited in the Protein Data Bank (PDB) with accession numbers 6N1A and 6N1B, respectively.

Glycan array screening. For the glycan array screening, 500 μg of *Fp*GalNAcDeAc_D2ext was labelled with fluorescein isothiocyanate (FITC) with a fluorescein/protein ratio of 1 using the Fluorotag FITC conjugation Kit (Sigma). Screening was performed at 5 and 50 μg ml^{-1} protein concentrations using the CFG's Protein-Glycan Interaction Core Facility with version 5.3 of the printed array, consisting of 600 glycans in replicates of 6. Analysis of binding motifs was performed with the webtool https://glycopattern.emory.edu/.

Active-site mutagenesis. On the basis of structural information (Fig. 4) and sequence alignment (Supplementary Fig. 11 and Supplementary Table 12), *Fp*GalNAcDeAc_D1min and *Fp*GalNase_truncA were mutated using the QuickChange protocol[54], using the primers noted in Supplementary Table 16. The mutants were purified via NiNTA and HIC columns as described above. The structural integrity of all mutants was checked via circular dichroism spectroscopy; all tested enzymes were structurally similar to their wild type (Supplementary Dataset, Tab 2). For mutants with relatively low activity, reactions were carried out under the same conditions used for full kinetic determinations; however, the substrate depletion method was used for determination of k_{cat}/K_M values as has been previously described[55]. In brief, at low concentrations of substrate where [Substrate] < K_M (equivalent to ~1/5–1/10 of K_M), the k_{cat}/K_M value can be approximated on nonlinear fitting of the reaction time course to a first-order curve and dividing by the enzyme concentration.

GH36 phylogenetic mapping. Reference sequences of GH36 were downloaded from the CAZy database using SACCHARIS's cazy_extract.pl script[56]. Phylogenetic-based protein profiling software, TreeSAPP (available at https://github.com/hallamlab/TreeSAPP), was used to both build the GH36 reference tree and place query sequences on the phylogenetic tree. Hidden Markov models (HMMs) from dbCAN were used to extract protein family domains from all full-length sequences downloaded from CAZy[57]. These sequences were then clustered at 70% sequence similarity using UCLUST to remove similar sequences and decrease the size of the tree for faster placement[58]. RAxML version 8.2.0 was used to build the reference tree of 738 proteins with the '—autoMRE' to decide when to quit bootstrapping before 1,000 replicates have been performed, and PROTGAMMAAUTO to select the optimal amino-acid substitution model[59,60].

TreeSAPP's treesapp.py was then used to map the query sequences onto the GH36 tree. Briefly, query protein sequences were aligned to the GH36 HMM profile built from the 738 reference sequences using hmmsearch and the aligned regions were extracted[61]. hmmalign was used to include the new query sequences in the reference multiple alignment and then TrimAl removed the non-conserved positions from the alignment file[62]. RAxML's evolutionary placement algorithm was used to identify optimal placement positions in the reference tree for each query sequence. Placements were filtered by likelihood weight ratio and concatenated into a single.Jplace file before being visualized in iTOL[63,64].

RBC conversion assays. Whole blood from healthy consenting donors was collected into a citrate Vacutainer using a protocol approved by the clinical ethics committee of the University of British Columbia. The tube was spun at 1,000 *g* for 5 min at room temperature, and RBCs were separated and washed 3 times with 1× PBS pH 7.4. For assays in the presence of 40 kDa dextran, washed RBCs (200 µl, 10% haematocrit) were placed in a tube, and the supernatant was partially removed and replaced with 1× PBS pH 7.4 (final concentration of 300 mg ml^{-1}) with and without 40 kDa dextran. In addition, some assays were performed in autologous platelet-poor plasma. RBCs were mixed carefully and placed on an orbital shaker for 30 s. Diluted enzyme solutions were then added to a final volume of 200 µl. The tubes were vortexed very gently, and placed on an orbital shaker for defined times at set temperatures.

MTS cards. After the reaction, RBCs were washed 3 times with an excess of 1× PBS pH 7.4 and analysed using MTS cards (ID-MTS Gel Cards from Ortho Clinical Diagnostics). For A typing, MTS Anti-A (Murine Monoclonal) Card Blood Grouping Reagents (prod. no. MTS080014) were used and for B typing MTS Anti-B (Murine Monoclonal) Card Blood Grouping Reagents (prod. no. MTS080015) were used. RBCs (12 µl, 5% haematocrit), suspended in diluent (MTS), were added carefully to the mini gel column, leaving a space between the blood and the contents of the mini gel. The MTS cards were centrifuged at 156 *g* for 6 min at room temperature using a Beckman Coulter Allegra X-22R centrifuge with a modified sample holder as recommended. The extent of antigen removal from the surface of the RBC was evaluated from the location of RBCs in the mini gel after spinning, according to the manufacturer's instructions. RBCs with a high surface antigen concentration agglutinated on interaction with the monoclonal antibody present in the gel column and could not penetrate (MTS score 4). RBCs with no surface antigens did not agglutinate and migrated to the bottom of the mini gel (MTS score 0). RBCs that underwent partial removal of surface antigens migrated to positions between these and were assigned scores between 0 (not present) and 4 (present) according to the manufacturer's instructions.

Agglutination assays for H antigen. To analyse the conversion of A antigen to H antigen after enzymatic treatment, washed A-antigen-cleaved RBCs (10% haematocrit) were mixed in equal parts with 2 µg ml^{-1} anti-H antibody (Anti-Blood Group H ab antigen antibody (97-I): cat no. ab24213 (Abcam)) and the appearance of agglutination within a 30 min time frame was monitored. RBCs that underwent agglutination with the anti-H antibody were assigned scores between 0 (no agglutination within 1,800 s) and 5 (agglutination within 120 s).

FACS. Enzyme-treated RBCs were washed twice with 1× PBS pH 7.4 and 1% haematocrit enzymatically converted RBCs were treated with 1/100 APC–anti-A antibody (Alexa Fluor 647 mouse anti-human blood group A: cat. no. 565384 (BD Pharmingen)) for 30 min at room temperature, and/or treated with 1/100 anti-H antibody (Anti-Blood Group H ab antigen antibody (97-I): cat no. ab24213 (Abcam)) for 30 min at 4 °C, and then washed once with 1× PBS pH 7.4. For detection of the anti-H antibody, a secondary Alexa Fluor 488-labelled antibody (goat anti-mouse IgM (heavy chain) cross-adsorbed secondary antibody, Alexa Fluor 488: cat no. A-21042 (Invitrogen)) in a 1:300 concentration was applied at room temperature for 30 min. Flow cytometry was performed after reconstitution into 1× PBS pH 7.4 (1% haematocrit) on a CytoFLEX flow cytometer using CytExpert Software 2.1 (Beckman Coulter). Native RBCs at 1% haematocrit, collected after centrifugation of whole blood at 1,000*g* for 5 min (Allegra X-22R centrifuge (Beckman Coulter)) for the complete removal of platelet-rich plasma, were used to set a gate for the data collection. The full gating strategy is presented in Supplementary Fig. 19.

Enzyme adsorption. The amount of remaining enzyme on A$^+$RBCs after washing was determined by residual A antigen cleavage activity. Therefore, A$^+$ RBCs from 1 donor were treated with 5 or 20 µg ml^{-1} *Fp*GalNAcDeAc and *Fp*GalNase, and incubated at 37 °C for 1 h. After the reaction, RBCs were washed 3 times with an excess of 1× PBS pH 7.4 and analysed. Analysis reaction was performed in 50 µl 1× PBS pH 7.4 with 0.1 mg ml^{-1} SpHex, AfcA, BgaA and 500 µM substrate; reactions were run as triplicates. The fluorescence signal (365/435 nm) resulting from MU release by hydrolysis was monitored on a Synergy H1 plate reader (BioTek).

Whole blood bag conversion. The whole blood bag was collected on 15 January 2018 (C0510 18 990033 A+) and mixed with SAGM to a final volume of 309 ml and 60% haematocrit. Then 12 ml each of 2 mg ml^{-1} *Fp*GalNAcDeAc and *Fp*GalNase were added and the bag was manually massaged for 10 min. The bag was incubated over a period of 72 h at 4 °C and 10 ml samples were taken at 24, 48 and 72 h. Samples were washed 7 times in 1× PBS pH 7.4 with centrifugation at 1,800 *g* for 5 min. A antigen removal was monitored using MTS anti-A (Murine Monoclonal) Card Blood Grouping Reagents (prod. no. MTS080014) and H antigen appearance was measured through agglutination with anti-H antibody (Anti-Blood Group H ab antigen antibody (97-I): cat. no. ab24213 (Abcam)).

Ethics statement. The collection of human faecal samples was approved by the Clinical Research Ethics Board of the University of British Columbia (ID no. H15-02967). The collection of human blood samples was approved by the Clinical Research Ethics Board of the University of British Columbia (ID no. H07-02198) and the Canadian Blood Services (REB no. 2017.029).

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

Data availability

The datasets generated and/or analysed during this study are available from the corresponding author on reasonable request. Structure datasets generated during the current study are available in the PDB repository under accession numbers 6N1A and 6N1B. The protein *Csp*GH36 identified during the current study is deposited in GenBank under the accession code MK500922.

Received: 27 December 2018; Accepted: 25 April 2019;
Published online: 10 June 2019

References

1. Daniels, G. The molecular definition of red cell antigens. *ISBT Sci. Ser.* **5**, 300–302 (2010).
2. Garratty, G. Modulating the red cell membrane to produce universal/stealth donor red cells suitable for transfusion. *Vox Sang.* **94**, 87–95 (2008).
3. 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).
4. 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).
5. Clausen, H. & Hakomori, S. Abh and related histo-blood group antigens - immunochemical differences in carrier isotypes and their distribution. *Vox Sang.* **56**, 1–20 (1989).
6. Liu, Q. P. et al. Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **25**, 454–464 (2007).
7. Anderson, K. M. et al. A clostridial endo-beta-galactosidase that cleaves both blood group A and B glycotopes. *J. Biol. Chem.* **280**, 7720–7728 (2005).
8. 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).
9. Handelsman, J. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. R.* **68**, 669–685 (2004).
10. Amann, R. I. et al. Combination of 16S ribosomal-RNA-targeted oligonucleotide probes with flow-cytometry for analyzing mixed microbial-populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
11. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **6**, 81 (2015).
12. Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
13. Konwar, K. M. et al. MetaPathwaysv2.5: quantitative functional, taxonomic and usability improvements. *Bioinformatics* **31**, 3345–3347 (2015).
14. Li, H. et al. The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.* **6**, 8292 (2015).
15. Yip, V. L. Y. et al. An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from *Thermotoga maritima*. *J. Am. Chem. Soc.* **126**, 8354–8355 (2004).
16. Chapanian, R. et al. Enhancement of biological reactions on cell surfaces via macromolecular crowding. *Nat. Commun.* **5**, 4683 (2014).
17. Comfort, D. A. et al. Biochemical analysis of *Thermotoga maritima* GH36 alpha-galactosidase (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. *Biochemistry* **46**, 3319–3330 (2007).
18. Fredslund, F. et al. Crystal structure of alpha-galactosidase from *Lactobacillus acidophilus* NCFM: insight into tetramer formation and substrate binding. *J. Mol. Biol.* **412**, 466–480 (2011).
19. Calcutt, M. J., Hsieh, H. Y., Chapman, L. F. & Smith, D. S. Identification, molecular cloning and expression of an alpha-*N*-acetylgalactosaminidase gene from *Clostridium perfringens*. *FEMS Microbiol. Lett.* **214**, 77–80 (2002).
20. Gerbal, A., Maslet, C. & Salmon, C. Immunological aspects of the acquired B antigen. *Vox Sang.* **28**, 398–403 (1975).
21. Judd, W. J. & Annesley, T. M. The acquired-B phenomenon. *Transfus. Med. Rev.* **10**, 111–117 (1996).
22. Marcus, D. M., Kabat, E. A. & Schiffman, G. Immunochemical studies on blood groups. XXXI. Destruction of blood group a activity by an enzyme from *Clostridium tertium* which deacetylates *N*-acetylgalactosamine in intact blood group substances. *Biochemistry* **3**, 437–443 (1964).
23. Yamamoto, H. & Iseki, S. Development of H-specificity in A substance by A-decomposing enzyme from *Clostridium tertium* A. *P. Jpn Acad.* **44**, 263 (1968).

24. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
25. Blair, D. E., Schuttelkopf, A. W., MacRae, J. I. & van Aalten, D. M. F. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl Acad. Sci. USA* **102**, 15429–15434 (2005).
26. Ficko-Blean, E. & Boraston, A. B. The interaction of a carbohydrate-binding module from a *Clostridium perfringens N*-acetyl-beta-hexosaminidase with its carbohydrate receptor. *J. Biol. Chem.* **281**, 37748–37757 (2006).
27. Cohen, M., Hurtado-Ziola, N. & Varki, A. ABO blood group glycans modulate sialic acid recognition on erythrocytes. *Glycobiology* **19**, 1349–1349 (2009).
28. Hyono, A. et al. Impacts of papain and neuraminidase enzyme treatment on electrohydrodynamics and IgG-mediated agglutination of type A red blood cells. *Langmuir* **25**, 10873–10885 (2009).
29. Varki, A. et al. Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323–1324 (2015).
30. Kwan, D. H., Ernst, S., Kötzler, M. P. & Withers, S. G. Chemoenzymatic synthesis of a type 2 blood group a tetrasaccharide and development of high-throughput assays enables a platform for screening blood group antigen-cleaving enzymes. *Glycobiology* **25**, 806–811 (2015).
31. Armstrong, Z., Rahfeld, P. & Withers, S. G. Discovery of new glycosidases from metagenomic libraries. *Methods Enzymol. Chem. Glycobiol. A* **597**, 3–23 (2017).
32. Lee, S. & Hallam, S. J. Extraction of high molecular weight genomic DNA from soils and sediments. *J. Vis. Exp.* **33**, e1569 (2009).
33. Lee, E., Chuang, H. Y., Kim, J. W., Ideker, T. & Lee, D. Inferring pathway activity toward precise disease classification. *PloS Comput. Biol.* **4**, e1000217 (2008).
34. Jeong, J. K. et al. Characterization of the *Streptococcus pneumoniae* BgaC protein as a novel surface beta-galactosidase with specific hydrolysis activity for the gal beta 1-3GlcNAc moiety of oligosaccharides. *J. Bacteriol.* **191**, 3011–3023 (2009).
35. Singh, A. K. et al. Unravelling the multiple functions of the architecturally intricate *Streptococcus pneumoniae* beta-galactosidase, BgaA. *PLoS Pathog.* **10**, e1004364 (2014).
36. Katayama, T. et al. Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2-alpha-L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95). *J. Bacteriol.* **186**, 4885–4893 (2004).
37. Williams, S. J. & Withers, S. G. Glycosynthases: mutant glycosidases for glycoside synthesis. *Aust. J. Chem.* **55**, 3–12 (2002).
38. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
39. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* https://arxiv.org/abs/1303.3997 (2013).
40. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
41. Treangen, T. J., Sommer, D. D., Angly, F. E., Koren, S. & Pop, M. Next generation sequence assembly with AMOS. *Curr. Protoc. Bioinformatics* **33**, 11.8.1–11.8.18 (2011).
42. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
43. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
44. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
45. Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* **71**, 982–994 (2008).
46. Palmier, M. O. & Van Doren, S. R. Rapid determination of enzyme kinetics from fluorescence: overcoming the inner filter effect. *Anal. Biochem.* **371**, 43–51 (2007).
47. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
48. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
49. Skubak, P. & Pannu, N. S. Automatic protein structure solution from weak X-ray data. *Nat. Commun.* **4**, 2777 (2013).
50. Potterton, L. et al. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **74**, 68–84 (2018).
51. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
52. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60**, 2184–2195 (2004).
53. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
54. Zheng, L., Baumann, U. & Reymond, J. L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* **32**, e115 (2004).
55. Vocadlo, D. J., Wicki, J., Rupitz, K. & Withers, S. G. Mechanism of *Thermoanaerobacterium saccharolyticum* ss-xylosidase: kinetic studies. *Biochemistry* **41**, 9727–9735 (2002).
56. Jones, D. R. et al. SACCHARIS: an automated pipeline to streamline discovery of carbohydrate active enzyme activities within polyspecific families and de novo sequence datasets. *Biotechnol. Biofuels* **11**, 27 (2018).
57. Yin, Y. B. et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451 (2012).
58. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
59. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
60. Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**, 758–771 (2008).
61. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
62. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
63. Matsen, F. A., Hoffman, N. G., Gallagher, A. & Stamatakis, A. A format for phylogenetic placements. *PLoS ONE* **7**, e31009 (2012).
64. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).

Acknowledgements

We thank H. Brumer for the use of his Dionex high-performance liquid chromatography system, H.-M. Chen for the synthesis of Gal-α-MU and GalNAc-α-MU and T. Lowary for a set of α-GalNAc-containing oligosaccharide substrates. We acknowledge the participation of the Protein-Glycan Interaction Resource of the CFG (supporting grant R24 GM098791) and the National Center for Functional Glycomics at Beth Israel Deaconess Medical Center, Harvard Medical School (supporting grant P41 GM103694). We thank L. Worrall and N. Cavaney for collecting the X-ray diffraction data and the Canadian Light source for access. The infrastructure at the Centre for Blood Research is supported by the Canada Foundation for Innovation and the British Columbia Knowledge Development Fund. J.N.K. is a recipient of a Michael Smith Foundation for Health Research Scholar award. P.R. was supported by a Leopoldina-Postdoc-Fellowship from the Deutsche Akademie der Naturforscher Leopoldina. Financial support from the Canadian Institutes for Health Research (CIHR grant MOP-136940) is gratefully acknowledged by J.N.K. and S.G.W. H.M. is supported by a Canadian Blood Services-MITACS fellowship and an NSERC CREATE Nanomat scholarship.

Author contributions

P.R. performed the majority of the screening and characterization work as well as preparing figures and tables and providing a first draft of the manuscript; L.S. performed all work related to three-dimensional structural analysis; H.M. performed the majority of the testing with RBCs, in conjunction with I.C., and contributed to the writing; C.M.-L. performed much of the bioinformatics analysis; S.J.H. provided expertise in metagenomic analysis and supervision of the bioinformatics; J.N.K. supervised all work related to testing with RBCs and edited the manuscript; S.G.W. conceived the project, supervised the enzymology and coordinated the writing of the manuscript.

Competing interests

The University of British Columbia has filed a patent related to this work on which P.R., J.N.K. and S.G.W. are named as authors.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0469-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.G.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Stephen G. Withers

Last updated by author(s): Apr 9, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
□	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
□	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	□	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	□	A description of all covariates tested
☒	□	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
□	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	□	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	□	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	□	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	□	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Sequencing was performed by the UBC Sequencing Centre (Vancouver, B.C., Canada) using a Illumina MiSeq system.
Data analysis	Data analysis was performed using the following software and codes: Trimmomatic, BWA, samtools, bam2fastq script, MEGAHIT, minimus2, Prodigal, MetaPathways v2.5 software package, SACCHARIS' cazy_extract.pl script, TreeSAPP, HMMs from dbCAN, RAxML version 8.2.0, UCLUST, PROTGAMMAAUTO, CytExpert Software 2.1, Python,CCP4i2 , XDS, Aimless, CRANK2, Coot, Refmac and iTOL. All scripts are available on GitHub.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data sets generated and/or analyzed during this study are available from the corresponding author on reasonable request. Structure datasets generated during the current study are available in the PDB repository under accession numbers 6N1A and 6N1B. Sequence data for CspGH36 is deposited as GenBank accession code "MK500922" . Raw data for CD spectroscopy and Glycan array are available in the Supplementary Data Set.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The created metagenomic library was based on the fecal material of one human participant.
Data exclusions	For the Anti-H antibody agglutination tests (shown in Supplementary Table 12) 7 samples from A+ RBC's were excluded. Those 7 samples did show already high agglutination scores of 4 to 5 against the Anti-H antibody in the control reaction which were saturating the detection limit of the assay. H antigens and A antigens are present on all A RBC's and their surface display profile is strongly donor dependent.
Replication	Replicates were successful for all experiments.
Randomization	Sample randomization is not relevant in our study.
Blinding	Blinding is not relevant in our study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☐	☒	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	1) Alexa Fluor 647 Mouse Anti-Human Blood Group A , BD Pharmingen, 565384, NaM87--1F6, Lot 7136920, Clone: M89D3, Dilution: 1/100 (stock conc 0.2 mg/mL) 2) Anti-Blood Group H ab antigen antibody, Abcam, ab24213, 97-I, Lot GR169458-15, Clone: 97-I, Dilution: 1/100 (stock conc 100 ug at 1mg/mL) 3) Goat anti-Mouse IgM (Heavy chain) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488, Invitrongen, A-21042, Lot 1964383, Clone: polyclonal , Dilution: 1/300 (stock conc 2 mg/mL) 4) RPE-conjugated Monoclonal Mouse Anti-Human CD235a, Glycophorin A, Dako, F7078, JC159, Lot 20007847, Clone: JC159, Dilution: 1/100 5) Alexa Fluor 647 Mouse IgG3, κ Isotype Control, BD Pharmingen, 560803, J606, Lot 7128550, Clone: J606, Dilution: 1/100 (stock conc 0.2 mg/mL) 6) Mouse IgM [B11/7] - Isotype Control, Abcam, ab91545, B11/7, Clone: b11/7, Dilution: 1/100 (stock conc 100 ug at 0.1 mg/mL) 7) Isotype Reagent mouse IgG1/RPE, Dako, X0928, Lot 20007084, Clone: DAK-G01, Dilution: 1/100
Validation	1) Alexa Fluor 647 Mouse Anti-Human Blood Group A, Mouse IgG3 κ, Flow Cyt, Relevant validation info can be found here: http://www.bdbiosciences.com/eu/reagents/research/antibodies-buffers/immunology-reagents/anti-human-antibodies/cell-surface-antigens/alexa-fluor-647-mouse-anti-human-blood-group-a-nam87-1f6/p/565384#, Citations: a) Blanchard D, Bruneau V, Bernard D, et al. Flow cytometry analysis of dual red blood cell populations after bone marrow transplantation. Br J Haematol. 1995; 89(4):741-747. b) David B, Bernard D, Navenot JM, Muller JY, Blanchard D. Flow cytometric monitoring of red blood cell chimerism after bone marrow transplantation. Transfus Med. 1999; 9(3):209-217. c) Walker RH, ed. Technical Manual of the American Association of Blood Banks. Arlington VA: 1990 2) Anti-Blood Group H ab antigen antibody, Mouse IgM, ELISA WB IP Flow Cyt, Relevant validation info can be found here: https://www.abcam.com/blood-group-h-ab-antigen-antibody-97-i-ab24213.html?productWallTab=ShowAll#top-306 3) RPE-conjugated Monoclonal Mouse Anti-Human CD235a, Glycophorin A, Mouse IgG1 κ, Flow Cyt, Relevant validation info can

be found here: https://www.agilent.com/store/productDetail.jsp?catalogId=R707801-2

Human research participants

Policy information about studies involving human research participants

Population characteristics

The metagenomic library created in this study was based on the fecal material of one healthy male Human research participant with the blood group AB+.The participant did not eat any food containing prebiotics or probiotics, nor did he receive any antibiotics 30 days before sample collection.

Recruitment

The Human research participant was chosen based on his blood type. The participation was voluntary. The participant was informed about the project though a "Participant Information and Consent Form", which he had to sign.

Ethics oversight

The collection of human fecal samples was approved by the Clinical Research Ethics Board of the University of British Columbia (ID: #H15-02967). The collection of human blood samples was approved by the Clinical Research Ethics Board of the University of British Columbia (ID: #H07-02198) and the Canadian Blood Services (REB: # 2017.029).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Enzyme treated RBCs were washed 2x with 1xPBS pH7.4. After diluting cells to 1% hematocrit, enzyme converted-RBCs were treated with 1/100 APC-anti-A antibody (Alexa Fluor® 647 Mouse Anti-Human Blood Group A: cat no. 565384 [BD Pharmingen]) for 30mins at RT, and/or treated with 1/100 anti-H antibody (Anti-Blood Group H ab antigen antibody [97-I]: cat no. ab24213 [Abcam]) for 30 mins at 4oC, then washed 1x with 1xPBS pH7.4. For detection of the anti-H antibody a secondary Alexa fluor 488-labelled antibody (Goat anti-Mouse IgM (Heavy chain) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488: cat no. A-21042 [Invitrogen]) in a 1/300 concentration was applied at RT for 30mins. Flow cytometry was performed after reconstituting enzyme converted RBCs into 1xPBS pH 7.4 (1% hematocrit) on a CytoFLEX flow cytometer using CytExpert Software 2.1 [Beckman Coulter]. Naive RBCs at 1% hematocrit, collected after centrifugation of whole blood at 1000g for 5mins (Allegra X-22R centrifuge [Beckman Coulter]) for the complete removal of PRP (platelet rich plasma), was used to set a gate for the data collection.

Instrument

CytoFLEX flow cytometer [Beckman Coulter]

Software

CytExpert Software 2.1 [Beckman Coulter]

Cell population abundance

Flow cytometer samples were prepared with RBCs that were collected after complete removal of platelet rich plasma after centrifugation of whole blood at 1000g for 5mins. 100% RBC population was used for the Flow cytometry assessment. To assure the purity of the cell population being assessed, naive RBCS were used to set a gate for the data collection. Data were collected only for the cell population defined by the gate.

Gating strategy

Naive RBCs at 1% hematocrit, collected after centrifugation of whole blood at 1000g for 5mins (Allegra X-22R centrifuge [Beckman Coulter]) for the complete removal of PRP (platelet rich plasma), was used to set a gate for the data collection. To confirm the cell population being assessed is RBCs, naive RBC samples were treated with 1/100 RPE-conjugated monoclonal mouse anti-Human CD235a, Glycophorin A, Clone JC159 (cat no: R7078 [Dako]) at 4oC for 30mins and assessed on a CytoFLEX flow cytometer [Beckman Coulter] to set a gate for the data collection. For 'positive' antigen A expressing samples, blood group A was used and for 'negative' antigen A expressing samples, blood group O was used as controls. 'Control' blood samples were typed using Micro Typing System (MTS) cards [ID-Micro Typing SystemTM (ID-MTSTM) Gel Cards from Ortho Clinical Diagnostics] according the the manufacturer's protocol prior to the flow cytometry assessment.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.