nature biotechnology

Bacterial glycosidases for the production of universal red blood cells

Qiyong P Liu[1,9], Gerlind Sulzenbacher[2,9], Huaiping Yuan[1], Eric P Bennett[3], Greg Pietz[1,3], Kristen Saunders[1], Jean Spence[1], Edward Nudelman[1], Steven B Levery[4], Thayer White[1], John M Neveu[5], William S Lane[5], Yves Bourne[2], Martin L Olsson[6,7], Bernard Henrissat[2] & Henrik Clausen[3,8]

Enzymatic removal of blood group ABO antigens to develop universal red blood cells (RBCs) was a pioneering vision originally proposed more than 25 years ago. Although the feasibility of this approach was demonstrated in clinical trials for group B RBCs, a major obstacle in translating this technology to clinical practice has been the lack of efficient glycosidase enzymes. Here we report two bacterial glycosidase gene families that provide enzymes capable of efficient removal of A and B antigens at neutral pH with low consumption of recombinant enzymes. The crystal structure of a member of the α-*N*-acetylgalactosaminidase family reveals an unusual catalytic mechanism involving NAD^+. The enzymatic conversion processes we describe hold promise for achieving the goal of producing universal RBCs, which would improve the blood supply while enhancing the safety of clinical transfusions.

The ABO blood group system, discovered in 1900 (ref. 1), is the most important blood group system to consider in transfusion medicine. This system is based on the presence or absence of the blood group antigens A and B. The corresponding blood group carbohydrate structures, designated ABH, are found at the termini of oligosaccharide chains on glycoproteins and glycolipids on the surface of erythrocytes and of endothelial and most epithelial cells[2,3]. The immunodominant monosaccharide that determines blood group A specificity is a terminal α-1,3-linked *N*-acetylgalactosamine (GalNAc), whereas the corresponding monosaccharide of blood group B specificity is an α-1,3-linked galactose (Gal) (**Fig. 1**). Group O cells lack both of these monosaccharides at the termini of their oligosaccharide chains, which instead are terminated with α-1,2-linked fucose (Fuc) residues and designated the H antigen (see review[2]). The genetic basis of ABO antigens has been elucidated; genes determining A and B blood types encode glycosyltransferases in which a few amino acid substitutions result in differing substrate specificities: the A enzyme for UDP-GalNAc and the B enzyme for UDP-Gal, whereas the gene for the O blood type encodes an inactive protein[4].

Plasma of blood group A individuals contains naturally occurring antibodies to the B antigen. Conversely, that of blood group B individuals contains antibodies to the A antigen. Blood group AB plasma has neither antibodies, and blood group O plasma has both antibodies. Individuals with anti-A and/or anti-B antibodies cannot receive a transfusion of blood containing the corresponding incompatible antigen(s). Inappropriate transfusion of ABO-incompatible RBCs can cause complement activation and lysis of the incompatible RBCs, often resulting in an acute intravascular hemolytic transfusion reaction. Inadvertent mismatching of RBCs for ABO blood groups remains one of the most common and persistent causes of serious and sometimes fatal adverse events following transfusion, according to hemovigilance monitoring systems[5,6]. Group O RBCs contain neither A nor B antigens and can be transfused safely into recipients of any ABO blood group, that is, group A, B, AB or O recipients. Thus, type O blood is considered 'universal' and may be used for all RBC transfusions. Hence, blood banks strive to maintain significant inventories of group O RBCs. In reality, however, there is a constant shortage of group O units as these are always in great demand for numerous reasons, including emergency situations in which the patient's blood group is unknown, shortages of blood from a particular blood group, pediatric transfusions, and lack of clarity regarding the patient's blood group (e.g., due to previous transfusions or allogeneic stem cell transplantation). Enzymatic removal of the immunodominant αGalNAc and αGal monosaccharides from group A and B structures of non-O RBCs offers an appealing approach to improve transfusion safety and the overall blood supply[7].

Goldstein originally envisioned enzymatic conversion of the A and B antigens on RBCs to H antigen by use of exoglycosidases for

[1]ZymeQuest Inc., 100 Cummings Center, Suite 436H, Beverly, Massachusetts 01915, USA. [2]Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS, Universités Aix-Marseille I & II, Case 932, 163 Avenue de Luminy, 13288 Marseille Cedex 9, France. [3]Departments of Cellular and Molecular Medicine and Oral Diagnostics, University of Copenhagen, Blegdamsvej, DK-2200 Copenhagen N, Denmark. [4]Department of Chemistry, University of New Hampshire, Durham, New Hampshire 03824, USA. [5]Harvard Microchemistry and Proteomics Analysis Facility, Harvard University, Cambridge, Massachusetts 02138, USA. [6]Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University and University Hospital Blood Center, SE-22185, Lund, Sweden. [7]Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA. [8]Hematology Division Brigham & Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [9]These authors contributed equally to this work. Correspondence should be addressed to G.S. (gerlind.sulzenbacher@afmb.univ-mrs.fr) or H.C. (hc@imbg.ku.dk).

Received 14 August 2006; accepted 4 February 2007; published online 1 April 2007; doi:10.1038/nbt1298

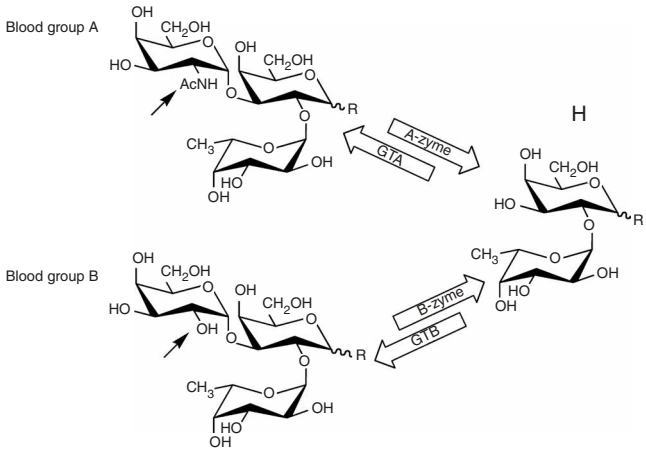

Figure 1 Structural basis of the ABO blood group antigens. The immunodominant A and B trisaccharide epitopes are formed from the common H disaccharide substrate by α1,3-*N*-acetylgalactosaminyltransferase (GTA), defined by the blood group A gene, and α-galactosyltransferase (GTB), defined by the blood group B gene, respectively. Conversely, the strategy used for enzymatic conversion of blood group A and B antigens to H involves exoglycosidases that specifically hydrolyze the α1,3GalNAc (α-*N*-acetylgalactosidase, A-zyme) or the α1,3galactose (α-galactosidase, B-zyme) to form the common H structure found on O RBCs. Black arrows indicate the different C-2 *N*-acetyl group of GalNAc and OH group of Gal in the immunodominant A and B epitopes, respectively. The immunodominant epitopes are positioned at the termini of oligosaccharide chains on glycolipids and glycoproteins as indicated by R. Increased complexity in ABH oligosaccharide structures are provided by the oligosaccharide carrier chain (R). On human RBCs most structures are based on type 2 polylactosamine chains with repeating Galβ1-4GlcNAc disaccharide units (both glycolipids and N-linked glycoproteins). A minor amount of type 1 chain ABH structures with Galβ1-3GlcNAc are found as glycolipids adsorbed from plasma[3]. A fraction of the blood group A glycolipids with type 2 chain core structures are extended by a βGal residue (Galβ1-3GalNAcα1-3(Fucα1-2)Galβ1-4GlcNAcβ1-R), which is further preferentially modified to an A-associated H epitope in subgroup A_2 RBCs (Fucα1-2Galβ1-3GalNAcα1-3(Fucα1-2)Galβ1-4GlcNAcβ1-R) and a repetitive A epitope (GalNAcα1-3(Fucα1-2)Galβ1-3GalNAcα1-3(Fucα1-2)Galβ1-4GlcNAcβ1-R) in subgroup A_1 RBCs (see **Fig. 4** for glycolipid structures). Cleavage of the repetitive A structure by an α-*N*-acetylgalactosaminidase will result in exposure of the H disaccharide epitope.

selective removal of αGalNAc and αGal residues of the immunodominant A and B trisaccharide antigens, respectively, to expose the underlying H antigen[8] (**Fig. 1**). For conversion of group B RBCs, he used an α-galactosidase derived from green coffee beans (*Coffea canephora*), which has a very low pH optimum (pH 3.5) and required as much as 1–2 g enzyme per unit of group B RBCs for adequate conversion at acidic pH 5.5. A phase 2 crossover trial was completed with full unit (~200–220 ml packed) transfusions of group B RBCs enzyme-converted to group O (B-ECO), in which the converted cells functioned and survived normally in group A and O patients[9]. Although this study clearly demonstrated that enzymatic conversion of RBCs is feasible and that B-ECO cells can function as well as ABO-matched untreated cells in transfusions, a major obstacle to clinical implementation of this method is the large quantities of enzyme required. Other investigators have explored different α-galactosidases and protocols for conversion of group B RBCs but without substantial improvements in enzyme consumption[10,11].

Successful conversion of group A RBCs has not yet been achieved. The biochemistry of A antigens is more complex than that of B antigens. Subgroup A_1 RBCs have ~5× more A antigens than subgroup A_2, and, in addition, subgroup A_1 cells have glycolipids with a repeated A structure, which in A_2 cells mainly exist as an A-associated H structure (see **Fig. 1** and **Table 1** for structures)[12,13]. Conversion of only the A_2 subgroup to group O was achieved with a

Table 1 Substrate specificities analyzed by TLC assay

		α-*N*-acetylgalactosaminidase	α-galactosidases		
Substrate	Blood group specificity	*E. meningosepticum* (pH 6.8)	Coffee bean[a] (pH 5.5)	*S. griseoplanus*[a] (pH 6.8)	*B. fragilis* (FragA) (pH 6.8)
GalNAcα1-2(3,4)Gal Structures					
GalNAcα1-2Galβ1-MCO		+	ND	ND	ND
GalNAcα1-3Galβ1-MCO	A-di	+	ND	ND	ND
GalNAcα1-3Galβ1-4GlcNAcβ1-MCO	Linear A-tri	+	ND	ND	ND
GalNAcα1-4Galβ1-4Glcβ1-MCO		+	ND	ND	ND
GalNAcα1-3(Fucα1-2)Gal	A-tri	+	ND	ND	–
GalNAcα1-3(Fucα1-2)Galβ1-3GalNAcβ1-AMC	A-tetra (type 3 chain)[b]	+ (0.25 U/mg)[c]	–	–	–
GalNAcα1-3(Fucα1-2)Galβ1-3GalNAcα1-3(Fucα1-2) Galβ1-4GlcNAcβ1-AMC	A-hepta (type 3 chain, repetitive A)	+ (0.25 U/mg)	ND	ND	ND
Galα1-3Gal Structures					
Galα1-3Gal	B-di	ND	ND	ND	–
Galα1-3Galβ1-4GlcNAc	Linear B	ND	+	–	–
Galα1-3(Fucα1-2)Gal	B-tri	–	+	+	+
Galα1-3(Fucα1-2)Galβ1-4GlcNAcβ1-AMC	B-tetra (type 2 chain)	–	+ (0.017 U/mg)	+	+ (6.6 U/mg)
Galα1-4Gal Structures					
Galα1-4Gal	P1	–	+	–	–
Galα1-4Galβ1-4Glc	P^k	–	+	–	–

[a]From published US patent application no. US 20,050,208,655. [b]Core structures of ABH antigens designated type 1–4, as described previously[3]. [c]Specific activity determined using quantification by fluorography of TLC analysis.
+, complete digestion of substrate using conditions that produce complete digestion of the AMC-labeled blood group A or B tetrasaccharide substrates. –, no detectable digestion of substrate under the same conditions even after extended reaction time (10×). ND, not determined.

chicken liver α-*N*-acetylgalactosaminidase using 1 g of enzyme per unit of RBCs at pH 5.5 (ref. 14). An α-*N*-acetylgalactosaminidase from *Clostridium perfringens* has also been used[15]. Purification of an α-*N*-acetylgalactosaminidase from *Arenibacter latericius* with neutral pH optimum, which appeared to convert both subgroup A_1 and A_2 cells after 24 h incubation, has been reported[16]. The efficiency of the conversion, however, was not confirmed by licensed typing methods, and the gene has not been identified yet. It is, therefore, not clear whether the difficulties with conversion of the A_1 subgroup relate to the complexity of A antigens or to the poor kinetic properties and acidic pH optima of the enzymes used to date[14,15].

In the past, α-*N*-acetylgalactosaminidases and α-galactosidases used for removing the immunodominant αGalNAc and αGal residues from the A and B antigens, respectively, have mainly been identified and characterized by enzymatic assays using simple monosaccharide *p*-nitrophenyl (pNP) substrates. We hypothesized that the use of this strategy would fail to identify enzymes with selective or exclusive substrate specificities for the more complex branched blood group A and B structures. We therefore screened a large panel of bacterial and fungal isolates to search for enzymes with the preferred substrate specificities and a neutral pH optimum. Here we describe two previously unknown prokaryotic glycosidase families, which contain enzymes with favorable or exclusive substrate specificities for the branched blood group structures, neutral pH optimum, and high efficiency and selectivity in enzymatic conversion of group A, B and AB RBCs to group O RBCs. The α-*N*-acetylgalactosaminidase functions with an unusual catalytic mechanism involving NAD^+ and the crystal structure of the enzyme reveals a structural fold that has not been found previously in glycosidases. The availability of these bacterial enzymes should allow efficient and cost-effective enzymatic conversion of blood group A, B and AB RBCs to universal RBCs.

RESULTS

Screening for prokaryotic exoglycosidases

To identify enzymes that might have preferred or exclusive specificity for blood group A and B structures and neutral pH optima, we screened a large panel of 2,500 fungal and bacterial isolates for enzymes capable of cleaving A and B oligosaccharide substrates. Most bacteria express high levels of known α-galactosidases that have acidic pH optima and high activity with the Galα-pNP substrate, and some also express α-*N*-acetylgalactosaminidases that have similarly high activity with the GalNAcα-pNP substrate. We therefore used differential screening with two assays; one with a blood group A or B tetrasaccharide substrate labeled with 7-amino-4-methyl-coumarin (A or B tetrasaccharide-AMC) substrate and one with the GalNAcα-pNP or Galα-pNP substrate (**Table 2**). Assays were performed at neutral pH to select enzymes with the preferred pH optimum. Five of the 2,500 isolates were found to display α-galactosidase activity with the B tetrasaccharide-AMC substrate and not with the monosaccharide Galα-pNP substrate, whereas two isolates showed α-*N*-acetylgalactosaminidase activity with the A tetrasaccharide-AMC substrate as well as the monosaccharide GalNAcα-pNP substrate (data not shown). Both activities exhibited neutral pH optima.

Improved enzymes for conversion of group A RBCs – family GH109

The α-*N*-acetylgalactosaminidase activities identified in the screen were similar to that of an *Elizabethkingia meningosepticum* (previously termed *Chryseobacterium meningosepticum*, ATCC no. 13253) α-*N*-acetylgalactosaminidase (GenBank accession no. AM039444), which was originally identified and cloned using the same blood group A tetrasaccharide-AMC substrate for screening[17]. The deduced protein sequence of the *E. meningosepticum* enzyme was used in GenBank searches to identify a family of enzymes found exclusively in prokaryotes, classified as GH109 in the carbohydrate-active enzyme database (http://www.CAZY.org/) (**Supplementary Fig. 1** online). The genes encoding five of these proteins were cloned and expressed (**Supplementary Methods** online). One of two genes identified in *Bacteroides fragilis* (AM039447) as well as those from *Shewanella oneidensis* (AM039445) and *Tannerella forsythensis* (AM039448) were found to encode functional α-*N*-acetylgalactosaminidases with properties similar to those of the *E. meningosepticum* enzyme (data not shown). The second gene identified in *B. fragilis* (AM039446) and the *Streptomyces coelicolor* gene (AM039449) did not appear to encode proteins with α-*N*-acetylgalactosaminidase activity.

A distant relationship of these enzymes to oxidoreductases was noted using the NCBI conserved domain search. We therefore considered the possibility that these enzymes functioned with NAD^+ and metal ions as cofactors. An N-terminal truncated construct (amino acids 18–444) of the *E. meningosepticum* was expressed efficiently in *Escherichia coli* with a yield of 1 g/l, purified to homogeneity and used for the studies reported here. The purified recombinant enzyme was not affected by EDTA/EGTA chelation of potential metal ions, and exogenous addition of metal ions did not enhance activity. Furthermore, addition of NAD^+ or NADH had no effect on enzyme activity (**Supplementary Fig. 2** online). As a result of these findings exogenous addition of metal ions and nucleotides were not used in standard enzyme assays in the study. However, the enzyme does in fact use NAD^+ as a cofactor, but NAD^+ is tightly bound, and the enzyme is saturated in its purified recombinant form.

The substrate specificity and kinetic properties of the recombinant *E. meningosepticum* α-*N*-acetylgalactosaminidase are presented in **Tables 1** and **2**. The enzyme has a neutral pH optimum with a relative high specificity for the blood group A AMC substrate (0.25 U/mg) (**Supplementary Figs. 3a** and **4a** online) compared with the monosaccharide GalNAcα-pNP substrate (>10 U/mg).

Table 2 Kinetic properties of the α-*N*-acetyl-galactosaminidase and α-galactosidase

	E. meningosepticum α-*N*-acetyl-galactosaminidase			*B. fragilis* α-galactosidase A (FragA)		
Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
Galα-pNP	0.026 ± 0.001	25.1 ± 1.4	0.0011	0.38 ± 0.01	0.70 ± 0.02	0.54
Galβ-pNP	0.010 ± 0.000	3.6 ± 0.3	0.0031	0.00[a]	NA	NA
GalNAcα-pNP	9.84 ± 0.16	0.077 ± 0.006	127.6	0.00	NA	NA
GalNAcβ-pNP	0.015 ± 0.000	0.23 ± 0.01	0.087	0.00	NA	NA

Data represent the average of triple or more measurements.
[a]No activity was found with 60×-longer reaction time than was required to readily detect the activity of Galα-pNP. NA, not applicable.

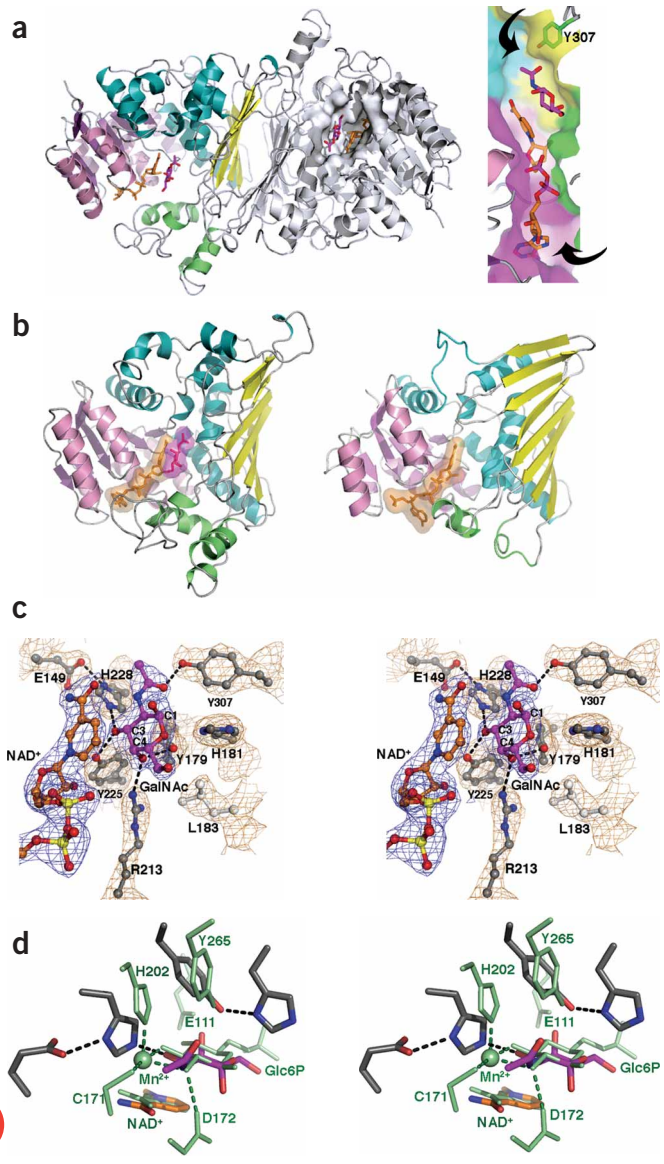

Figure 2 Overall structure of the *E. meningosepticum* α-*N*-acetylgalactosaminidase. (**a**) Left panel. Cartoon representation of the overall structure of the α-*N*-acetylgalactosaminidase dimer. In the left subunit, the dinucleotide binding domain (residues 19–148) is shown in pink, the α-helices of the C-terminal domain (residues 149–444) in cyan, the β-sheet forming the dimer interface in yellow, and the α-helical bundle covering the NAD^+-binding tunnel in green. In the right subunit, a partial surface representation shows the entry of the NAD^+ binding tunnel and the active-site pocket. NAD^+ and GalNAc are shown as orange and red sticks, respectively. Right panel. Close-up and clipped view of the NAD^+ binding tunnel created at the interface of the dinucleotide binding domain (pink) and the α-helical bundle (green). (**b**) Side-by-side representation of α-*N*-acetylgalactosaminidase (left) and *Zymomonas mobilis* glucose-fructose oxidoreductase (PDB 1OFG) (right,) using the same coloring convention as in **a**. The NAD^+ molecule and GalNAc bound to the α-*N*-acetylgalactosaminidase and the $NADP^+$ cofactor bound to 1OFG are highlighted. (**c**) Stereo view of the active site of α-*N*-acetylgalactosaminidase oriented as in **b**. The GalNAc residue, part of the NAD^+ and the surrounding amino acid residues described in the text are shown as ball-and-stick models. Maximum-likelihood/σA-weighted $2F_o$-F_c electron density (contoured at ~1.5 σ) is shown in blue for GalNAc and NAD^+ and in orange for protein residues. (**d**) Stereo view of the superimposition of the active sites of α-*N*-acetylgalactosaminidase (gray) and *B. subtilis* 6-phospho-α-glucosidase (green) (PDB 1NRH).

Surprisingly, low but detectable activity was also observed with GalNAcβ-pNP as well as with Galα and Galβ-pNP substrates (**Supplementary Fig. 5** online). Interestingly, the apparent K_m value for GalNAcα-pNP is similar to that of GalNAcβ-pNP, although k_{cat} values differ more than 1,000-fold. In contrast, the K_m value for Galβ-pNP is over tenfold higher than that for the GalNAc counterpart. Although glycoside hydrolases are usually extremely specific for the anomeric configuration of the glycosidic bond, a member of glycosidase family GH4, namely *Fusobacterium mortiferum* phospho-α-glucosidase, has been reported to be able to cleave both α- and β-anomers of the nonnatural *p*-NP 6-phosphoglucosides substrates[18]. We did not observe any cleavage of the natural globoside (Gb_4) glycolipid structure with a terminal β1-3GalNAc residue; however, low but detectable cleavage of the disaccharide substrate GalNAcβ1-4Galβ1-$O(CH_2)_8COOCH_3$ was observed (data not shown). Monitoring the hydrolysis of GalNAcα-pNP by NMR showed that the α-*N*-acetylgalactosaminidase functions with a mechanism leading to overall retention of the anomeric configuration (**Supplementary Results** online). The combined findings of sequence relatedness to oxidoreductases, lack of similarity to known glycosidases and the unusual enzymatic properties motivated us to determine the structure of the enzyme by X-ray crystallography.

Structural architecture of the α-*N*-acetylgalactosaminidase

The crystal structure of the *E. meningosepticum* α-*N*-acetylgalactosaminidase was solved by the multiple anomalous dispersion (MAD) method, taking advantage of the anomalous dispersion of selenium, and refined against native data to 2.3 Å resolution (**Supplementary Table 1** online). The structure of a complex with GalNAc was refined to 2.4 Å resolution. The enzyme is present as a homodimer in the crystal and appears to form a dimer in solution, as evidenced by dynamic light-scattering analysis. However, analysis by size-exclusion chromatography suggests that the enzyme exists as a monomer or a dimer, depending on buffer conditions (data not shown). The final structure of the enzyme comprises the entire mature protein (residues 18–444) and a molecule of NAD^+. Although the distant sequence relatedness with oxidoreductases suggested that the enzyme could function with NAD^+ as a cofactor, similarly to enzymes of the GH4 family[19–21], this was a surprise because we had observed no stimulation of activity with exogenous addition of NAD^+ or NADH in our studies of the kinetic properties of the enzyme (**Supplementary Fig. 2** online).

Structurally, the protein consists of two independent and closely associated domains, forming a narrow tunnel in which the NAD^+ molecule is embedded (**Fig. 2a**). The N-terminal domain forms a classical dinucleotide-binding Rossmann fold[22], whereas the C-terminal domain consists of an α/β domain with a large open-faced, seven-stranded, anti-parallel β-sheet, adjacent to nine α-helices and an α-helical bundle covering the dinucleotide-binding tunnel. The open-faced β-sheet, with a surface area of 2,800 $Å^2$, representing 15% of the total surface of a monomer, makes up the dimer interface. The NAD^+ molecule, copurified with the recombinant enzyme, is anchored within a 10 Å long, narrow tunnel, almost entirely shielded from the solvent; the nicotinamide ring is located in a narrow pocket situated at the bottom of a large depression at the surface of the protein. Because exogenous addition of NAD^+ did not stimulate the enzyme, we developed a method for depletion of NAD^+ by precipitating the protein with ammonium sulfate under acidic

conditions. The release of NAD^+ was confirmed by high-performance liquid chromatography analysis using standards (data not shown). The acid ammonium-sulfate–treated enzyme was inactive after resolubilization; complete activity was restored by exogenous addition of NAD^+, but not by related compounds including NADH (**Supplementary Fig. 6** online). The same dependence of NAD^+ was also found for the activity with GalNAcβ-pNP, confirming that this is an intrinsic activity of the enzyme (**Supplementary Fig. 7** online).

The overall fold of the α-*N*-acetylgalactosaminidase is unique among the more than 370 glycosidases with known three-dimensional structures (carbohydrate-active enzyme database at http://www.CAZY.org/). A Dali search[23] for structural homologs within the nonredundant set of protein structures from the Protein Data Bank revealed that the closest structural relatives belong to the Gfo/Idh/MocA oxidoreductase family (Z-score of 29.4 and r.m.s. deviation of 3.0 Å for 329 equivalent Cα-atoms for the closest structural neighbor, namely *Zymomonas mobilis* glucose-fructose oxidoreductase, 1OFG[24]) (**Fig. 2b**). More distant structural homologs are all identified by means of the classical Rossmann fold. The structural similarity extends to the level of the active-site architecture, where the spatial arrangement of the cofactor and several key residues is conserved, suggesting a common ancestor that has evolved its NAD^+-based molecular mechanism to adapt to diverse metabolic requirements.

Active site of the α-*N*-acetylgalactosaminidase

The complex with the reaction product, GalNAc, showed the carbohydrate inside a narrow pocket situated in the middle of a large crater. The wide opening of the crater appears suitable for the accommodation of large aglycon structures such as the blood group A antigen. The GalNAc moiety is found close to NAD^+ with the C-3 atom of the sugar 3 Å away from the C-4 atom of the nicotinamide ring (**Fig. 2c**). Four residues, Tyr307, Tyr225, His228 and Glu149, which appear to be important for the interactions with the carbohydrate (**Supplementary Fig. 8** online), are invariant within the GH109 family (**Supplementary Fig. 9b** online).

What governs the preference of the enzyme for GalNAc over Gal? Replacement of the C-2 acetamido group by a hydroxyl substantially impairs enzyme-substrate interaction, as suggested by a 2,000-fold reduction in k_{cat} for the hydrolysis of Galα-pNP compared with GalNAcα-pNP and by a more than tenfold increase in K_m (**Table 2**). The structure of α-*N*-acetylgalactosaminidase in complex with GalNAc shows that the interaction of Tyr307 with the carbonyl oxygen of the acetamido group of the substrate is probably responsible for the preference for GalNAc substrates.

Our finding that the second family member from *B. fragilis* (AM039446) did not exhibit α-*N*-acetylgalactosaminidase activity suggests that the family includes enzymes acting on different substrates. Alternative substrate specificity might originate from chemical variability at C-4, C-5 and/or C-6, as NAD^+ and the invariant residues impose invariant interactions with the C-2 and C-3 substituents of the carbohydrate. Indeed, in the complex structure, OH-4 establishes hydrogen bonds with the conserved Tyr179 and residue Arg213, whereas no hydrogen bonding partner could be observed for OH-6. This exo-cyclic hydroxyl is surrounded by residues Leu183, Val186, Glu209 and Arg213. Whereas Leu183 is strictly conserved in all members of GH109, the three latter residues are conserved only in the branch of the evolutionary tree comprising *E. meningosepticum* and the enzymes from *Flavobacterium sp.* MED217 and *Robigitalea biformata* HTCC2501 (**Supplementary Figs. 1a** and **9a** online). These three enzymes are also characterized by a ~15-residue insertion at the tip of the α-helical bundle.

The catalytic mechanism of the α-*N*-acetylgalactosaminidase

Although the α-*N*-acetylgalactosaminidase and the glycosidases from family GH4 differ in their overall protein folds, their active-site architectures show striking similarities (**Fig. 2d**), thereby constituting a beautiful example of convergent evolution. The glycoside hydrolase family GH4 groups together α- and β-glycosidases that have been shown recently to use an NAD^+-dependent mechanism that is highly unusual for glycosidic bond cleavage[19–21]. The structural superimposition of the active site of α-*N*-acetylgalactosaminidase onto those of *Bacillus subtilis* phospho-α-glucosidase, GlvA[20] and *Thermotoga maritima* 6-phospho-β-glycosidase, BglT[25], reveals a perfect overlap of the dinucleotide cofactors, the sugar rings and a tyrosine residue.

Family GH4 enzymes contain a divalent metal ion coordinated by a conserved cysteine, a conserved histidine and the nicotinamide group and thought to stabilize an enodiolate-reaction intermediate[20]. Notably, the crystal structure of the α-*N*-acetylgalactosaminidase is devoid of any metal ion, in agreement with the observation that the enzyme does not display any metal ion requirements (**Supplementary Fig. 2** online). Monitoring of the reaction catalyzed by α-*N*-acetylgalactosaminidase indicated that the enzyme proceeds with a mechanism that leads to retention of the anomeric configuration and concomitant exchange of the GalNAc H-2 atom for a solvent proton (**Supplementary Results** online). These results parallel exactly those of family GH4 glycosidases and therefore indicate that GH109 enzymes most likely operate by a similar mechanism. In this mechanism, an 'onboard' NAD^+ molecule oxidizes the substrate at C-3, thereby acidifying the proton at C-2. Deprotonation of C-2 by an enzymatic base with concomitant elimination of the glycosidic oxygen generates a 1,2-unsaturated intermediate. The reaction is completed by addition of water to the Michael-like acceptor and reduction of the resulting ketone by the onboard NADH molecule, which returns to the initial NAD^+ state, ready for another catalytic cycle (**Supplementary Fig. 10** online). This mechanism, found first in family GH4 enzymes, allows cleavage of thioglycosides[26], a feature that is extremely rare among 'classical' glycosidases. Thin layer chromatography (TLC) analysis showed that α-*N*-acetylgalactosaminidase is indeed able to cleave benzyl-1-thio-α-*N*-acetylgalactosaminide (data not shown).

During the reaction catalyzed by GH4 glycosidases, departure of the aglycone is thought to be acid catalyzed[20,25], but the role and identity of the acid catalyst is still uncertain. Interestingly, the structure of the α-*N*-acetylgalactosaminidase does not reveal any suitable residue to protonate the interosidic atom and suggests that this step is perhaps unnecessary in family GH109. This is perfectly compatible with the observed activity with benzyl-1-thio-α-*N*-acetylgalactosaminide and provides a structural explanation of the hydrolysis of both α- and β-anomers of pNP substrates; indeed only a minor adjustment of the side chain of His181 is necessary to accommodate a β-linked substrate.

After family GH4, family GH109 constitutes the second glycosidase family that displays structural and mechanistic relatedness to oxidoreductases. There is nothing in the α-*N*-acetylgalactosaminidase sequence that suggested hydrolytic activity on glycosides. In consequence it is conceivable that other distant homologs of known oxidoreductases may also correspond to glycosidases and the repertoire of this category of enzymes might further develop. We note a very distant but intriguing similarity between family GH109 members and a recently discovered *T. forsythensis* sialidase[27].

Improved enzymes for conversion of group B RBCs – family GH110

Of the five isolates found to express α-galactosidase activity with the B tetrasaccharide-AMC at neutral pH, one isolate (no. 2,357) from a

Table 3 Blood group typing of ECO RBCs with enzyme dose titrations

			ABO typing results[a]							
			Ortho anti-A		ES-15		Ortho anti-B		Diagast anti-B	
ECO RBCs	A-zyme dose/ml pRBCs (mg)	B-zyme dose/ml pRBCs (mg)	IS	4 °C	IS	4 °C	IS	4 °C	IS	4 °C
A_1-ECO RBCs	0.73		0	0	0	0				
	0.60		0	0	0	0				
	0.50		0	0	0	0				
	0.37		0	0	0	0				
	0.23		0	0	Vw+	0				
	0.13		0	0	W+	W+				
	0.03		0	0	1+	1+				
B-ECO RBCs		0.02					0	0	0	0
		0.01					0	0	0	0
		0.005					0	0	vw	w
		0.0025					0	w	1	2
		0.0013					0	2	1	3
		0.0006					w	2	2	3
A_1B-ECO RBCs	0.73	0.02	0	0	0	0	0	0	0	0
	0.60	0.01	0	0	0	0	0	0	0	0
	0.50	0.005	0	0	0	0	0	0	0	0
	0.37	0.0025	0	0	0	0	0	vw+	w+	w+
	0.23	0.0013	0	0	0	0	0	w+	w+	1+
	0.13	0.0006	0	0	vw+	w+	0	1+	1+	1+
	0.03	0.0003	1+	2+	2+	2+	vw+	2+	1+	2+

[a]Typing with licensed ABO typing reagents and methods as indicated using agglutination score 0 to 4+.
IS, immediate spin reactions; pRBCs, packed RBCs.

Streptomyces griseoplanus strain exhibited the highest enzyme activity in lysates and was found to secrete the α-galactosidase enzyme into the medium. Initial attempts to purify the enzyme from the medium failed because of an insufficient amount of enzyme, and we therefore screened different nutrient carbon sources using minimal medium in an attempt to enhance secretion. A 10- to 40-fold induction of secretion was achieved with galactose and lactose (**Supplementary Methods** online), and an enzyme with a molecular weight of ~75 kDa was isolated from the medium. A number of peptide sequences (S1-S7) were obtained by Edman degradation and tandem mass spectrometry, and database searches identified weak sequence similarities with a putative protein from *Streptomyces avermitilis*. Further searches with the *S. avermitilis* sequence identified four additional genes with substantial similarity from *Bacteriodes fragilis* and *Bacteriodes thetaiotaomicron* (**Supplementary Figs. 1b** and **9b** and **Supplementary Methods** online). All five identified genes were cloned by PCR from their corresponding genomic DNA, expressed in *E. coli* as His_6-tagged proteins, and shown to encode α-galactosidases when analyzed with a B tetrasaccharide-AMC substrate (data not shown). The lack of sequence similarity between the newly identified α-galactosidases and any known glycosidase in the CAZY database allows us to define the glycosidase family GH110. One of the two α-galactosidase homologs found in *B. fragilis* NCTC 9343 (designated FragA, EMBL Nucleotide Sequence Database accession no. AM109955) was expressed in *E. coli* without its N-terminal signal sequence (amino acids 23–605) with a yield of 500 mg/l, purified to homogeneity, and used for further studies.

The substrate specificity and kinetic properties of the recombinant FragA α-galactosidase are presented in **Tables 1** and **2**. The enzyme has a broad pH optimum between 5 and 7.5 with the B tetrasaccharide-AMC substrate (**Supplementary Fig. 3b** online). Interestingly, the recombinant FragA enzyme exhibited low activity with the Galα-pNP substrate (0.3 U/mg) compared to the blood group B substrate (6.6 U/mg) (**Table 2**), whereas the purified endogenous enzyme from *S. griseoplanus* did not exhibit detectable activity with the Galα-pNP substrate (data not shown). The substrate specificity was remarkably stringent for α-1,3-linked galactose in the branched blood group B structure (**Table 1** and **Supplementary Fig. 4b** online). The enzyme cleaved neither α4Gal linkages found in P1 and P^k blood group antigens nor the α3Gal linkage in the linear B structure without fucose (the so-called Galili antigen). The FragA α-galactosidase has more than 300-fold higher activity with the blood group B tetrasaccharide-AMC substrate than the coffee bean enzyme, whereas the inverse (100-fold) is the case for the Galα-pNP substrate (**Tables 1** and **2**). One α-galactosidase with neutral pH optimum has been isolated previously from the marine bacterium *Pseudoalteromonas sp.* (KMM 701) with a specific activity of 9.8 U/mg with the Galα-pNP substrate[11], but given the different properties reported, this enzyme is probably unrelated to the genes described here. We are attempting to crystallize the α-galactosidase to further elucidate the catalytic mechanism of this enzyme family.

Enzymatic conversion of blood group A, B and AB RBCs at neutral pH

The two glycosidase families described here contain enzymes that have highly suitable kinetic properties for enzymatic removal of A and B antigens from the surface of RBCs. We first developed a buffer system for enzymatic conversion of RBCs with the recombinant *E. meningosepticum*. The recombinant enzyme has an isoelectric point of 7.6 as shown by isoelectrofocusing, and analysis of the electrostatic surface of the crystal structure indicated that the protein surface near the catalytic region is slightly positive. We therefore hypothesized that

the enzyme at neutral pH could interact with the strong negative charges of primarily sialic acids on the surface of RBCs leading to local concentration of the enzyme at substrate sites. Enzyme conversion was performed at low ionic strength in glycine buffer with varying concentrations of NaCl. Complete removal of A antigens was achieved in 200 mM glycine (pH 6.8) with 3 mM NaCl as evaluated by agglutination typing with routine licensed reagents and methods as well as by sensitive fluorescence-activated cell sorting (FACS) analysis (**Table 3** and **Fig. 3**). In this buffer system >90% of the enzyme was associated with the RBCs after centrifugation. In contrast, increasing the NaCl concentration to 10 mM or more resulted in release of the enzyme from the RBCs. In agreement with our hypothesis, increased NaCl also negatively affected the conversion efficiency, showing that the interaction is important for efficiency (data not shown). Importantly, this feature also allowed a simple and efficient wash-removal process of the enzyme involving three wash cycles in isotonic PBS. The residual enzyme load associated with washed cells was consistently less than 40 ng/ml of packed RBCs as measured with a sensitive capture enzyme-linked immunosorbent assay using total detergent lysates of cells (data not shown). The conversion process is linear with enzyme dose and time, and has a low enzyme consumption of 300 μg/ml packed RBCs for 60 min of incubation. RBCs from more than 200 group A_1 and A_2 donor units have been analyzed and all were converted efficiently. Approximately 60 mg enzyme is required to convert one unit (~200 ml packed RBCs) of blood group A_1 cells. Conversion of the weak blood group A_2 subgroup, however, requires only 15 mg enzyme. The conversion process does not affect physiologic and metabolic parameters of RBCs, including osmotic fragility, levels of 2,3-DPG, ATP and methemoglobin (data not shown).

We selected the recombinant FragA α-galactosidase because it has a slightly basic isoelectric point similar to the *E. meningosepticum* enzyme; using the same glycine buffer system as for group A RBCs, we effectively produced, by enzymatic conversion of blood group B RBCs with the FragA enzyme, B-ECO RBCs that type as group O by licensed reagents and methods used in clinical blood bank practice (**Table 3**). Similar to our findings with the A conversion process, the low ionic strength is required for efficient conversion, and increasing NaCl concentration in the glycine buffer or use of other buffer systems with higher ionic strength such as PBS and phosphate-citrate decreases the efficiency. The conversion process is linear with enzyme dose and time, and has a very low enzyme consumption of 10 μg/ml packed RBCs for 60 min incubation period. RBCs from more than 50 group B donors have been analyzed in repeat studies and all were converted efficiently with 10 μg/ml packed RBCs (data not shown). Thus, less than 2 mg enzyme are required to convert one unit of blood group B cells. Because the two enzymes used for conversion of A and B RBCs were selected for similar properties and function efficiently in the same conversion buffer system, we were able to convert blood group AB RBCs efficiently both by sequential enzyme digest as well as with a one-pot enzyme digest with the α-*N*-acetylgalactosaminidase and α-galactosidase (**Table 3**).

The efficiency of the conversion processes was evaluated by sensitive FACS analysis using anti-A and anti-B reagents (**Fig. 3**). FDA-licensed anti-A blood grouping reagents showed complete removal of A antigens on a homogenous population of RBCs. The high sensitivity of the assay was demonstrated by clear detection of the low amount of A antigen present on the weak A_x subgroup RBCs. We also included a highly sensitive anti-A monoclonal antibody, ES-15, that detects the minimal amount of A antigen, which is present naturally on group B RBCs[28]. ES-15 clearly detected both the A antigen on A_x subgroup RBCs and on group B RBCs. The ES-15 reactivity with B RBCs was sensitive to treatment with the α-*N*-acetylgalactosaminidase, confirming that the antibody does detect the minor amounts of A antigens on B RBCs. The ES-15 reagent, but not licensed anti-A reagents, did show a tendency toward a shift in the histogram curve (**Fig. 3a**, middle panel) of A-ECO RBCs, indicating that miniscule amounts of A antigens at a level considerably lower than on group B RBC could be detected on A_1-ECO but not on A_2-ECO RBCs. Similarly, FDA-licensed anti-B reagents showed a homogeneous B-ECO cell population almost nonreactive with anti-B. The high sensitivity of this assay was demonstrated by including AB_{weak} subgroup cells, on which the weakly expressed B antigens were readily detected. Several licensed anti-B typing reagents were tested (data not shown), and one CE-approved anti-B blood grouping reagent, Diagast, was found to exhibit higher sensitivity with the AB_{weak} cells and a minor proportion of B-ECO cells were weakly labeled by the Diagast anti-B antibody, although at a level lower than the weak AB_{weak} subgroup cells. Analysis with an anti-H antibody was included to demonstrate appearance of the appropriate H antigen product after removal of either αGalNAc or αGal terminal residues, and both A-ECO and B-ECO RBCs expressed H similar to group O RBCs as expected (**Fig. 3**, lower panels).

Further biochemical analysis of ECO cells included TLC of glycolipids as well as SDS-PAGE western blot of total membrane (glyco)-protein extracts. TLC of glycolipid extracts revealed essentially complete conversion of blood group A and B glycolipid species to the corresponding H species, including the major monosialyl A and B gangliosides designated G9A and G9B (**Fig. 4**). This was further confirmed by mass spectrometric analysis of the glycolipid fractions migrating in the region of the most abundant A-active glycolipid species, designated A^a and A^b, which showed a complete absence of the A glycolipid species in extracts from converted RBCs and the presence of only the corresponding H_1 and H_2 products (data not shown). The blood group A antigens are believed to be more complex than B antigens in that a fraction of the A active glycolipids have a repeated A structure[13]. This structure has been found only on glycolipids and not on glycoproteins. Cleavage of the terminal GalNAc residue of these glycolipids will result in exposure of an H structure extended from an internal A structure in H-A^a and H-A^b glycolipids (**Fig. 4**). Both subgroups A_1 and A_2 RBCs contain these glycolipids, and after enzyme conversion the A_1 subgroup RBCs will contain more than the A_2 subgroup. This H antigen resembles the mucin-type H structure (Fucα1-2Galβ1-3GalNAcα1-O-Ser/Thr) expressed in epithelial tissues of all ABO groups[29] and is not reactive with typing anti-A antibodies. It is therefore our hypothesis that these A-associated H-A structures found only on glycolipids do not represent antigens recognized by human anti-A immunity, but clinical studies are required to address this. The analysis of glycolipids from converted cells also showed that the *E. meningosepticum* α-*N*-acetylgalactosaminidase does not cleave the major glycolipid globoside (Gb_4, blood group P antigen terminating by a GalNAcβ1-3 linkage) in RBC membranes (**Fig. 4**), despite the finding that the enzyme has low but detectable activity with GalNAcβ-pNP (**Table 2**). The glycolipid analysis clearly demonstrated the effectiveness of the process as well as how selective the process is in digesting only the blood group A and B terminating glycolipids. Western blot with the sensitive ES-15 anti-A and the Diagast anti-B monoclonal antibodies revealed very faint diffuse staining of membrane glycoproteins of A-ECO and B-ECO RBCs, respectively (data not shown), which is in agreement with the findings that FACS analysis with the most sensitive monoclonal antibody reagents reveal minute amounts of residual A and B antigens (**Fig. 3**).

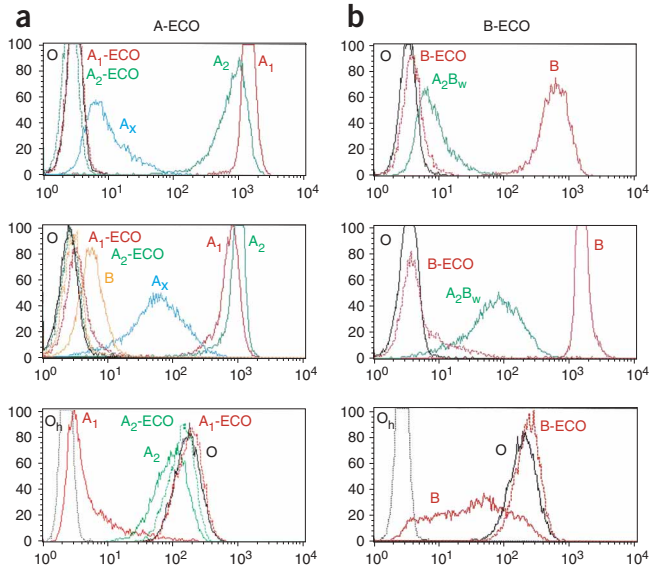

Figure 3 FACS of native and ECO RBCs. Histograms show the degree of staining of RBCs with different blood grouping reagents. The *x*-axis represents the fluorescence intensity on a logarithmic scale whereas the *y*-axis shows the number of RBCs evaluated. Solid lines represent untreated native RBCs and dashed lines in the same color are the enzyme-treated RBCs from the same individual (see color code below). The dotted black line in the H-antigen staining histogram (lower panels) represents the negative control, RBCs from a Bombay (O_h) donor lacking any A, B or H antigen on RBCs. To facilitate interpretation of the histograms, enzyme-treated group O and subgroup RBC were not included. However, their curves completely overlap the result for native group O (data not shown). (**a**) RBCs were stained with licensed monoclonal anti-A blood grouping reagent (upper panel), the monoclonal ES-15 anti-A research reagent (middle panel) and direct-conjugated monoclonal anti-H (lower panel). Red, blood group A_1; green, A_2; blue, A_x; orange, B; black, O. The weak A_x subgroup control was from a serologically and genetically characterized test RBC donor with the A^x-1 O^1 genotype. Group B RBCs were tested only with ES-15 as licensed anti-A reagents do not detect any A antigen on their surface. (**b**) RBCs were stained with the monoclonal licensed anti-B blood grouping reagents from Ortho (US FDA-licensed) (upper panel), Diagast (CE-approved in European Union) (middle panel), and direct-conjugated monoclonal anti-H (lower panel). Red, blood group B; green, subgroup A_2B_w; black, O.

DISCUSSION

The present study identified two prokaryotic glycosidase gene families, not previously described, containing α-*N*-acetylgalactosaminidase (GH109) and α-galactosidase (GH110) activities with neutral pH optima and high activity with blood group A and B substrates, respectively. The GH109 α-*N*-acetylgalactosaminidase family represents to our knowledge the only known source of enzymes suitable for efficient enzymatic conversion of blood group A RBCs. The GH110 α-galactosidase family has unprecedented restricted specificity for blood group B substrates. The enzymes are expressed with high yields in *E. coli* and because they have similar properties, a single common conversion buffer system and process can be used to remove A and B antigens and produce ECO RBCs from A, B and AB RBCs that type as blood group O with routine licensed typing reagents and methods. Extensive FACS and biochemical analyses confirm the efficient removal of the immunodominant A and B antigens and exposure of the underlying H antigens. The current process, which is performed manually at neutral pH, is scalable to automated full-unit conversions, and ECO cells produced by this method are predicted to survive and function in a manner equivalent to native group O RBCs in non-ABO matched individuals as reported previously for B-ECO RBCs[9]. The process has a projected consumption of ~60 mg (A-ECO) and 2 mg (B-ECO) recombinant enzyme with 60-min enzyme treatment per unit RBCs. This is ~30- (A-ECO) and 1,000-fold less (B-ECO) enzyme than the conversion protocol developed for group B RBCs with the Coffee bean α-galactosidase[9]. Accordingly, we believe that automated cost-effective processes can be developed for practical use in transfusion medicine.

Glycosidases are grouped in over 100 sequence-based families that have been shown to correlate with the 3-D structure and molecular mechanisms of catalysis[30,31]. The glycosidase families identified in this study exhibit no substantial sequence similarity to any other glycosidase families, including families GH4, GH27, GH36 and GH57 (CAZY database at http://afmb.cnrs-mrs.fr/CAZY/), which contain all previously reported α-*N*-acetylgalactosaminidases and α-galactosidases, including the two most studied enzymes for blood group B red-cell conversion derived from green coffee and from soy beans[32,33], as well as enzymes used previously to convert weak blood group A_2 RBCs[14,34].

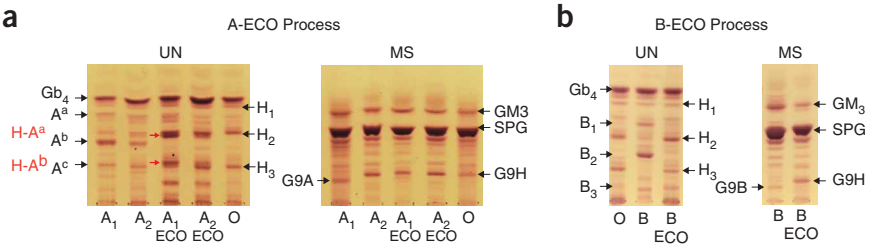

Figure 4 TLC analysis of native and ECO RBC glycolipids. (**a**,**b**) Total upper neutral (UN) and monosialyl (MS) glycolipid fractions from RBC's were separated on high-performance thin layer chromatography plates. Arrows indicate mobility of major blood group ABH glycolipids as well as Gb_4 (globoside) and SPG (sialyl-paragloboside, NeuAcα2-6Galβ1-4GlcNAcα1-3Galβ1-4Glcβ1-Cer). The nomenclature of ABH active glycolipids used for type 2 chain ABH active glycolipids is $H_1/B_1/A^a$, $H_2/B_2/A^b$ and $H_3/B_3/A^c$ with increasing number of Galβ1-4GlcNAc disaccharide units[3]. Although, glycolipid bands separated by TLC may contain multiple species, the H_1 (Fucα1-2Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ1-Ceramide), H_2 (as H_1 with one lactosamine unit Galβ1-4GlcNAcβ1-3 inserted) and H_3 (as H_1 with two lactosamine units Galβ1-4GlcNAcβ1-3 inserted) bands in O UN clearly represent major distinct bands with different mobilities. The corresponding A (**a**) and B (**b**) structures A^a/B_1, A^b/B_2 and A^c/B_3 GalNAc/Galα1-3 added to the penultimate βGal residues in H structures in A and B UN are similarly distinct bands. Comparison with the profile of ECO reveals disappearance of A and B variants and appearance of the corresponding H bands. This includes the branched G9A/B glycolipid (GalNAc/Galα1-3[Fucα1-2]Galβ1-4GlcNAcβ1-3[NeuAcα2-3Galβ1-4GlcNAcβ1-6]Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ1-Ceramide), which is a branched structure with one branch terminated by sialic acid and one branch with an ABH determinant. Group A active glycolipids are more complex than B in that they include the internal A structures A-associated H (designated H-A^a and H-A^b), which migrate just below the H_2 and H_3 glycolipids, respectively. The A^b glycolipid band contains a mixture of type 2 chain A^b and the repetitive A-A^a glycolipid (GalNAcα1-3(Fucα1-2)Galβ1-3GalNAcα1-3(Fucα1-2)Galβ1-4GlcNAcβ1-4Galβ1-4Glcβ1-Ceramide), whereas the A^c band contains type 2 chain A^c and the repetitive A-A^a glycolipid[3]. The loading corresponds to glycolipid fractions derived from 4 ml of packed RBCs.

The human lysosomal α-galactosidase and α-*N*-acetylgalactosaminidase are close homologs to each other and are both members of family GH27. All of these enzymes are characterized by having exclusively a classical catalytic machinery and molecular mechanism leading to retention of the anomeric configuration[35,36], with the exception of family GH4 enzymes that use NAD^+ and Mn^{2+} as cofactors[19–21]. The α-*N*-acetylgalactosaminidase family described here is unusual in that it uses an NAD^+-based mechanism, similarly to family GH4 enzymes. Interestingly, the proposed mechanism of the α-*N*-acetylgalactosaminidase is not prone to transglycosylation, a known side reaction of retaining glycosidases, which can fortuitously give rise to nonnatural structures at high substrate concentration.

Preferred properties of an exoglycosidase suitable for enzymatic conversion of RBCs include the following parameters: (i) high substrate specificity for the blood group antigens to restrict the reaction to the immunodominant blood group A and B antigens; (ii) reaction conditions suitable for maintenance of RBC integrity and functions; (iii) high efficiency in cleavage of antigens on the RBC surface to minimize residual antigens and enzyme consumption; and (iv) properties to facilitate enzyme removal from the RBCs by routine cell-washing techniques. The glycosidases presented in this study offer all of these characteristics. In particular, the α-galactosidases represent, to our knowledge, the only α3-linkage–specific exo-galactosidases reported and their specificity for the branched blood group B trisaccharide structure suggests restricted substrate specificity. This has practical implications as human RBCs express other αGal-terminating structures, including α4Gal residues in the P1 and P^k blood group antigens, which are cleaved by the α-galactosidases used in previous studies[32,33]. The two enzyme families described here perform efficiently in conversion of RBCs; ECO cells type as group O with all licensed reagents; and sensitive FACS and glycolipid analyses confirm efficient removal of A and B antigens (**Figs. 3** and **4**). Finally, enzymes from both families are slightly basic and associate with the negatively charged RBCs through ionic interactions, thereby enabling efficient removal with isotonic buffer solutions, such as PBS, used for cell washing.

The availability of enzymes from these glycosidase families has resulted in the development of a simple and efficient process for producing universal RBCs that type as blood group O. Clinical translation of this approach may allow improvement of the blood supply and enhancement of patient safety in transfusion medicine.

METHODS

Enzyme assays. Structures of substrates are shown in **Tables 1** and **2.** The AMC derivatives were custom synthesized by Alberta Chemical Research Council. The 8-methoxycarbonyloctyl ($(CH_2)_8COOCH_3$, MCO) oligosaccharides were obtained from the same source, and other substrates were obtained from Sigma-Aldrich, V-LABS and Oxford Glyco Sciences. Assays with oligosaccharide structures were performed at 26 °C in reaction mixtures of 10 μl. Product formation was analyzed by TLC using silica gel–coated TLC plates (Merck). The AMC-labeled tetrasaccharides were used for screening and one unit of enzyme activity was defined as the amount of enzyme required to cleave 1 μmole of A or B tetrasaccharide-AMC substrate per minute using a 10-μl reaction volume with 100 mM $NaPO_4$, pH 6.8, 50 mM NaCl, 0.25 mg/ml of BSA and 0.1 mM substrate. Importantly, calculation of the specific activities with these substrates was based on complete cleavage of the substrate in the reaction volume. Assays with chromogenic pNP substrates were carried out at 26 °C in reaction mixtures of 400 μl with 100 mM $NaPO_4$, pH 6.8, 50 mM NaCl and 2.5 mM substrate. Reactions were terminated by addition of 600 μl 1 M Na_2CO_3 and pNP formation was quantified at 405 nm ($\varepsilon = 18{,}300\ cm^{-1}M^{-1}$ at pH > 10). In the chromogenic assays all results were based on reactions with consumption of <10% of the substrate. Determination of kinetic parameters was performed using substrate concentrations of 0.5 to 3–10 times the obtained K_m. Reactions with the α-*N*-acetylgalactosiminidase and Galα-pNP, however, were carried out using 1–12 mM substrate, which is substantially lower that the apparent K_m.

Crystallization and data collection. Before crystallization the purification buffer was extensively exchanged with 10 mM HEPES buffer at pH 7.5. Recombinant α-*N*-acetylgalactosaminidase was crystallized using the hanging drop vapor diffusion method by mixing equal amounts of protein at 21 mg/ml with a solution consisting of 46–50% 2-methyl-2,4-pentanediol (MPD) and 75 mM sodium citrate at pH 6.25. Rod-shaped crystals appeared within 1 d and belong to space group $P6_522$ with cell dimensions of 88 × 88 × 300 Å and one molecule per asymmetric unit. A MAD data set for a selenomethionylated protein crystal, as well as a native data set was collected at beam line ID23-EH1 (ESRF, Grenoble) and a data set for a complex with GalNAc was collected at beam line ID14-EH2 (ESRF, Grenoble). Before data collection, crystals were flash-frozen in a nitrogen gas stream at 100 °K. Data were indexed and integrated with MOSFLM[37] and scaled with the program SCALA[38]. Data collection statistics are summarized in **Supplementary Table 1** online.

Structure solution and refinement. Data were prepared for phasing with XPREP (Bruker). The selenium substructure was solved with SHELXD[39], taking advantage of data up to the highest resolution shell, and phasing combined with solvent flattening was carried out with SHELXE[40], leading to a pseudo-free correlation coefficient of 70.5%. Starting from experimental phases, an initial model comprising 361 residues (out of 429), 257 of which docked into sequence, was automatically built with the ARP/wARP package[41] and completed manually using TURBO-FRODO[42]. The resulting model was refined against native data extending to 2.3-Å resolution with REFMAC[43]. A random set of 4% (1,231) reflections was set aside for cross validation purposes. Automated solvent building was performed with ARP/wARP[41]. For the structure solution of the enzyme-GalNAc complex, the final native model was subjected to rigid body refinement with REFMAC[43] and further refinement was carried out as described above, with 4.1% (1,111) of reflections, taken over from the native data set, set apart for cross validation. The final protein structures encompass residues Lys19 to Tyr444 and have correct stereochemical properties, with all residues contained in the allowed regions of the Ramachandran plot, as verified with WHATCHECK[44]. Refinement and structure quality statistics are listed in **Supplementary Table 1** online. **Figure 1** was drawn with ChemDraw, CambridgeSoft Corporation, and **Figure 2** was generated with PYMOL (DeLano, W.L. The PyMOL Molecular Graphics System (2002) on http://www.pymol.org/).

Enzymatic conversion of RBCs with glycosidases at neutral pH. Standard enzymatic conversion reactions were performed in 1 ml reaction mixtures containing 200 mM glycine, pH 6.8 and 3 mM NaCl with 30% packed red blood cells and enzyme as indicated. Fresh whole blood was obtained from Oklahoma Blood Institute and buffy coat removed. RBCs were prewashed 1:1 and 1:4 vol/vol in conversion buffer before addition of enzyme. The conversion reaction was incubated for 60 min with gentle mixing at 26 °C, followed by four repeat washing cycles with 1:4 vol/vol of saline by centrifugation at 1,000 r.p.m. The washed enzyme-treated ECO RBCs were ABO-typed according to standard blood banking techniques using licensed monoclonal antibody reagents[45]. Murine monoclonal Anti-A reagent obtained from Ortho. Undiluted, concentrated murine monoclonal anti-A typing reagent supplied for further manufacturing purposes by Celliance. Murine monoclonal Anti-B reagents obtained from Ortho and Diagast.

Flow cytometry. Flow cytometry analysis of native and ECO RBCs was performed using a FACScan flow cytometer (Becton Dickinson) with monoclonal antibody reagents and phycoerythrin-labeled rat-anti-mouse kappa immunoglobulin (Becton Dickinson). An FITC-conjugated monoclonal anti-H (BRIC 231 from BITS/IBGRL) was also used. Briefly, 500,000 RBCs were suspended in PBS in 96-well plates in a 50-μl reaction volume. Cells were fixed for 10 min in 0.1% glutaraldehyde in PBS to prevent agglutination of antigen-positive cells. After washing 3× in 150 μl PBS, cells were again suspended in 50 μl PBS, after which 5 μl of undiluted primary antibody was added and incubated for 10 min. After three washes and resuspension in 50 μl PBS, 5 μl of undiluted secondary antibody was added. Cells were then analyzed

after another three washes (as above) and resuspension in 300 µl PBS. All incubations were carried out in darkness at 20 °C under gentle agitation. A total of 10,000 events were evaluated and an RBC gate corresponding to ~90% of all cells was set during analysis of the data.

Glycolipid analysis. Glycosphingolipids were prepared essentially as previously described[46]. Briefly, RBC membranes were extracted by homogenization in ten volumes of isopropanol/hexane/water (55:25:20, vol/vol/vol), filtered, evaporated and partitioned by Folch extraction method. The upper phase was collected, evaporated, dialyzed and applied to a DEAE Sephadex column. The pass-through fraction contained the total upper neutral lipids, and the monosialyl fraction was eluted with 0.05 M ammonium acetate, evaporated and dialyzed against water to remove acetate. Total upper neutral (UN) and monosialyl (MS) glycolipid fractions were analyzed by high-performance thin-layer chromatography. TLC plates (Merck) were developed using the solvent system chloroform/methanol/0.5 M $CaCl_2$ in water (50:40:10, vol/vol/vol). Glycolipids were stained by heating with 0.05% orcinol in 0.5 M H_2SO_4.

Accession codes. EMBL Nucleotide Sequence Database: the nucleotide sequences of the cloned genes reported in this paper were deposited with accession nos. AM039444, AM039445, AM039446 AM039447, AM039448 and AM039449 for the α-*N*-acetylgalactosaminidases and homologs; AM109953, AM109954, M109955, AM109956 and AM109957 for the α-galactosidases. Protein Data Bank: coordinates of the crystal structures have been deposited with accession reference nos. 2IXA and 2IXB.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Margot Kruskall. We thank Phil Robbins for his invaluable help throughout this work. We are grateful to Annika Hult at the Blood Center in Lund for technical assistance with FACS analysis and to Etienne Danchin, AFMB, for preparing **Supplementary Fig. 1** online. This work was supported by ZymeQuest Inc. and the Centre National de la Recherche Scientifique. Work performed in M.L.O.'s laboratory was supported by the Swedish Research Council (project no. K2005-71X-14251), governmental ALF research grants to Lund University Hospital, the Inga and John Hain Foundation for Medical Research and Region Skåne, Sweden. The European Synchrotron Radiation Facility (ESRF) is acknowledged for beam time allocation.

AUTHOR CONTRIBUTIONS

Q.P.L. contributed to screening and purification of enzymes, planning and design of the project, and manuscript writing; G.S. contributed to the X-ray crystallography and manuscript writing; H.Y. contributed to purification of enzymes; E.P.B. and K.S. contributed to cloning of genes; G.P. contributed to cloning of genes; J.S. and K.S. contributed to development of enzyme conversion protocol; E.N. contributed to glycolipid analysis; S.B.L. contributed with NMR analysis; T.W. contributed to planning and design of the project; J.M.N. and W.S.L. contributed to sequencing of purified enzyme; Y.B. contributed to manuscript writing; M.L.O. contributed to FACS analysis, planning and design of the project and manuscript writing; B.H. contributed to planning and manuscript writing; H.C. contributed to planning and design of the project, and manuscript writing.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturebiotechnology.

Published online at http://www.nature.com/naturebiotechnology
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Landsteiner, K. Agglutination phenomena of normal human blood. *Wien. Klin. Wochenschr.* **113**, 768–769 (2001).
2. Watkins, W.M. Biochemistry and Genetics of the ABO, Lewis, and P blood group systems. *Adv. Hum. Genet.* **10**, 1–136, 379–185 (1980).
3. 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).
4. Yamamoto, F., Clausen, H., White, T., Marken, J. & Hakomori, S. Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**, 229–233 (1990).
5. Sazama, K. Transfusion errors: scope of the problem, consequences, and solutions. *Curr. Hematol. Rep.* **2**, 518–521 (2003).
6. Stainsby, D. *et al. Serious Hazards of Transfusion Annual Report 2004* (Serious Hazards of Transfusion Office, Manchester Blood Centre, Manchester, UK; 2005).
7. Olsson, M.L. *et al.* Universal red blood cells–enzymatic conversion of blood group A and B antigens. *Transfus. Clin. Biol.* **11**, 33–39 (2004).
8. 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).
9. 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).
10. Vosnidou, N.C. *et al.* Seroconversion of type B to O erythrocytes using recombinant Glycine max α-D-galactosidase. *Biochem. Mol. Biol. Int.* **46**, 175–186 (1998).
11. Bakunina, I.Y. *et al.* Alpha-galactosidase of the marine bacterium *Pseudoalteromonas* sp. KMM 701. *Biochemistry (Mosc.)* **63**, 1209–1215 (1998).
12. Clausen, H., Levery, S.B., Kannagi, R. & Hakomori, S. Novel blood group H glycolipid antigens exclusively expressed in blood group A and AB erythrocytes (type 3 chain H). I. Isolation and chemical characterization. *J. Biol. Chem.* **261**, 1380–1387 (1986).
13. Clausen, H., Levery, S.B., Nudelman, E., Tsuchiya, S. & Hakomori, S. Repetitive A epitope (type 3 chain A) defined by blood group A_1-specific monoclonal antibody TH-1: chemical basis of qualitative A_1 and A_2 distinction. *Proc. Natl. Acad. Sci. USA* **82**, 1199–1203 (1985).
14. Zhu, A., Monahan, C., Wang, Z.K. & Goldstein, J. Expression, purification, and characterization of recombinant α-*N*-acetylgalactosaminidase produced in the yeast *Pichia pastoris*. *Protein Expr. Purif.* **8**, 456–462 (1996).
15. Hsin-Yeh, H., Chapman, L.F., Calcutt, M.J. & Smith, D.S. Recombinant *Clostridium perfringens* α-*N*-acetylgalactosaminidase blood group A_2 degrading activity. *Artif. Cells Blood Substit. Immobil. Biotechnol.* **33**, 187–199 (2005).
16. Bakunina, I.Y. *et al.* α-N-acetylgalactosaminidase from marine bacterium *Arenibacter latericius* KMM 426T removing blood type specificity of A-erythrocytes. *Biochemistry (Mosc.)* **67**, 689–695 (2002).
17. Landry, D. Isolation and composition of a novel glycosidase from chryseobacterium. US patent 6,458,525. (2002).
18. Pikis, A., Immel, S., Robrish, S.A. & Thompson, J. Metabolism of sucrose and its five isomers by *Fusobacterium mortiferum*. *Microbiology* **148**, 843–852 (2002).
19. Yip, V.L. *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).
20. Rajan, S.S. *et al.* Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD^+/Mn^{2+}-dependent phospho-alpha-glucosidase from *Bacillus subtilis*. *Structure* **12**, 1619–1629 (2004).
21. Yip, V.L. & Withers, S.G. Mechanistic analysis of the unusual redox-elimination sequence employed by *Thermotoga maritima* bglt: a 6-phospho-beta-glucosidase from glycoside hydrolase family 4. *Biochemistry* **45**, 571–580 (2006).
22. Lesk, A.M. NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* **5**, 775–783 (1995).
23. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
24. Kingston, R.L., Scopes, R.K. & Baker, E.N. The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure* **4**, 1413–1428 (1996).
25. Varrot, A. *et al.* NAD^+ and metal-ion dependent hydrolysis by family 4 glycosidases: structural insight into specificity for phospho-beta-D-glucosides. *J. Mol. Biol.* **346**, 423–435 (2005).
26. Yip, V.L. & Withers, S.G. Family 4 glycosidases carry out efficient hydrolysis of thioglycosides by an alpha,beta-elimination mechanism. *Angew. Chem. Int. Edn. Engl.* **45**, 6179–6182 (2006).
27. Ishikura, H., Arakawa, S., Nakajima, T., Tsuchida, N. & Ishikawa, I. Cloning of the *Tannerella forsythensis* (*Bacteroides forsythus*) siaHI gene and purification of the sialidase enzyme. *J. Med. Microbiol.* **52**, 1101–1107 (2003).
28. Goldstein, J., Lenny, L., Davies, D. & Voak, D. Further evidence for the presence of A antigen on group B erythrocytes through the use of specific exoglycosidases. *Vox Sang.* **57**, 142–146 (1989).
29. Clausen, H., Holmes, E. & Hakomori, S. Novel blood group H glycolipid antigens exclusively expressed in blood group A and AB erythrocytes (type 3 chain H). II. Differential conversion of different H substrates by A_1 and A_2 enzymes, and type 3 chain H expression in relation to secretor status. *J. Biol. Chem.* **261**, 1388–1392 (1986).
30. Davies, G.J., Gloster, T.M. & Henrissat, B. Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr. Opin. Struct. Biol.* **15**, 637–645 (2005).
31. Henrissat, B. & Davies, G.J. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol.* **124**, 1515–1519 (2000).
32. Zhu, A. *et al.* Characterization of recombinant alpha-galactosidase for use in seroconversion from blood group B to O of human erythrocytes. *Arch. Biochem. Biophys.* **327**, 324–329 (1996).
33. Davis, M.O., Hata, D.J., Johnson, S.A., Walker, J.C. & Smith, D.S. Cloning, expression and characterization of a blood group B active recombinant α-D-galactosidase from soybean (Glycine max). *Biochem. Mol. Biol. Int.* **39**, 471–485 (1996).
34. Calcutt, M.J., Hsieh, H.Y., Chapman, L.F. & Smith, D.S. Identification, molecular cloning and expression of an α-N-acetylgalactosaminidase gene from *Clostridium perfringens*. *FEMS Microbiol. Lett.* **214**, 77–80 (2002).
35. Henrissat, B. Glycosidase families. *Biochem. Soc. Trans.* **26**, 153–156 (1998).
36. Rye, C.S. & Withers, S.G. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573–580 (2000).

37. Leslie, A.G.W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26** (1992).
38. CCP4. The CCP4 Suite. Programs for protein crystallography. *Acta Crystallogr D Biol. Crystallogr.* **50**, 760–763 (1994).
39. Schneider, T.R. & Sheldrick, G.M. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002).
40. Sheldrick, G.M. Macromolecular phasing with SHELXE. *Z. Kristallogr.* **217**, 644–650 (2002).
41. Perrakis, A., Morris, R. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463 (1999).
42. Roussel, A. & Cambillau, C. TURBO-FRODO. in *Silicon Graphics Geometry Partners Directory* (eds. Silicon Graphics Committee) 77–78 (Silicon Graphics, Mountain View, California, 1989).
43. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240–255 (1997).
44. Hooft, R.W., Vriend, G., Sander, C. & Abola, E.E. Errors in protein structures. *Nature* **381**, 272 (1996).
45. Technical Manual (AABB, Bethesda, Maryland, USA; 2005).
46. Clausen, H., Levery, S.B., McKibbin, J.M. & Hakomori, S. Blood group A determinants with mono- and difucosyl type 1 chain in human erythrocyte membranes. *Biochemistry* **24**, 3578–3586 (1985).