The catalytic acid-base in GH109 resides in a conserved GGHGG loop and allows for comparable α-retaining and β-inverting activity in an *N*-acetylgalactosaminidase from *Akkermansia muciniphila*

David Teze,[†,§] Bashar Shuoker,[†,⊽,§] Evan Kirk Chaberski[‡], Sonja Kunstmann,[†] Folmer Fredslund,[‡] Günther H.J. Peters,[‡] Eva Nordberg Karlsson,[⊽] Ditte Hededam Welner,[*,‡] Maher Abou Hachem[*,†]

[†]Department of Biotechnology and Biomedicine, Technical University of Denmark, DK-2800 Lyngby, Denmark
[‡]Enzyme Engineering and Structural Biology, The Novo Nordisk Center for Biosustainability, Kemitorvet, building 220, DK-2800 Lyngby, Denmark
[⊽]Biotechnology, Department of Chemistry (KILU), Lund University, Post Office Box 124, 221 00 Lund, Sweden
[‡]Department of Chemistry, Technical University of Denmark, DK-2800 Lyngby, Denmark

[§]Authors contributed equally to the study.

ABSTRACT: Enzymes active on glycosidic bonds are defined according to the stereochemistry of both substrates and products of the reactions they catalyse. The CAZy classification further assigns these enzymes into sequence-based families sharing a common stereochemistry for substrates (either α- or β-) and products, *i.e.* inverting or retaining mechanism. Enzymes from glycoside hydrolase family GH109 are thus described as α-retaining *N*-acetylgalactosaminidases. Here we describe *Am*GH109A from the human gut symbiont *Akkermansia muciniphila*, which displays both α-retaining and β-inverting activities with comparable efficiencies on natural disaccharides, a dual specificity that could provide an advantage in targeting a broader range of host-derived glycans. We rationalise this discovery through bioinformatics, structural, mutational, and computational studies, unveiling a histidine residing in a conserved GGHGG motif as the elusive catalytic acid-base of the GH109 family.

KEYWORDS: glycoside hydrolase, human gut microbiota, inverting, mechanism, MD simulations, mucin, retaining, structure.

INTRODUCTION

The human gut microbiota (HGM) exerts a profound impact on human health and plays a key role in the metabolic and immune homeostasis of the host.[1-2] Specific signatures of this complex microbial community are associated with a variety of disorders including colorectal cancer[3] and inflammatory bowel diseases.[4-5] Importantly, the HGM is also associated with insulin resistance[6] and obesity,[7] both of which are growing lifestyle diseases.

The implication of the abundant (1−4%) human gut symbiont *Akkermansia muciniphila* in protection from obesity and impact on other aspects of human health attracts increasing attention. Indeed, a strong inverse correlation between obesity and *A. muciniphila* abundance is observed in humans[8-9] and it has been shown that a single *A. muciniphila* outer membrane protein mediates positive effect on the metabolism of obese mice.[10]*A. muciniphila*, which is the sole representative of the phylum Verrucomicrobiota in the HGM, is a specialist degrader of mucin.[11] Mucin is a collective name for a family of high molecular mass heavily-decorated (about 80 % w/w) *O*-glycoproteins that coat the surfaces of enterocytes. Mucin is an important physical barrier and a site of adhesion for distinct bacteria including *A. muciniphila*, which adheres strongly to human epithelial colonic cell lines, strengthens enterocyte monolayer integrity *in vitro*[12] and restores the thickness of the mucin layer in obese mice.[13] *A. muciniphila* also induces adaptive immune response,[14] consistent with the intimate association and cross-talk between this symbiont and the human host. The genome of *A. muciniphila*[15] encodes a substantial battery of Carbohydrate Active enZymes (CAZymes, http://www.cazy.org)[16] that mostly targets host-derived glycoconjugates, thereby supporting the ecological specialization of this bacterium.[17] α-Glycosidic linkages in such glycans occur frequently at the non-reducing ends or as side chains (*e.g.* fucosyl or sialyl units), whereas β-linkages are commonly present within the glycan chains[17-18].

Enzymes within a glycoside hydrolase (GH) family, defined based on sequence similarity, share a common structural fold, substrate stereoselectivity and a catalytic mechanism that governs the stereochemistry of the products.[16] Thus, a GH family is either inverting or retaining and usually either α- or β-active. Pseudo-exceptions are enzymes acting on structurally similar substrates that differ in both their D/L and α/β configurations (*e.g.* substrates with non-reducing α-L-arabinopyranosyl and β-D-galactopyranosyl are hydrolysed by GH42 enzymes,[19] likewise β-L-arabinopyranosidase and α-D-galactopyranosidase activities are found in some GH27 members).

True exceptions to the GH classification are found in GH4, GH97 and GH109. Thus, GH97 encompasses either α-retaining or α-inverting enzymes, both obeying the typical GH general acid/base mechanism, but using different catalytic residues.[20] By contrast, GH4 employs a non-canonical NAD^+- and Mn^{2+}-dependent mechanism and harbours members that are either α-retaining[21] or β-retaining.[22-23] Enzymes of GH109 also display a redox mechanism assisted by a NAD^+ cofactor, akin to GH4, but lacking the Mn^{2+} cofactor, which in GH4 is required for the formation of an alkoxide at position 3.[24-25] The mechanism of GH109 has not been fully elucidated, in particular no catalytic acid/base has been proposed to activate a nucleophilic water molecule and protonate the glycosidic bond oxygen, which otherwise would lead to the energetically unfavourable departure of an alkoxide group.[24] Currently, GH109 enzymes are described as solely α-*N*-acetylgalactosaminidases, discovered in the quest for enzymatic conversion of the blood

group A antigen to the universal O-type by releasing terminal *N*-acetyl galactosaminide (GalNAc) units.[25-26] Only a single enzyme from this family is kinetically and structurally characterized, namely the enzyme from *Elizabethkingia meningosepticum* (NagA).[24-25, 27] The GH109 classification was based on NagA that is mainly active on 4-nitrophenyl 2-acetamido-2-deoxy-α-D-galactopyranoside (α-*p*NPGalNAc), with about 1500-fold lower activity on the β-linked anomer (4-nitrophenyl 2-acetamido-2-deoxy-β-D-galactopyranoside, β-*p*NPGalNAc).[25] The dominance of a single stereoselectivity is the hallmark of the present GH paradigm, as no single GH has been reported to catalyse the hydrolysis of both α- and β- at comparable and relevant activity levels.

Here, we describe two members of the GH109 family from *A. muciniphila* that notably have β-inverting activity alongside the expected α-retaining mechanism ascribed to the family. The kinetic signatures of both enzymes were markedly different from the previously described NagA. Strikingly, the first enzyme, *Am*GH109A, showed higher specificity (k_{cat}/K_M) towards β-*p*NPGalNAc (7.3 $s^{-1}\cdot mM^{-1}$) than for α-*p*NPGalNAc (2.4 $s^{-1}\cdot mM^{-1}$). Uniquely, *Am*GH109A displayed similar, relevant activities towards non-activated α- and β-GalNAc(1→3)Gal disaccharides. The second enzyme *Am*GH109B displayed about 18.4-fold lower efficiency on β-*p*NPGalNAc as compared to the α- counterpart, but similar K_m values for both substrates. Kinetic, bioinformatics, structural, mutational and computational investigations allowed us to establish that a histidine (*Am*GH109A H404) in a conserved and flexible GGHGG motif acts as the catalytic acid-base in GH109 for both α-retaining and β-inverting activities.

RESULTS AND DISCUSSION

The *N*-acetyl galactosaminidase *Am*GH109A displays unprecedented β-inverting and α-retaining activities at similar levels. The kinetic parameters of GH109 have only been reported for a single enzyme (NagA) against the activated substrate analogues α-*p*NPGalNAc and β-*p*NPGalNAc, assigning GH109 as an α-specific family based on the about 1500-fold lower β-activity.[25] Kinetic analysis was carried out on the *A. muciniphila* enzymes *Am*GH109A and *Am*GH109B towards aryl substrate analogues, which demonstrated that both enzymes display unusually high β-activities (Table 1, Figure 1a). The catalytic efficiency of *Am*GH109B was only about 18-fold lower on the β-aryl substrate, mainly due to lower k_{cat}. Strikingly, the efficiency of *Am*GH109A was about 3-fold higher on the β-analogue, owing to ≈7-fold lower K_m and only 40% lower k_{cat}. The observed α/β-activities were reproducible, excluding contamination.

Table 1. Catalytic parameters of GH109 *N*-acetyl galactosaminidases towards *p*NPGalNAc.

Enzyme	Substrate stereochemistry	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1}\cdot mM^{-1}$)
*Am*GH109A	α	2.6 ± 0.2	1.1 ± 0.1	2.4
	β	1.1 ± 0.03	0.15 ± 0.01	7.3
*Am*GH109B	α	16.5 ± 0.3	0.39 ± 0.02	42.3
	β	0.9 ± 0.01	0.38 ± 0.02	2.3
NagA[a]	α	9.84 ± 0.16	0.077 ± 0.006	127.6
	β	0.015 ± 0.000	0.23 ± 0.01	0.087

[a]Values from Liu *et al*[25]

This dual activity is more conceivable in GH109 than in classical (non NAD^+-dependent) GHs, as the leaving group departure precedes the nucleophilic attack on the anomeric C1 (concomitant in classical GHs). The pK_a of *para*-nitrophenyl (7.24) makes departure as a nitrophenolate plausible without acid catalysis. Accordingly, the reaction towards β-*p*NPGalNAc could proceed without a catalytic proton donor on the GalNAc β-face, while the presence of the base catalyst on the GalNAc α-face would still be needed to activate the nucleophilic water molecule. This reaction scheme would lead to an inversion of stereochemistry and the release of α-GalNAc from β-*p*NPGalNAc, which is indeed confirmed by our NMR analysis (Supporting Information Figure S1).

Non-activated substrates are also likely to require a catalytic acid on the GalNAc β-face to assist the departure of the carbohydrate unit from the +1 subsite (pK_a >12 for a glycosyl leaving group). Thus, we investigated the activity of *Am*GH109A towards the α- and β-GalNAc(1→3)Gal motifs found in the blood group A antigen and in the globo antigen series, respectively. Surprisingly, *Am*GH109A exhibited similar activities against these non-activated disaccharides with a lower K_m for the β-GalNAc(1→3)Gal as compared to the substrate α-form (Figure 1b). A conservative estimate from the initial rates against 2.5 mM substrate, yields a k_{cat} of 10 min^{-1} and 20 min^{-1} at 25 °C for β-GalNAc(1→3)Gal and α-GalNAc(1→3)Gal, respectively. While no GH109 has yet been kinetically characterized on natural sugars, GH109 members are considered as "slow" enzymes.[26] The only other GH family adopting a related mechanism is GH4. This family has maximal reported turnover rates in the order of 10 min^{-1} for non-activated disaccharides, indicating that the nature of the oxidative mechanism in GH4/109 may be limiting to slower rates than classical (*i.e.* non-oxidative) GHs.[28] Importantly, these experiments were consistent with the presence of an acid/base catalyst, which has not been previously identified.

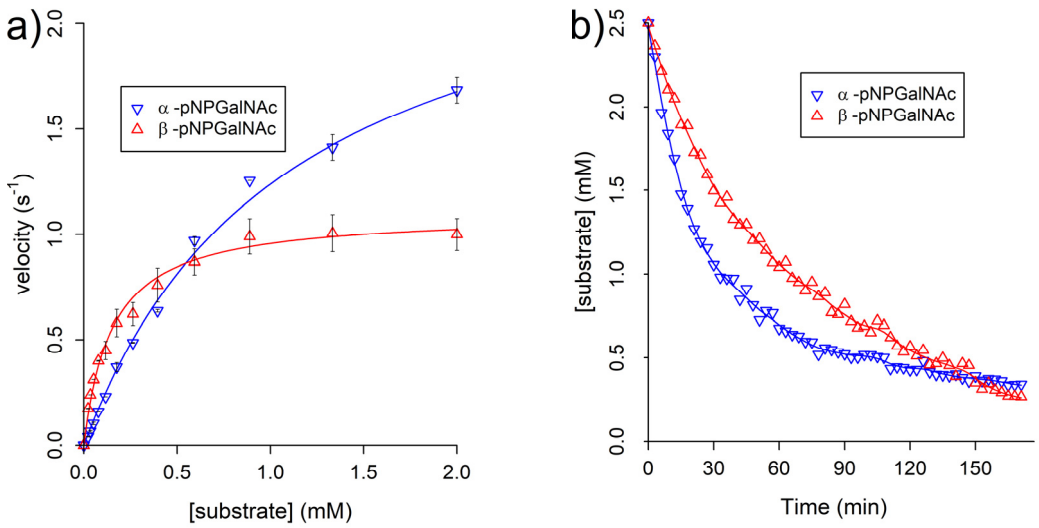

Figure 1. The dual activity of *Am*GH109A on substrates with a terminal non-reducing α- or β-GalNAc. **(a)** Michaelis-Menten plot of *p*NP-GalNAc hydrolysis. The data markers are the mean of three replicates with standard deviations and the solid lines are the fits of the Michaelis-Menten expression to the initial rate data. **(b)** NMR monitoring of GalNAc(α/β1,3)Gal hydrolysis by 3.2 μM *Am*GH109A. Reactions were performed at 25°C, in 20 mM HEPES pH or pD 6.6.

Structural analysis. To elucidate the missing acid/base catalyst in the mechanism of GH109 and to discern the structural elements behind the β-inverting activity, we determined the structure of *Am*GH109A (PDB: 6T2B). The crystal structure of *Am*GH109A was solved by molecular replacement using NagA as search model (PDB: 2IXA, 36.5% sequence identity) and refined to 2.13 Å resolution (Supporting Information Table S1). Four *Am*GH109A molecules are observed in the asymmetric unit, each with a GalNAc and a NAD^+ molecule bound in the active site. Analysis of the structure using the PISA server[29] indicates that *Am*GH109A forms a homodimer, which is organized in the same fashion as the NagA dimer[24] (Supporting Information Figure S2a).

The overall structure (Figure 2a) of *Am*GH109A closely resembles that of NagA (RMSD of 0.393 Å for 874 C_α atoms) and comprises an N-terminal Rossmann domain and a C-terminal α/β domain.[25] One notable difference relative to NagA, is the more open and shallow active site in *Am*GH109A, in particular the solvent accessible NAD^+ binding groove (Supporting Information Figure S2b,c). The difference in active site architecture is partly due to shortening of the two loops comprising residues F188-S208, and D322-G331 in NagA (Figure 2a, Supporting Information, Figure 2Sb,c). The GalNAc binding site in *Am*GH109A is similar to that of NagA with hydrogen bonds to Y226, R244, Y256, H259, Y339 (Supporting Information Figure S2d, Fig. 3a), equivalent to Y179, R213, Y225, H228, Y307 in NagA.

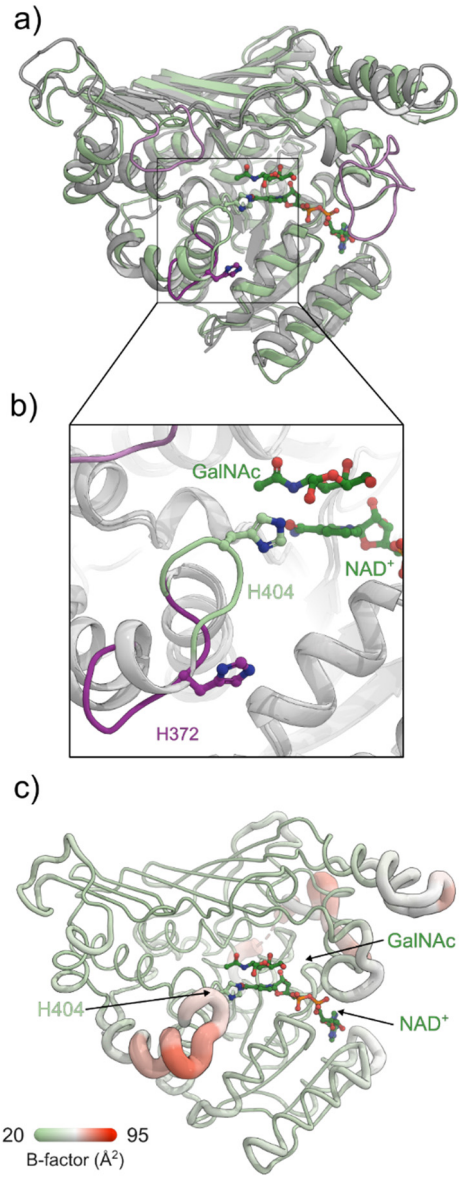

Figure 2. Structural analysis of *Am*GH109A. **(a)** A cartoon depicting the overall structure of *Am*GH109A (PDB: 6T2B, green) superimposed on NagA (PDB: 2IXA, grey). The main differences in the active site region are the two elongated loop regions (light purple) F188-S208, and D322-G331 in NagA and a loop that presents a conserved histidine residue in GH109 (NagA H372) and adopts a conformation away from the active site in NagA (magenta), whereas the corresponding loop is located in the active site of *Am*GH109A, which positions the conserved H404 at a hydrogen-bonding distance from the anomeric C1-OH group of the subsite –1 bound GalNAc unit. **(b)** A close-up of the flexible loop described in (a), which shows the two different conformations adopted by this loop in NagA (magenta) and in *Am*GH109A (green). The histidine located in the highly conserved GGHGG motif within this loop is shown in sticks in both enzymes and coloured the same as the flexible loop. The position of this histidine and its polar contact with the C1-OH of GalNAc highlighted this residue as a candidate for the acid/base catalyst, which has not been identified in GH109. **(c)** The structure of *Am*GH109A depicted as a cartoon B-factor putty, which highlights the loop encompassing the conserved GGHGG motif as a highly flexible region in the enzyme.

One striking difference is the additional contact to H404 in *Am*GH109A, which is not observed in NagA (Supporting Information Figure S2b-d). The α-anomeric oxygen is at a hydrogen bonded to

H404, located in a highly flexible glycine-rich loop (GGHGG in *Am*GH109A and GAGHGG in NagA). The computed pK_a (6.2)[30] and the hydrogen-bond to the anomeric proton (Figure 3a) highlight H404 as a plausible catalytic acid-base candidate in the mechanism. The assignment of the corresponding histidine in NagA (H372) was obscured as this residue was located >12 Å from the GalNAc unit due to a different loop conformation (Figure 2a,b, Supporting Information Figure S2c). Additional support for a functionally important role of this histidine stems from the conservation of the GHGG motif in 95 % of 3049 protein sequences sharing 20–65 % pairwise identities. An additional glycine preceding this motif is also present in more than 80% of the sequences, as observed for *Am*GH109A and *Am*GH109B.

H404 is the catalytic acid/base in GH109. To assess the role of H404 in the catalytic mechanism of GH109 enzymes, we analysed the effects of the H404A and H404F mutations. The activity of the H404F variant was detectable but not measurable ($k_{cat}<10^{-3}$ min^{-1}), while H404A could be chemically rescued towards hydrolysis of both α- and β-*p*NPGalNAc up to 3% of the wild-type activity upon imidazole addition (Figure 3b). Taken together, the conservation of the (G)GHGG motif, the structural data showing a relevant hydrogen bonding distance from the anomeric carbon and the mutational data establish that H404 is the acid/base catalyst in GH109, allowing us to complete the mechanism previously proposed for this enzyme family[24] (Figure 3c).

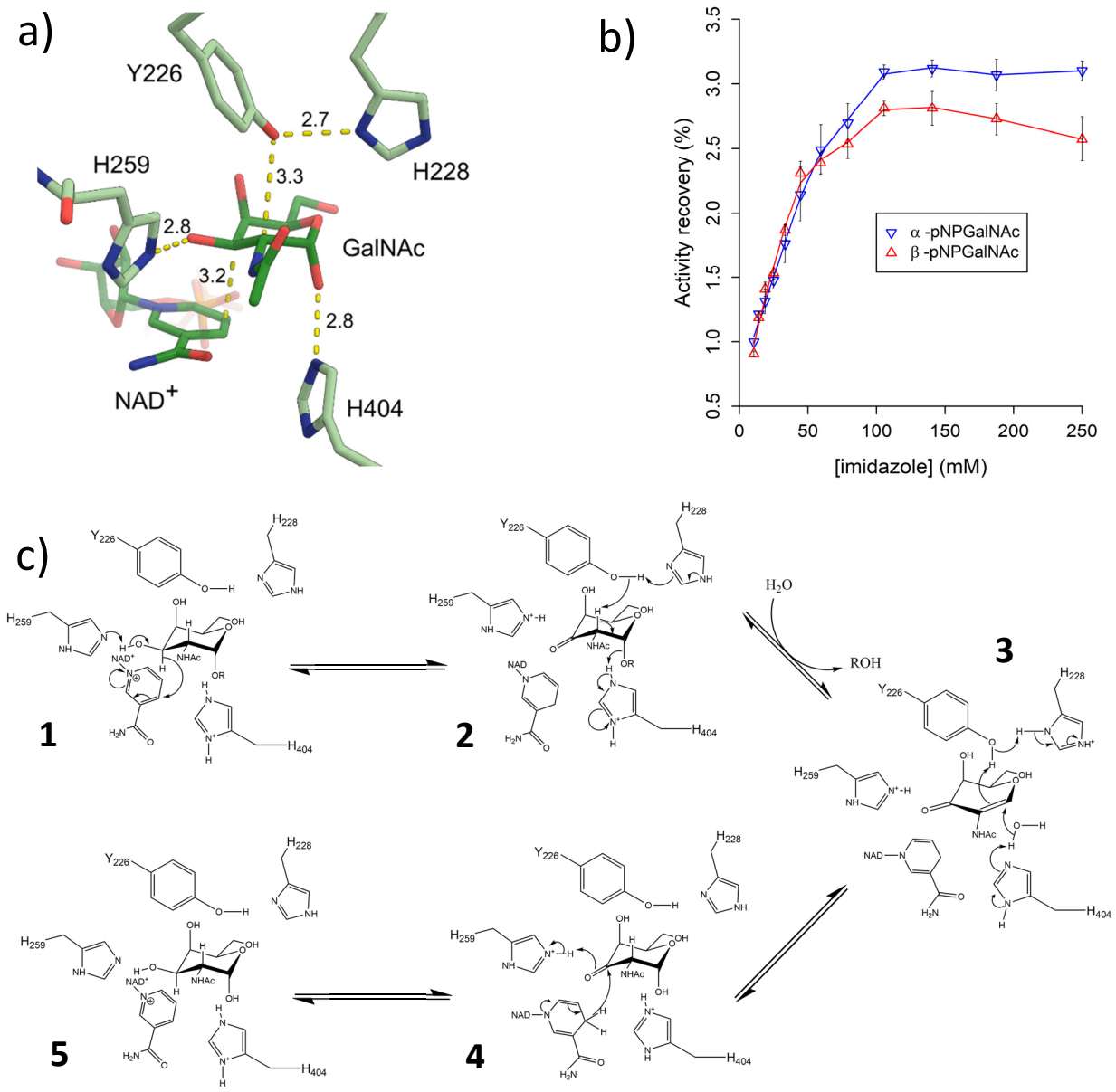

Figure 3. A conserved histidine corresponding to *Am*GH109A H404 is the acid/base catalyst in the mechanism of GH109 enzymes. **(a)** The catalytic residues Y226, H228, H259 and H404, as well as the GalNAc and the NAD^+ cofactor are represented in sticks. The distances between atoms that will exchange protons or hydride along the catalytic cycle are shown in yellow dotted lines. **(b)** Chemical rescue by imidazole of the H404A mutant activity on *p*NP-GalNAc at 400 µM, compared to the wild-type enzyme. **(c)** Scheme of the proposed revised α-retaining mechanism for GH109 enzymes, residues numbered according to the *Am*GH109A sequence.

Molecular dynamics simulations. The molecular determinants of the dual α/β specificity were explored by means of molecular dynamics (MD) simulations. The *Am*GH109A structure was simulated in free form and in complex with the two α- or β-GalNAc(1→3)Gal disaccharide substrates. The disaccharides, in an energy minimized conformation, were added to the active site by superimposition to the GalNAc present in *Am*GH109A (PDB: 6T2B). In the course of 400 ns simulations, the ligand free form and the complex simulations resulted in comparable protein flexibilities and conformations (Supporting Information Figure S3a). Both disaccharides mainly populate two conformational states based on the dihedral angles Φ (C2'-C1'-O1'-C3) and Ψ (C1'-O1'-C3-C4) (Supporting Information Figure S3b), a major state populated by 64 % and 95 % for the α-disaccharide and β-disaccharide, respectively, as well as a less populated set of conformations (36% for α, 5% for β). Interestingly, the most populated conformation for each disaccharide was shifted from the GalNAc in the –1 subsite of the crystal structure (Figure 4a,b). In this most-visited configuration, the α-disaccharide seemed to be not appropriately positioned for the first step of the reaction, with 5.9 Å distance to H259. On the other hand, Y226 and H404 were well-positioned for catalysing the E2 elimination of the second reaction step (Figure 3c, step 2). Conversely, the β-disaccharide appears better positioned for the first step of the reaction, and not for the second one, with both distance and positioning between Y226 and C3' being unsuitable for the reaction. While no reactivity assessment can be directly drawn from MD simulations, one can note that the kinetic data are fully consistent with a faster first step (Figure 3c, step 1) followed by a slower second step for the β-species with both lower k_{cat} and K_m. The observed kinetics are also consistent with the E2 elimination chemistry for the α-disaccharide (Figure 3c, step 2) being faster than then the stepwise E1cB mechanism, which is likely to take place for the β-disaccharide (Figure 4e). Note that instead of the E2 elimination displayed in Figure 3c, the β-anomer would likely react through an E1cB elimination mechanism, with a transient carbanion on a C2 stabilised by the carbamyl at this position. Indeed, in a β configuration, the C2-H2 and the C1-O1 bonds cannot meet the antiperiplanar requirement of an E2 elimination (Figure 4e).

The four distances represented in Figure 4a,b between C3' and the NAD^+, C2' and the phenolic oxygen of Y226, the H259 Nε and O3', and between the H404 Nε and O1' were monitored along the simulation (Figure 4c,d) as proton or hydride exchange between these atoms is crucial for the reaction. The α-disaccharide distances vary more than the β-counterpart (Figure 4c,d) due to a more populated and more frequently visited subpopulation (36% versus 5%, Supporting Information Figure S3b). Notably, the glycosidic bond oxygen O1' shows the least fluctuation in both disaccharide forms over the course of the simulations, and occupies a similar position despite the opposite configurations (Supplement Figure S3c). Thus, the H404 $N_ε$ is located within 3.5–4.5 Å of O1' for both the β- and α-disaccharides in most of simulations, which is supportive of the well-positioning of this residue to be the acid catalyst for both α- and β- substrates (Figure 4c,d).

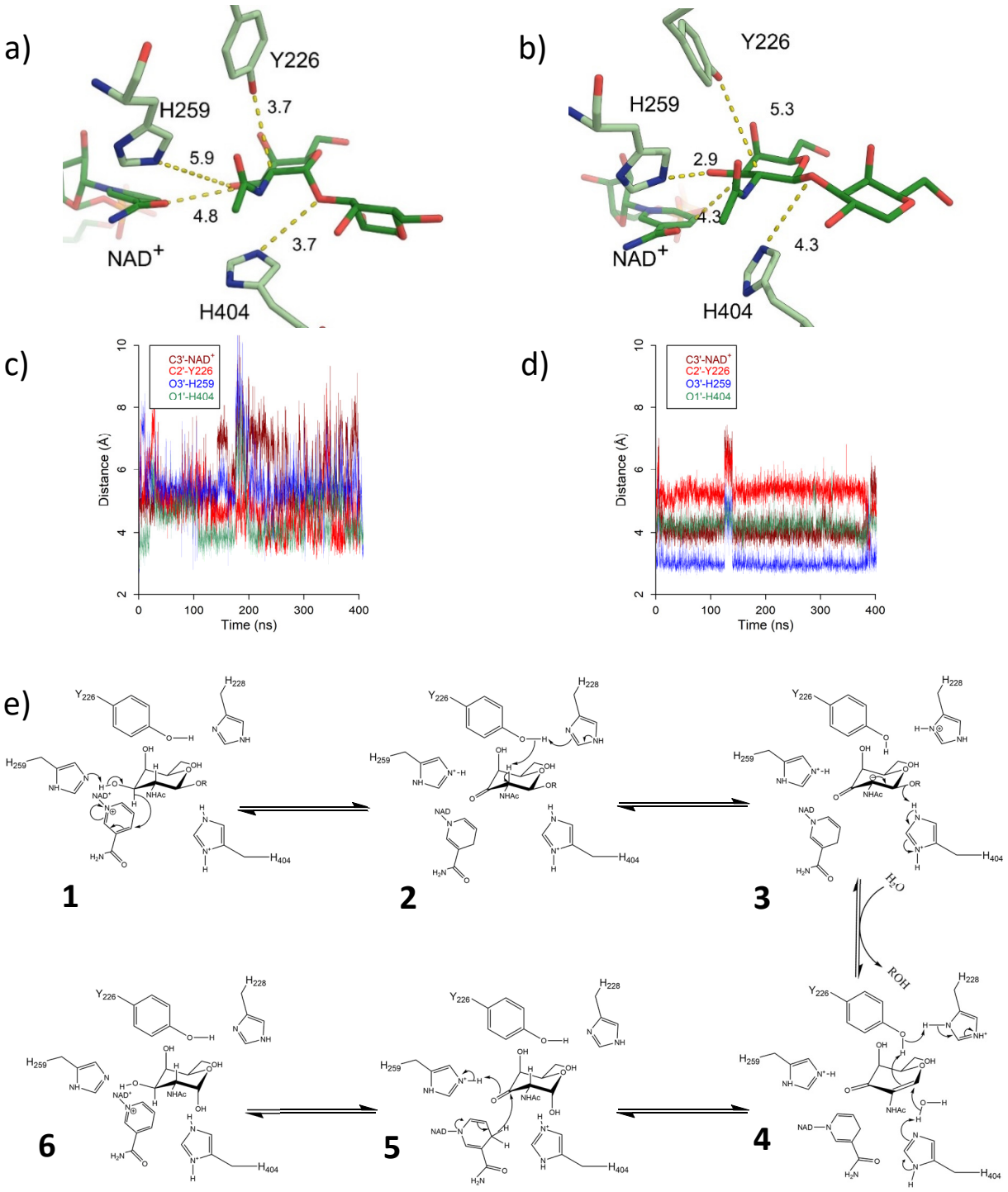

Figure 4. Molecular dynamics simulations of *Am*GH109A bound to the α- and β-GalNAc(1→3)Gal disaccharides. **(a), (b)** Disaccharide (α- in (a) and β (b)), catalytic residues Y226, H259, H404 and NAD^+ cofactor are shown in sticks for the most populated conformation of the MD simulations (frames 8.4 ns and 13.4 ns, produced by *gmx cluster* with a 0.5 Å cut-off). Distances between atoms that will exchange protons or hydride along the catalytic cycle are shown as dotted lines. **(c)** and **(d)** Evolution of the distances between C3' and the NAD, C2' and the phenolic oxygen of Y226, the H259 Nε and O3', and between the H404 Nε and O1' along the simulations for the

α- and β-disaccharide, respectively. **(e)** Scheme of the proposed β-retaining mechanism in GH109 based on *Am*GH109A.

CONCLUSION

This work unambiguously shows that distinct enzymes from GH109 display dual α-retaining and β-inverting activities. Moreover, these activities were observed on natural disaccharides and at similar levels. Bioinformatics and structural analyses provided a rationale for this dual activity by identifying a flexible loop harbouring the GGHGG conserved motif within GH109. This loop provides the catalytic acid/base, which is identified in the present study and corroborated using mutational analysis and MD simulations. A practical implication of this is that GH109 enzymes should be systematically assayed against β-substrates as well as their α-counterparts. Formerly, only a single GH4, MalH from *F. mortiferum*, has shown similar levels of activity towards both stereochemistry, but at very low levels ($k_{cat} \approx 10^{-3}$ min^{-1}) and against artificial, activated substrates (*p*NP-α-D-Glc-6-phosphate and *p*NP-β-D-Glc-6-phosphate).[31]

We speculate that this dual functionality could be an evolutionary adaptation of the specialised host-glycan degrading genus *Akkermansia* that inhabits a mucin- and host glycoconjugate-rich ecological niche. This adaptation may reduce the number of enzymes needed to confer access to a broader diversity of α- and β-linked substrates from mucin and/or glycans that decorate shed epithelial enterocytes.

MATERIALS AND METHODS

Chemicals and carbohydrate substrates. *N*-Acetylgalactosamine (GalNAc), α-*p*NPGalNAc and β-*p*NPGalNAc were from Sigma. The disaccharides α-GalNAc(1→3)Gal and β-GalNAc(1→3)Gal were from Carbosynth (Berkshire, UK). All other chemicals were of analytical grade, unless stated otherwise.

Cloning, expression and purification of the glycoside hydrolase family 109 (GH109) enzymes from *Akkermansia muciniphila* ATCC BAA-835. The gene fragment of locus tags Amuc_0017 (GenBank: ACD03864.1, here *Am*GH109B) and Amuc_0920 (GenBank: ACD04752.1, here *Am*GH109A), which encode the GH109 mature peptides lacking the signal peptides (amino acid residues 1-26 and 1-29, respectively as predicted by SignalP 4.0[32], were amplified from *Akkermansia muciniphila* ATCC BAA-835 (DSM 22959) genomic DNA using the primer (sense 5′-**AGGAGATATACCATG**CAGGAAGTAGCCCCCCTGG-3', antisense 5′-**GGTGGTGGTGCTCGA**GCTGGACGATGCCCAGCG-3') and (sense 5′-**AGGAGATATACCATG**CTCCCTGGGAAGGCTGTG-3', antisense 5′-**GGTGGTGGTGCTCGA**GTGAAACCACGGCCAGCG-3') respectively to generate amplicons flanked by sequences for homologous recombination (in bold) with the pET28a(+) vector (Novagen, Madison, WI). Infusion cloning (Clonetech/takara, CA, USA) was used to clone these amplicons into the NcoI and XhoI sites of the above vector. The resulting recombinant plasmids pET28a-*Am*GH109A pET28a-*Am*GH109B, were transformed into *Escherichia coli* DH5α and transformants were selected on LB plates supplemented with kanamycin (50 μg·mL^{-1}). The synthetic genes encoding the single mutant variants *Am*GH109A H404A and *Am*GH109A

H404F, both cloned in the same vector were purchased from Biomatik (Ontario, Canada) and transformed in *E. coli* DH5α.

The wild type and mutant enzyme variants were produced in *E. coli* Rosetta(DE3) (Novagen, Madison, WI) grown in 2 L LB medium at 30 °C to $OD_{600}\approx0.5$, followed by cooling the medium on ice for 30 min, then the expression was induced by the addition of 200 µM isopropyl β-D-1-thiogalactopyranoside and growth was continued overnight at 18 °C. The cells were harvested by centrifugation (10,000 *g*, 30 min), re-suspended in 10 ml of the purification buffer A (20 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10% (v/v) glycerol, pH 7.5) and lysed by an homogenizer (SPCH-1, Stansed Fluid Power, Essex, UK), followed by incubation for 30 min on ice with 5 µL benzonase nuclease (Sigma). The lysate was then centrifuged for 20 min 45000 *g* at 4°C and the supernatant was filtered (0.45 µm) and loaded onto a HisTrap HP column (GE Healthcare, Uppsala, Sweden). Then, the bound protein was washed (13 column volumes, CV) and eluted with the same buffer using an imidazole gradient from 10 to 400 mM in 15 CV. Pure protein fractions based on SDS-PAGE analysis were collected and concentrated to 5 mg mL^{-1}, applied onto a HiLoad 16/600 Superdex 75 prep grade column (GE healthcare), and eluted by 1.2 CV of 20 mM HEPES, 150 mM NaCl, pH 6.8, and concentrated with 10 kDa Amicon ultracentrifugal filters (Millipore, Darmstadt, Germany). The fractions were analysed by SDS-PAGE and pure fractions were pooled, concentrated, and the protein concentration was determined using a Nanodrop (Thermo, Waltham, MA) using the theoretically predicted extinction coefficients $\varepsilon_{280\ nm}$ 91330 $M^{-1}cm^{-1}$ and 70860 $M^{-1}cm^{-1}$ for *Am*GH109A and *Am*GH109B, respectively as determined using the ProtParam tool (http://web.expasy.org/protparam). Finally, NaN_3 (0.005% w/v) was added to the enzyme stocks that were stored at 4°C for further use.

Enzymatic analyses. Spectrophotometry-based kinetics experiments were performed in a PowerWave XS microtiter plate reader monitored by the Gen5 software (Bio-Tek Instruments, Inc.) by measuring OD_{405} each 20 s for 20 min at 25 °C. Reactions well of 200 µL contained *p*NPGalNAc 0.0025–10 mM (twelve concentrations), HEPES buffer 50 mM pH 6.6, and 100 nM of *Am*GH109A/B, whereas the *Am*GH109A H404A and H404F were evaluated at 1.5 µM. Initial rates, calculated from slopes of *p*NP formation *vs* time, were fitted to the Michaelis-Menten equation $v = \frac{k_{cat}\cdot[S]\cdot[E]}{K_M+[S]}$ using OriginPro 2015 (OriginLab, USA) to obtain k_{cat} and K_M values. NMR spectra were recorded on an 800 MHz Bruker Avance III (799.85 MHz for ^1H) equipped with a 5 mm TCI cryoprobe using ^1H with presaturation. Reaction mixtures of 600 µL containing 3.2 µM *Am*GH109A and 2.5 mM disaccharides in a 50 mM deuterated HEPES buffer pD=6.6 (corresponding to a measured "pH"=6.2)[33] were analysed. Time course experiments were obtained using pseudo-2D kinetics experiments, with spectra recorded every 3 min. Integration was performed on peaks at 4.16, 4.11 and 3.96 ppm (α-disaccharide) or 4.15, 4.04 and 3.93 ppm (β-disaccharide).

Bioinformatics. Sequences were retrieved from the non-redundant protein database,[34] using *Am*GH109A as a query. Protein BLAST searches were performed on the NCBI server (www.ncbi.nlm.nih.gov), using default options but for the "Max target sequences" parameter, set at 20 000. On July 3rd, 2018, 18464 sequences were obtained and clustered to limit pairwise sequence identity at 65 % by iterative cd-hit runs.[35] Iterative multiple sequence alignments using Clustal Ω[36] were performed to increase the minimum pairwise sequence identity by 5 % increments until convergence is reached and all sequences share ≥20% pairwise sequence identity.

Crystallography. A stock solution of *Am*GH109A (6 mg·mL^{-1}) in 20 mM HEPES, 150 mM NaCl, pH 7.1 was supplemented with *N*-Acetylgalactosamine to a final concentration of 10 mM one hour prior to setting up sitting drops using the crystal Gryphon liquid handling robot (Art Robbins Instruments, Sunnyvale, USA). Drops containing equal volumes (150 nL) of protein and reservoir solution (15% w/v PEG 2000 MME and 0.1 mM sodium citrate) were equilibrated against 60 µL reservoir solution. Crystals with a maximum dimension of approximately 100 µm formed after 1 day. MicroMount loops (MiTeGen) were used to harvest crystals that were cryoprotected in the reservoir solution with the addition of 15% (w/v) PEG 400 before being flash frozen in liquid nitrogen. A data set of 3,800 frames (0.1 degree per frame) was collected at the BioMAX beamline (MaxIV, Lund, Sweden) with a detector distance of 220.399 mm and X-ray wavelength of 1.3799 Å. The data were processed automatically at the beam line using autoPROC[37] to a resolution of 2.13 Å using autoPROC's default cutoff criteria. Data were phased with molecular replacement using Phaser[38] in the Phenix software package,[39] with the *N*-acetylgalactosaminidase from *Elizabethkingia meningoseptica* (PDB: 2IXA) as search model. An initial model was built using Phenix.autobuild[39] and completed with alternating manual rebuilding in Coot[40] and automatic Phenix.refine refinement.[41] Ligands were placed with Phenix.LigandFit.[42]

Molecular dynamics simulations. The structure of *Am*GH109A (PDB: 6T2B) was protonated in the H++ Server with pH 7.0 and 0.15 M NaCl ionic strength[43-45] and parametrized with AMBER14 for proteins.[46] The *Am*GH109A chain B was chosen as the most complete structure. Parameters for NAD^+ were generated based on published data.[47-48] Disaccharides D-GalNAc*p*-α(1-3)-D-Gal*p*-β-OH (a13) and D-GalNAc*p*-β(1-3)-D-Gal*p*-β-OH (b13) were parametrized and minimized with GLYCAM06[49] and superimposed in the protein crystal structures with the solved D-GalNAc*p*. The simulation complex was placed in an almost cubic box with the dimensions ≈85x83x84 Å.^{3} Simulations were run with the TIP3P water model[50] and charges were equalized with the *gmx genion* routine.[51] MD simulations were carried out using the GROMACS2018 program package.[51-52] After energy minimization to a maximum force smaller than 100 kJ·mol^{-1}·nm^{-1} (steep descent), the system was equilibrated in two simulations with 1 ns each. All simulations were performed under isothermal-isobaric (NPT) ensemble conditions with Parrinello-Rahman barostat coupling[53] (reference pressure of 1.0 bar and coupling time constant of 1 ps) and a V-rescale thermostat[54] (reference temperature 300 K and coupling time constant 0.2 ps). Simulation time step was 2 fs. Hydrogen bonds were constrained using the LINCS algorithm.[55] All systems were simulated for at least 400 ns. Simulations were analysed with *gmx rms*, *gmx rmsf* and *gmx distance* and *gmx cluster* packages from GROMACS2018.[51-52]

■ ASSOCIATED CONTENT

Supporting Information

Appended below

ABBREVIATIONS

HGM, Human gut microbiota; GH109, Glycoside hydrolase family 109; α-*p*NPGalNAc, 4-nitrophenyl 2-acetamido-2-deoxy-α-d-galactopyranoside; β-*p*NPGalNAc, 4-nitrophenyl 2-acetamido-2-deoxy-β-d-galactopyranoside; MD, molecular dynamics; *Am*GH109A, GH109 from *Akkermansia muciniphila* encoded by the locus Amuc_0920, *Am*GH109B, GH109 from *Akkermansia muciniphila* encoded by the locus Amuc_0017.

AUTHOR INFORMATION

Corresponding Authors

*Maher abou Hachem (maha@bio.dtu.dk) and *Ditte Hededam Welner (diwel@dtu.dk)

Author Contributions

The research was conceived by BS, MAH, DT and ENK. BS did the initial cloning and activity screening of the GH109 enzymes. DT performed the NMR experiments and the kinetic analysis of the wild-type and the mutant enzymes. EKC, FF and DHW carried out the structural characterization of the enzyme. SK performed the MD analyses. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

Ministry of higher education and scientific research of Iraq for a PhD stipend for BS. The Novo Nordisk Foundation for a postdoctoral fellowship for DT (NNF17OC0025660). European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie postdoc fellowship agreement No 713683 for SK. (NNF17OC0025660). The Novo Nordisk Foundation for the grants NNF10CC1016517 and NNF16OC0019088 for DW. Biochemical instrumentation at DTU Bioengineering was supported by grants from the Independent Research Fund Denmark, the Danish Strategic Research Council and the Carlsberg Foundation. X-ray data collection was supported by Danscatt. The NMR spectra were recorded at the NMR Center DTU, supported by the Villum Foundation. We thank Tine Sofie Nielsen for protein production and purification. We thank the staff of the synchrotrons ALS and MAX IV for technical assistance.

REFERENCES

1. Gensollen, T.; Iyer, S. S.; Kasper, D. L.; Blumberg, R. S., How colonization by microbiota in early life shapes the immune system. *Science* **2016,** *352* (6285), 539-544.
2. Sonnenburg, J. L.; Bäckhed, F., Diet-microbiota interactions as moderators of human metabolism. *Nature* **2016,** *535* (7610), 56-64.
3. Garrett, W. S., Cancer and the microbiota. *Science* **2015,** *348* (6230), 80-86.
4. Takahashi, K.; Nishida, A.; Fujimoto, T.; Fujii, M.; Shioya, M.; Innaeda, H.; Inatomi, O.; Bamba, S.; Andoh, A.; Sugimoto, M., Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease. *Digestion* **2016,** *93* (1), 59-65.
5. Lloyd-Price, J.; Arze, C.; Ananthakrishnan, A. N.; Schirmer, M.; Avila-Pacheco, J.; Poon, T. W.; Andrews, E.; Ajami, N. J.; Bonham, K. S.; Brislawn, C. J.; Casero, D.; Courtney, H.; Gonzalez, A.; Graeber, T. G.; Hall, A. B.; Lake, K.; Landers, C. J.; Mallick, H.; Plichta, D. R.; Prasad, M.; Rahnavard, G.; Sauk, J.; Shungin, D.; Vazquez-Baeza, Y.; White, R. A., 3rd; Braun, J.; Denson, L. A.; Jansson, J. K.; Knight, R.; Kugathasan, S.; McGovern, D. P. B.; Petrosino, J. F.; Stappenbeck, T. S.; Winter, H. S.; Clish, C. B.; Franzosa, E. A.; Vlamakis, H.; Xavier, R. J.; Huttenhower, C., Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **2019,** *569* (7758), 655-662.
6. Pedersen, H. K.; Gudmundsdottir, V.; Nielsen, H. B.; Hyotylainen, T.; Nielsen, T.; Jensen, B. A.; Forslund, K.; Hildebrand, F.; Prifti, E.; Falony, G.; Le Chatelier, E.; Levenez, F.; Dore, J.; Mattila, I.; Plichta,

D. R.; Poho, P.; Hellgren, L. I.; Arumugam, M.; Sunagawa, S.; Vieira-Silva, S.; Jørgensen, T.; Holm, J. B.; Trost, K.; Kristiansen, K.; Brix, S.; Raes, J.; Wang, J.; Hansen, T.; Bork, P.; Brunak, S.; Oresic, M.; Ehrlich, S. D.; Pedersen, O., Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **2016,** *535* (7612), 376-81.
7. Turnbaugh, P. J.; Ley, R. E.; Mahowald, M. A.; Magrini, V.; Mardis, E. R.; Gordon, J. I., An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **2006,** *444* (7122), 1027-31.
8. Dao, M. C.; Everard, A.; Aron-Wisnewsky, J.; Sokolovska, N.; Prifti, E.; Verger, E. O.; Kayser, B. D.; Levenez, F.; Chilloux, J.; Hoyles, L.; Dumas, M.-E.; Rizkalla, S. W.; Dore, J.; Cani, P. D.; Clement, K.; Consortium, M. I.-O., *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut* **2016,** *65* (3), 426-436.
9. Karlsson, C. L. J.; Onnerfalt, J.; Xu, J.; Molin, G.; Ahrne, S.; Thorngren-Jerneck, K., The microbiota of the gut in preschool children with normal and excessive body weight. *Obesity* **2012,** *20* (11), 2257-2261.
10. Plovier, H.; Everard, A.; Druart, C.; Depommier, C.; Van Hul, M.; Geurts, L.; Chilloux, J.; Ottman, N.; Duparc, T.; Lichtenstein, L.; Myridakis, A.; Delzenne, N. M.; Klievink, J.; Bhattacharjee, A.; van der Ark, K. C.; Aalvink, S.; Martinez, L. O.; Dumas, M. E.; Maiter, D.; Loumaye, A.; Hermans, M. P.; Thissen, J. P.; Belzer, C.; de Vos, W. M.; Cani, P. D., A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat. Med.* **2016,** *3* (1), 107-113.
11. Derrien, M.; Vaughan, E. E.; Plugge, C. M.; de Vos, W. M., *Akkermansia muciniphila* gen. nov., sp nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* **2004,** *54*, 1469-1476.
12. Reunanen, J.; Kainulainen, V.; Huuskonen, L.; Ottman, N.; Belzer, C.; Huhtinen, H.; de Vos, W. M.; Satokari, R., *Akkermansia muciniphila* Adheres to enterocytes and strengthens the integrity of the epithelial cell layer. *Appl. Environ. Microbiol.* **2015,** *81* (11), 3655-3662.
13. Everard, A.; Belzer, C.; Geurts, L.; Ouwerkerk, J. P.; Druart, C.; Bindels, L. B.; Guiot, Y.; Derrien, M.; Muccioli, G. G.; Delzenne, N. M.; de Vos, W. M.; Cani, P. D., Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (22), 9066-9071.
14. Ansaldo, E.; Slayden, L. C.; Ching, K. L.; Koch, M. A.; Wolf, N. K.; Plichta, D. R.; Brown, E. M.; Graham, D. B.; Xavier, R. J.; Moon, J. J.; Barton, G. M., *Akkermansia muciniphila* induces intestinal adaptive immune responses during homeostasis. *Science* **2019,** *364* (6446), 1179-1184.
15. van Passel, M. W. J.; Kant, R.; Zoetendal, E. G.; Plugge, C. M.; Derrien, M.; Malfatti, S. A.; Chain, P. S. G.; Woyke, T.; Palva, A.; de Vos, W. M.; Smidt, H., The genome of *Akkermansia muciniphila*, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *PLoS One* **2011,** *6* (3), 8.
16. Lombard, V.; Ramulu, H. G.; Drula, E.; Coutinho, P. M.; Henrissat, B., The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014,** *42* (D1), D490-D495.
17. Tailford, L. E.; Crost, E. H.; Kavanaugh, D.; Juge, N., Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **2015,** *6*, No. 81.
18. Holmen Larsson, J. M.; Thomsson, K. A.; Rodriguez-Pineiro, A. M.; Karlsson, H.; Hansson, G. C., Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2013,** *305* (5), G357-63.
19. Viborg, A. H.; Katayama, T.; Arakawa, T.; Abou Hachem, M.; Lo Leggio, L.; Kitaoka, M.; Svensson, B.; Fushinobu, S., Discovery of α-L-arabinopyranosidases from human gut microbiome expands the diversity within glycoside hydrolase family 42. *J. Biol. Chem.* **2017,** *292* (51), 21092-21101.

20. Gloster, T. M.; Turkenburg, J. P.; Potts, J. R.; Henrissat, B.; Davies, G. J., Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora. *Chem. Biol.* **2008,** *15* (10), 1058-67.
21. Yip, V. L.; Thompson, J.; Withers, S. G., Mechanism of GlvA from *Bacillus subtilis*: a detailed kinetic analysis of a 6-phospho-α-glucosidase from glycoside hydrolase family 4. *Biochemistry* **2007,** *46* (34), 9840-52.
22. Yip, V. L.; Varrot, A.; Davies, G. J.; Rajan, S. S.; Yang, X.; Thompson, J.; Anderson, W. F.; Withers, S. G., An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 β-glycosidase from *Thermotoga maritima*. *J. Am. Chem. Soc.* **2004,** *126* (27), 8354-5.
23. Yip, V. L. Y.; Withers, S. G., Mechanistic analysis of the unusual redox-elimination sequence employed by *Thermotoga maritima* bglt: A 6-phospho-β-glucosidase from glycoside hydrolase family 4. *Biochemistry* **2006,** *45* (2), 571-580.
24. Sulzenbacher, G.; Liu, Q. P.; Bennett, E. P.; Levery, S. B.; Bourne, Y.; Ponchel, G.; Clausen, H.; Henrissat, B., A novel α-*N*-acetylgalactosaminidase family with an NAD(+)-dependent catalytic mechanism suitable for enzymatic removal of blood group A antigens. *Biocatal. Biotransform.* **2010,** *28* (1), 22-32.
25. Liu, Q. P.; Sulzenbacher, G.; Yuan, H.; Bennett, E. P.; Pietz, G.; Saunders, K.; Spence, J.; Nudelman, E.; Levery, S. B.; White, T.; Neveu, J. M.; Lane, W. S.; Bourne, Y.; Olsson, M. L.; Henrissat, B.; Clausen, H., Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **2007,** *25* (4), 454-64.
26. Rahfeld, P.; Sim, L.; Moon, H.; Constantinescu, I.; Morgan-Lang, C.; Hallam, S. J.; Kizhakkedathu, J. N.; Withers, S. G., An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat. Microbiol.* **2019**.
27. Chakladar, S.; Shamsi Kazem Abadi, S.; Bennet, A. J., A mechanistic study on the α-*N*-acetylgalactosaminidase from *E. meningosepticum*: a family 109 glycoside hydrolase. *MedChemComm* **2014,** *5* (8), 1188-1192.
28. Hall, B. G.; Pikis, A.; Thompson, J., Evolution and biochemistry of family 4 glycosidases: implications for assigning enzyme function in sequence annotations. *Mol. Biol. Evol.* **2009,** *26* (11), 2487-97.
29. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007,** *372* (3), 774-97.
30. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004,** *32*, W665-W667.
31. Pikis, A.; Immel, S.; Robrish, S. A.; Thompson, J., Metabolism of sucrose and its five isomers by *Fusobacterium mortiferum*. *Microbiology* **2002,** *148* (3), 843-852.
32. Petersen, T. N.; Brunak, S.; von Heijne, G.; Nielsen, H., SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011,** *8* (10), 785-786.
33. Li, N. C.; Tang, P.; Mathur, R., Deuterium isotope effects on dissociation constants and formation constants1. *J. Phys. Chem. B* **1961,** *65* (6), 1074-1076.
34. Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S., UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **2004,** *32* (Database issue), D115-9.
35. Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W., CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **2010,** *26* (5), 680-2.
36. Sievers, F.; Higgins, D. G., Clustal Omega. *Curr. Protoc. Bioinformatics* **2014,** *48* (1), 3.13.1-3.13.16.

37. Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G., Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2011,** *67* (Pt 4), 293-302.
38. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.
39. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2010,** *66*, 213–221.
40. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2010,** *66* (Pt 4), 486-501.
41. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **2002,** *58* (Pt 11), 1948-54.
42. Terwilliger, T. C.; Klei, H.; Adams, P. D.; Moriarty, N. W.; Cohn, J. D., Automated ligand fitting by core-fragment fitting and extension into density. *Acta crystallographica. Section D, Biological crystallography* **2006,** *62* (Pt 8), 915-922.
43. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V., H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012,** *40* (Web Server issue), W537-41.
44. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **2005,** *33* (Web Server issue), W368-71.
45. Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A., A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* **2006,** *63* (4), 928-38.
46. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-713.
47. Pavelites, J. J.; Gao, J.; Bash, P. A.; Mackerell Jr., A. D., A molecular mechanics force field for NAD+ NADH, and the pyrophosphate groups of nucleotides. *J. Comput. Chem.* **1997,** *18* (2), 221-239.
48. Walker, R. C.; de Souza, M. M.; Mercer, I. P.; Gould, I. R.; Klug, D. R., Large and Fast Relaxations inside a Protein: Calculation and Measurement of Reorganization Energies in Alcohol Dehydrogenase. *J. Phys. Chem. B* **2002,** *106* (44), 11658-11665.
49. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29* (4), 622-55.
50. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.
51. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.
52. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005,** *26* (16), 1701-1718.

53. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
54. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.
55. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472.

Supporting information for:

The catalytic acid-base in GH109 resides in a conserved GGHGG loop and allows for comparable α-retaining and β-inverting activity in an *N*-acetylgalactosaminidase from *Akkermansia muciniphila*

David Teze,[†,§] Bashar Shuoker,[†, ₸,§] Evan Kirk Chaberski[‡], Sonja Kunstmann,[†] Folmer Fredslund,[‡] Günther H.J. Peters,[‡] Eva Nordberg Karlsson,[₸] Ditte Hededam Welner,[*‡] Maher Abou Hachem[*,†]

[†]Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, 2800, Denmark
[‡]Enzyme Engineering and Structural Biology, Novo Nordisk Center for Biosustainability, Kemitorvet, building 220 2800 Lyngby, Denmark
[₸]Biotechnology, Department of Chemistry (KILU), Lund University, Post Office Box 124, 221 00 Lund, Sweden
[‡]Department of Chemistry, Technical University of Denmark, DK-2800 Lyngby, Denmark

[§]Authors contributed equally to the study.

Correspondence: [*]Maher abou Hachem (maha@bio.dtu.dk) and [*]Ditte Hededam Welner (diwel@dtu.dk)

Content

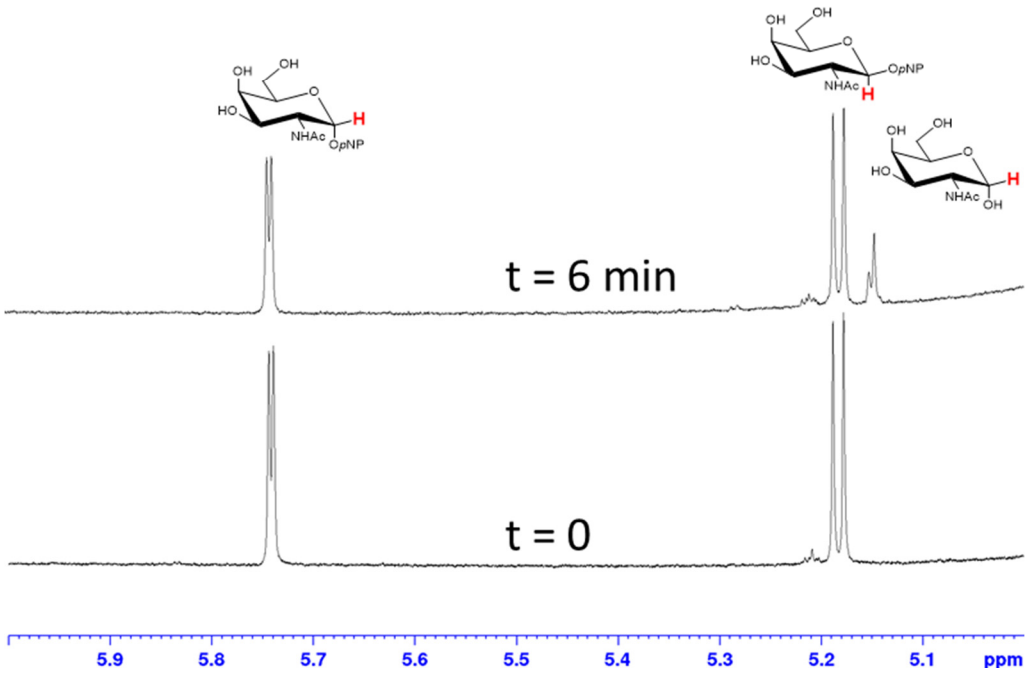

Figure S1. Analysis of the seterochemical outcome of *Am*GH109A-catalysed reactions using NMR. To investigate *Am*GH109A mechanism, we monitored by ^{1}H NMR the conversion of of α- and β-*p*NP-GalNAc (500 µM each) to GalNAc in 20 mM phosphate buffer, pD = 6.6 at 298 K. At t = 6 min (above), only the H1 from the α-GalNAc anomer is visible at δ = 5.15 ppm. Traces of the β-GalNAc H1 are visible at δ = 4.56 ppm after t = 18 min due to mutarotation.

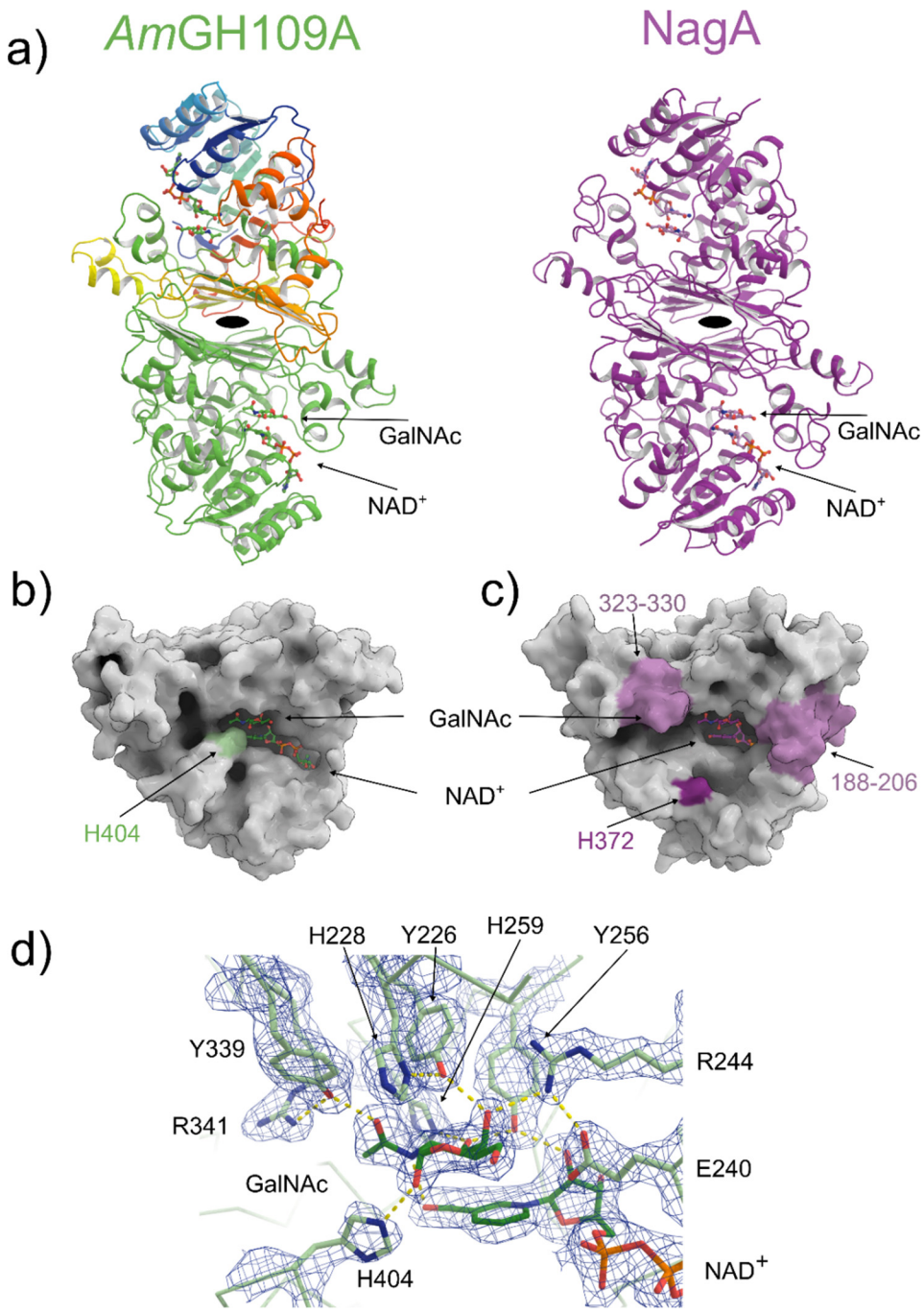

Figure S2. Structural comparison of NagA and *Am*GH109A. **(a)** Cartoon representation of the *Am*GH109A dimer (one subunit (chain) is coloured from blue to red, corresponding to N- and C-terminal respectively, and the other subunit is in green) and NagA (shown in purple) with the bound GalNAc and NAD^+ ligands shown as ball and stick. The two-fold axis of the dimer is shown. **(b)** Surface representation of *Am*GH109A showing the location of H404 and a relatively water accessible NAD^+. **(c)** Surface representation of NagA showing the two extended loops (188-206 and 323-330) as well as H372. Loop 323-330 shields the NAD^+ from the solvent. **(d)** The final $2mF_o$-dF_c electron density map at 1 σ at the active site of *Am*GH109A. Yellow dots indicating hydrogen bond network around the GalNAc ligand.

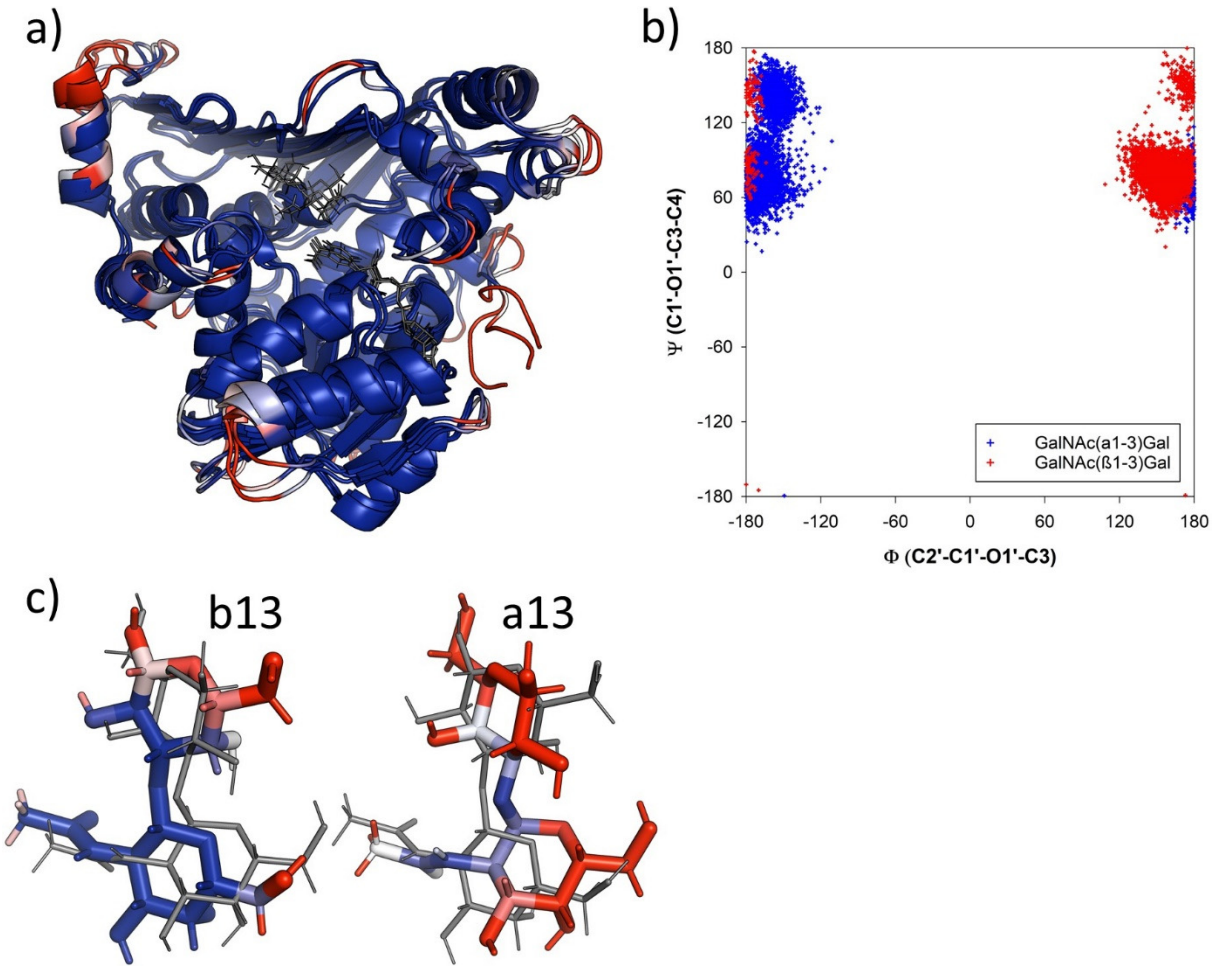

Figure S3. Molecular dynamics simulation ligand conformations. **(a)** Superimposition of *Am*GH109A after simulation for 400 ns with the most populated conformation of the protein without ligand and with α- and β-disaccharide. The protein is depicted as cartoon coloured based on B-factors derived from the simulation (blue_white_red spectrum, minimum=20, maximum=50). **(b)** Ramachandran Plot of the glycosidic linkage angles from the α- (blue) and β- (red) disaccharides. **(c)** The most populated conformation of the α- (**a13**, right) and β-disaccharide (**b13**, left) superimposed vice-versa (in grey thin sticks) coloured by B factors (spectrum as in a).

Table S1. Data collection and refinement statistics of *Am*GH109A. Statistics for the highest-resolution shell are in parentheses.

Beamline / Wavelength (Å)	MaxIV-Biomax / 1.3799
PDB	6T2B
Resolution range (Å)	70.87-2.13 (2.20-2.13)
Space group	$P\ 2_1$
Unit cell *a, b, c* (Å), β (°)	66.11, 182.64, 70.87, 90.139
Total reflections	660109 (42892)
Unique reflections	92335 (7967)
Multiplicity	7.1 (5.4)
Completeness (%)	93.60 (73.20)
Mean I/sigma(I)	11.61 (2.22)
Wilson B-factor	28.91
R_{merge}	0.112 (0.58)
R_{meas}	0.121 (0.645)
R_{pim}	0.044 (0.274)
CC1/2	0.998 (0.714)
CC*	0.999 (0.913)
Reflections used in refinement	88157 (6927)
Reflections used for *R*-free	1896 (146)
R_{work}	0.164 (0.234)
R_{free}	0.195 (0.304)
CC(work)	0.971 (0.853)
CC(free)	0.955 (0.739)
Non-hydrogen atoms	15130
macromolecules	13622
ligands	262
solvent	1246
Protein residues	1750
RMS bonds (Å)	0.003
RMS angles (˚)	0.64
Ramachandran favored (%)	97.58
Ramachandran allowed (%)	2.42
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.91
Clash score	2.22
Average B-factor	34.03
macromolecules	33.87
ligands	30.68
solvent	36.45
Number of TLS groups	12

Table S2. Molecular dynamics simulations of GH109 disaccharide complexes of *Akkermansia muciniphila*.

Simulation	*Am*GH109A-No Ligand	*Am*GH109A-a13	*Am*GH109A-b13
PDB ID	6T2B	6T2B	6T2B
Ligand	-	a13	b13
Atoms system	53553	62241	62241
Atoms protein	6837	6837	6837
Atoms NAD^+	70	70	70
Atoms ligand	-	51	51
Number water	15548	18427	18427
Number Ions	2	2	2

a13: D-Gal*p*NAc-α(1-3)- D-Gal*p*-β-OH, **b13**: D-Gal*p*NAc-β(1-3)- D-Gal*p*-β-OH