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**

Teze, David; Shuoker, Bashar; Chaberski, Evan Kirk; Kunstmann, Sonja; Fredslund, Folmer; Nielsen, Tine Sofie; Stender, Emil G. P.; Peters, Günther H.J.; Nordberg Karlsson, Eva ; Welner, Ditte Hededam

Total number of authors:
11

Published in:
A C S Catalysis

Link to article, DOI:
10.26434/chemrxiv.9989102.v1
10.1021/acscatal.9b04474

Publication date:
2020

Document Version
Peer reviewed version

Article

The Catalytic Acid-Base in GH109 Resides in a Conserved GGHGG Loop and Allows for Comparable #-Retaining and #-Inverting Activity in an *N*-Acetylgalactosaminidase from *Akkermansia muciniphila*

David Teze, Bashar Shuoker, Evan Kirk Chaberski, Sonja Kunstmann, Folmer Fredslund, Tine Sofie Nielsen, Emil G. P. Stender, Günther H. J. Peters, Eva Nordberg Karlsson, Ditte Hededam Welner, and Maher Abou Hachem

ACS Catal., **Just Accepted Manuscript** • DOI: 10.1021/acscatal.9b04474 • Publication Date (Web): 11 Feb 2020

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,[‡] Tine Sofie Nielsen[†], Emil G.P. Stender[†], 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. Here we describe the *N*-acetylgalactosaminidases *Am*GH109A and *Am*GH109B from the human gut symbiont *Akkermansia muciniphila*. Notably, *Am*GH109A displays α-retaining and β-inverting *N*-acetylgalactosaminidase activities with comparable efficiencies on natural disaccharides. This dual specificity 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, GH4, human gut microbiota, inverting, mechanism, MD simulations, mucin, retaining.

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-glycosylated (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 the adaptive immune response,[14] consistent with the intimate association and cross-talk between this symbiont and the human host. A marked decrease in the abundance of *A. muciniphila* has been shown in inflammatory bowel disease and ulcerative colitis patients, correlating with an overall increase in the total mucosa associated bacteria, especially *Ruminococcus torques* and *Ruminococcus gnavus*.[15] Accordingly, the perturbation of the intricate balance of the mucolytic community (and thereby mucin homeostasis) is correlated to inflammation. For example, the presence of *A. muciniphila* has been shown to exacerbate gut inflammation induced by *Salmonella enterica* subsp. *enterica* serovar Typhimurium in gnotobiotic mice harbouring an eight membered gut microbiota mock community.[16] To date, insight into the enzymatic apparatus that confers the growth of *A. muciniphila* on host glycans at the mucin barrier remains limited. The genome of *A. muciniphila*[17] encodes a substantial battery of Carbohydrate Active enZymes (CAZymes, http://www.cazy.org)[18] that mostly targets host-derived glycoconjugates, thereby supporting the ecological specialization of this bacterium.[19] Both α-glycosidic linkages, *e.g.* fucosyl or sialyl and non-reducing ends of the A and B blood group antigens and β-linkages are commonly present in host-derived glycans.[19-20]

Enzymes within a glycoside hydrolase (GH) family share a common structural fold, substrate stereoselectivity and catalytic mechanism.[18] 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,[21]

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.[22] By contrast, GH4 employs a non-canonical NAD^+- and Mn^{2+}-dependent mechanism and harbours members that are either α-retaining[23] or β-retaining.[24-25] Enzymes of GH109 also display a redox mechanism assisted by a NAD^+ cofactor, akin to GH4, but lacking the Mn^{2+} cofactor, required in GH4 for the formation of an alkoxide at position 3.[26-27] 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.[26] Currently, GH109 enzymes are described as 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 galactosamine (GalNAc) units.[27-28] A single enzyme from this family is both kinetically and structurally characterized, namely the enzyme from *Elizabethkingia meningosepticum* (NagA).[26-27, 29] 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).[27] 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 potentially target GalNAc units present in a variety of host-derived glycans, and 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 $s^{-1}\cdot mM^{-1}$). *Am*GH109A displayed similar, relevant activities towards non-activated α- and β-GalNAc(1→3)Gal disaccharides, which has not been reported to date. The second enzyme *Am*GH109B displayed about 18.4-fold lower efficiency on β-*p*NPGalNAc as compared to its 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 β-inverting and α-retaining activities at similar levels. The kinetic parameters of GH109 have firstly been reported for 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.[27] The recombinant *A. muciniphila* GH109 enzymes *Am*GH109A and *Am*GH109B were expressed in soluble form and possessed relatively high thermal stabilities with unfolding temperatures (T_m) of 51.6 and 63.4 °C, respectively (Figure S1). The addition of 1 mM NAD^+ had only a minor effect on the T_m of both

enzymes suggesting that the enzymes were already saturated by the cofactor that is likely to be bound with high affinity (Figure S2). This was consistent with the modest effect of added NAD^+ on activity of both enzymes (Figure S3). The pH profiles for both enzymes were examined towards β-*p*NPGalNAc and the highest activity was observed at pH 6.6 and 6.9 for *Am*GH109A and *Am*GH109B, respectively (Figure S4). The pH profiles of both these enzymes appeared broader than the counterpart for NagA that is mainly active between pH 7-8.[29]

Kinetic analyses were carried out on *Am*GH109A and *Am*GH109B towards aryl glycoside 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 glycoside compared to the α-counterpart, 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*[27]

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*-nitrophenol (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 would lead to an inversion of stereochemistry and the release of α-GalNAc from β-*p*NPGalNAc. Indeed, only signals with chemical shifts corresponding to α-GalNAc (δ = 5.14 ppm) were observed immediately upon enzyme addition, while signals corresponding to the β-anomer (δ = 4.58 ppm) appeared later due to mutarotation. Moreover, the anomeric signals appear as pseudo-singlets stemming from the D/H exchange on C-2 (Figure 1c,d).[29]

Non-activated substrates are, however, 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, as evident from the time-course NMR monitoring of the hydrolysis reactions of the disaccharides (Figure 1c,d).

Conservative estimates from the initial rates against 2.5 mM substrate yield 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. Catalytic efficiencies, but not k_{cat} values, of NagA and three other additional GH109 enzymes have been recently reported on methylumbelliferyl derivatives of A-antigens.[28] The only other GH family adopting a related mechanism is GH4, which 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 dictate slower rates than in classical (*i.e.* non-oxidative) GHs.[30] Importantly, these experiments were consistent with the presence of an acid/base catalyst, which has not been previously identified in either family.

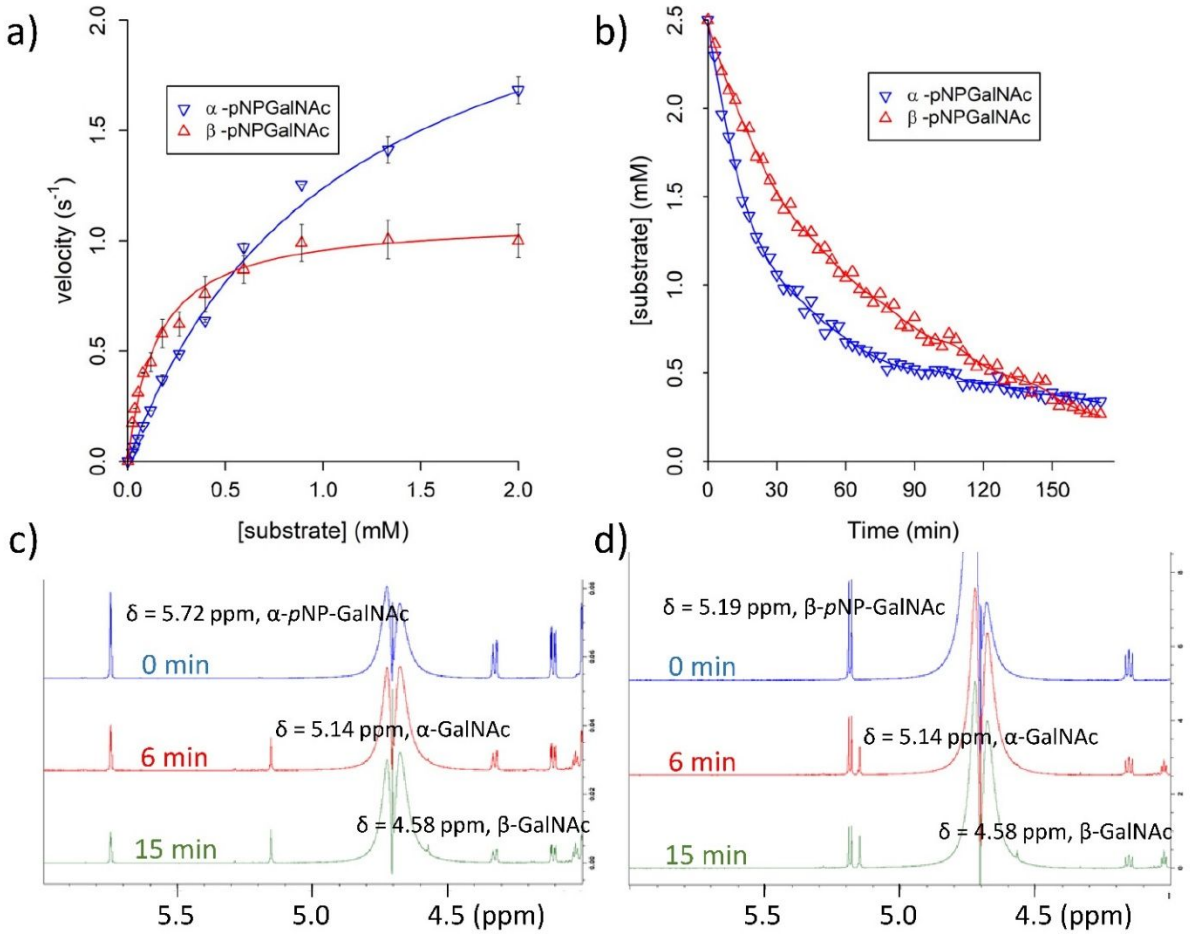


Figure 1. The dual activity of *Am*GH109A on substrates with a terminal non-reducing α- or β-GalNAc. **(a)** Michaelis-Menten plot of *p*NPGalNAc hydrolysis. The data markers are the mean of three replicates with standard deviations and the solid lines in are the fits of the Michaelis-Menten expression 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 298 K, in 20 mM HEPES pD 6.6. The data that could not be modelled with a Michaelis-Menten model and the solid lines are not fits to that model. **(c)** and **(d)** NMR monitoring of the hydrolysis by 15 $\mu g \cdot mL^{-1}$ *Am*GH109A of 2 mM α-*p*NPGalNAc and β-*p*NPGalNAc, respectively. Reactions were performed at 298 K, in 20 mM phosphate 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 the highest resolution NagA model (PDB: 2IXA, 36.5% sequence identity) and refined to 2.13 Å resolution (Table S1). Four *Am*GH109A molecules are observed in the asymmetric unit, each with a GalNAc and an NAD^+ molecule bound in the active site. Analysis of the structure using the PISA server[31] indicates that *Am*GH109A forms a homodimer, which is organized in the same fashion as the NagA dimer[24] (Figure S5a).

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.[27] 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 (Figure S5b,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, Figure S5b,c). The GalNAc binding site in *Am*GH109A is similar to that of NagA with hydrogen bonds to Y226, R244, Y256, H259, Y339, equivalent to Y179, R213, Y225, H228, Y307 in NagA (Figure 3a, Figure S5d-e).

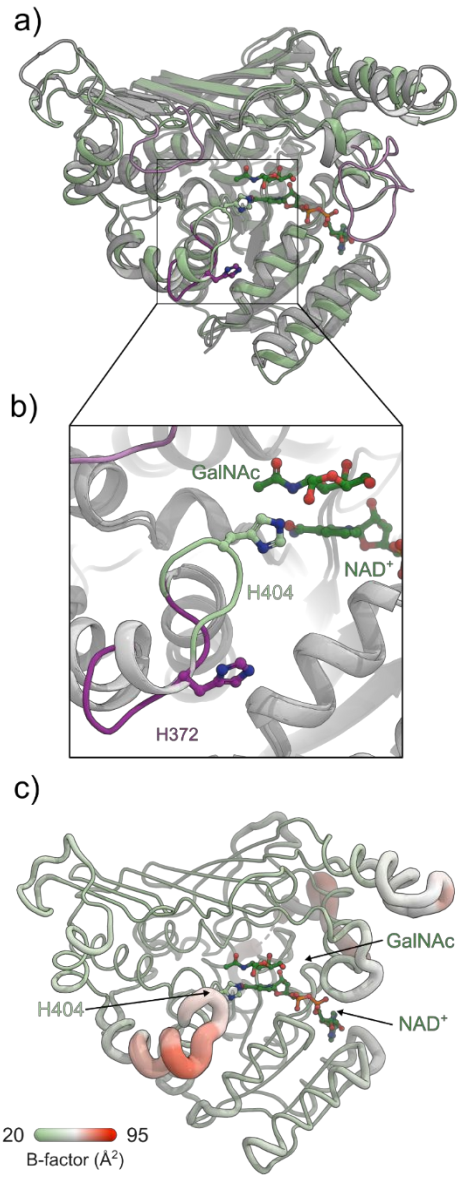


Figure 2. Structural analysis of *Am*GH109A. **(a)** The overall structure of *Am*GH109A (PDB: 6T2B, green) superimposed on NagA in complex with GalNAc (PDB: 2IXB, grey). The positions of the GalNAc in both structures are identical, but only the one from *Am*GH109A is shown for clarity. The main differences 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 histidine residue in GH109 (*Am*GH109A H404 and NagA H372) within a GGHGG conserved motif. This loop adopts a conformation away from the active site in NagA (magenta), whereas it is located in the active site of *Am*GH109A, positioning the conserved H404 at a hydrogen-bonding distance from the anomeric C1-OH group of the subsite –1 bound GalNAc. **(b)** A close-up showing the loop described in (a) adopting two different conformations in NagA (magenta) and in *Am*GH109A (green). The conserved histidine within this loop is shown in sticks. The polar contact of this histidine with the C1-OH of GalNAc unveiled this residue as a candidate for the acid/base catalyst, which has not been identified in GH109. **(c)** The structure of *Am*GH109A scaled to the value of the B-factor (Pymol B-factor putty representation), which indicated the flexibility of the loop encompassing the conserved GGHGG motif.

One striking difference between NagA and *Am*GH109A, is that the GalNAc α-anomeric oxygen is at a hydrogen bonding distance to H404 in *Am*GH109A, while the corresponding residue (H372)

is over 12 Å away in NagA (Figure 2a,b, Figure S5b-d). These histidines are located within highly flexible, glycine-rich loops (GGHGG in *Am*GH109A and GAGHGG in NagA). The computed pK_a (6.2)[32] and the hydrogen-bond to the anomeric proton in *Am*GH109A (Figure 3a) highlight H404 as a plausible catalytic acid-base candidate in the mechanism. 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 (Supporting Information Files 2 and 3). An additional glycine preceding this motif is also present in more than 80% of the sequences, as observed for *Am*GH109A and *Am*GH109B, but not in NagA. Given that GH109 and GH4 members share a NAD^+-dependent mechanism and harbour both α- and β-active enzymes, we also performed a comparative structural analysis of these two GH families. The active site region in GH4 displays a variable architecture, with a distinct clade resembling GH109 (Figure S6 and related discussion).

H404 is the catalytic acid/base in GH109. To assess the role of H404 in the catalytic mechanism of GH109 enzymes, we produced and characterized the *Am*GH109A single mutants H404A and H404F. 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, Figure S7).

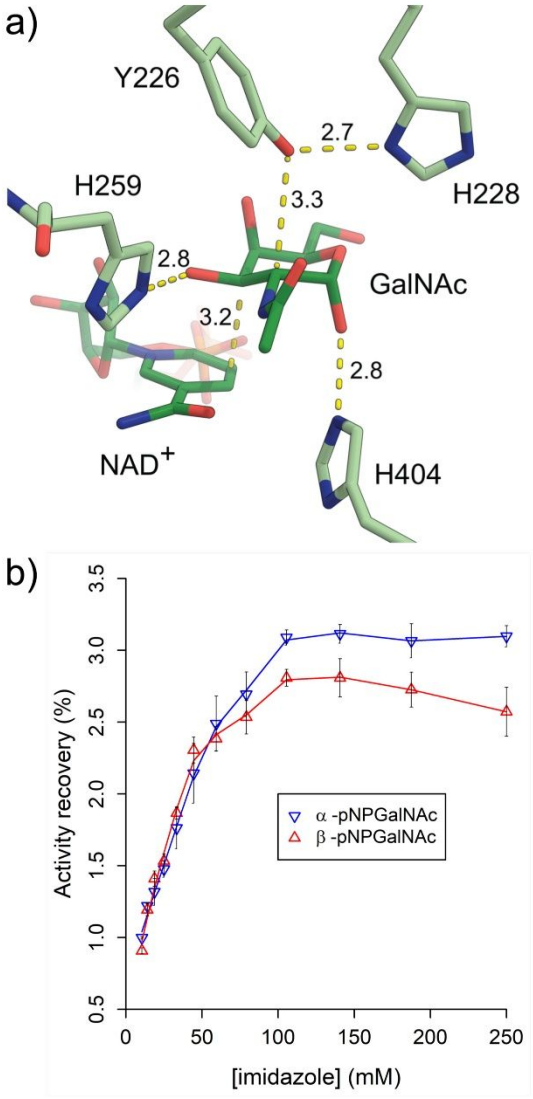
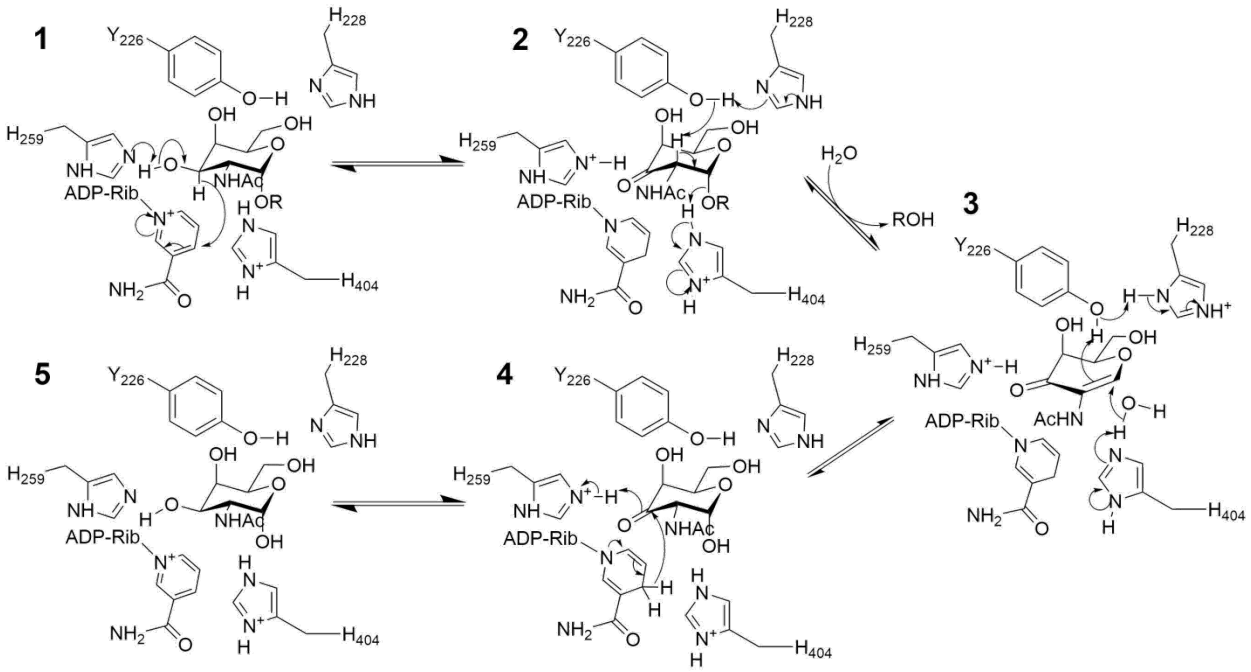


Figure 3. A conserved histidine corresponding to *Am*GH109A H404 is the acid/base catalyst in the mechanism of GH109 enzymes. **(a)** The catalytic residues Y226, H228, H259 and H404, as well as the GalNAc and the NAD^+ cofactor are shown in sticks. The distances between atoms that will exchange protons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

Neither the stability nor the NAD^+-binding affinity of this mutant were markedly affected compared to *Am*GH109A (Figure S1-S2). The *Am*GH109A-H404A activity in the presence of 250 mM imidazole was unaffected by the addition of up to 2.5 mM NAD^+ (Figure S3). Moreover, our NMR analyses confirmed that imidazole acts as an exogenous acid/base catalysts and that the product of the rescue reaction is α-GalNAc (Figure S7).

Taken together, the conservation of the GGHGG motif, the relevant hydrogen bonding distance from the anomeric carbon in the *Am*GH109A structure and the mutational data establish that H404

is the acid/base catalyst in GH109, allowing us to complete the α-retaining mechanism previously proposed[26] for this enzyme family (Scheme 1).



Scheme 1. Proposed revised α-retaining mechanism for GH109 enzymes, residues numbered according to the *Am*GH109A sequence.

Molecular dynamics simulations support the flexibility of the loop bearing the histidine catalytic acid/base in *Am*GH109A and NagA. The molecular determinants of the dual α/β specificity were further explored by molecular dynamics (MD) simulations. The *Am*GH109A structure was simulated in free form and in complex with either α- or β-GalNAc(1→3)Gal disaccharide substrates, 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 with the protonation state suggested for the first step of the mechanisms (protonated H404, deprotonated H259 and H228, see Schemes 1 and 2, Step 1; Figure S8a). The four distances represented in Figure 4a,b between H3' and the NAD^+, H2' and the phenolic oxygen of Y226, the H259 $N\varepsilon_2$ and HO3', and between the H404 $N\varepsilon_2$ and O1' were monitored along the simulations, as proton or hydride exchanges between these atoms are crucial for the catalysed reactions (Figure 4c,d).

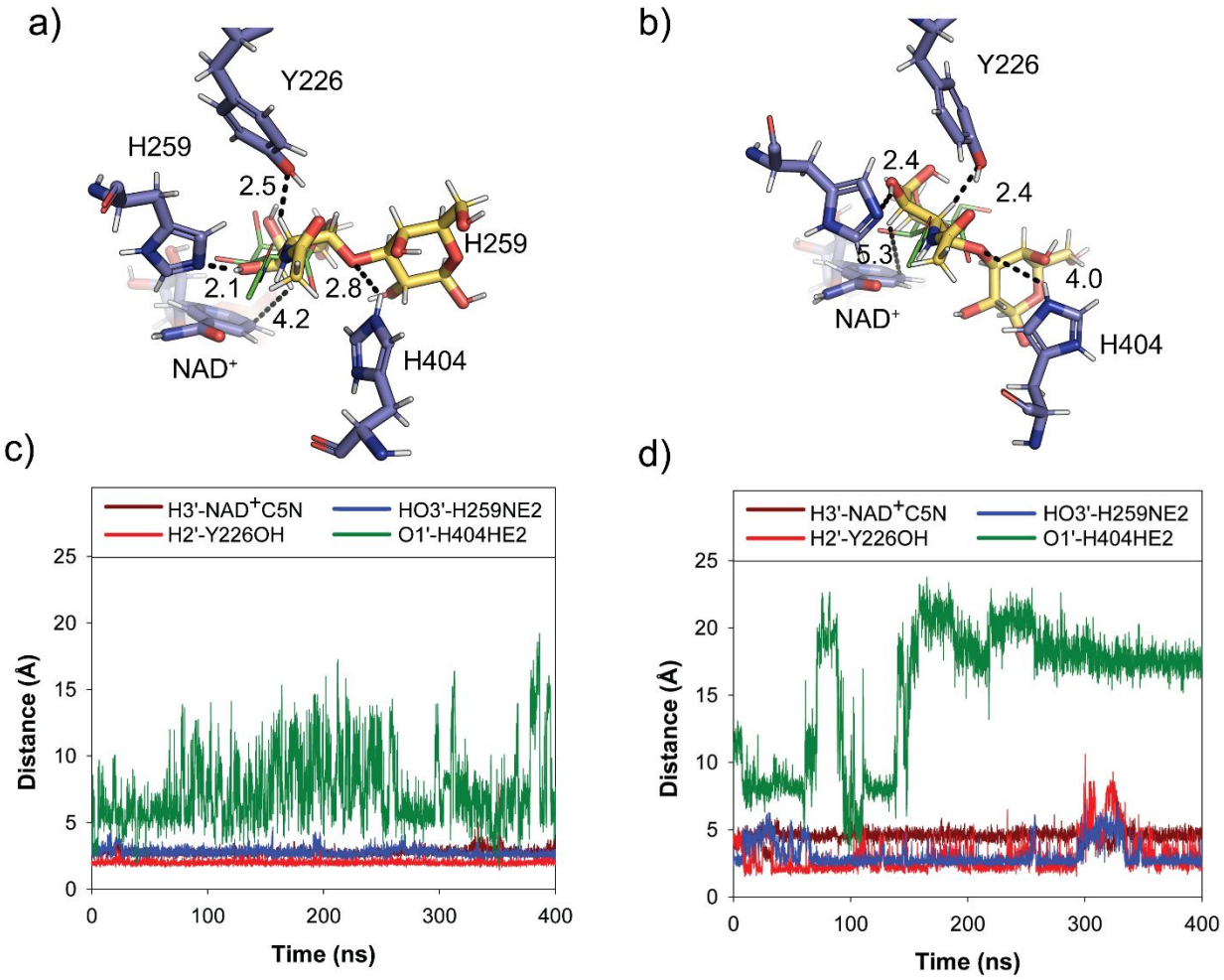
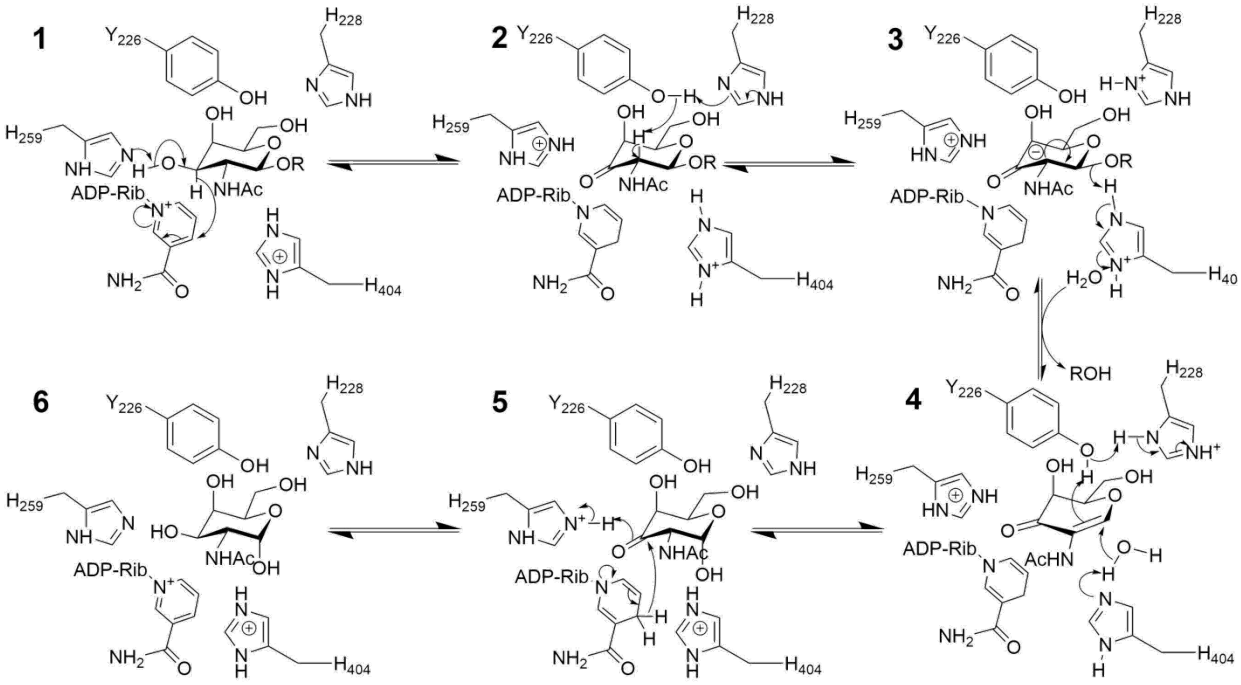


Figure 4. Molecular dynamics simulations of *Am*GH109A bound to the α- and β-GalNAc(1→3)Gal disaccharides. **(a)**, **(b)** are the α- and β-disaccharides, respectively, catalytic residues Y226, H259, H404 and NAD^+ cofactor are shown in blue sticks for the closest conformation to the ligand (yellow sticks) (frames 346.4 ns and 99.6 ns, respectively). Distances between atoms that will exchange protons or hydride along the catalytic cycle are shown as dotted lines and the crystallographic GalNAc is shown as green sticks. **(c)** and **(d)** Evolution of the distances between

H3' and the NAD^+-C5N, H2' and the phenolic oxygen of Y226-OH, HO3' and H259-Nε_2 as well as between the O1' and H404 Hε_2 along the simulations for the α- and β-disaccharide, respectively.

Importantly, H404 $N_{\varepsilon 2}$ is located within 3.0–4.5 Å of O1' for both the β- and α-disaccharides for 20 ns or in most of the 400 ns simulation, respectively, supporting that H404 can occupy catalytically competent conformations as acid catalyst for both α- and β- substrates (Figure 4c,d). This led us to propose H404 as the acid catalyst for the β-inverting mechanism through an E1cB elimination (Scheme 2).



Scheme 2. The proposed β-inverting E1cB mechanism in GH109 based on *Am*GH109A residue numeration.

The k_{cat} values of *Am*GH109A are in the same range for the four assayed substrates (0.5−2.5 s^{-1}), despite large differences in both leaving groups and stereochemistries. Thus, it is tempting to propose that the rate-limiting step takes place after the formation of the glycal-ketone intermediate, and is either the hydration (Scheme 2, step 4) or the reduction reaction (Scheme 2, step 5). However, this cannot be generalised to the GH109 family, as both *Am*GH109B and NagA display strongly different k_{cat} depending on the stereochemistry, which likely indicates that the E1cB mechanism is the rate limiting step for the β-*p*NPGalNAc in these enzymes. By analyzing rates of various α-aryl-GalNAc, Chakladar et al. proposed that the hydration step (Scheme 1, step 3) might be rate-limiting for the hydrolysis of these compounds by NagA.[29] Simulations with histidine protonation states corresponding to the reaction step 3 were also performed, resulting in a lower loop flexibility and an increase in the α-disaccharide flexibility (Figures S8-S9 and related discussion).

The MD simulations of NagA showed that the corresponding loop that harbours the identified catalytic residue is also flexible, moving by over 12 Å to reach catalytically competent conformations of the conserved histidine (H372) towards the β-disaccharide (Figure S10a,b and

related discussion). These findings strongly support the mechanistic importance of the flexibility of this loop and its role as harbouring the His catalytic acid/base in GH109, rather than being a peculiarity of *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 stereochemistries, but at very low levels ($k_{cat} \approx 10^{-3}$ min^{-1}) and against activated substrate analogues (*p*NP-α-D-Glc-6-phosphate and *p*NP-β-D-Glc-6-phosphate).[33]

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 would 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 epithelial enterocytes. It may also be a more generic feature shared by other GH109 enzymes.

■ 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,[34] were amplified from *Akkermansia muciniphila* ATCC BAA-835 (DSM 22959) genomic DNA using the primers (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 ℃ 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 ℃. The cells were harvested by centrifugation (10000 *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 a STANSTED high pressure homogenizer (SPCH-1, Homogenising Systems, Essex, UK), followed by incubation for 30 min on ice with 5 μL benzonase nuclease (Sigma). The lysate was centrifuged for 20 min (45000 *g*, 4 °C) and the supernatant was filtered (0.45 μm) and loaded onto a HisTrap HP column (GE Healthcare, Uppsala, Sweden). 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 $\varepsilon_{280\ nm}$ 70860 M^{-1}·cm^{-1} and 91330 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.

Enzyme stability. *Am*GH109A, AmGH109A H404A and *Am*GH109B and were dialysed against 3x100 volumes of 50 mM Na_2HPO_4/NaH_2PO_4 pH 6.6. The unfolding temperature of (T_m) was determined using a Prometheus NT.48 NanoDSF (Nano Temper, Germany). The fluorescent emission intensity at λ = 330 nM and λ = 350 nM and scattering were recorded with an excitation at λ = 280 nM at 10 % intensity. The samples were scanned at 1 °C·min^{-1}. The T_m was determined from the peak maximum of the first derivative of the fraction $\frac{\text{fluorescent intensity at 350 nm}}{\text{fluorescent intensity at 330 nm}}$ using PR.stabilityAnalysis (Nano Temper, Germany). NAD^+ dissolved in the same buffer as above was added to a final concentration of 1 mM 20 min prior to analysis. The experiments were performed in triplicates, and the data were reported as means with standard deviations.

Enzymatic analyses. The pH-activity profiles of the *Am*GH109A and *Am*GH109B (both at 100 nM) were determined in Britton-Robinson universal buffer adjusted to 0.1 M ionic strength in the pH range 5.0−9.1 (12 pH values) in 150 μL reaction towards 2 mM β-*p*NPGalNAc, which was added to initiate the reactions. The experiments were performed in triplicates using a PowerWave XS microtiter plate reader monitored by the Gen5 software (Bio-Tek Instruments, Inc.) by monitoring absorbance at 405 nm (A_{405}) every 40 s for 10 min at 25 °C. The absorbance values were corrected at the different pH values by calculating the phenolate concentration using the phenol pK_a=7.24. Spectrophotometry-based kinetics experiments were performed similarly, but the absorbance at 405 nm (A_{405}) was monitored every 20 s for 20 min at 25 °C. The reactions

(200 µL), which were carried out in 96-wells, contained (α/β)-*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. Chemical rescue experiments were performed on *Am*GH109A H404A in the same buffer systems as above in the presence of 10−250 mM imidazole as an exogenous acid/base. 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. To assess the effect of NAD^+ additions, the activity of *Am*GH109A (100 nM) towards 1.2 mM β-*p*NPGalNAc was measured in the presence or absence of either 0.5 mM NAD^+ in a similar assay as above, but with a reaction volume of 150 µL. The activity of *Am*GH109A H404A (5 µM) was measured similarly in the presence of 250 mM imidazole similar to the concentration used in the chemical rescue experiments in the presence and absence of 0.5 mM or 2.5 mM NAD^+. The initial rates were determined from the slopes of the linear part of the progress curves. The experiments were performed in triplicates and the data are reported as means with standard deviations.

NMR analyses. 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 at 298 K. 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)[35] 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). Stereochemical outcomes were measured in a 50 mM deuterated HEPES buffer pD=6.6 in presence of 2 mM *p*NPGalNAc and 15 $\mu g \cdot mL^{-1}$ (*Am*GH109A) or 2 $mg \cdot mL^{-1}$ (*Am*GH109A-H404A) enzyme.

Bioinformatics. Sequences were retrieved from the non-redundant protein database,[36] 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.[37] Iterative multiple sequence alignments using Clustal Ω[38] were performed to increase the minimum pairwise sequence identity by 5 % increments until convergence is reached and all sequences share ⩾20% pairwise sequence identity using a previously described script.[39]

Crystallography. A stock solution of *Am*GH109A (6 $mg \cdot mL^{-1}$) 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[40] to a resolution of 2.13 Å using autoPROC's default cutoff criteria. Data were phased with molecular replacement using Phaser[41] in the Phenix software package,[42] with the *N*-acetylgalactosaminidase from *Elizabethkingia meningoseptica* (PDB: 2IXA) as search model. An initial model was built using Phenix.autobuild[42] and completed with alternating manual rebuilding in Coot[43] and automatic Phenix.refine refinement.[44] Ligands were placed with Phenix.LigandFit.[45]

Molecular dynamics simulations. The structure of *Am*GH109A (PDB: 6T2B) and NagA (PDB: 2IXB) was protonated in the H++ Server with pH 7.0 and 0.15 M NaCl ionic strength, catalytic residue protonation was adjusted according to the mechanism of step 1[46-48] and the proteins were parametrized with the AMBER14 forcefield for proteins.[49] The *Am*GH109A chain B was chosen as the most complete structure. Parameters for NAD^+ were generated based on published data.[50-51] Disaccharides D-Gal*p*NAc-α(1-3)-D-Gal*p*-β-OH (a13) and D-GalNAc-β(1-3)-D-Gal-β-OH (b13) were parametrized and minimized with GLYCAM06[52] and superimposed in the protein crystal structures with the solved D-GalNAc (Table S2). The simulation complex was placed in an almost cubic box with the dimensions ≈85x83x84 Å.[3] Simulations were run with the TIP3P water model[53] and charges were equalized with the *gmx genion* routine.[54] MD simulations were carried out using the GROMACS2018 program package.[54-55] After energy minimization to a maximum force smaller than 100 $kJ \cdot mol^{-1} \cdot nm^{-1}$ (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[56] (reference pressure of 1.0 bar and coupling time constant of 1 ps) and a V-rescale thermostat[57] (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.[58] 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.[54-55] The details of the simulations and the protonation states are enlisted in Tables S2-S3.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.
Supporting Information File 1: Enzyme stability data, activity dependence of NAD^+, pH activity profiles, structural comparison of GH109 and between GH109 and GH4, NMR analysis of chemical rescue, Molecular dyanmics and structure refinement and data collection.
Supporting Information Files 2 and 3: multiple sequence alignment and the sequence conservation of GH109.

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 analyses 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 in collaboration with GP. TSN produced and purified the *A. muciniphila* GH109 enzymes for the revision, performed the pH dependence profiles and the NAD^+ dependence experiments. EGPS performed the stability analyses of the enzymes. 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 the staff of the synchrotron MAX IV for technical assistance. The Prometheus NT.48 is funded by a Novo Nordisk Foundation grant (NNFSA170028392) to DTU-Bioengineering to recruit Professor Alexander K. Buell.

■ 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), 1027-1031.

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. Png, C. W.; Linden, S. K.; Gilshenan, K. S.; Zoetendal, E. G.; McSweeney, C. S.; Sly, L. I.; McGuckin, M. A.; Florin, T. H. J., Mucolytic Bacteria with Increased Prevalence in IBD Mucosa Augment *In Vitro* Utilization of Mucin by Other Bacteria. *Am. J. Gastroenterol.* **2010,** *105* (11), 2420-2428.

16. Ganesh, B. P.; Klopfleisch, R.; Loh, G.; Blaut, M., Commensal *Akkermansia muciniphila* Exacerbates Gut Inflammation in *Salmonella typhimurium*-Infected Gnotobiotic Mice. *PLoS One* **2013,** *8* (9), e74963.

17. 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.

18. 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.

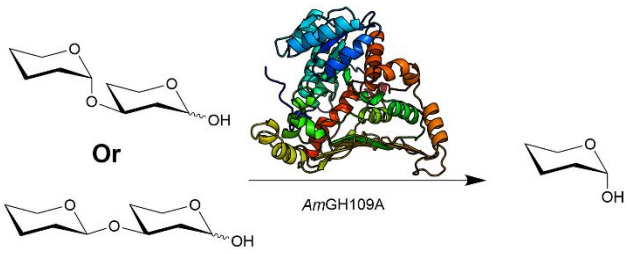
19. Tailford, L. E.; Crost, E. H.; Kavanaugh, D.; Juge, N., Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **2015,** *6*, 81.

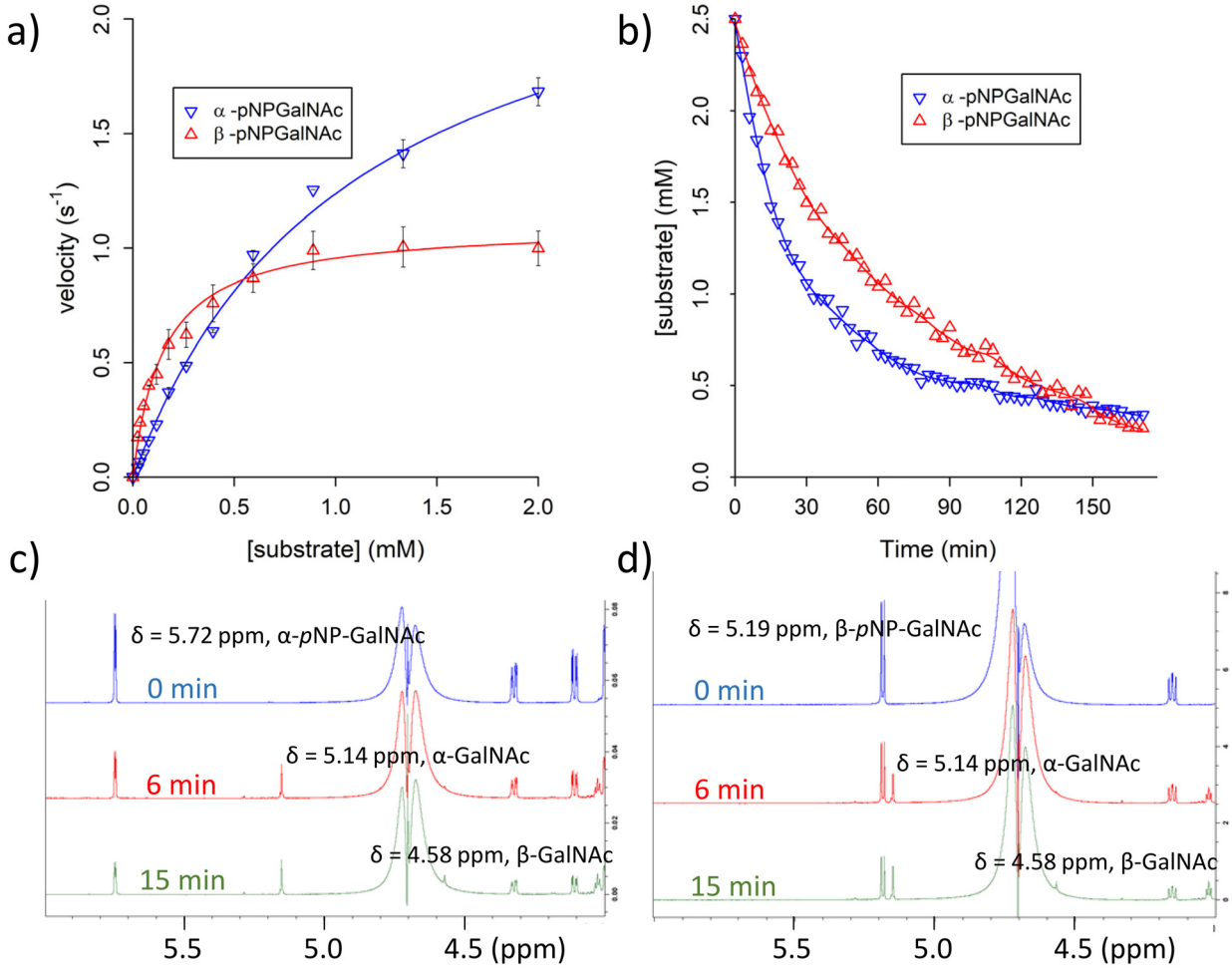
20. Holmen Larsson, J. M.; Thomsson, K. A.; Rodriguez-Pineiro, A. M.; Karlsson, H.; Hansson, G. C., Studies of Mucus in Mouse Stomach, Small Intestine, and Colon. III. Gastrointestinal Muc5ac and Muc2 Mucin *O*-Glycan Patterns Reveal a Regiospecific Distribution. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2013,** *305* (5), G357-363.
21. 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.
22. 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), 1058-1067.
23. Yip, V. L.; Thompson, J.; Withers, S. G., Mechanism of GlvA from *Bacillus subtilis*: A Detailed Kinetic Analysis of A 6-Phospho-A-Glucosidase from Glycoside Hydrolase Family 4. *Biochemistry* **2007,** *46* (34), 9840-9852.
24. 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.
25. 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.
26. 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.
27. 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), 454-464.
28. 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,** *4* (9), 1475-1485.
29. 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.
30. 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), 2487-2497.
31. Krissinel, E.; Henrick, K., Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **2007,** *372* (3), 774-797.
32. 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.
33. 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.
34. 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.
35. 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.

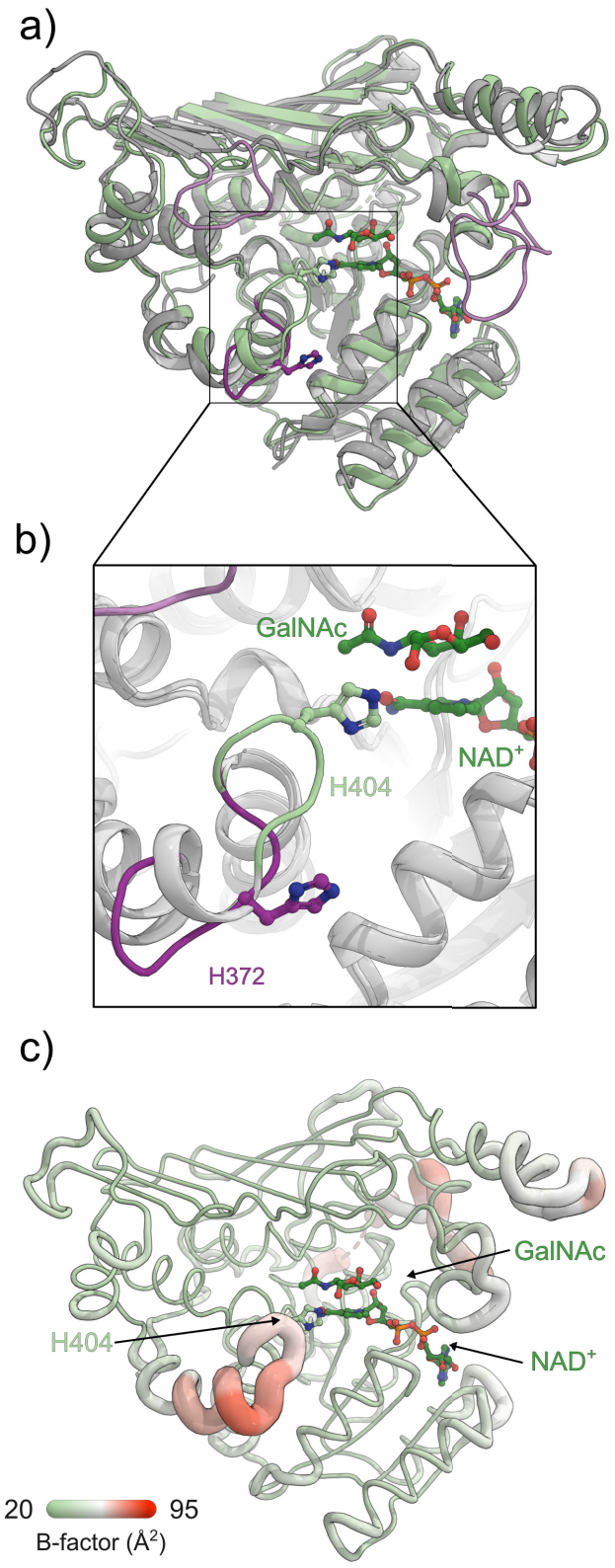
36. Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S., UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* **2004,** *32* (Database issue), D115-9.
37. 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), 680-682.
38. Sievers, F.; Higgins, D. G., Clustal Omega. *Curr. Protoc. Bioinformatics* **2014,** *48* (1), 3.13.1-3.13.16.
39. Teze, D.; Jiao, Z.; Wiemann, M.; Gulshan Kazi, Z.; Lupo, R.; Rønne, M. E.; Carlström, G.; Duus, J. Ø.; Sanejouand, Y.-H.; O'Donohue, M. J.; Nordberg Karlsson, E.; Régis, F.; Stålbrand, H.; Svensson, B., Rational Enzyme Design Without Structural Knowledge: A Sequence-Based Approach for Effi-Cient Generation of Glycosylation Catalysts. *ChemRxiv.* **2020,** *Preprint* (https://doi.org/10.26434/chemrxiv.11538708.v1)
40. 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.
41. 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.
42. 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.
43. 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.
44. 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), 1948-1954.
45. 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.
46. 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), W537-541.
47. 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), W368-371.
48. 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.
49. 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), 3696-3713.
50. 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.
51. 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.

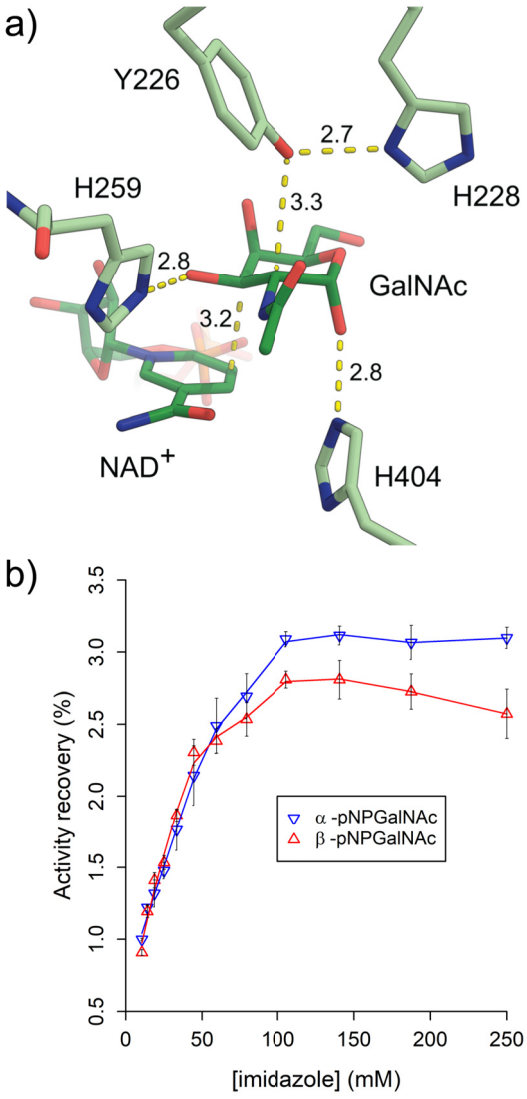
52. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008,** *29* (4), 622-655.
53. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions For Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.
54. 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.
55. 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.
56. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
57. Bussi, G.; Donadio, D.; Parrinello, M., Canonical Sampling Through Velocity Rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.
58. 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.

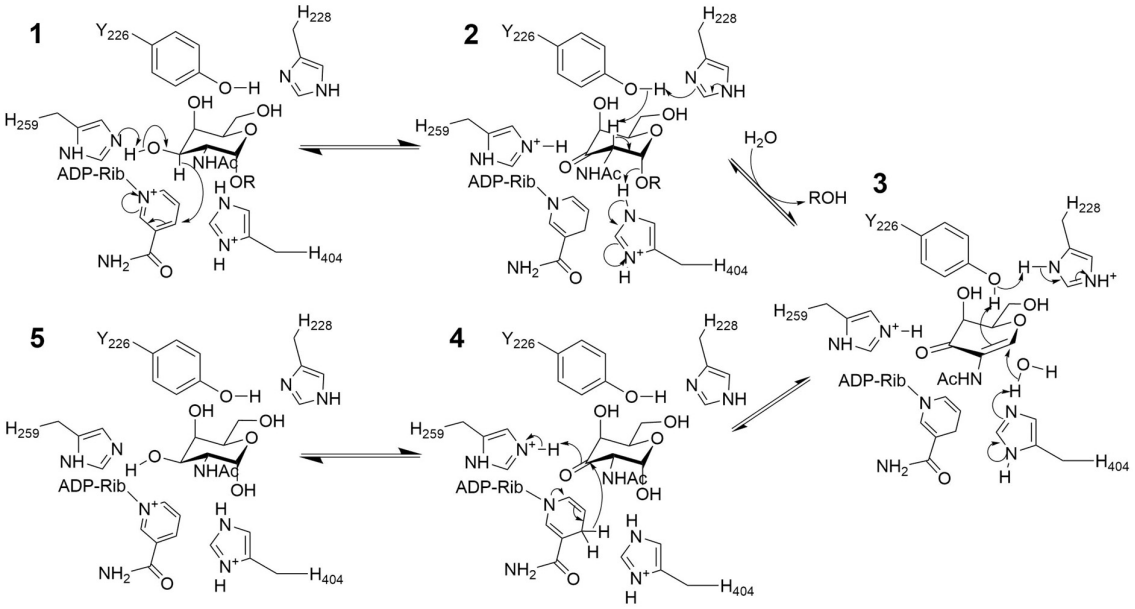
For Table of Contents Only

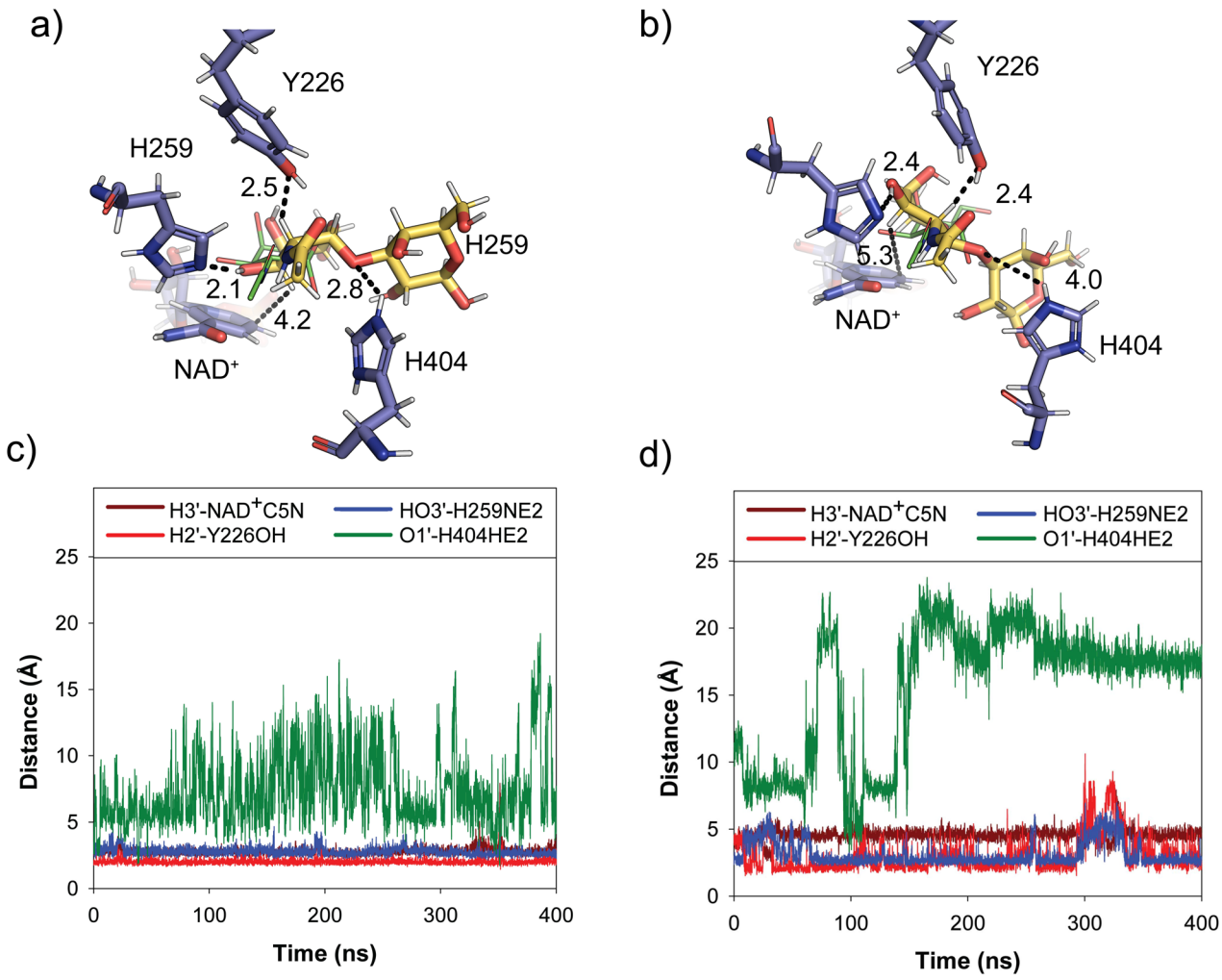


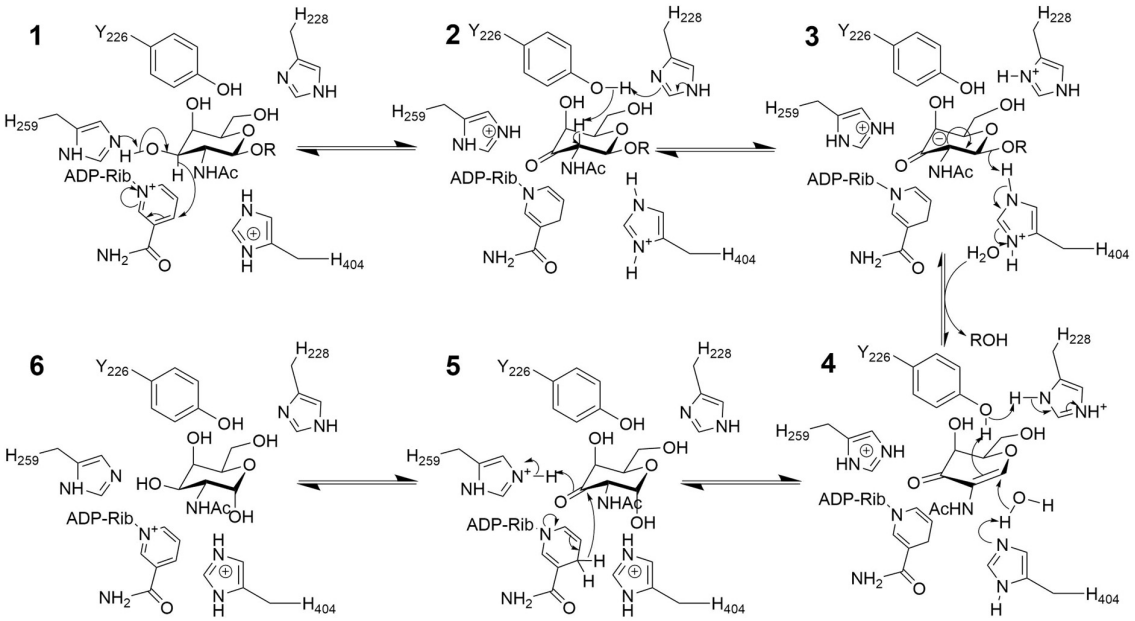

a)
α -pNPGalNAc
β -pNPGalNAc
velocity (s^{-1})
[substrate] (mM)
b)
α -pNPGalNAc
β -pNPGalNAc
[substrate] (mM)
Time (min)
c)
δ = 5.72 ppm, α-pNP-GalNAc
0 min
δ = 5.14 ppm, α-GalNAc
6 min
δ = 4.58 ppm, β-GalNAc
15 min
5.5
5.0
4.5 (ppm)
d)
δ = 5.19 ppm, β-pNP-GalNAc
0 min
δ = 5.14 ppm, α-GalNAc
6 min
δ = 4.58 ppm, β-GalNAc
15 min
5.5
5.0
4.5 (ppm)

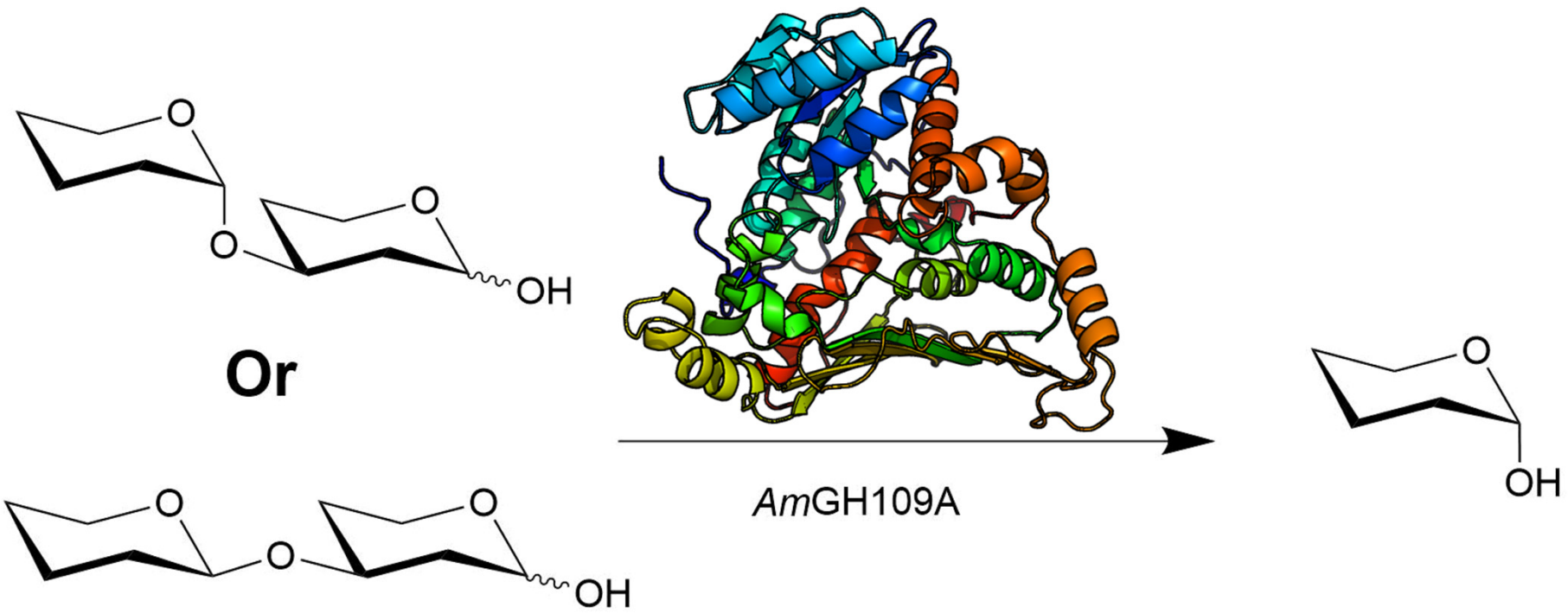

a)
b)
GalNAc
NAD+
H404
H372
c)
GalNAc
H404
NAD+
20
95
B-factor (Å2)


a)
Y226
2.7
H228
H259
3.3
2.8
GalNAc
3.2
2.8
NAD+
H404
b)
Activity recovery (%)
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
50
100
150
200
250
[imidazole] (mM)
α -pNPGalNAc
β -pNPGalNAc


1
2
3
4
5
Y226
H228
H259
H404
ADP-Rib
NHAc
AcHN
OR
ROH
H2O
NH2


a)
Y226
H259
2.5
H259
2.1
2.8
4.2
NAD+
H404
b)
Y226
2.4
2.4
5.3
4.0
NAD+
H404
c)
H3'-NAD+C5N
HO3'-H259NE2
H2'-Y226OH
O1'-H404HE2
Distance (Å)
Time (ns)
d)
H3'-NAD+C5N
HO3'-H259NE2
H2'-Y226OH
O1'-H404HE2
Distance (Å)
Time (ns)


1
2
3
4
5
6
ROH
ADP-Rib
NHAc
AcHN


O
O
O
OH
Or
O
O
O
OH
AmGH109A
O
OH