Structural and mechanistic insights into the substrate specificity and hydrolysis of GH31 α-N-acetylgalactosaminidase

Biochimie 195 (2022) 90–99

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Structural and mechanistic insights into the substrate specificity and hydrolysis of GH31 α-*N*-acetylgalactosaminidase

Takatsugu Miyazaki [a, b, *], Marina Ikegaya [b], Santiago Alonso-Gil [c, **]

[a] Research Institute of Green Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan
[b] Department of Bioscience, Graduate School of Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan
[c] Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Vienna, Austria

ARTICLE INFO

Article history:
Received 8 September 2021
Received in revised form
15 November 2021
Accepted 22 November 2021
Available online 23 November 2021

Keywords:
Gut bacteria
Mucin
N-acetylgalactosamine
O-glycan
Quantum mechanics

ABSTRACT

Glycoside hydrolase family 31 (GH31) is a diversified family of anomer-retaining α-glycoside hydrolases, such as α-glucosidase and α-xylosidase, among others. Recently, GH31 α-*N*-acetylgalactosaminidases (Nag31s) have been identified to hydrolyze the core of mucin-type *O*-glycans and the crystal structure of a gut bacterium *Enterococcus faecalis* Nag31 has been reported. However, the mechanisms of substrate specificity and hydrolysis of Nag31s are not well investigated. Herein, we show that *E. faecalis* Nag31 has the ability to release *N*-acetylgalactosamine (GalNAc) from *O*-glycoproteins, such as fetuin and mucin, but has low activity against Tn antigen. Mutational analysis and crystal structures of the Michaelis complexes reveal that residues of the active site work in concert with their conformational changes to act on only α-*N*-acetylgalactosaminides. Docking simulations using GalNAc-attached peptides suggest that the enzyme mainly recognizes GalNAc and side chains of Ser/Thr, but not strictly other peptide residues. Moreover, quantum mechanics calculations indicate that the enzyme preferred *p*-nitrophenyl α-*N*-acetylgalactosaminide to Tn antigen and that the hydrolysis progresses through a conformational itinerary, $^4C_1 \rightarrow {}^1S_3 \rightarrow {}^4C_1$, in GalNAc of substrates. Our results provide novel insights into the diversification of the sugar recognition and hydrolytic mechanisms of GH31 enzymes.

1. Introduction

Glycoside hydrolases (GHs) are widely distributed in nature and catalyze the hydrolysis of the glycosidic bonds of carbohydrates. These enzymes are involved in various physiological processes, including carbohydrate digestion, cell wall reconstruction, and glycoconjugate synthesis. Based on their amino acid sequence homology and according to the CAZy database (http://www.cazy.org), there are more than 160 GH families [1]. Their catalytic mechanisms are largely divided into anomer-retaining and -inverting mechanisms. These mechanisms usually employ a pair of carboxylate residues, such as Asp or Glu, as catalysts (refer to CAZypedia, http://www.cazypedia.org/ [2]). The retaining hydrolytic mechanism, which is the Koshland double-replacement mechanism, is achieved in two steps (glycosylation and deglycosylation) *via* a covalent glycosyl-enzyme intermediate (GEI) [3]. In contrast, the inverting hydrolytic reaction proceeds through a single-displacement mechanism, such that a nucleophilic water molecule activated by a base catalyst attacks an anomeric carbon in glycosides. Almost all steps in both reaction mechanisms progress through oxocarbenium-ion-like transition states. Stereochemistry of substrate sugars at subsite −1 (subsites were defined according to Davies et al. [4]) are changed in each step during the hydrolytic reaction by GHs. The changes, also referred to as conformational itinerary, vary among substrate sugars and GHs [5]. An understanding of the conformational itineraries of GHs is important, considering that mimics of transition states in each GH catalysis can be utilized as inhibitors and probes [6].

GH family 31 (GH31) is one of the largest families, which comprise GHs that are active on α-glycosidic linkages in carbohydrates. Many of the characterized GH31 enzymes are α-glucosidases originating from prokaryotic and eukaryotic origins [7,8]. GH31 also contains *exo*-acting α-glycosidases, such as mannosyl-oligosaccharide α-1,3-glucosidases [9], α-xylosidases [10,11], α-galactosidases [12], sulfoquinovosidases [13], and α-*N*-

* Corresponding author. Research Institute of Green Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan.
** Corresponding author.
E-mail addresses: miyazaki.takatsugu@shizuoka.ac.jp (T. Miyazaki), santiago.alonso.gil@univie.ac.at (S. Alonso-Gil).

https://doi.org/10.1016/j.biochi.2021.11.007

Abbreviations

BSM	bovine submaxillary mucin
GalNAc	*N*-acetylgalactosamine
GalNAcα-pNP	4-nitrophenyl α-*N*-acetylgalactosaminide
GEI	glycosyl-enzyme intermediate
GH	glycoside hydrolase
GH31	glycoside hydrolase family 31
MC	Michaelis complex
MES	2-(*N*-morpholino)ethanesulfonic acid
Nag 31	GH31 α-*N*-acetylgalactosaminidase
PC	product complex
PCR	Polymerase chain reaction
PDB	Protein Data Bank
pNP	*para*-nitrophenyl
QM	quantum mechanics
TLC	thin-layer chromatography
WT	wild-type

acetylgalactosaminidases [14–16]. Since GH31 enzymes possess the retaining hydrolytic mechanism, some of them can catalyze transglycosylation with several α-glucosidic linkages on acceptor substrates. An *endo*-acting dextranase and cycloalternan-degrading enzyme are also found in GH31 [17–21]. Despite the variety of substrate specificity, the conformational itinerary of sugar ring at subsite −1 in GH31 enzymes was proposed to be $^4C_1 \rightarrow (^4H_3)^{\ddagger} \rightarrow {}^1S_3 \rightarrow (^4H_3)^{\ddagger} \rightarrow {}^4C_1$ by X-ray crystallographic analyses of the Michaelis complexes (MCs) and covalent intermediates in several GH31 enzymes [9–11,13,21,22]. All these enzymes are active on substrates consisting of *gluco*- or *xylo*-configured sugars; however, the conformational itinerary of GH31 enzymes that hydrolyze C4-epimeric *galacto*-configured substrates, i.e., α-galactosidase and α-*N*-acetylgalactosaminidase, are yet to be analyzed.

α-*N*-Acetylgalactosaminidases (EC 3.2.1.49) belonging to GH31 (Nag31s) family of enzymes were initially identified by screening metagenomic DNA libraries from the human gut microbiome [14]. Nag31s from *Bacteroides caccae* and *Phocaeicola plebeius* (formerly *Bacteroides plebeius*) hydrolyzed synthetic substrates, including 4-nitrophenyl α-*N*-acetylgalactosaminide (GalNAcα-pNP), and 4-methylumbelliferyl α-*N*-acetylgalactosaminide and released *N*-acetylgalactosamine (GalNAc) from an *O*-glycoprotein fetuin and *O*-glycopeptide. Nag31s were also identified in the pathogen *Clostridium perfringens* [14], the gut bacterium *Enterococcus faecalis* [15], and the domestic silkworm *Bombyx mori* [16]. We previously reported the crystal structure of the catalytic domain of Nag31 from *E. faecalis* (EfNag31A) in apo form and in complex with GalNAc [15]. Based on the comparison of the structures, unlike other known GH31 enzymes, the active site of EfNag31A was found to have two distinct conformations: open and closed forms (Fig. 1). The catalytic nucleophile Asp455 is located away from subsite −1 in the open form (apo), while this residue gets closer to C1 atom of GalNAc in the closed form (complex with GalNAc, hereafter WT/GalNAc). However, the GalNAc molecule found in the complex structure adopts a β-form and it remains unclear whether such conformational change in the active site occurs in the case of binding substrates. In this study, we constructed and analyzed several mutants

Table 1
Data collection and refinement statistics.

	D455A/GalNAcα-pNP	D455N/GalNAcα-Ser
Data collection		
Beamline	PF AR NW12A	PF AR NW12A
Wavelength (Å)	1.0000	1.0000
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	81.9, 83.4, 146.3	82.7, 83.0, 148.6
Resolution range (Å)	50–1.42 (1.50–1.42)	50–1.63 (1.72–1.63)
Measured reflections	2,459,929	1,669,991
Unique reflections	184,729	127,951
Completeness (%)	98.0 (94.7)	100 (100)
Redundancy	13.3 (13.5)	13.1 (12.1)
Mean $I/\sigma(I)$	24.5 (2.7)	23.8 (2.6)
R_{merge}	0.067 (1.044)	0.068 (0.961)
$CC_{1/2}$	(0.766)	(0.809)
Refinement statistics		
R_{work}/R_{free}	0.174/0.196	0.177/0.200
RMSD		
Bond lengths (Å)	0.014	0.011
Bond angles (°)	1.726	1.629
Number of atoms		
Protein	7124	7044
Ligand/Ion	76	80
Water	792	558
Average *B* (Å^2)		
Protein	23.5	28.8
Ligands	28.9	34.8
Water	27.4	31.8
Ramachandran plot		
Favored (%)	96.6	96.6
Outliers (%)	0.2	0.3
Clashscore	3.61	2.01
MolProbity score	1.41	1.18
PDB codes	7F7Q	7F7R

Values for the highest resolution shells are given in parentheses.

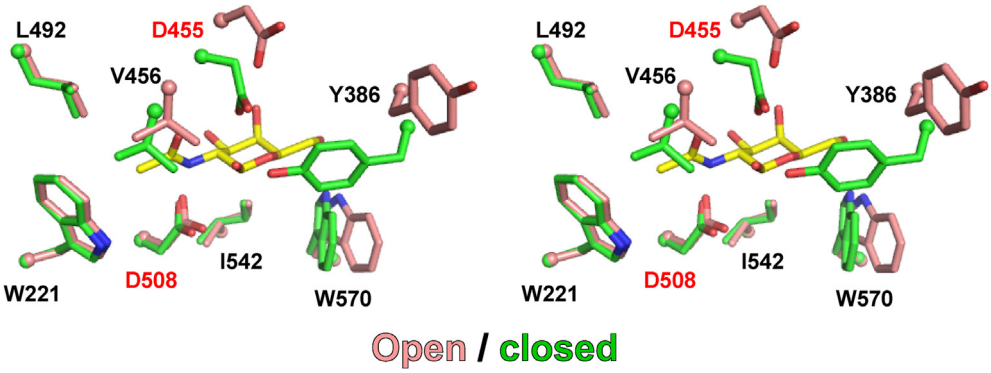

Fig. 1. Stereo view of the open and closed forms of EfGH31 active site. The active site residues in unliganded EfGH31 (PDB 6M76, *pink*) and in GalNAc-bound EfGH31 (PDB 6M77, *green*) are shown as *stick* models, while their Cα atoms are shown as *spheres*. The GalNAc molecule is indicated by *yellow*. Asp455 and Asp508 are the catalytic nucleophile and acid/base, respectively.

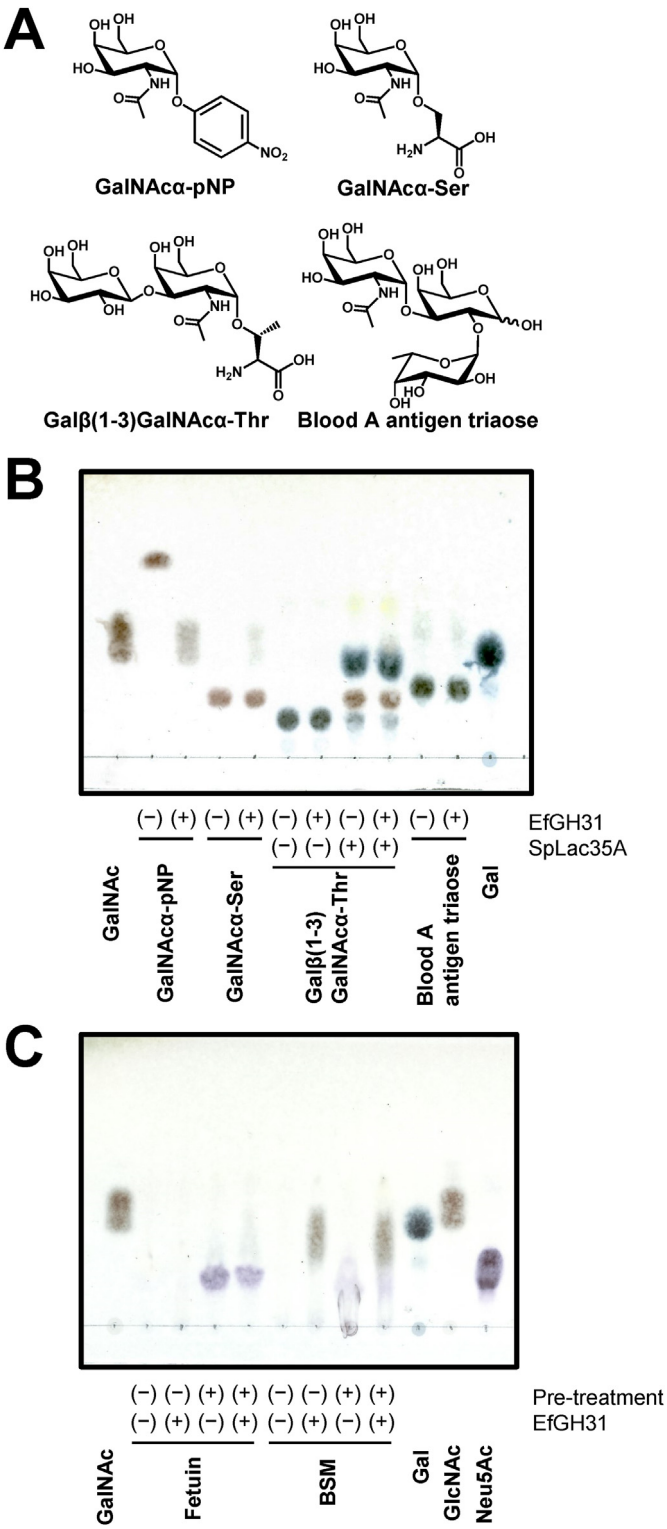

Fig. 2. Hydrolytic activity of EfGH31 toward various naturally occurring α-GalNAc-containing carbohydrates. (A) Chemical structures of the glycosides tested. **(B)** TLC analysis of hydrolytic reaction for GalNAcα-pNP and natural α-GalNAc-containing glycosides by EfGH31 with or without SpLac35A for 12 h. **(C)** TLC analysis of hydrolytic reaction for *O*-glycoproteins (fetuin and BSM) by EfGH31 for 20 h with or without 20-h pretreatment of *exo*-glycosidases (see Materials and Methods). Standard sugars are also developed on TLC: GalNAc, *N*-acetylgalactosamine; Gal, galactose; GlcNAc, *N*-acetylglucosamine; Neu5Ac, *N*-acetylneuraminic acid.

of EfNag31A to investigate how the enzyme recognizes and hydrolyzes substrate sugars. The crystal structures of EfNag31A catalytic domain (named EfGH31) in complex with two substrates were determined. Combined with computational analysis, we propose the substrate recognition mechanism and conformational itinerary of Nag31s.

2. Materials and methods

2.1. Materials

GalNAcα-pNP, Tn antigen (GalNAcα-Ser), Galβ(1–3)GalNAcα-Thr, and blood type A antigen triaose GalNAcα(1–3)[Fucα(1–2)]Gal were procured from Cayman Chemical (Ann Arbor, MI, USA), Dextra Laboratories (Reading, UK), Tokyo Chemical Industry (Tokyo, Japan), and Elicityl (Crolles, France), respectively. Fetuin from bovine serum and mucin from bovine submaxillary glands (BSM) were procured from Merck (Darmstadt, Germany). All other reagents were of analytical grade and were procured from FUJIFILM Wako Pure Chemical Industry (Osaka, Japan) or Nacalai Tesque (Kyoto, Japan), unless otherwise stated.

2.2. Construction and preparation of mutants

Site-directed mutagenesis was performed by inverse polymerase chain reaction (PCR) with the desired primers (Table S1) and a pET28a expression plasmid harboring the DNA that encodes the catalytic domain of EfNag31A (residues 43–984 of total 1866, EfGH31) [15] was used as a template. KOD-Plus-Neo DNA polymerase (Toyobo, Osaka Japan) was used for the PCR. All constructed plasmids were confirmed by DNA sequencing. All recombinant enzymes used in this study contained an N-terminal His tag (MGSSHHHHHHSSGLVPRGSHM) and were expressed and purified in the same manner as the recombinant wild-type (WT) EfGH31 described previously [15]. Protein purity was determined by SDS-PAGE. Protein concentration was measured at 280 nm based on theoretical molar absorption coefficients calculated using the ExPASy ProtParam server (https://web.expasy.org/protparam/).

2.3. Enzymatic assay and kinetic study

For substrate specificity assay, reaction mixtures containing a substrate (3.5 mM for GalNAcα-pNP or 10 mM for others), 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES)–NaOH buffer (pH 6.0), and 100 μg/mL EfGH31 were incubated at 37 °C for 12 h. *Streptococcus pneumoniae* β-galactosidase 35A (SpLac35A, NZYTech, Lisbon, Portugal) was also incubated with Galβ(1–3)GalNAcα-Thr, which was used as a substrate. For fetuin and BSM, 4 mg of the glycoproteins were pre-treated with 10 μg/mL SpLac35A, 500 unit/mL of *Clostridium perfringens* neuraminidase (New England Biolabs, Ipswich, MA, USA), 50 unit/mL of *Streptomyces plicatus* β-*N*-acetylhexosaminidase f (New England Biolabs), and 20 mM MES-NaOH buffer (pH 6.0) in 1 mL of the reaction mixture at 30 °C for 20 h. The reaction mixtures were concentrated by centrifugal evaporation and the glycoproteins were precipitated twice with 80% ethanol to remove the released sugars. The precipitate was dissolved in buffer and incubated with 100 μg/mL EfGH31 at 30 °C for 20 h, followed by centrifugation to remove proteins, and the supernatant was concentrated by centrifugal evaporation. Reaction products were monitored by thin-layer chromatography (TLC) on Silica Gel 60 plates (Merck) with a developing solvent of 1-butanol/acetic acid/water (2:1:1 by volume) and visualized using diphenylamine/aniline/phosphoric acid reagent [23].

For kinetic assay of EfGH31 mutants, GalNAcα-pNP was used as a substrate and the reaction condition was the same as described

previously [20], except for enzyme concentrations: W221N, 470 nM; Y386A, 940 nM; Y386F, 940 nM; D455A, 3.8 μM; D455N, 3.8 μM; V456A, 94 nM; L492R, 940 nM; D508N, 3.8 μM; I542F, 470 nM; and W570A, 470 nM. Kinetic parameters were calculated by fitting to the Michaelis–Menten equation using non-linear regression analysis by KaleidaGraph software (Synergy Software, Reading PA, USA).

2.4. Structural analysis

D455A and D455N (16 mg/mL) were crystallized at 20 °C using the hanging-drop vapor diffusion method, where 1.0 μL of protein solution was mixed with an equal volume of a crystallization reservoir solution. Their crystals appeared in the same condition as WT EfGH31 [15]. Crystals of D455A and D455N were soaked with reservoir solutions containing 10 mM GalNAcα-pNP (for 1 h) and 10 mM GalNAcα-Ser (for 5 min), respectively. All crystals were cryoprotected with the reservoir solution supplemented with ethylene glycol at a final concentration of 20% (v/v) and then flash-frozen in liquid nitrogen. Diffraction data were collected at AR-NW12A beamline (Photon Factory, Tsukuba, Japan). Data were processed using XDS [24]. Structures were solved by molecular replacement method using MOLREP [25] in CCP4 program suite [26] using the coordinate of native EfGH31 (PDB 6M76) as a search model. Refinement and manual model building were performed with REFMAC5 [27] and COOT [28], respectively. The data collection and refinement statistics are summarized in Table 1. Molecular docking simulations were performed using AutoDock 4.2.6 [29] using the coordinate of the closed form of EfGH31 (WT/GalNAc, PDB 6M77). GalNAc-attached peptide ligands were prepared from the coordinates of PDB 5T5L and 5T5P. Molecular images were prepared with PyMOL (Schrödinger LLC, New York, NY, USA).

2.5. Circular dichroism spectroscopy

Circular dichroism (CD) spectra were obtained for the recombinant enzymes (0.1 mg/mL in 10 mM sodium phosphate buffer, pH 7.0) using a 1-mm path-length quartz cell in a J-820 spectropolarimeter (JASCO Co., Tokyo, Japan) at room temperature. Four spectra from 190 to 260 nm were taken at a scanning speed of 50 nm/min and averaged for each sample.

2.6. Full-QM cluster model calculations

Starting from the available coordinates in PDB 6M77 [15], two models of the active site of the GH31 enzyme were constructed by adding *p*-nitrophenyl (GH31pNP model, Fig. S1A) and neutral serine (GH31Ser model, Fig. S1B) at the α position of the C1 of the GalNAc ring. The GH31Ser model is formed by 270 atoms, including the substrate, three crystallographic water (one of them is required for the deglycosylation step of the hydrolysis), protonated catalytic acid-base (A/B) agent (Asp455), non-protonated catalytic nucleophilic (Nuc) agent (Asp508), two non-truncated amino acids (Asp384 and Tyr386), and ten truncated amino acids (Trp221, Lys453, Leu492, Trp505, Asp538, Ile542, Phe543, Met567 and Trp570). The GH31pNP model is formed by 271 atoms that only differ from the first model in the exchange of the +1 molecule. In both models, the position of 16 atoms of the truncated residues remains fixed in order to maintain the original topology of the enzymatic cavity. The models have a total charge of −3.

All calculations were performed using quantum mechanics (QM) Turbomole 7.4 code [30,31]. The functional and basis set used to construct the electronic density of the model was PBE/def2-SVP [32,33]. A conductor-like screening model (COSMO) [34] was implemented to mimic the solvent effect of an enzymatic environment ($\varepsilon = 10$). The energies and geometries along the reaction coordinate(s) were obtained from scan calculations with an optimization threshold criterion of 10^{-6} a.u. (the transition states presented in this work are considered as transition state-like).

3. Results and discussion

3.1. Substrate specificity on natural glycosides

B. caccae and *P. plebeius* Nag31 proteins (BcGH31 and BpGH31) exhibited activity that resulted in the release of GalNAc from fetuin and *O*-glycopeptide [14], whereas EfGH31 activity had been previously investigated using only the synthetic substrate GalNAcα-pNP [15]. We determined whether naturally occurring sugars containing an α-GalNAc residue were hydrolyzed by EfGH31 (Fig. 2A). Interestingly, GalNAcα-Ser was a poor substrate for EfGH31 compared with GalNAcα-pNP, since most GalNAcα-Ser was retained despite the overnight incubation (Fig. 2B). Galβ(1–3)GalNAcα-Thr and blood A antigen were not hydrolyzed, but incubation together with β-1,3-galactosidase from *Streptococcus pneumoniae* β-galactosidase 35A (SpLac35A) enabled EfGH31 to release a trace of GalNAc from Galβ(1–3)GalNAcα-Thr. Similarly, BcGH31, BpGH31, and a lepidopteran xenolog BmNag31 were reported not to hydrolyze blood A antigen. Afterward, we investigated the activity of EfGH31 toward *O*-glycoproteins (Fig. 2C). No product was observed when fetuin was treated with only EfGH31, whereas fetuin pretreated with SpLac35A and neuraminidase resulted in the release of GalNAc. This result supports the previous study where bacterial Nag31s removed *O*-linked GalNAc from Ser or Thr residue in fetuin [14]. Furthermore, EfGH31 released GalNAc from BSM, regardless of pretreatment with SpLac35A, β-*N*-acetylhexosaminidase, and neuraminidase. The activity toward mucin had not ever been reported in the other Nag31s. The difference in the reactivity toward

Table 2
Kinetic parameters for the hydrolysis of GalNAcα-pNP by EfGH31 and its mutants.

Enzyme	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m (%)
EfGH31 (WT) [15]	6.54 ± 0.12	158 ± 12	41	100
W221N	3.79 ± 0.35	(3.86 ± 0.50) × 10^3	1.0	2.4
Y386A	(5.06 ± 0.72) × 10^{-2}	(2.36 ± 0.53) × 10^3	0.021	0.051
Y386F	(8.51 ± 0.04) × 10^{-2}	60.0 ± 1.9	1.4	3.4
D455A (nucleophile)	(1.00 ± 0.02) × 10^{-2}	73.4 ± 10.3	0.14	0.34
D455N (nucleophile)	(2.39 ± 0.03) × 10^{-3}	23.2 ± 3.6	0.10	0.24
V456A	1.40 ± 0.03	101 ± 10	14	34
L492R	(1.75 ± 0.03) × 10^{-2}	79.4 ± 6.3	0.22	0.54
D508N (acid/base)	(2.31 ± 0.01) × 10^{-3}	17.2 ± 5.5	0.13	0.32
I542F	2.04 ± 0.05	(2.23 ± 0.09) × 10^3	0.92	2.2
W570A	2.89 ± 0.12	(4.44 ± 0.25) × 10^3	0.65	1.6

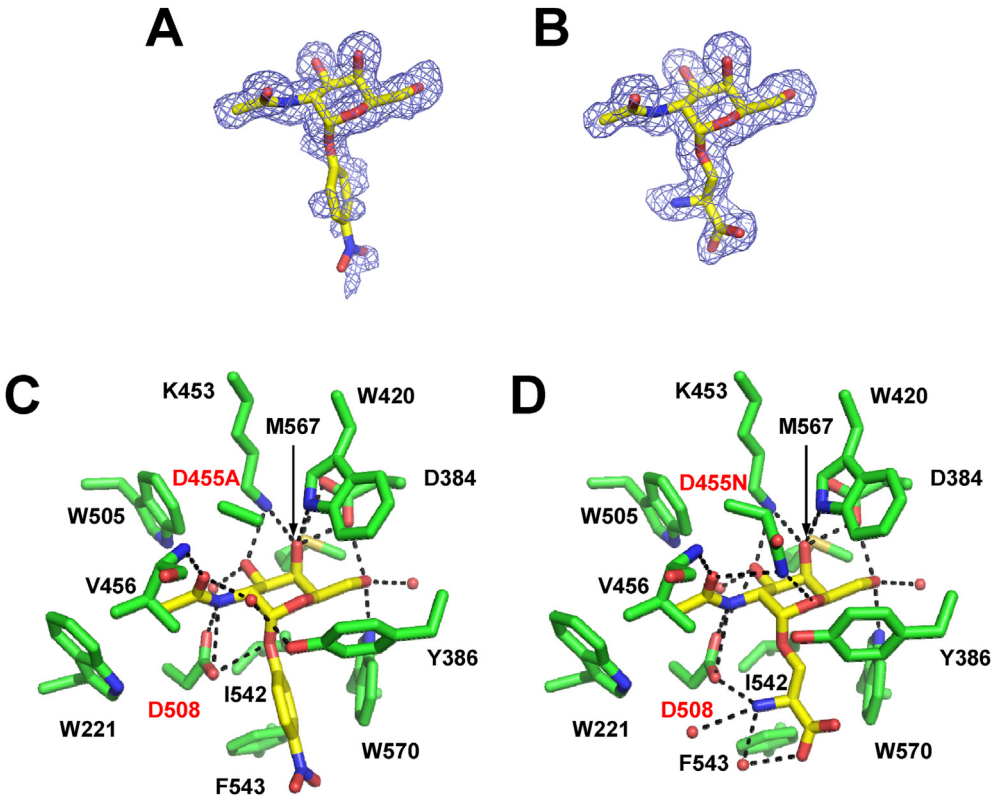

Fig. 3. Substrate binding in EfGH31 active site. (A and B) F_o-F_c electron density maps (contoured at 2.0 σ) for GalNAcα-pNP (A) and GalNAcα-Ser (B). **(C and D)** The active site residues of D455A/GalNAcα-pNP (C) and of D455N/GalNAcα-Ser (D). The side chains of amino acid residues interacting with substrates and the main chain of Val456 are shown in *green stick* models, while the ligands are shown in *yellow*. Hydrogen bonds are shown as *dashed lines*, while water molecules are shown as *red spheres*. The catalytic residues are labeled in *red*.

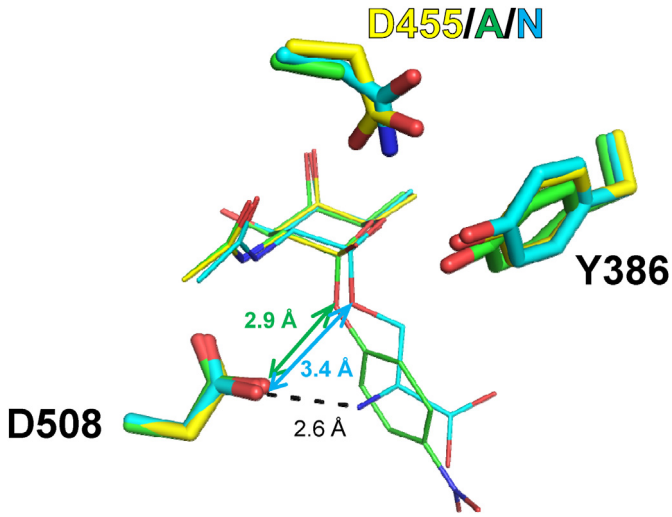

Fig. 4. Structural comparison of active sites of EfGH31 complexed with ligands. The crystal structures of WT/β-GalNAc (*yellow*), D455A/GalNAcα-pNP (*green*), and D455N/GalNAcα-Ser (*cyan*) are superimposed. Distances between the glycosidic bonds and the catalytic acid/base Asp508 (*green* and *cyan arrows*) and between the amide group of the serine and Asp508 (*black dashed line*) are described.

fetuin and BSM may be caused by the contents of a variety of *O*-glycans, i.e., mono-GalNAc and longer glycans, in these glycoproteins [35,36]. *Bifidobacterium bifidum* has an intracellular GH129 α-*N*-acetylgalactosaminidase (NagBb), which showed higher activity toward GalNAcα-Ser than GalNAcα-pNP and was proposed to be involved in mucin degradation [37]. Although it is not clear whether NagBb can directly degrade *O*-glycopeptides or *O*-glycoproteins, recombinant EfGH31 exhibited a different substrate preference than NagBb. EfNag31A has a predicted signal peptide and a C-terminal membrane-anchoring region [15]. Therefore, EfNag31A was suggested to be involved in the extracellular degradation of host mucin-type *O*-glycoproteins and *E. faecalis* may degrade mucin *via* a different series of enzymatic steps than *B. bifidum*.

3.2. Site-directed mutagenesis in the active site

Based on the crystal structure of EfGH31 WT in complex with GalNAc (WT/GalNAc), Asp455 and Asp508 are identified as catalytic nucleophiles and catalytic acid/base, respectively, and several residues are involved in GalNAc recognition with conformational changes (Fig. 1), which are not observed in the other known GH31 enzymes. Considering the difference between Nag31s and other GH31 enzymes (Fig. S2), ten mutants of EfGH31 were constructed and their hydrolytic activity was determined using GalNAcα-pNP to clarify their specificity for α-*N*-acetylgalactosaminides. The CD spectra of these mutants suggested that all the substitutions did not significantly affect protein folding (Fig. S3). As expected, the catalytic residue mutants (D455A, D455N, and D508N) remarkably lost activity (<0.4% compared to the k_{cat}/K_m value of WT) (Table 2). The mutants W221N, Y386F, V456A, and I542F showed moderate activity (2.2–34%). Hydrolytic activities of the other mutants (Y386A, L492R, and W570A) decreased to less than 2% of WT.

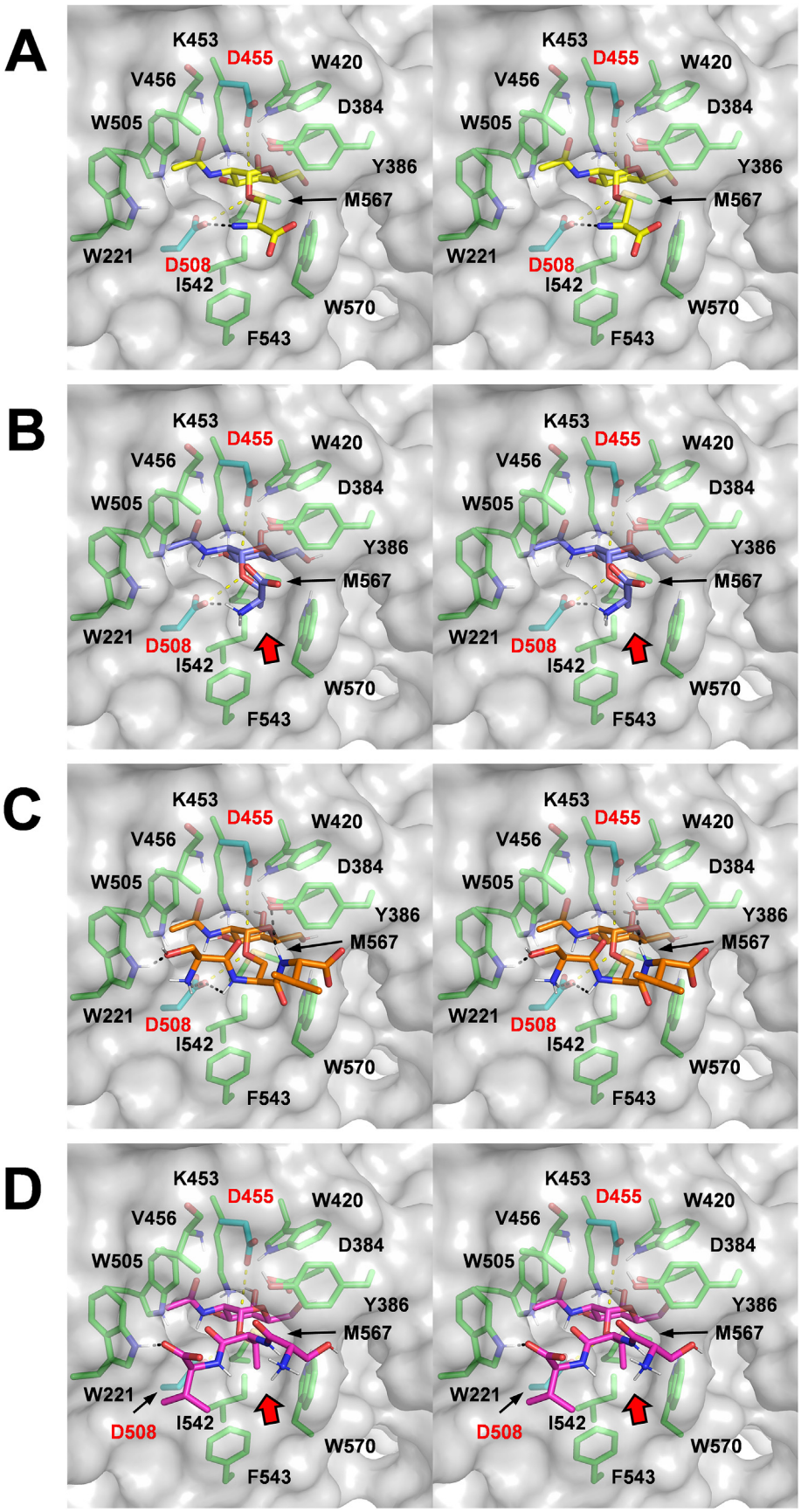

Fig. 5. Molecular docking of GalNAcα-Thr and *O*-GalNAc peptides. (A) Superimposition of GalNAcα-Ser from D455N/GalNAcα-Ser into the active site of the closed form of EfGH31 (WT/GalNAc, PDB 6M77). **(B–D)** Molecular docking of GalNAcα-Thr (B), Ser-Ser(GalNAc)-Val (C), and Ser-Thr(GalNAc)-Val (D) into the closed form (see Materials and Methods).

Trp221 is located on a loop in the N-domain, but not in the catalytic A-domain. The length of the corresponding loops in GH31 enzymes vary. An Asn residue was located at a similar position in *C. japonicus* α-xylosidase [11]. Trp221 was predicted to form a hydrophobic environment with Val456 and Leu492 in order to accept the acetamide group of GalNAc and the Trp221→Asn substitution was suggested to reduce the affinity for GalNAc and as expected resulted in a high K_m value (Table 2). The k_{cat} of V456A was lower than one-fourth of that of WT, while their K_m values were similar, thus suggesting that interaction between Val456 and the substrate is likely to affect the conformation of the loop where Val456 and the catalytic nucleophile Asp455 are located. Leu492 is completely conserved among Nag31s, whereas it is substituted for arginine in all other known GH31 enzymes, including α-galactosidases (Fig. S2) [12]. The Leu492→Arg substitution was suggested to obstruct the hydrolysis of GalNAcα-pNP; however, L492R showed no activity toward *p*-nitrophenyl α-galactopyranoside as seen for WT (data not shown), thus suggesting that the mechanisms of substrate recognition and concomitant conformational changes are complicated. Ile542 surrounds the α-face of GalNAc C4 atom with Met567 and Trp570, while other structure-determined GH31 enzymes have tyrosine or phenylalanine residues at this position. Trp570 forms a hydrogen bond with GalNAc O6 atom. The face might be destroyed by Ile542→Phe and Trp570→Ala mutations, considering that the K_m values of I542F and W570A for GalNAcα-pNP increased by 14- and 28-fold, respectively. The mutations at Tyr386, whose conformations are most different between the open forms and closed forms, resulted in a remarkable decrease in the k_{cat} values. Moreover, the K_m value of Y386A was 39-fold higher than that of Y386F. These results suggest that the aromatic group of Tyr386 is important for the binding of a substrate and that the hydroxy group may be involved in fixing the substrate orientation to properly catalyze hydrolysis and/or stabilization of intermediates during the hydrolytic reaction. All these residues, except for Leu492, which is replaced with valine in lepidopteran BmNag31 [16], are completely conserved among the characterized Nag31s. Although what directly triggers the closed form is not clear, the residues in the active sites of Nag31s were found to work in concert to act on only α-*N*-acetylgalactosaminides.

3.3. Complex structures with substrates and docking simulation

Crystals of the catalytic nucleophile mutants D455A and D455N were soaked with GalNAcα-pNP for few minutes or hours and analyzed by X-ray crystallography. In the crystal structure of D455A soaked with GalNAcα-pNP for 1 h (D455A/GalNAcα-pNP), an electron density map for GalNAcα-pNP was observed in the active site (Fig. 3A). By contrast, no ligand or only GalNAc was found in the other soaked crystals, which is probably due to insufficient soaking time or hydrolysis by its slightly remaining activity, respectively. The crystal structure of D455N in complex with GalNAcα-Ser (D455N/GalNAcα-Ser) was also determined. The electron density for pNP was poor in D455A/GalNAcα-pNP, while those for α-GalNAc in both structures and the serine residue in D455N/GalNAcα-Ser were well defined (Fig. 3B). Their active sites adopted the closed form like the previously determined WT/GalNAc (PDB 6M77) and the orientation of GalNAc at subsite −1 (GalNAc −1) and interacting amino acid residues are identical (Fig. 3C and D). No direct hydrogen bond with pNP and the serine residue, except between the amino group of the serine and OD2 atom of the catalytic acid/base Asp508, was identified. Superimposition of D455A/GalNAcα-pNP and D455N/GalNAcα-Ser reveals that a distance between the glycosidic bond of GalNAcα-Ser and Asp508 (3.4 Å) is slightly longer than that between the glycosidic bond of GalNAcα-pNP and Asp508 (2.6 Å) (Fig. 4). However, it remains unclear whether this distance and interaction between Asp508 and the serine amino group caused the low hydrolytic activity toward GalNAcα-Ser (the reason of the low activity is discussed below). The difference in the reactivities on GalNAcα-Ser between EfGH31 and GH129 NagBb is likely to be caused by the difference in their active site architectures in spite of their similar retaining hydrolytic mechanisms (Fig. S4) [38].

To investigate the substrate recognition mechanism of EfGH31 for *O*-glycoproteins, molecular docking simulations were performed. First, using the coordinates of the closed form (WT/GalNAc), we simulated GalNAcα-Thr, which is another core structure in *O*-glycan and found that the model in which an α-GalNAc residue fits into subsite −1 was most stable (Fig. 5B). Moreover, it was suggested that the methyl group of the threonine could fit into the narrow pocket formed by the hydrophobic residues Ile542, Phe543, and Trp570. On the other hand, docking analysis using *O*-glycopeptides, Ser-Ser(GalNAc)-Val, and Ser-Thr(GalNAc)-Val, showed that the GalNAcα-Ser and GalNAcα-Thr moieties of the ligands are located in similar position with the free GalNAcα-Ser and GalNAcα-Thr (Fig. 5A and B) and the tripeptide portions are located outside the active site pocket (Fig. 5C and D). The methyl group of the threonine of Ser-Thr(GalNAc)-Val was surrounded by Ile542, Phe543, and Trp570, as with the free GalNAc-Thr (Fig. 5B and D). The only polar interactions between both tripeptides and EfGH31 are hydrogen bonds with Trp221 and Asp508. Both peptides have opposite orientations. These observations suggest that peptide parts of *O*-GalNAcylated proteins may not be strictly recognized by EfGH31. The distances between GalNAc C1 in each substrate and the nucleophile Asp455 (3.0 Å) and the distances between glycosidic oxygen in each substrate and the acid/base Asp508 (2.7–3.0 Å) can enable catalysis. These results suggest that EfGH31 active site specifically recognizes and hydrolyzes GalNAc-attached Ser or Thr of glycopeptides.

3.4. Conformational itinerary

To decipher the conformational itinerary of GalNAc along the hydrolysis in the active site of EfGH31, we constructed two active site models, namely GH31pNP and GH31Ser, where GalNAcα-pNP and GalNAcα-Ser were put in the active site, respectively, and we performed quantum mechanical calculations. Following the potential energy surface of the hydrolysis reaction in both models (Fig. 6A), the formation of the GEI requires an activation energy of ~18 kcal mol^{-1} for both substrates; however, the process is endergonic in case of using serine as a leaving group (+4.7 kcal mol^{-1}) and exergonic in case of the presence of pNP (−2.9 kcal mol^{-1}). This fact leads to a change in the transition state energy of the deglycosylation step (TS_{deg}), which makes the reaction difficult in the GH31Ser model (23.7 kcal mol^{-1}), but favor the reaction in the GH31pNP model (9.6 kcal mol^{-1}). Due to the aromatic nature of pNP, the protonation state of the leaving groups differs between GH31pNP and GH31Ser in the GEI*/GEI minima (Fig. 6 B and C). The presence of the conjugate base of pNP leads into stabilization of the system after the glycosylation step. While the effect of changing the leaving group over a hydrolytic process is

Colors of stick models are follows: active site residues, *green*; catalytic residues, *cyan*; GalNAcα-Ser, *yellow*; GalNAcα-Ser, *slate blue*; Ser-Ser(GalNAc)-Val, *orange*; Ser-Thr(GalNAc)-Val, *magenta*. Black dashed lines are indicated as hydrogen bonds between EfGH31 and ligand amino acid residues, while yellow dashed lines are indicated as distances between the C1 of GalNAc and the nucleophile Asp455 and between the glycosidic O and the acid/base Asp508. The pocket where the methyl groups of GalNAc-attached Thr enter is indicated by *red arrows*.

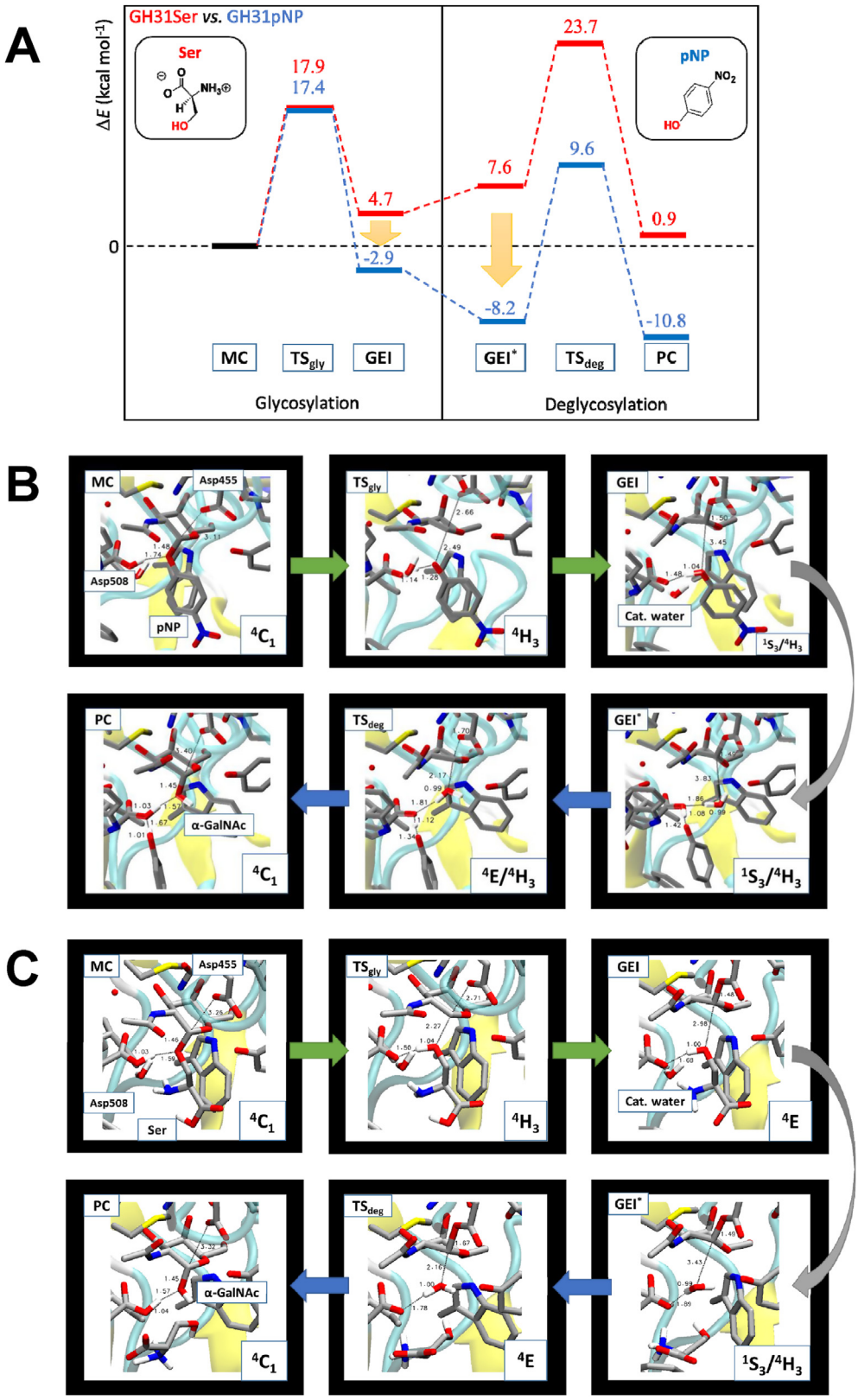

Fig. 6. Quantum mechanical calculations along the proposed hydrolytic reaction pathway. (A) Potential energy surface of the hydrolysis toward GalNAcα-Ser (GH31Ser, *red*) and GalNAcα-pNP (GH31pNP, *blue*) models. **(B)** Representative structures along the hydrolysis pathway (glycosylation, *up*–deglycosylation, *down*) in the GH31pNP model. **(C)** Representative structures along the hydrolysis pathway (glycosylation, *up*–deglycosylation, *down*) in the GH31Ser model. Abbreviations: MC, Michaelis complex; TS_{gly}, transition state in the glycosylation step; GEI, glycosyl-enzyme intermediate; TS_{deg}, transition state in the deglycosylation step; PC, product complex.

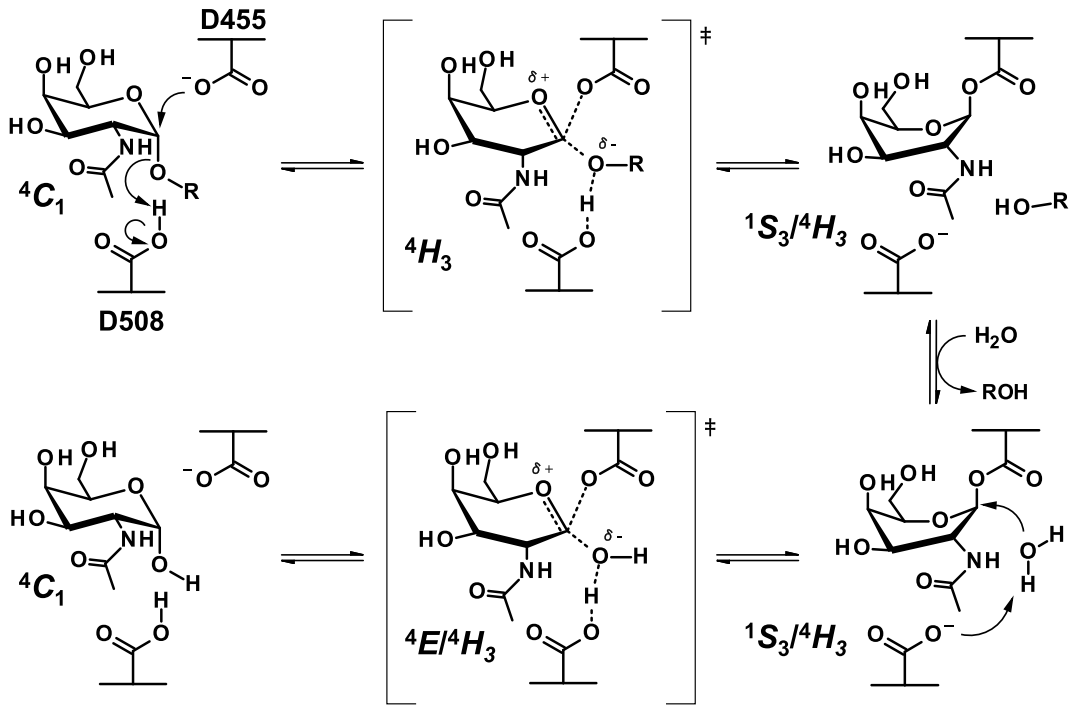

Fig. 7. Proposed mechanism of EfGH31 hydrolysis toward α-*N*-acetylgalactosaminide.

not experimentally known, we can talk about a "leaving group-assisted" mechanism. The leaving group is just affecting the stability of the GEI/GEI*, while the barrier GEI* → TS_{gly} → PC remains similar between both systems (~16–18 kcal mol^{-1}). Finally, the formation of the final product complex (PC) is slightly endergonic with serine (+0.9 kcal mol^{-1}) and exergonic with *p*-nitrophenol (−10.8 kcal mol^{-1}). These results agree with the enzymatic studies where the GH31 enzyme preferred GalNAcα-pNP to GalNAcα-Ser.

The conformational itineraries, followed by GalNAc −1 moiety along the reaction in both models, are shown in Fig. 5B and C. Cremer and Pople [39] puckering coordinates' evolution is shown in Table S2. In agreement with the conformation observed in the WT/GalNAc crystal structure, the MCs present no distortion in the sugar ring ($^4C_1$). In the GH31pNP model whose potential energy surface demonstrates the feasibility of the reaction, the itinerary follows $^4C_1 \rightarrow (^4H_3)^{\ddagger} \rightarrow {}^1S_3/^4H_3$ conformational region for glycosylation but follows $^1S_3/^4H_3 \rightarrow (^4E/^4H_3)^{\ddagger} \rightarrow {}^4C_1$ for deglycosylation. The GH31 conformational itinerary $^4C_1 \rightarrow (^4H_3)^{\ddagger} \rightarrow {}^1S_3 \rightarrow (^4H_3)^{\ddagger} \rightarrow {}^4C_1$ was initially proposed based on the crystal structure of the covalent intermediate of *E. coli* α-xylosidase YicI with 5-fluoro-xylosyl fluoride, which adopted $^1S_3$ skew boat conformation [10]. The $^1S_3$ covalent intermediates have been reported in some other GH31 enzymes that are active on α-glucosides [9,21,22], α-xylosides [11], and α-sulfoquinovosides [13]. Therefore, the result in the present study matches the proposed conformational itinerary (Fig. 7), although the crystal structure of Nag31 covalent intermediate has not been investigated. GH31 is classified into clan GH-D together with GH27 and GH36 based on the structural similarity and conservation of their catalytic machinery, where two aspartic acid residues act as catalysts through the retaining mechanism [40]. The MCs and PCs of human GH27 α-galactosidase and α-*N*-acetylgalactosaminidase adopted a stable $^4C_1$ form, while the GEIs trapped using trinitrophenyl-2,2-difluoro-α-galactopyranoside showed a $^1S_3$ conformation of 2,2-difluorogalactose covalently bound to their catalytic nucleophiles [41,42]. These conformations were supported by the QM/MM metadynamics using human α-galactosidase [43]. The covalent intermediate of *Thermus thermophilus* GH36 α-galactosidase was trapped using the synthetic 2-deoxy-2-fluorocarbagalactosyl substrate, showing a similar sugar distortion ($^4H_3/^1S_3$) [44]. Collectively, the conformational itineraries of substrates are highly conserved in the clan GH-D regardless of the substrate structures.

4. Conclusions

In this study we investigated the substrate specificity of EfGH31 α-*N*-acetylgalactosaminidase toward natural substrates and found that the enzyme released GalNAc from bovine submaxillary mucin. Based on analysis of the crystal structures complexed with substrates, mutational analyses, and docking simulations, EfGH31 was suggested to recognize mainly a GalNAcα-Ser/Thr moiety, exhibiting substrate recognition mechanisms that are different from those of known GH31 enzymes, apart from Nag31s. The QM calculations indicate that the conformational itinerary is conserved among GH31 enzymes and clan GH-D enzymes, regardless of the substrate type. These findings will provide insight on the design of specific inhibitors and probes to detect this activity in the gut. Also, it will also enable the development of tools for analyzing *O*-glycan structures.

Author contributions

Conception of the work, TM, SAG; collection and analysis of data, TM, MI, SAG; writing of manuscript, TM, SAG.

Funding

This work was supported by Japan Society for the Promotion of Science KAKENHI [grant No. 19K15748] (to TM). SAG acknowledges the VIP^2 fellowship 2020 funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement [No. 847548].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The coordinates of D455A/GalNAcα-pNP and D455N/GalNAcα-Ser were deposited at the Worldwide Protein Data Bank (http://www.wwpdb.org/), with the accession codes 7F7Q and 7F7R, respectively.

Acknowledgments

We thank the staff of the Photon Factory for their help in the X-ray data collection. This research was performed under the approval of the Photon Factory Program Advisory Committee (Proposal No. 2019G097). We also thank Enago (www.enago.jp) for the English language review.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biochi.2021.11.007.

References

[1] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) D490–D495.
[2] CAZypedia Consortium, Ten years of CAZypedia: A living encyclopedia of carbohydrate-active enzymes, Glycobiology 28 (2018) 3–8.
[3] D.E. Koshland Jr., Stereochemistry and the mechanism of enzymatic reactions, Biol. Rev. 28 (1953) 416–436.
[4] G.J. Davies, K.S. Wilson, B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, Biochem. J. 321 (1997) 557–559.
[5] G.J. Davies, A. Planas, C. Rovira, Conformational analyses of the reaction coordinate of glycosidases, Acc. Chem. Res. 45 (2012) 308–316.
[6] L. Wu, Z. Armstrong, S.P. Schröder, C. de Boer, M. Artola, J.M. Aerts, H.S. Overkleeft, G.J. Davies, An overview of activity-based probes for glycosidases, Curr. Opin. Chem. Biol. 53 (2019) 25–36.
[7] H.A. Ernst, L. Lo Leggio, M. Willemoës, G. Leonard, P. Blum, S. Larsen, Structure of the *Sulfolobus solfataricus* α-glucosidase: implications for domain conservation and substrate recognition in GH31, J. Mol. Biol. 358 (2006) 1106–1124.
[8] K. Tan, C. Tesar, R. Wilton, L. Keigher, G. Babnigg, A. Joachimiak, Novel α-glucosidase from human gut microbiome: substrate specificities and their switch, Faseb. J. 24 (2010) 3939–3949.
[9] A.T. Caputo, D.S. Alonzi, L. Marti, I.B. Reca, J.L. Kiappes, W.B. Struwe, A. Cross, S. Basu, E.D. Lowe, B. Darlot, et al., Structures of mammalian ER α-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals, Proc. Natl. Acad. Sci. U. S. A 113 (2016) E4630–E4638.
[10] A.L. Lovering, S.S. Lee, Y.W. Kim, S.G. Withers, N.C. Strynadka, Mechanistic and structural analysis of a family 31 α-glycosidase and its glycosyl-enzyme intermediate, J. Biol. Chem. 280 (2005) 2105–2115.
[11] J. Larsbrink, A. Izumi, F.M. Ibatullin, A. Nakhai, H.J. Gilbert, G.J. Davies, H. Brumer, Structural and enzymatic characterization of a glycoside hydrolase family 31 α-xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification, Biochem. J. 436 (2011) 567–580.
[12] T. Miyazaki, Y. Ishizaki, M. Ichikawa, A. Nishikawa, T. Tonozuka, Structural and biochemical characterization of novel bacterial α-galactosidases belonging to glycoside hydrolase family 31, Biochem. J. 469 (2015) 145–158.
[13] G. Speciale, Y. Jin, G.J. Davies, S.J. Williams, E.D. Goddard-Borger, YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids, Nat. Chem. Biol. 12 (2016) 215–217.
[14] P. Rahfeld, J.F. Wardman, K. Mehr, D. Huff, C. Morgan-Lang, H.M. Chen, S.J. Hallam, S.G. Withers, Prospecting for microbial α-*N*-acetylgalactosaminidases yields a new class of GH31 *O*-glycanase, J. Biol. Chem. 294 (2019) 16400–16415.
[15] T. Miyazaki, E.Y. Park, Crystal structure of the *Enterococcus faecalis* α-*N*-acetylgalactosaminidase, a member of the glycoside hydrolase family 31, FEBS Lett. 594 (2020) 2282–2293.
[16] M. Ikegaya, T. Miyazaki, E.Y. Park, Biochemical characterization of *Bombyx mori* α-*N*-acetylgalactosaminidase belonging to the glycoside hydrolase family 31, Insect Mol. Biol. 30 (2021) 367–378.
[17] Y. Gozu, Y. Ishizaki, Y. Hosoyama, T. Miyazaki, A. Nishikawa, T. Tonozuka, A glycoside hydrolase family 31 dextranase with high transglucosylation activity from *Flavobacterium johnsoniae*, Biosci. Biotechnol. Biochem. 80 (2016) 1562–1567.
[18] S.H. Light, L.A. Cahoon, A.S. Halavaty, N.E. Freitag, W.F. Anderson, Structure to function of an α-glucan metabolic pathway that promotes *Listeria monocytogenes* pathogenesis, Nat. Microbiol. 2 (2016) 16202.
[19] T. Tagami, E. Miyano, J. Sadahiro, M. Okuyama, T. Iwasaki, A. Kimura, Two novel glycoside hydrolases responsible for the catabolism of cyclobis-(1→6)-α-nigerosyl, J. Biol. Chem. 291 (2016) 16438–16447.
[20] K. Tsutsumi, Y. Gozu, A. Nishikawa, T. Tonozuka, Structural insights into polysaccharide recognition by *Flavobacterium johnsoniae* dextranase, a member of glycoside hydrolase family 31, FEBS J. 287 (2020) 1195–1207.
[21] J. Larsbrink, A. Izumi, G.R. Hemsworth, G.J. Davies, H. Brumer, Structural enzymology of *Cellvibrio japonicus* Agd31B protein reveals α-transglucosylase activity in glycoside hydrolase family 31, J. Biol. Chem. 287 (2012) 43288–43299.
[22] H.J. Rozeboom, S. Yu, S. Madrid, K.H. Kalk, R. Zhang, B.W. Dijkstra, Crystal structure of α-1,4-glucan lyase, a unique glycoside hydrolase family member with a novel catalytic mechanism, J. Biol. Chem. 288 (2013) 26764–26774.
[23] K. Anderson, S.C. Li, Y.T. Li, Diphenylamine-aniline-phosphoric acid reagent, a versatile spray reagent for revealing glycoconjugates on thin-layer chromatography plates, Anal. Biochem. 287 (2000) 337–339.
[24] W. Kabsch, X.D.S. W, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125–132.
[25] A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement, J. Appl. Crystallogr. 30 (1997) 1022–1025.
[26] Collaborative Computational Project, Number 4, the CCP4 suite: programs for protein crystallography, Acta Crystallogr. D Biol. Crystallogr. 50 (1994) 760–763.
[27] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr, D Biol. Crystallogr. 53 (1997) 240–255.
[28] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 486–501.
[29] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791.
[30] R. Ahlrichs, M. Baer, M. Haeser, H. Horn, C. Koelmel, Electronic structure calculations on workstation computers: the program system TURBOMOLE, Chem. Phys. Lett. 162 (1989) 165–169.
[31] M. Von Arnim, R. Ahlrichs, Performance of parallel TURBOMOLE for density functional calculations, J. Comput. Chem. 19 (1998) 1746–1757.
[32] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868. J.P. Perdew, K. Burke, M. Ernzerhof, Errata: Generalized gradient approximation made simple, Phys. Rev. Lett. 78 (1997) 1396–1396.
[33] F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy, Phys. Chem. Chem. Phys. 7 (2005) 3297–3305.
[34] A. Klamt, G. Schüürmann, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, J. Chem. Soc. Perkin. Trans. 2 (1993) 799–805.
[35] M. Windwarder, F. Altmann, Site-specific analysis of the *O*-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry, J. Proteomics. 108 (2014) 258–268.
[36] J. Kim, C. Ryu, J. Ha, J. Lee, D. Kim, M. Ji, C.S. Park, J. Lee, D.K. Kim, H.H. Kim, Structural and quantitative characterization of mucin-type *O*-glycans and the identification of *O*-glycosylation sites in bovine submaxillary mucin, Biomolecules 10 (2020) 636.
[37] M. Kiyohara, T. Nakatomi, S. Kurihara, S. Fushinobu, H. Suzuki, T. Tanaka, S.I. Shoda, M. Kitaoka, T. Katayama, K. Yamamoto, et al., α-*N*-Acetylgalactosaminidase from infant-associated bifidobacteria belonging to novel glycoside hydrolase family 129 is implicated in alternative mucin degradation pathway, J. Biol. Chem. 287 (2012) 693–700.
[38] M. Sato, D. Liebschner, Y. Yamada, N. Matsugaki, T. Arakawa, S.S. Wills, M. Hattie, K.A. Stubbs, T. Ito, T. Senda, H. Ashida, S. Fushinobu, The first crystal structure of a family 129 glycoside hydrolase from a probiotic bacterium reveals critical residues and metal cofactors, J. Biol. Chem. 22 (2017) 12126–12138.
[39] D. Cremer, J.A. Pople, General definition of ring puckering coordinates, J. Am. Chem. Soc. 97 (1975) 1354–1358.
[40] B. Henrissat, A. Bairoch, Updating the sequence-based classification of glycosyl hydrolases, Biochem. J. 316 (1996) 695–696.
[41] N.E. Clark, S.C. Garman, The 1.9 Å structure of human α-*N*-acetylgalactosaminidase: the molecular basis of Schindler and Kanzaki diseases, J. Mol. Biol. 393 (2009) 435–447.
[42] A.I. Guce, N.E. Clark, E.N. Salgado, D.R. Ivanen, A.A. Kulminskaya, H. Brumer 3rd, S.C. Garman, Catalytic mechanism of human α-galactosidase, J. Biol. Chem. 285 (2010) 3625–3632.
[43] X.L. Pan, W. Liu, J.Y. Liu, Mechanism of the glycosylation step catalyzed by human α-galactosidase: a QM/MM metadynamics study, J. Phys. Chem. B 117 (2013) 484–489.
[44] W. Ren, R. Pengelly, M. Farren-Dai, S. Shamsi Kazem Abadi, V. Oehler, O. Akintola, J. Draper, M. Meanwell, S. Chakladar, K. Świderek, et al., Revealing the mechanism for covalent inhibition of glycoside hydrolases by carbasugars at an atomic level, Nat. Commun. 9 (2018) 3243.