Biochemical characterization of Bombyx mori α-N-acetylgalactosaminidase belonging to the glycoside hydrolase family 31

Biochemical characterization of *Bombyx mori* α-*N*-acetylgalactosaminidase belonging to the glycoside hydrolase family 31

Marina Ikegaya[1], Takatsugu Miyazaki[1,2*], and Enoch Y. Park[1,2]

[1] Department of Agriculture, Graduate School of Integrated Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan.

[2] Green Chemistry Research Division, Research Institute of Green Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan.

***Correspondence:** Takatsugu Miyazaki, Green Chemistry Research Division, Research Institute of Green Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan; Tel: +81-54-238-4886; E-mail: miyazaki.takatsugu@shizuoka.ac.jp

Running Title: *Bombyx mori* GH31 α-*N*-acetylgalactosaminidase

Acknowledgments

We thank Dr. Jun Takeuchi for helping in the ITC experiment. This work was supported in part by the Japan Society for the Promotion of Science (KAKENHI grant No. 19K15748 to TM). We also thank Enago (www.enago.jp) for the English language review. The authors have no conflicts of interest to declare.

Abstract

Horizontal gene transfer is an important evolutionary mechanism not only for bacteria but also for eukaryotes. In the domestic silkworm *Bombyx mori*, a model species of lepidopteran insects, some enzymes are known to have been acquired by horizontal transfer; however, the enzymatic features of protein BmNag31, belonging to glycoside hydrolase family 31 (GH31) and whose gene was predicted to be transferred from *Enterococcus* sp., are unknown. In this study, we reveal that the transcription of *BmNag31* increases significantly during the prepupal to pupal stage, and decreases in the adult stage. The full-length BmNag31 and its truncated mutants were heterologously expressed in *Escherichia coli* and characterized. Its catalytic domain exhibits α-*N*-acetylgalactosaminidase activity and the carbohydrate-binding module family 32 domain shows binding activity toward *N*-acetylgalactosamine, similar to the *Enterococcus faecalis* homolog, EfNag31A. Gel filtration chromatography and blue native polyacrylamide gel electrophoresis analyses indicate that BmNag31 forms a hexamer whereas EfNag31A is monomeric. These results provide insights into the function of lepidopteran GH31 α-*N*-acetylgalactosaminidase.

Keywords: α-*N*-acetylgalactosaminidase, *Bombyx mori*, glycoside hydrolase family 31, horizontal gene transfer, hexamer

Abbreviations: αGalNAcase, α-*N*-acetylgalactosaminidase; CBM, carbohydrate-binding module; FN3, fibronectin type 3; GalNAc, *N*-acetylgalactosamine; GH, glycoside hydrolase; GH31, glycoside hydrolase family 31; GlcNAc, *N*-acetylglucosamine; HGT, horizontal gene transfer; ITC, isothermal titration calorimetry; MBP, maltose-binding protein

Introduction

Horizontal gene transfer (HGT) is a common means of obtaining new genes in bacteria. There is also a considerable number of genes in animals, including insects, that have been derived from HGT. In insects, some horizontally transferred genes encode functional proteins that may have played an important role in their evolution (Sieber et al., 2017).

Bombyx mori is a model lepidopteran species whose genome has been repeatedly sequenced by several projects (Goldsmith et al., 2005; Kawamoto et al., 2019; Lu et al., 2020). The resulting genomic information has accelerated the pace of research in Lepidoptera, for example, bioinformatic analysis has revealed that *B. mori* and other lepidopteran species possess several genes obtained by HGT (Li et al., 2011; Sun et al., 2013; Zhu et al., 2011). Some of these genes encode functional enzymes, e.g., chitinase BmChi-h (Daimon et al., 2003; Liu et al., 2017), β-fructofuranosidase BmSUC1 (Daimon et al., 2008; Miyazaki et al., 2020), and 4,5-DOPA dioxygenase (Wang et al., 2019). BmChi-h and BmSUC1 are glycoside hydrolases (GHs) belonging to the glycoside hydrolase families GH18 and GH32, respectively, based on their amino acid sequences according to the CAZy database (http://www.cazy.org) (Lombard et al., 2014). BmChi-h is a chitinase expressed in the epidermis and midgut during the molting process, a time when the chitin cuticle undergoes degradation. BmSUC1 is a β-fructofuranosidase that is expressed in the midgut and silk glands to digest sucrose. In *B. mori*, *BmNag31* encodes a glycoside hydrolase belonging to glycoside hydrolase family 31 (GH31). Phylogenetic analysis indicates that the ancestral *BmNag31* was acquired *via* HGT from *Enterococcus* sp., most likely the bacterium *Enterococcus faecalis* (Wheeler et al., 2013). *E. faecalis*, one of the major gut bacteria in the human intestine, is also found in the midgut of silkworms (Chen et al., 2018; Qin et al., 2010). In the silkworm, *BmNag31* is located on the 28th chromosome as a single-copy gene with no introns, thus supporting the HGT-mediated acquisition hypothesis (Li et al., 2011; Wheeler et al., 2013).

GH31 is a large family comprising diverse enzymes with various substrate specificities and reaction mechanisms. It includes GHs such as α-glucosidase (Kashiwabara et al., 2000), α-xylosidase (Lovering et al., 2005), and α-galactosidase (Miyazaki and Park, 2020); and it also includes transglycosidases (Aga et al., 2002) and α-glucan lyases (Rozeboom et al., 2013). Aside from BmNag31, most of the other GH31 proteins encoded on the *B. mori* genome share

homology with metabolic enzymes in animals, such as endoplasmic reticulum α-glucosidase II and acid α-glucosidase, which are involved in *N*-glycan processing (Watanabe et al., 2013; D'Alessio and M. Dahms, 2015) and lysosomal degradation of glycogen (Roig-Zamboni et al., 2017), respectively. However, because of the divergence among the members of GH31, it is difficult to predict the enzymatic activities and substrate specificities of other uncharacterized GH31 enzymes that share low sequence homologies with characterized enzymes.

A recent report indicates that the GH31 proteins, BpGH31 and BcGH31, from the human gut bacteria *Phocaeicola plebeius* (formerly *Bacteroides plebeius*) and *Bacteroides caccae*, respectively, exhibit α-*N*-acetylgalactosaminidase (αGalNAcase) activity, removing the *O*-linked α-*N*-acetylgalactosamine (GalNAc) residue from a glycopeptide but not the blood type A antigen (Rahfeld et al., 2019). Moreover, we recently determined the crystal structure of GH31 αGalNAcase (EfNag31A) from *E. faecalis* (Miyazaki and Park, 2020). The sequence of its xenolog, BmNag31, is 39.9%–52.5% identical to them. However, its physiological role and enzymatic activity remain unclear. To elucidate the function of BmNag31, we cloned the *BmNag31* gene, investigated the properties of the enzyme, and compared them with the properties of EfNag31A.

Results

Primary structure of Nag31 and comparison with orthologs in other lepidopteran species

BmNag31 consists of three domains: the GH31 catalytic domain, FN3 domain, and CBM32 domain. A bacterial homolog, EfNag31A, shares 52.5% sequence identity with BmNag31, and has a similar domain organization, except that the dockerin, cohesin, and repeating FIVAR domains, and the transmembrane regions follow the CBM32 domain at the C terminus. It should be noted that EfNag31A possesses an N-terminal secretion signal peptide, whereas BmNag31 has no signal peptide based on SignalP prediction server (Fig. 1). In addition, DeepLoc-1.0 and SilkDB predicted that the BmNag31 is localized in cytoplasm (Fig. S1). These features imply that BmNag31 is not a secretory protein, thus it differs from EfNag31A, which may be secreted and works on the cell surface of *E. faecalis* (Miyazaki and Park, 2020).

Our BLAST searches against the InsectBase and NCBI protein databases yielded 19 sequences which we aligned with the EfNag31A sequence. All lepidopteran BmNag31 homologs (Nag31s), as well as BmNag31, were predicted to have no signal peptide at their N termini (Fig. S2). The catalytic residues (predicted nucleophile Asp407 and general acid/base Asp460 in BmNag31) and residues involved in GalNAc recognition were mostly conserved in all available lepidopteran Nag31 proteins, except for the Nag31 protein from *Manduca sexta* (Fig. S2). All available lepidopteran Nag31s have no additional C-terminal domain following the CBM32 domain. By comparing the gene sequences between *B. mori* and *Danaus plexippus*, Wheeler et al. estimated that the HGT between a bacterium and their common ancestor lepidopteran insect occurred at least 65 million years ago (Wheeler et al., 2013). Many lepidopteran genome analyses showed that Nag31 genes are widely distributed in many lepidopteran insects. The estimated divergence time of the most phylogenetically distant species, *Plutella* and *Bombyx*, is about the middle of the Cretaceous (Kawahara et al., 2019), suggesting that HGT occurred 120 million years ago or earlier.

Expression analysis of BmNag31

The transcription levels of *BmNag31* in the Malpighian tubule, testis, and ovary were 3–5 times higher than that in the fat body of day-3 fifth-instar larvae (Fig. 2A). The transcription levels of *BmNag31* remained stable in the whole bodies of the first- to fifth-instar larvae. Then, the expression level increased remarkably (up to 60 times higher than that of the day-3 fifth-instar larva) from the prepupal stage until day-4 pupa. The expression level then significantly decreased at the adult stage (Fig. 2B). Therefore, the transcription levels of BmNag31 were suggested to depend on the developmental stages rather than the organs. Because *BmNag31* gene is predicted to have been derived from the midgut symbiont *Enterococcus* sp., the expression levels of *BmNag31* in the midgut during different stages were also examined. In the midgut, the expression level of *BmNag31* also increased during the pupal stage (up to 11 times higher than that of the day-3 fifth-instar larva), but this change was minor compared to that of the whole body (Fig. 2C).

Oligomerization of BmNag31

The recombinant proteins, maltose-binding protein-fused BmNag31 (MBP-BmNag31) and the GH31 catalytic domain containing a fibronectin type 3 (FN3) domain (named BmGH31) were successfully expressed in *E. coli* and purified by affinity chromatography (Fig. S3), whereas the full-length BmNag31 lacking the MBP tag was difficult to obtain due to the protein's insolubility. Therefore, MBP-BmNag31 and BmGH31 were used for this study. The calculated molecular masses of MBP-BmNag31 and BmGH31 (based on the amino acid sequences) are 166 and 106 kDa, respectively. However, the molecular mass of BmGH31 as determined by gel filtration is 655 kDa, suggesting that BmGH31 forms a hexamer in buffer solution (Figs. 3A and S4). MBP-BmNag31 eluted at a point lower than the highest molecular weight marker protein (thyroglobulin, 660 kDa), thus the molecular weight of MBP-BmNag31 could not be determined accurately but was estimated to be more than 787 kDa. Moreover, the blue native PAGE analysis supported the hexameric state based on the result of gel filtration chromatography (Fig. 3B). In the case of the catalytic domain of EfNag31A (EfGH31), gel filtration results suggest that the molecular masses of its forms with and without the CBM32 domain (theoretical molecular weights = 122 kDa and 107 kDa, respectively) are 117 and 99.7 kDa, respectively, indicating that EfGH31 is monomeric.

Kinetic analysis and substrate specificity

The hydrolytic activities of both MBP-BmNag31 and BmGH31 were tested against various *para*-nitrophenyl (pNP) glycosides. The enzymes hydrolyzed *p*-nitrophenyl *N*-acetyl-α-D-galactosaminide (GalNAcα-pNP) but not GlcNAc-β-1,3-GalNAcα-pNP and the other substrate, indicating that BmNag31 is an exo-acting αGalNAcase, which matches the activity of its bacterial homologs (Cao et al., 2020; Rahfeld et al., 2019). In addition, BmGH31 was active toward bovine submaxillary mucin (Fig. S5) whereas blood type A antigen triaose (GalNAcα1-3(Fucα1-2)Gal) and Tn antigen (GalNAcα-Ser) were not hydrolyzed by the enzyme (data not shown). The k_{cat} values of MBP-BmNag31 and BmGH31 against GalNAcα-pNP are 3.23 ± 0.10 and 4.81 ± 0.11 s^{-1}, respectively; while their K_m values are 610 ± 10 and 440 ± 5 μM, respectively. Thus, the truncation of CBM32 domain did not significantly affect the hydrolytic activity of BmNag31. The k_{cat}/K_m values of BmNag31 are 3–8 times lower than that of the Nag31 bacterial homologs due to the latter's relatively high K_m values (Table 1). For both

recombinant enzymes, the optimum pH and temperature for the hydrolysis of GalNAcα-pNP were 6.0–6.5 and 45°C–50ºC, respectively. Both recombinant enzymes were stable (>90% residual activity) at pH values ranging from 4 to 10 at 4°C for 20 hours. Similarly, they were stable up to 45°C after 30 minutes of incubation (Fig. S6).

Binding activity of BmCBM32 toward monosaccharides

The binding activities of a carbohydrate-binding module family 32 (CBM32) domain (BmCBM32) and a CBM32 domain of EfNag31A (EfCBM32) toward GalNAc, D-galactose, L-fucose, and *N*-acetylglucosamine (GlcNAc) were examined by ITC. BmCBM32 and EfCBM32 bound GalNAc, with association constants (K_a) of 260 ± 11.9 M^{-1} and 266 ± 24.6 M^{-1}, respectively (Fig. 4 and Table 2). Both proteins did not bind to the other sugars tested (Fig. 4). The K_a value of BmCBM32 is 10 times lower than the values reported for CBM32s shown in Table 2 (Ficko-Blean and Boraston, 2009, 2006; Grondin et al., 2017), except for CpCBM32-1 (Table 2). Compared to various characterized CBM32s, the amino acid sequences of BmCBM32 and EfCBM32 match most closely with CBM32-1 from *Clostridium perfringens* (CpCBM32-1) (26.1% for EfCBM32 and 25.2% for BmCBM32). In summary, our results indicate that the binding activity of BmCBM32 to GalNAc is almost the same as EfCBM32 and is similar to CpCBM32-1, consistent with their sequence homology.

Homology models of BmGH31 and BmCBM32

Using the structure of EfGH31 as a template, we used the SWISS-MODEL server to model the BmGH31 structure. EfGH31 consists of the following five domains: N-domain (β-sandwich), catalytic A-domain ($(\beta/\alpha)_8$ barrel), proximal C-domain (β-sandwich), distal C-domain (β-sandwich), and FN3 domain. The A'-subdomain is inserted within the A-domain (Cao et al., 2020). The homology model of BmGH31 shows the same domain architecture as EfGH31 (Fig. 5A). In contrast to the domain architecture, the surface electrostatic potentials differ significantly between BmGH31 and EfGH31 (Fig. 5B and C). The electrostatic potentials of the A-, N-, and proximal C-domains lying on the same side as the active site are all positively charged on BmGH31. This electron bias is absent in the corresponding area of EfGH31. On the opposite side the protein, most of the surface area is either positively charged or non-charged

199 in EfGH31. In contrast, many negatively charged residues are present in the corresponding area
200 of BmGH31 (Fig. 5B). Because surface electrostatic potential is an important factor
201 determining oligomerization (Ali and Imperiali, 2005), this electron potential bias of protein
202 surface may explain the difference between the oligomeric states of EfNag31A and BmNag31.
203 The catalytic residues, Asp407 and Asp460, in the A-domain are configured such that they
204 can interact with GalNAc in the substrate-binding pocket (Fig. 6A). The residues interacting
205 with GalNAc except for Val444 are structurally conserved between BmGH31 and EfGH31.
206 This observation supports the strict specificity for α-*N*-acetylgalactosaminide substrate. The
207 homology model of BmCBM32 was also constructed by the SWISS-MODEL server using the
208 crystal structure of CpCBM32-1, a domain of *C. perfringens* putative Nag31 (Grondin et al.,
209 2017). Although the structure of CpCBM32-1 in complex with GalNAc was not determined,
210 13 amino acid residues were identified as candidate amino acid residues involved in GalNAc
211 recognition based on nuclear magnetic resonance spectroscopy (Grondin et al., 2017). The
212 amino acid residues involved in binding in CBM32 proteins are diverse and difficult to predict
213 from their primary structure due to their low sequence homology (Fig. S7). Superimposition of
214 the homology models and the crystal structures indicates that only three amino acid residues
215 corresponding to His990, Asn1022, and Phe1085 in CpCBM32-1 were conserved in
216 BmCBM32 and EfCBM32 (Fig. 6B). Compared with another domain CpCBM32-3 of *C.*
217 *perfringens* Nag31 in complex with GalNAc, only two residues, His961 and Asn1003, of
218 BmCBM32 are conserved but Phe1059 is substituted to tyrosine in CpCBM32-3.
219
220 **Discussion**
221 This study demonstrates the hydrolytic activity of BmNag31 toward GalNAcα-pNP with
222 similar efficiency as the homolog EfNag31A, supported by conservation of the active site
223 residues identified by the primary sequence alignment and the structural homology modeling.
224 In contrast, the BmCBM32 showed the affinity toward GalNAc that is similar to those of
225 EfCBM32 and CpCBM32-1 but is lower than those of CpCBM32-2 and the other GalNAc-
226 recognizing CBM32 proteins. The variety in residues involved in GalNAc-binding may cause
227 the difference of binding affinity to GalNAc. These results indicate that *BmNag31* encodes a
228 functional protein with similar characters as EfNag31A.

One notable difference between BmNag31 and EfNag31A is the oligomeric state of the former, which was observed for both recombinant MBP-BmNag31 and BmGH31 proteins. This observation indicates that the C-terminal CBM32 domain is not necessary for hexamer formation and the addition of the N-terminal MBP tag (42.5 kDa) does not affect oligomerization. Therefore, the N terminus probably locates on the surface of the hexamer. Three prokaryotic GH31 enzymes, α-xylosidase YicI from *E. coli* (Lovering et al., 2005), α-glucosidase MalA from *Sulfolobus solfataricus* (Ernst et al., 2006), and α-glucosidase from *Bacillus thermoamyloliquefaciens* (Kashiwabara et al., 2000), exist in hexameric form. YicI is a dimer consisting of trimers and MalA is a trimer consisting of dimers. Both form a cage-like hexamer in which the active sites face the inner side of the hexamer. One possible contribution of the hexamerization is protein stabilization (Ali and Imperiali, 2005). Another one is that the multiple CBM32 domains would make BmNag31 stay in the vicinity of GalNAc-exposed O-glycoprotein substrates. Details on the hexamer structure of BmNag31 will require further investigation using X-ray crystallography, small-angle X-ray scattering, and cryo-electron microscopy.

The Nag31s (BcGH31 and BpGH31) in bacteria isolated from human feces work with other carbohydrate-active enzymes in cleaving peptide-linked GalNAc for uptake (Rahfeld et al., 2019). The lumen of the mammalian intestine is covered by a highly developed mucosal layer that is mainly composed of mucins, i.e., heterologous *O*-GalNAc glycans attached to proteins (Bergstrom and Xia, 2013). *O*-GalNAc glycans are also present in insects, where they are structurally less complex than their mammalian counterpart, but are essential in many biological processes (Li et al., 2020; Walski et al., 2017). In *B. mori*, *N*-acetylgalactosaminyltransferase transfers GalNAc from uridine diphosphate GalNAc to a specific residue of a polypeptide in the Golgi apparatus during the first step of *O*-glycosylation. The enzyme has been identified and biochemically characterized previously (Xu et al., 2018); however, many aspects of *O*-glycosylation in insects are still unclear.

Here, we reveal that the expression level of *BmNag31* increases during the pupal stage and then decreases at the adult stage, suggesting that the transcription level of *BmNag31* might be controlled by transcription factors related to pupation. When the silkworm transforms from larva to pupa, its organs are reconstructed dynamically. One possible biological role of

BmNag31 is the degradation of peptide-linked *O*-glycans. In pupation, cell components are degraded by lysosomal enzymes during phagocytosis or the fusion of autophagosomes with lysosomes (Tettamanti and Casartelli, 2019). The pH-optimum of lysosomal enzymes is usually acidic (Winchester, 2005), but the optimum pH of BmNag31 is 6–6.5, and its hydrolytic activity decreased by 50% in a pH 5 buffer solution. Moreover, BmNag31 has no signal sequence related to transportation to the endoplasmic reticulum. To our knowledge, α-*N*-acetylgalactosaminidase from other insects have not been reported, but insects including *B. mori* have genes for proteins belonging to the GH27 family, which contains vertebrate lysosomal αGalNAcase (Wang et al., 1990). Further investigation will be required to reveal the activity of these proteins and their relationship in insects.

In conclusion, *BmNag31* is remarkably expressed during the pupal stage of *B. mori* and encodes a functional αGalNAcase with the CBM32 domain that showed affinity toward GalNAc. Furthermore, considering together with the fact that Nag31 homologs are widely distributed among lepidopteran insects, Nag31 proteins may play a role mainly in pupation process. This is the first report of a eukaryotic GH31 αGalNAcase and a eukaryotic hexameric GH31 enzyme and will contribute to understanding the metabolism of carbohydrates in Lepidoptera.

Experimental Procedures

Chemicals and strains

p-Nitrophenyl α-D-galactopyranoside, *p*-nitrophenyl β-D-fucopyranoside, *p*-nitrophenyl β-D-glucuronide, *p*-nitrophenyl β-D-xylopyranoside, *p*-nitrophenyl β-cellobioside, and GlcNAc-β-1,3-GalNAcα-pNP were purchased from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). *p*-Nitrophenyl α-L-fucopyranoside, *p*-nitrophenyl α-D-glucopyranoside, *p*-nitrophenyl *N*-acetyl-β-D-glucosaminide, *p*-nitrophenyl *N*-acetyl-β-D-galactosaminide, *p*-nitrophenyl α-D-mannopyranoside were from Merck (Darmstadt, Germany). *N*-Acetylgalactosamine, *p*-nitrophenyl β-D-glucopyranoside, and *p*-nitrophenyl β-D-galactopyranoside were from FUJIFILM Wako Pure Chemical Co. (Osaka, Japan). *p*-

Nitrophenyl *N*-acetyl-α-D-glucopyranoside, *p*-nitrophenyl α-L-rhamnopyranoside, and *p*-nitrophenyl α-D-xylopyranoside were from Carbosynth (Berksher, UK). Fucose and galactose were from Nacalai Tesque (Kyoto, Japan). *p*-Nitrophenyl β-D-mannopyranoside (Megazyme USA), GalNAcα-pNP (Cayman Chemical, Michigan, USA), blood type A antigen triaose (Elicityl, Crolles, France), and Tn antigen (Dextra Laboratories Ltd., Thames Valley Science Park, UK) were also used in this study. Silkworms (*B. mori* F1 hybrid Fuyo × Tsukubane) purchased from Ehime Sanshu Inc. (Ehime, Japan) were raised on an artificial diet, Silkmate S2 (Nohsan Corporation, Yokohama, Japan), in an incubator at 25°C.

In silico analysis

Using the amino acid sequence of BmNag31 from SilkBase (http://silkbase.ab.a.u-tokyo.ac.jp/cgi-bin/index.cgi) we searched for similar sequences in the NCBI server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and InsectBase (http://www.insect-genome.com/) database (Livak and Schmittgen, 2001). The resulting lepidopteran sequences and the sequence of EfNag31A (GenBank id EOK08638.1) were aligned using the ClustalW program running within MEGA X software (Kumar et al., 2018) and the alignment diagram was generated using ESPript 3.0 (Robert and Gouet, 2014). SignalP 5.0 (http://www.cbs.dtu.dk/services/SignalP/) was used to predict the location of the signal peptide. Subcellular localization of BmNag31 was predicted by DeepLoc-1.0 (http://www.cbs.dtu.dk/services/DeepLoc) (Armenteros et al., 2017). Homology models of BmNag31, BmCBM32, and EfCBM32 were generated by the SWISS-MODEL server (https://swissmodel.expasy.org) using the coordinate of EfGH31 (PDB 6M77) or the CBM32-1 domain from *Clostridium perfringens* (PDB 4LPL) as a template. The surface electron potential was calculated using the APBD electro-statistics plugin of PyMOL software (https://pymol.org).

Transcriptional analysis

Day-3 fifth-instar silkworm larvae were dissected; their tissues were washed in phosphate-buffered saline (PBS) and immediately frozen in liquid nitrogen, then stored at −80°C. Total RNA was extracted from the whole body or individual tissues of silkworms using

TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA), followed by treatment with DNase I (Nippon Gene, Tokyo, Japan). First-strand cDNA was synthesized from 500 ng of RNA using PrimeScript RT reagent kit (Takara Bio, Shiga, Japan). Specific genes were quantified by quantitative PCR (qPCR) using the primers listed in Table S1 and the SYBR Green qPCR reagent (Thermo Fisher Scientific). The expression levels of *BmNag31* were normalized relative to that of the glyceraldehyde-3-phosphate dehydrogenase (BmGAPDH) gene (Guo et al., 2016) and the expression ratio of *BmNag31* was calculated by the $2^{-\Delta Ct}$ method (Pfaffl, 2001).

Plasmid construction

Genomic DNA of a fifth-instar silkworm larva was extracted using the ISOGENOME (Nippon Gene) genome extraction reagent. The DNA fragment encoding BmNag31 was amplified from the genomic DNA using the primers BmGH31_pET21a_F and BmGH31_pET21a_R (Table S1). A pET21a vector (Merck) was linearized by NdeI and XhoI restriction enzymes and then ligated with the fragment encoding BmNag31 using the In-Fusion cloning kit (Takara Bio). To construct the plasmid used to express MBP-BmNag31, first, the DNA fragment encoding BmNag31 with a C-terminal hexahistidine tag was amplified using the primer pair BmGH31_NcoI_F and Ct-His6-SalI_R, and BmNag31-harboring pET21a as a template. The amplicons were then ligated into the pMAL-c5x vector (New England Biolabs, Ipswich, MA, USA). We then constructed plasmids expressing BmGH31 (residues 1–925) and BmCBM32 domain (residues 926–1069). First, DNA fragments encoding the domains were amplified from plasmid pET21a harboring BmNag31 using the primer pairs BmGH31_pET21a_F and BmGH31ΔCBM32_XhoI_R, and BmCBM32_NheI_F and BmCBM32_XhoI_R. The resulting amplicons were then ligated into pET28a vectors (Merck) using NdeI and XhoI restriction sites, and NheI and XhoI sites, respectively. The expression plasmid for EfCBM32 (residues 981–1126) was generated by inverse PCR using the EfNag31A-harboring pET28a as a template, which was constructed in a previous study (Miyazaki and Park, 2020). The identities of all DNAs were confirmed by sequencing and the nucleotide sequence of *BmNag31* was submitted to the DDBJ/EMBL/GenBank databases with the accession number LC581276.

Recombinant expression and purification

E. coli BL21 (DE3) harboring the desired plasmid was grown at 37°C to an optical density (600 nm) of 0.6 in 1 L of Luria-Bertani medium containing 50 μg/mL of kanamycin or 50 μg/mL of carbenicillin. After cooling the medium to 20°C on ice, expression was induced by the addition of 0.1 mM IPTG at 20°C for 20 hours. The cells were harvested by centrifugation for 10 minutes (4°C 5,000 × g) and stored at −20°C. The cell pellet was resuspended in two types of purification buffers: PBS for MBP-BmNag31 or 50 mM sodium phosphate buffer (pH 8.0) containing 300 mM NaCl and 20 mM imidazole for the other His-tagged proteins. Following resuspension, the cells were disrupted by ultrasonication and then centrifuged to remove insoluble materials. To purify MBP-BmNag31, the supernatant was loaded onto an amylose resin (New England Biolabs) column pre-equilibrated with PBS. The column was then washed with PBS and the protein was eluted with PBS containing 10 mM maltose. Using 30-kDa cutoff Amicon Ultra centrifugal units (Merck), we replaced the purified MBP-BmNag31 buffer with 50 mM sodium phosphate buffer (pH 8.0) containing 300 mM NaCl and 20 mM imidazole. Purified MBP-BmNag31, along with other His-tagged proteins in the cell lysate supernatant, were loaded onto a Ni-NTA Agarose (Qiagen, Manchester, UK) column, and the unbound proteins were washed with the same buffer containing 20–50 mM imidazole. Proteins were eluted with the same buffer containing 100 mM imidazole and then concentrated by ultrafiltration using Amicon Ultra centrifugal units and 50 mM sodium phosphate buffer (pH 6.5) containing 300 mM NaCl. MBP-BmNag31, BmCBM32, and EfCBM32 were then dialyzed with 50 mM 4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid (HEPES) buffer (pH 7.0). Protein purity was confirmed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Protein concentration was calculated from the absorbance at 280 nm using the theoretical extinction coefficients predicted by the ProtParam tool (https://web.expasy.org/protparam/) based on their amino acid sequences.

Molecular weight determination

Gel filtration chromatography was performed using the ÄKTA explorer 10S system (GE Healthcare, Chicago, IL, USA). Purified MBP-BmNag31 and BmGH31 were concentrated to

5 mg/mL with Amicon Ultra centrifugation filter units, then the concentrated proteins were applied onto a HiLoad 16/60 Superdex 200 prep grade column and eluted by 1.2 column volumes of 20 mM sodium phosphate buffer (pH 7.0) containing 300 mM NaCl. EfNag31A and EfGH31 were also analyzed by gel filtration chromatography according to previously published procedures (Miyazaki and Park, 2020). For blue native PAGE, 1 mg/mL of each protein and marker protein solution (HMW Native Marker Kit, GE Healthcare) were individually mixed with an equal volume of sample buffer [1% (w/v) Coomassie Brilliant Blue G-250, 20 mM 6-aminocaproic acid, 20 mM Bis-Tris–HCl (pH 7.0), 20% (v/v) glycerol] and then incubated on ice for 5 minutes. Ten microliters of each sample was applied on 3%–10% PAGEL native PAGE gels (ATTO, Tokyo, Japan) and electrophoresed using 50 mM Bis-Tris–HCl (pH 7.0) as the anode buffer and 0.02% Coomassie Brilliant Blue G-250, 50 mM Tricine, 15 mM Bis-Tris–HCl (pH 7.0) as the cathode buffer. After electrophoresis, the gel was destained using a 25% methanol–7.5% acetic acid solution.

Enzyme assays

Hydrolytic activity toward various *p*-nitrophenyl glycosides was measured in 50 μL reaction mixtures containing 25.6 μg/mL MBP-BmNag31 or 12.8 μg/mL BmGH31, 1 mM of a substrate, and 20 mM sodium phosphate buffer (pH 7.0) at 30°C. To examine the effect of pH on hydrolytic activity, reaction mixtures containing 11 μg/mL of MBP-BmNag31 or 6.4 μg/mL BmGH31 and 1 mM GalNAcα-pNP were prepared with McIlvaine (sodium citrate–phosphate) buffer at pH 4.0–8.0 or in glycine-HCl buffer at pH 9.0–10. The mixtures were incubated for 10 min at 30°C. The effect of temperature on hydrolytic activity was examined using 50 mM sodium phosphate buffer (pH 6.5) containing 100 mM NaCl and 1 mM GalNAcα-pNP. Mixtures were incubated at temperatures ranging from 20°C–60°C. We tested the enzymes' pH stability by incubating 400 μg/mL of both enzymes for 20 hours in McIlvaine buffer (pH 4.0–8.0) or glycine-HCl buffer (pH 9.0–10). We tested the enzymes' thermal stability by incubating 150 μg/mL of MBP-BmNag31 and 150 μg/mL of BmGH31 at 4°C–60°C for 30 minutes in 50 mM sodium phosphate buffer containing 300 mM NaCl (pH 6.5). The residual activities for pH stability and thermostability were measured at 30°C with the reaction mixtures containing 9 μg/mL and 15 μg/mL of each protein, respectively, and 50 mM sodium phosphate buffer (pH

409 7.0) containing 1 mM of GalNAcα-pNP. The mixtures were incubated at 30°C for experiments
410 testing pH stability and for thermostability. The initial hydrolytic reaction velocities of
411 GalNAcα-pNP were determined using 50 mM sodium phosphate buffer (pH 6.0) containing
412 five concentrations (0.1–2.0 mM) of GalNAcα-pNP at 40°C. All reactions above were
413 performed in triplicate and quenched by adding 100 μL of 1 M Na_2CO_3. The amount of released
414 *p*-nitrophenol was measured at 405 nm. Kinetic parameters were calculated by fitting to the
415 Michaelis–Menten equation using nonlinear regression analysis by KaleidaGraph software
416 (Synergy Software, Reading PA, USA). The hydrolytic activity of the enzymes toward blood
417 group A antigen triaose was evaluated in 10-μL reaction mixtures containing 10 mM of the
418 oligosaccharide and 150 μg/mL of BmGH31 in 50 mM sodium phosphate buffer (pH 7.0). For
419 the Tn antigen, the reaction mixture contained 6 mM Tn antigen and 100 μg/mL of BmGH31
420 or EfGH31. All reaction mixtures were incubated at 30°C overnight, followed by thin-layer
421 chromatography (TLC) using TLC Silica Gel 60 F_{254} TLC plates (Merck). To visualize the
422 carbohydrates, the TLC plate was sprayed with 10% sulfuric acid in methanol and then baked.
423 5 mg of mucin from bovine submaxillary gland (Merck) was pre-treated with 250 units/mL of
424 *Clostridium perfringens* neuraminidase, 25 units/mL of *Streptomyces plicatus* β-*N*-
425 acetylhexosaminidase (New England Biolabs), and 25 μg/mL of *Streptococcus pneumoniae* β-
426 galactosidase 35A (NZYTech, Lisbon, Portugal) in 1 mL of reaction mixture for 20 hours at
427 30 °C, and then the reaction mixture was concentrated by centrifugal evaporation and the mucin
428 was precipitated twice with 80% ethanol to remove the released sugars. The precipitate was
429 dissolved in water and incubated with BmGH31 for 20 hours at 30 °C. The reaction mixture
430 was mixed with four volumes of ethanol, followed by centrifugation to remove proteins, and
431 then supernatant was concentrated by centrifugal evaporation. Reaction products were analyzed
432 by TLC and developed in a mixture of 1-butanol/acetic acid/water (2 : 1 : 1 by volume).
433 Released sugars were visualized using diphenylamine/aniline/phosphoric acid reagent
434 (Anderson et al., 2000).
435
436 **Isothermal titration calorimetry**
437 The binding affinities of CBM32 domains toward GalNAc, D-galactose, L-fucose, and
438 GlcNAc were analyzed by isothermal titration calorimetry (ITC) using a MicroCal iTC200

(Malvern Panalytical Ltd, Enigma Business Park, UK). BmCBM32 and EfCBM32 were dialyzed against 50 mM HEPES buffer (pH 7.0) and filtered using 0.45 μm filters (Merck). Titrations were performed at 25°C by injecting 2 μL aliquots of 10 mM ligand dissolved in the 50 mM HEPES buffer (pH 7.0) into a cell containing 0.1 mM CBM32 protein. The resulting heat release was recorded and the titration data were analyzed using MicroCal Origin ITC software (Malvern Panalytical Ltd). Thermodynamic parameters were obtained by nonlinear least-squares fitting of experimental data using the one set sites binding model.

Author contributions

TM conceived and supervised the study; MI and TM performed experiments; all authors analyzed the data; MI and TM wrote and revised the manuscript; all authors read and approved the manuscript.

References

Aga, H., Maruta, K., Yamamoto, T., Kubota, M., Fukuda, S., Kurimoto, M., Tsujisaka, Y., (2002) Cloning and Sequencing of the Genes Encoding Cyclic Tetrasaccharide-synthesizing Enzymes from *Bacillus globisporus* C11. *Bioscience, Biotechnology, and Biochemistry*, **66**, 1057–1068. https://doi.org/10.1271/bbb.66.1057

Anderson, K., Li, S., Li, Y., (2000) Diphenylamine–Aniline–Phosphoric Acid Reagent, a Versatile Spray Reagent for Revealing Glycoconjugates on Thin-Layer Chromatography Plates. *Analytical Biochemistry*, **287**, 337-339, https://doi.org/10.1006/abio.2000.4829

Armenteros, J.J.A., Sønderby, C. K., Sønderby, S. K., Nielsen, H., Winther, O., (2017) DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*, **33**, 3387–3395. https://doi.org/10.1093/bioinformatics/btx431

Ali, M.H., Imperiali, B., (2005) Protein oligomerization: How and why. *Bioorganic & Medicinal Chemistry*, **13**, 5013–5020. https://doi.org/10.1016/j.bmc.2005.05.037

Bergstrom, K.S.B., Xia, L., (2013) Mucin-type *O*-glycans and their roles in intestinal homeostasis. *Glycobiology*, **23**, 1026–1037. https://doi.org/10.1093/glycob/cwt045

Cao, H., Walton, J.D., Brumm, P., Phillips, G.N., (2020) Crystal structure of α-xylosidase from *Aspergillus niger* in complex with a hydrolyzed xyloglucan product and new insights in accurately predicting substrate specificities of GH31 family glycosidases. *ACS Sustainable Chemistry and Engineering*, **8**, 2540–2547. https://doi.org/10.1021/acssuschemeng.9b07073

Chen, B., Yu, T., Xie, S., Du, K., Liang, X., Lan, Y., Sun, C., Lu, X., Shao, Y., (2018) Comparative shotgun metagenomic data of the silkworm *Bombyx mori* gut microbiome. *Scientific Data*, **5**, 180285. https://doi.org/10.1038/sdata.2018.285

Daimon, T., Hamada, K., Mita, K., Okano, K., Suzuki, M.G., Kobayashi, M., Shimada, T., (2003) A *Bombyx mori* gene, *BmChi-h*, encodes a protein homologous to bacterial and baculovirus chitinases. *Insect Biochemistry and Molecular Biology*, **33**, 749–759. https://doi.org/10.1016/S0965-1748(03)00084-5

Daimon, T., Taguchi, T., Meng, Y., Katsuma, S., Mita, K., Shimada, T., (2008) β-Fructofuranosidase genes of the silkworm, *Bombyx mori*: Insights into enzymatic adaptation of *B. mori* to toxic alkaloids in mulberry latex. *Journal of Biological Chemistry*, **283**, 15271–15279. https://doi.org/10.1074/jbc.M709350200

D'Alessio, C., M. Dahms, N., (2015) Glucosidase II and MRH-Domain Containing Proteins in the Secretory Pathway. *Current Protein & Peptide Science*, **16**, 31–48. https://doi.org/10.2174/1389203716666150213160438

Ernst, H.A., lo Leggio, L., Willemoës, M., Leonard, G., Blum, P., Larsen, S., (2006) Structure of the *Sulfolobus solfataricus* α-Glucosidase: Implications for Domain Conservation and Substrate Recognition in GH31. *Journal of Molecular Biology,* 358, 1106–1124. https://doi.org/10.1016/j.jmb.2006.02.056

Ficko-Blean, E., Boraston, A.B., (2009) *N*-Acetylglucosamine Recognition by a Family 32 Carbohydrate-Binding Module from *Clostridium perfringens* NagH. *Journal of Molecular Biology*, **390**, 208–220. https://doi.org/10.1016/j.jmb.2009.04.066

Ficko-Blean, E., Boraston, A.B., (2006) The Interaction of a Carbohydrate-binding Module from a *Clostridium perfringens N*-Acetyl-β-hexosaminidase with Its Carbohydrate Receptor. *Journal of Biological Chemistry*, **281**, 37748–37757. https://doi.org/10.1074/jbc.M606126200

Goldsmith, M.R., Shimada, T., Abe, H., (2005) The genetics and genomics of the silkworm, *Bombyx mori*. *Annual Review of Entomology*, **50**, 71–100. https://doi.org/10.1146/annurev.ento.50.071803.130456

Grondin, J.M., Duan, D., Kirlin, A.C., Abe, K.T., Chitayat, S., Spencer, H.L., Spencer, C., Campigotto, A., Houliston, S., Arrowsmith, C.H., Allingham, J.S., Boraston, A.B., Smith, S.P., (2017) Diverse modes of galacto-specific carbohydrate recognition by a family 31 glycoside hydrolase from *Clostridium perfringens*. *PLoS ONE*, **12**. https://doi.org/10.1371/journal.pone.0171606

Guo, H., Jiang, L., Xia, Q., (2016) Selection of reference genes for analysis of stress-responsive genes after challenge with viruses and temperature changes in the silkworm *Bombyx mori*. *Molecular Genetics and Genomics*, **291**, 999–1004. https://doi.org/10.1007/s00438-015-1125-4

Kashiwabara, S., Azuma, S., TsudukiI, M., Suzuki, Y., (2000) The Primary Structure of the

Subunit in *Bacillus thermoamyloliquefaciens* KP1071 Molecular Weight 540,000 Homohexameric α-Glucosidase II Belonging to the Glycosyl Hydrolase Family 31. *Bioscience, Biotechnology, and Biochemistry*, **64**, 1379–1393. https://doi.org/10.1271/bbb.64.1379

Kawahara, A.Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E.F.A., Donath, A., Gimnich, F., Frandsen, P.B., Zwick, A., dos Reis, M., Barber, J.R., Peters, R.S., Liu, S., Zhou, X., Mayer, C., Podsiadlowski, L., Storer, C., Yack, J.E., Misof, B., Breinholt, J.W., (2019) Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 22657-22663. https://doi.org/10.1073/pnas.1907847116

Kawamoto, M., Jouraku, A., Toyoda, A., Yokoi, K., Minakuchi, Y., Katsuma, S., Fujiyama, A., Kiuchi, T., Yamamoto, K., Shimada, T., (2019) High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, **107**, 53–62. https://doi.org/10.1016/j.ibmb.2019.02.002

King, B.R., Vural, S., Pandey, S., Barteau, A., Guda, C., (2012) ngLOC: software and web server for predicting protein subcellular localization in prokaryotes and eukaryotes. *BMC Research Notes*, **5**, 351. https://doi.org/10.1186/1756-0500-5-351

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, **35**, 1547–1549. https://doi.org/10.1093/molbev/msy096

Li, W., de Schutter, K., van Damme, E.J.M., Smagghe, G., (2020) Synthesis and biological roles of O-glycans in insects. *Glycoconjugate Journal*, **37**, 47–56. https://doi.org/10.1007/s10719-019-09867-1

Li, Z.W., Shen, Y.H., Xiang, Z.H., Zhang, Z., (2011) Pathogen-origin horizontally transferred genes contribute to the evolution of Lepidopteran insects. *BMC Evolutionary Biology*, **11**, 356. https://doi.org/10.1186/1471-2148-11-356

Liu, T., Chen, L., Zhou, Y., Jiang, X., Duan, Y., Yang, Q., (2017) Structure, catalysis, and inhibition of OfChi-h, the lepidoptera-exclusive insect chitinase. *Journal of Biological Chemistry*, **292**, 2080–2088. https://doi.org/10.1074/jbc.M116.755330

Livak, K.J., Schmittgen, T.D., (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods*, **25**, 402–408. https://doi.org/10.1006/meth.2001.1262

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, **42**, D490–D495. https://doi.org/10.1093/nar/gkt1178

Lovering, A.L., Lee, S.S., Kim, Y.-W., Withers, S.G., Strynadka, N.C.J., (2005) Mechanistic and Structural Analysis of a Family 31 α-Glycosidase and Its Glycosyl-enzyme Intermediate. *Journal of Biological Chemistry*, **280**, 2105–2115.

https://doi.org/10.1074/jbc.M410468200

Lu, F., Wei, Z., Luo, Y., Guo, H., Zhang, G., Xia, Q., Wang, Y., (2020) SilkDB 3.0: Visualizing and exploring multiple levels of data for silkworm. *Nucleic Acids Research*, **48**, D749–D755. https://doi.org/10.1093/nar/gkz919

Miyazaki, T., Park, E.Y., (2020) Crystal structure of the *Enterococcus faecalis* α-*N*-acetylgalactosaminidase, a member of the glycoside hydrolase family 31. *FEBS Letters*, **594**, 2282–2293. https://doi.org/10.1002/1873-3468.13804

Miyazaki T., Oba N., Park. E.Y., (2020) Structural insight into the substrate specificity of *Bombyx mori* β-fructofuranosidase belonging to the glycoside hydrolase family 32. *Insect biochemistry Molecular Biology*, **127,** 103494

Pfaffl, M.W., (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research,* **29**, 45e–445. https://doi.org/10.1093/nar/29.9.e45

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, Shaochuan, Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, Shengting, Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, Songgang, Qin, N., Yang, H., Wang, Jian, Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D., Wang, Jun, (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* **464**, 59–65. https://doi.org/10.1038/nature08821

Rahfeld, P., Wardman, J.F., Mehr, K., Huff, D., Morgan-Lang, C., Chen, H.-M., Hallam, S.J., Withers, S.G., (2019) Prospecting for microbial α-*N*-acetylgalactosaminidases yields a new class of GH31 *O*-glycanase. *Journal of Biological Chemistry,* **294**, 16400–16415. https://doi.org/10.1074/jbc.RA119.010628

Robert, X., Gouet, P., (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research,* **42**, W320–W324. https://doi.org/10.1093/nar/gku316

Roig-Zamboni, V., Cobucci-Ponzano, B., Iacono, R., Ferrara, M.C., Germany, S., Bourne, Y., Parenti, G., Moracci, M., Sulzenbacher, G., (2017) Structure of human lysosomal acid α-glucosidase–a guide for the treatment of Pompe disease. *Nature Communications*, **8**, 1111. https://doi.org/10.1038/s41467-017-01263-3

Rozeboom, H.J., Yu, S., Madrid, S., Kalk, K.H., Zhang, R., Dijkstra, B.W., (2013) Crystal structure of α-1,4-Glucan lyase, a unique glycoside hydrolase family member with a novel catalytic mechanism. *Journal of Biological Chemistry*, **288**, 26764–26774. https://doi.org/10.1074/jbc.M113.485896

Sieber, K.B., Bromley, R.E., Dunning Hotopp, J.C., (2017) Lateral gene transfer between

prokaryotes and eukaryotes. *Experimental Cell Research*, **358**, 421–426. https://doi.org/10.1016/j.yexcr.2017.02.009

Sun, B.F., Xiao, J.H., He, S.M., Liu, L., Murphy, R.W., Huang, D.W., (2013) Multiple ancient horizontal gene transfers and duplications in lepidopteran species. *Insect Molecular Biology*, **22**, 72–87. https://doi.org/10.1111/imb.12004

Tettamanti, G., Casartelli, M., (2019) Cell death during complete metamorphosis. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **374**, 20190065. https://doi.org/10.1098/rstb.2019.0065

Walski, T., de Schutter, K., van Damme, E.J.M., Smagghe, G., (2017) Diversity and functions of protein glycosylation in insects. *Insect Biochemistry and Molecular Biology,* **83**, 21–34. https://doi.org/10.1016/j.ibmb.2017.02.005

Wang, A. M., Bishop, D. F., Desnick, R. J., (1990) Human alpha-N-acetylgalactosaminidase-molecular cloning, nucleotide sequence, and expression of a full-length cDNA. Homology with human alpha-galactosidase A suggests evolution from a common ancestral gene. Journal of Biological Chemistry, **265**, 21859-21866, https://doi.org/10.1016/S0021-9258(18)45818-8

Wang, C. -F., Sun, W., Zhang, Z., (2019) Functional characterization of the horizontally transferred 4,5-DOPA extradiol dioxygenase gene in the domestic silkworm, *Bombyx mori*. *Insect Molecular Biology,* **28**, 409–419. https://doi.org/10.1111/imb.12558

Watanabe, S., Kakudo, A., Ohta, M., Mita, K., Fujiyama, K., Inumaru, S., (2013) Molecular cloning and characterization of the α-glucosidase II from *Bombyx mori* and *Spodoptera frugiperda*, *Insect Biochemistry and Molecular Biology*, **43**, 319-327, https://doi.org/10.1016/j.ibmb.2013.01.005

Wheeler, D., Redding, A.J., Werren, J.H., (2013) Characterization of an Ancient Lepidopteran Lateral Gene Transfer. *PLoS ONE,* **8**, e59262. https://doi.org/10.1371/journal.pone.0059262

Winchester, B., (2005) Lysosomal metabolism of glycoproteins. *Glycobiology,* **15**, 1R-15R. https://doi.org/10.1093/glycob/cwi041

Xu, J., Morio, A., Morokuma, D., Nagata, Y., Hino, M., Masuda, A., Li, Z., Mon, H., Kusakabe, T., Lee, J.M., (2018) A functional polypeptide *N*-acetylgalactosaminyltransferase (PGANT) initiates *O*-glycosylation in cultured silkworm BmN4 cells. *Applied Microbiology and Biotechnology,* **102**, 8783–8797. https://doi.org/10.1007/s00253-018-9309-6

Zhu, B., Lou, M.M., Xie, G.L., Zhang, G.Q., Zhou, X.P., Li, B., Jin, G.L., (2011) Horizontal gene transfer in silkworm, *Bombyx mori*. *BMC Genomics,* **12**, 248. https://doi.org/10.1186/1471-2164-12-248

Table 1. Kinetic parameters for the hydrolysis of GalNAcα-pNP by BmNag31 recombinant enzymes and related enzymes.

Enzyme	k_{cat} (s-1)	K_m (μM)	k_{cat}/K_m (s^{-1} mM^{-1})	Reference
MBP-BmNag31	3.23 ± 0.10	610 ± 10	5.30	This study
BmGH31	4.81 ± 0.11	440 ± 5	10.8	This study
EfGH31-CBM32 [a]	4.45 ± 0.04	136 ± 6	33	(Miyazaki and Park, 2020)
EfGH31 [b]	6.54 ± 0.12	158 ± 12	41	(Miyazaki and Park, 2020)
BcGH31 [c]	3.6 ± 0.2	110 ± 10	32 ± 4	(Rahfeld et al., 2019)
tBcGH31 [b]	2.1 ± 0.1	140 ± 20	16 ± 2	(Rahfeld et al., 2019)
BpGH31 [c]	7.2 ± 0.3	270 ± 20	27 ± 2	(Rahfeld et al., 2019)
tBpGH31 [b]	2.22 ± 0.09	140 ± 20	16 ± 2	(Rahfeld et al., 2019)

[a] Truncated mutant containing the putative GH, FN3, and CBM32 domains.

[b] Truncated mutant containing the putative GH and FN3 domains.

[c] Full-length enzyme.

630 **Table 2. Binding affinity of CBM32 proteins toward monosaccharides.**

CBM	GH catalytic domain	Sequence identity[a]	K_d (M)	K_a (M^{-1})	Ligand	Method	Reference
BmCBM32	GH31	100%		260 ± 11.9	GalNAc	ITC	This study
EfCBM32	GH31	40.8%		266 ± 24.6	GalNAc	ITC	This study
CpCBM32-1	GH31	25.2%	6 ± 2	167 [b]	GalNAc	NMR	(Grondin et al., 2017)
CpCBM32-2	GH31	20.4%	0.9 ± 0.4	1×10^3 [b]	GalNAc	NMR	(Grondin et al., 2017)
CpCBM32-3	GH31	19.4%	0.9 ± 0.3	1×10^3 [b]	GalNAc	NMR	(Grondin et al., 2017)
CpCBM32	GH84 (CpGH84C)	20.8%		0.98 (±0.17) $\times 10^3$	Galactose	UV difference	(Ficko-Blean and Boraston, 2006)
				0.86 (±0.12) $\times 10^3$	GalNAc	UV difference	(Ficko-Blean and Boraston, 2006)
NagHCBM32-2	GH84 (CpGH84A)	16.5%		1.88 (± 0.02) $\times 10^3$	GlcNAc	ITC	(Ficko-Blean and Boraston, 2009)

631 [a]Amino acid sequence identity with BmNag31.

632 [b]Calculated from the reported K_d values.

633

Figures

Figure 1. Modular architectures of native and recombinant proteins of BmNag31 and EfNag31A.

BmNag31 and EfNag31A refer to the native form of these enzymes, and the other names refer to the recombinant proteins that were expressed in *E. coli* and used in this study. Colors represent the following: signal peptide, *blue*; GH31 catalytic domain, *pale blue*; fibronectin type 3 (FN3) domain, *yellow*; carbohydrate-binding module 32 (CBM32) domain, *green*; Type III cohesin-like domain (COH) and dockerin-like domain (DOC), *gray*; FIVAR domain, *pink*; transmembrane, *purple*; His tag, *orange*; and maltose binding protein (MBP) tag, *light pink*.

Figure 2. Expression *BmNag31* gene in different organs and developmental stages.

Relative expression levels of *BmNag31* in each organ in day-3 fifth-instar larvae (A), whole body of first instar larvae to adult stages (B), and midgut from day-3 fifth-instar larvae to day-4 pupae (C) were analyzed using quantitative RT-PCR. The housekeeping gene *BmGAPDH* was used as the internal control. Abbreviations: L1 to L5, first- to fifth-instar larval stages; PP, prepupal stage; P, pupal stage; D, day; W, wandering stage (day-7 fifth-instar larval stage). All experiments were performed in triplicate.

Figure 3. Oligomerization analysis of BmNag31.

The oligomeric state of BmNag31 was examined by gel filtration chromatography (A) and blue native PAGE (B). The experimental conditions are described in the Materials and Methods section. (A) Colors and lines of chromatograms are follows: MBP-BmNag31, *blue* line; BmGH31, *red* line; Blue dextran, *gray* line; molecular weight marker, *gray* dashed line. (B) A 3%–10 % gradient acrylamide gel was used for blue native PAGE. Lane M, marker; lane 1, MBP-BmNag31; lane 2, BmGH31. MBP-BmNag31 and BmGH31 were highlighted by arrows.

Figure 4. Binding activity of CBM32.

Binding activity of BmNag31 (A) and EfNag31A (B) toward GalNAc, D-galactose, GlcNAc, and L-fucose. The results of binding analysis of CBM32 and toward GalNAc show the raw ITC

data (upper panels) and integrated heat of binding obtained from the raw data (lower panels). For the other results using D-galactose, GlcNAc, and L-fucose as ligands, only the raw ITC data are shown.

Figure 5. Surface electrostatic potential of BmGH31 and EfGH31.

(A) The homology model of BmGH31 was generated using the SWISS-MODEL server with the EfGH31 coordinate complexed with GalNAc (PDB 6M77). The ribbon models of the N-domain (10–230), A-domain (231–566), A'-subdomain (263–302), proximal C-domain (567–657), distal C-domain (656–832), and FN3 domain (833–916) are shown in *blue*, *red*, *cyan*, *orange*, and *pink*, respectively. (B and C) The surface electrostatic potential of BmGH31 (B) and EfGH31 (C).

Figure 6. Homology model showing the active site of BmNag31 and binding site of CBM32s.

Homology model of the active sites of BmNag31, BmCBM32, and EfCBM32 were generated using the SWISS-MODEL server. (A) The BmGH31 homology model (*magenta*) is superimposed on the EfGH31 complexed with GalNAc (PDB 6M77) (*cyan*). Side chains of amino acid residues interacting with GalNAc are shown in stick model and GalNAc is shown in yellow stick model. The conserved residues of BmGH31 are labeled and the residue (Leu492) of EfGH32 corresponding to Val444 of BmGH31 is highlighted in *cyan*. N and A/B mean nucleophilic and acid/base catalytic residues, respectively. (B) The homology models of BmCBM32 (*magenta*) and EfCBM32 (*cyan*) are superimposed on the crystal structures of CpCBM32-1 (PDB 4LPL, *red*) and CpCBM32-3 in complex with GalNAc (4P5Y, *orange*). The residues predicted to bind GalNAc (*yellow*) are labeled with the same colors as the stick models.

688 **Supporting Information**

689

690 **Figure S1. Predicted subcellular localization of BmNag31.** Subcellular localization of
691 BmNag31 predicted by ngLOC (King et al., 2012) and visualized by SilkDB3.0 web server
692 (https://silkdb.bioinfotoolkits.net) is shown in (A). Hierarchical tree (B) and protein regions
693 important for the subcellular localization (C) were generated by DeepLoc-1.0 web server
694 (http://www.cbs.dtu.dk/services/DeepLoc).

695

696 **Figure S2. Multiple sequence alignment of Nag31 proteins from Lepidoptera and**
697 ***Enterococcus faecalis*.** Sequence alignment was performed using the ClustalW program, and
698 the figure was generated using ESPript 3.0. Catalytic residues and amino acid residues directly
699 interacting with GalNAc were predicted from the crystal structure of EfNag31A and are
700 highlighted by *blue stars* and *green triangles*, respectively. The signal peptide of EfNag31A is
701 indicated by the *black square*.

702

703 **Figure S3. SDS-PAGE analysis of the recombinant proteins expressed in *E. coli*.** (A) SDS-
704 PAGE of MBP-BmNag31. M, protein marker; lane 1, crude extract; lane 2, first affinity
705 chromatography with amylose resin; lane 3, second affinity chromatography with Ni-NTA
706 agarose. (B) SDS-PAGE of BmGH31. M, protein marker; lane 1, crude extract; lane 2,
707 BmGH31 purified with Ni-NTA agarose.

708

709 **Figure S4. Chromatogram of gel filtration.** Proteins are applied to a HiLoad 16/60 Superdex
710 200 prep grade column and eluted with 20 mM sodium phosphate buffer (pH 7.0) containing
711 300 mM NaCl. Colors and lines of chromatograms are follows: MBP-BmNag31, *blue line*;
712 BmGH31, *red line*; blue dextran 2000, *gray line*; molecular weight marker (thyroglobulin,
713 ferritin, aldolase, conalbumin, and ovalbumin), *gray dashed line*.

714

715 **Figure S5. Hydrolytic activity of BmGH31 toward bovine submaxillary mucin.** Bovine
716 submaxillary mucin (BSM) with or without pretreatment of neuraminidase, β-galactosidase,
717 and β-*N*-acetylhexosaminidase was incubated with BmGH31, and then analyzed by TLC.

Released *N*-acetylneuraminic acid (Neu5Ac) and GalNAc were indicated by arrows.

Figure S6. Effects of temperature and pH on the hydrolysis of GalNAcα-pNP by MBP-BmNag31 and BmGH31. pH dependence (A), temperature dependence (B), pH stability (C), and thermostability (D) of MBP-BmNag31 and BmGH31. pH dependence was measured in McIlvaine buffer at pH 4.0–8.0 or in glycine-HCl buffer at pH 9.0–10.0. Temperature dependence was examined in 50 mM sodium phosphate buffer (pH 6.5) at 20°C–60°C. pH stability was measured at 30°C after incubating for 20 hours in McIlvaine buffer at pH 4.0–8.0 or in glycine-HCl buffer at pH 9.0–10.0 and 4°C. Thermostability was measured at 30°C after incubation in 50 mM sodium phosphate buffer containing 300 mM NaCl (pH 6.5) at 4°C–60°C. Symbols used are as follows: MBP-BmNag31, *filled circle*; BmGH31, *open circle*. All experiments were performed in triplicate.

Figure S7. Multiple sequence alignment of CBM32 proteins showing GalNAc binding activity. Sequence alignment was performed using the ClustalW program, and the figure was generated using ESPript 3.0. Among amino acid residues involved in GalNAc recognition of CpCBM32-1 (denoted by asterisks), amino acid residues that are also conserved among BmCBM32, EfCBM32, and CpCBM32s are highlighted in yellow.

Table S1. Primers used in this study.

Ikegaya et al. Fig. 1

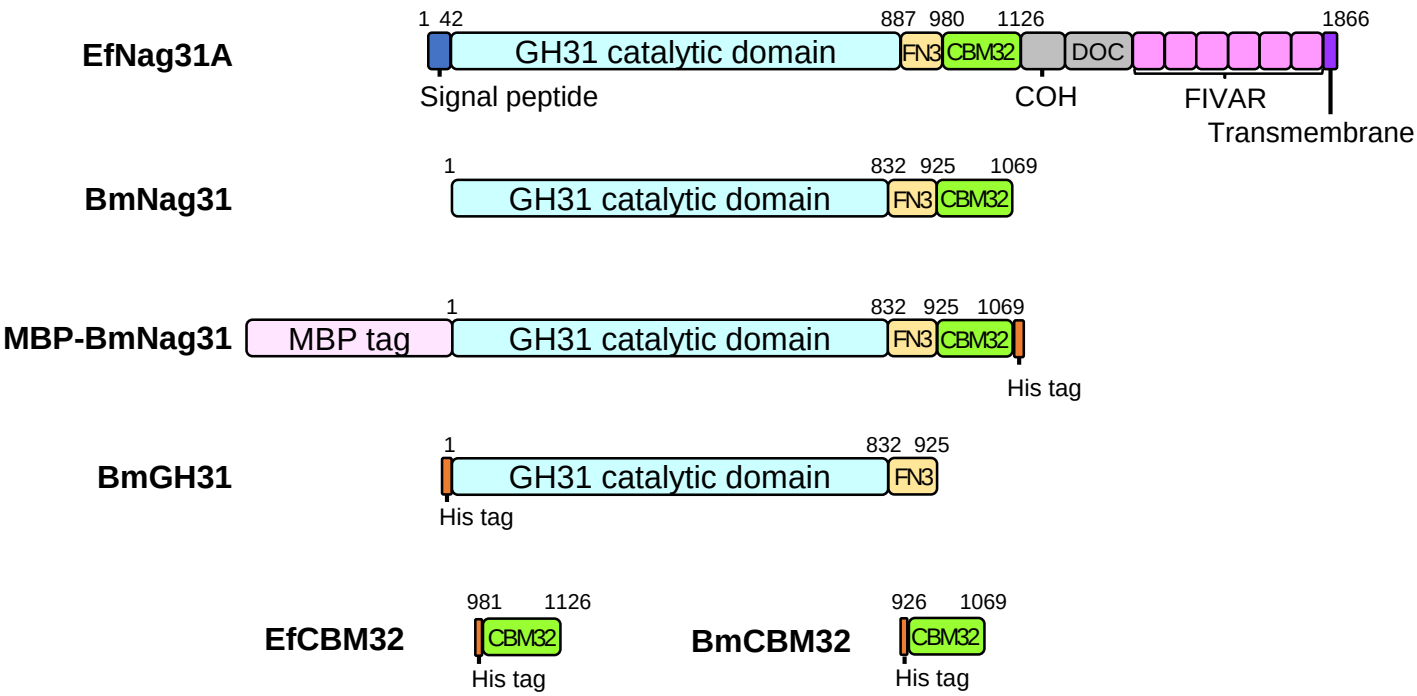

Figure 1. Modular architectures of native and recombinant proteins of BmNag31 and EfNag31A.

BmNag31 and EfNag31A refer to the native form of these enzymes, and the other names refer to the recombinant proteins that were expressed in *E. coli* and used in this study. Colors represent the following: signal peptide, *blue*; GH31 catalytic domain, *pale blue*; fibronectin type 3 (FN3) domain, *yellow*; carbohydrate-binding module 32 (CBM32) domain, *green*; Type III cohesin-like domain (COH) and dockerin-like domain (DOC), *gray*; FIVAR domain, *pink*; transmembrane, *purple*; His tag, *orange*; and maltose binding protein (MBP) tag, *light pink*.

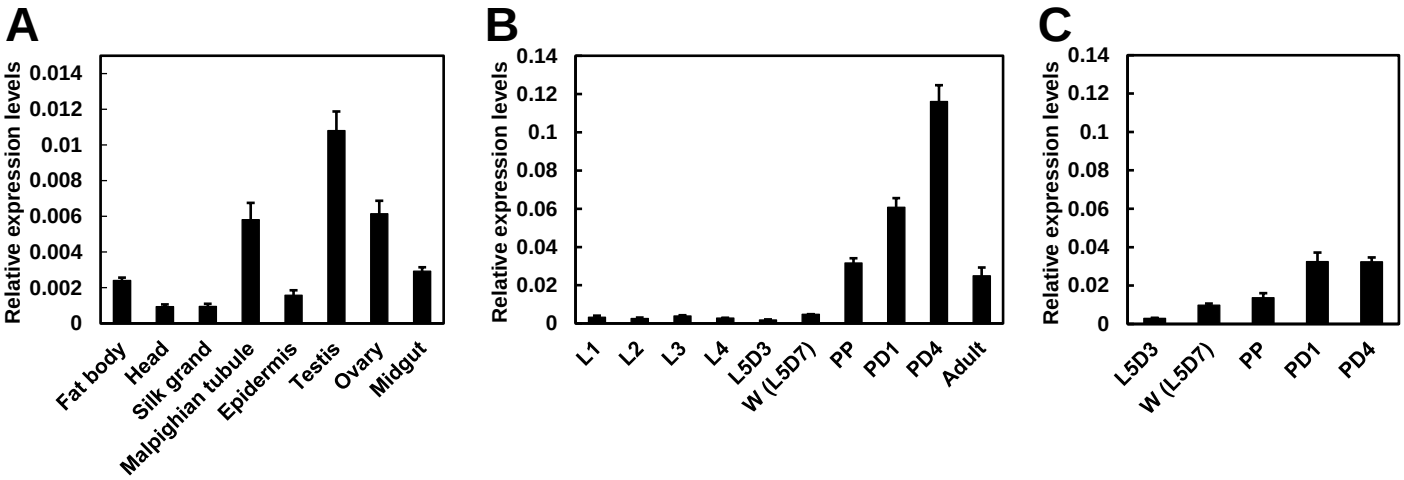

Figure 2. Expression *BmNag31* gene in different organs and developmental stages.

Relative expression levels of *BmNag31* in each organ in day-3 fifth-instar larvae (A), whole body of first instar larvae to adult stages (B), and midgut from day-3 fifth-instar larvae to day-4 pupae (C) were analyzed using quantitative RT-PCR. The housekeeping gene *BmGAPDH* was used as the internal control. Abbreviations: L1 to L5, first- to fifth-instar larval stages; PP, prepupal stage; P, pupal stage; D, day; W, wandering stage (day-7 fifth-instar larval stage). All experiments were performed in triplicate.

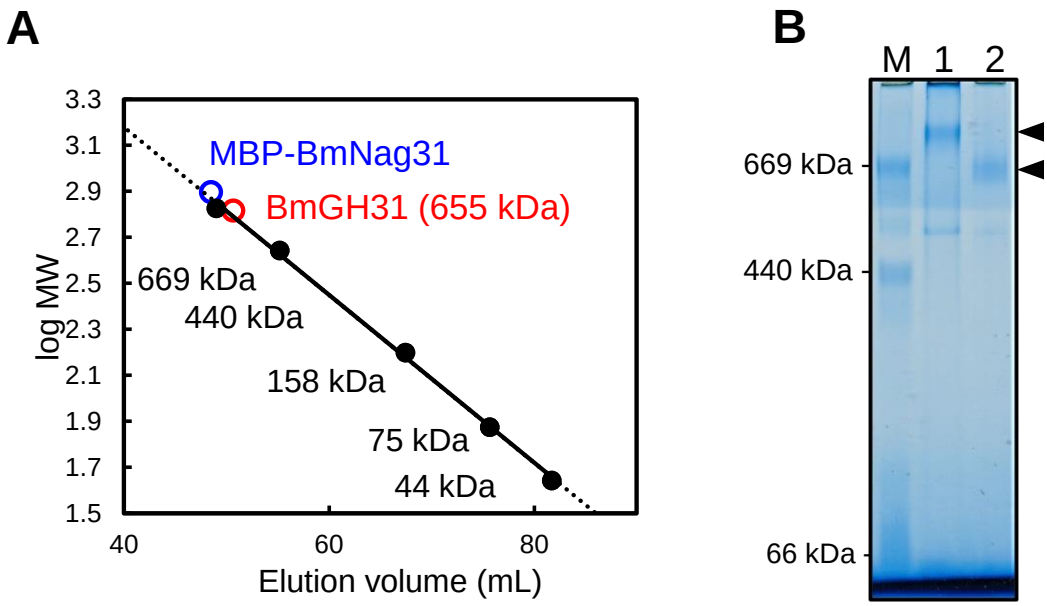

Figure 3. Oligomerization analysis of BmNag31.

The oligomeric state of BmNag31 was examined by gel filtration chromatography (A) and blue native PAGE (B). The experimental conditions are described in the Materials and Methods section. (A) Colors and lines of chromatograms are follows: MBP-BmNag31, *blue* line; BmGH31, *red* line; Blue dextran, *gray* line; molecular weight marker, *gray* dashed line. (B) A 3%–10 % gradient acrylamide gel was used for blue native PAGE. Lane M, marker; lane 1, MBP-BmNag31; lane 2, BmGH31. MBP-BmNag31 and BmGH31 were highlighted by arrows.

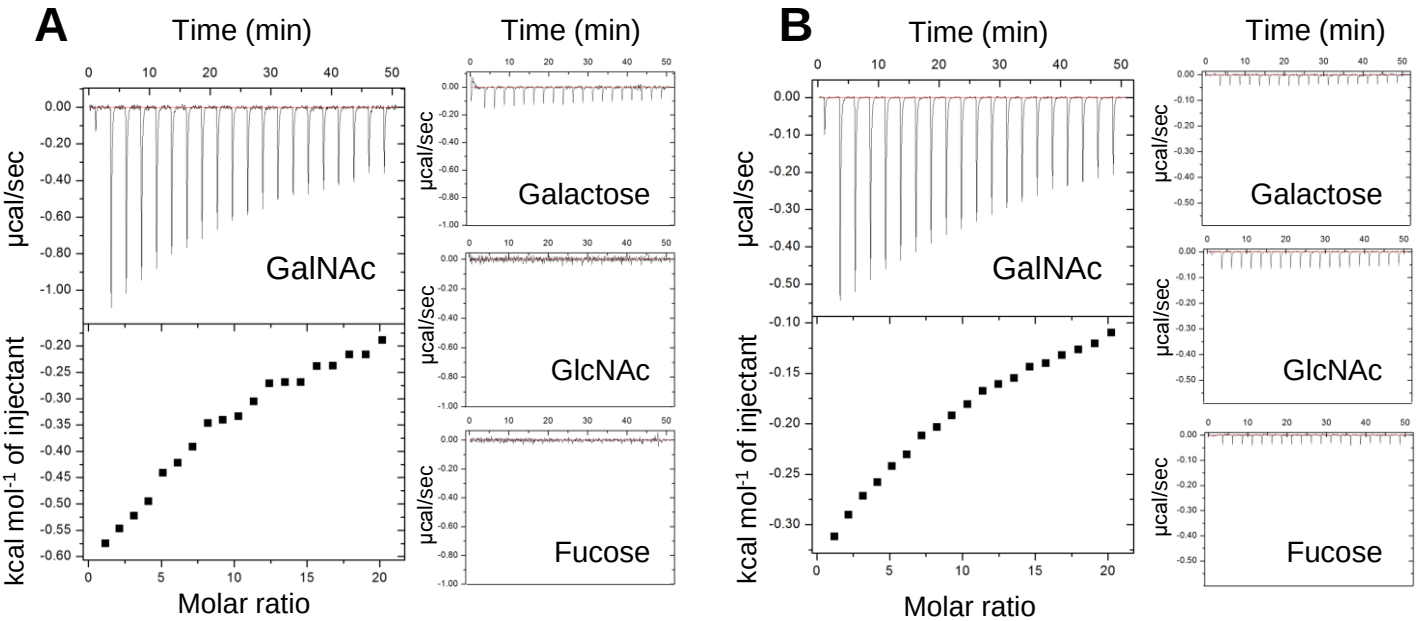

Figure 4. Binding activity of CBM32.

Binding activity of BmNag31 (A) and EfNag31A (B) toward GalNAc, D-galactose, GlcNAc, and L-fucose. The results of binding analysis of CBM32 and toward GalNAc show the raw ITC data (upper panels) and integrated heat of binding obtained from the raw data (lower panels). For the other results using D-galactose, GlcNAc, and L-fucose as ligands, only the raw ITC data are shown.

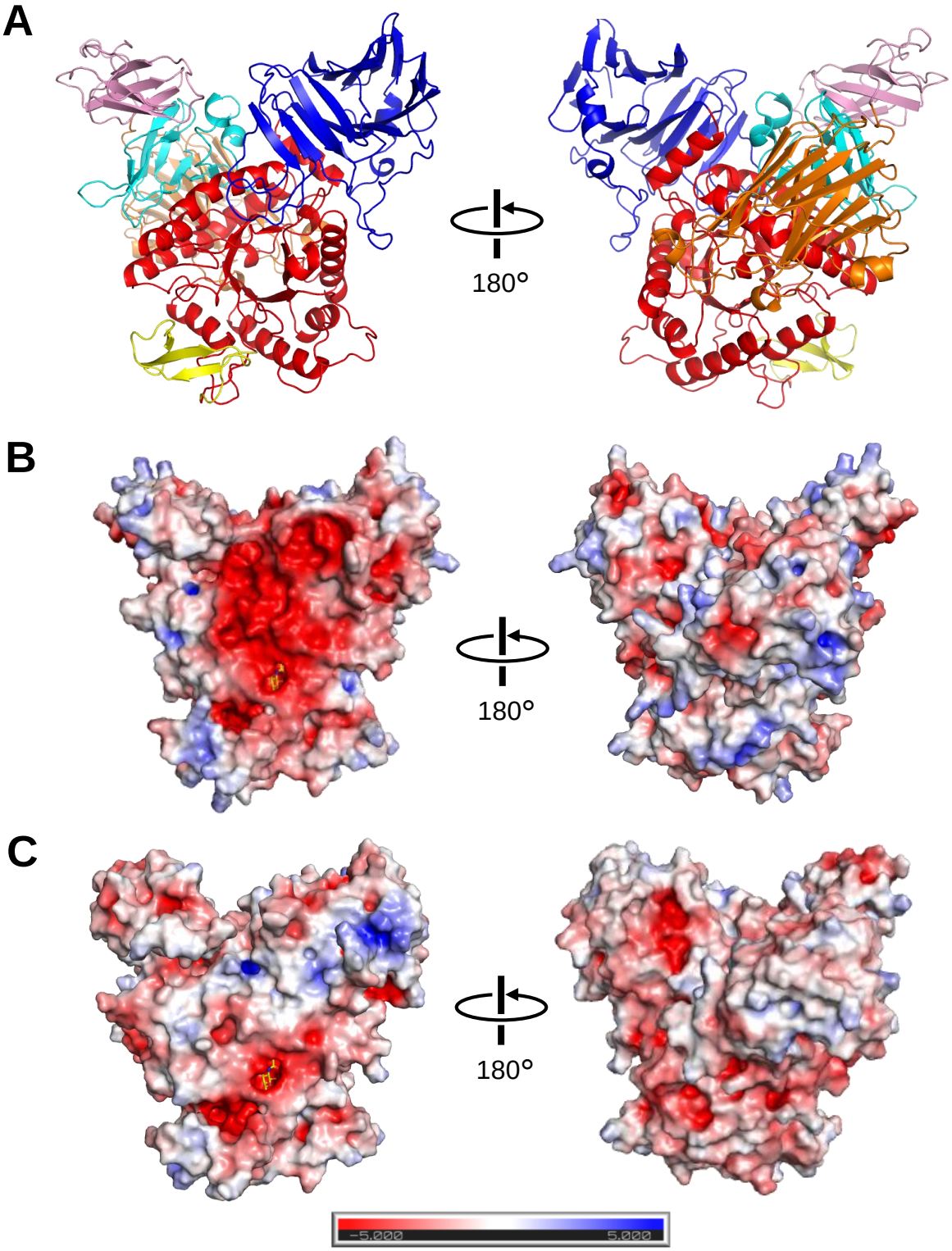

Figure 5. Surface electrostatic potential of BmGH31 and EfGH31.

(A) The homology model of BmGH31 was generated using the SWISS-MODEL server with the EfGH31 coordinate complexed with GalNAc (PDB 6M77). The ribbon models of the N-domain (10–230), A-domain (231–566), A'-subdomain (263–302), proximal C-domain (567–657), distal C-domain (656–832), and FN3 domain (833–916) are shown in *blue*, *red*, *cyan*, *orange*, and *pink*, respectively. (B and C) The surface electrostatic potential of BmGH31 (B) and EfGH31 (C).

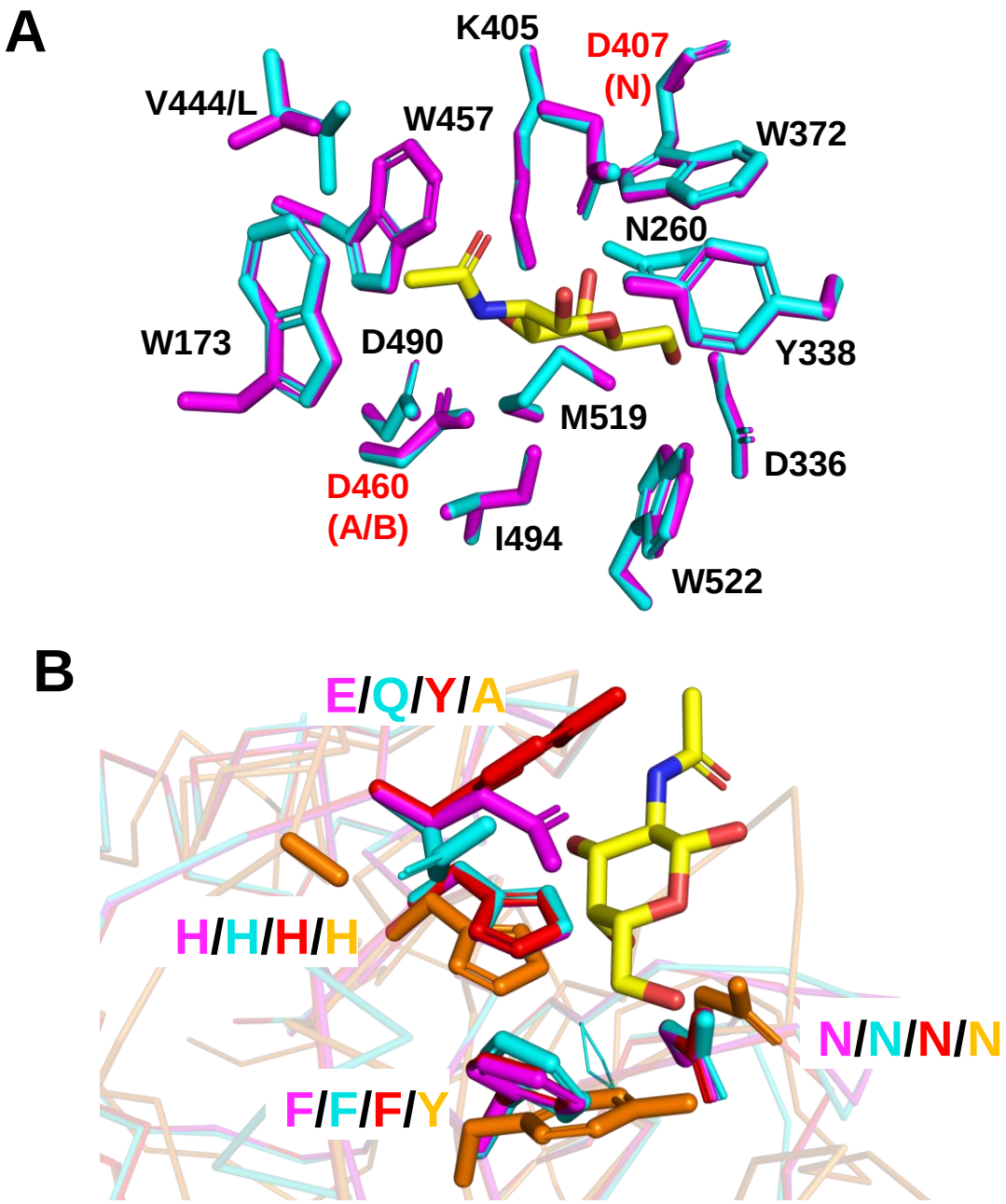

Figure 6. Homology model showing the active site of BmNag31 and binding site of CBM32s.

Homology model of the active sites of BmNag31, BmCBM32, and EfCBM32 were generated using the SWISS-MODEL server. (A) The BmGH31 homology model (*magenta*) is superimposed on the EfGH31 complexed with GalNAc (PDB 6M77) (*cyan*). Side chains of amino acid residues interacting with GalNAc are shown in stick model and GalNAc is shown in yellow stick model. The conserved residues of BmGH31 are labeled and the residue (Leu492) of EfGH32 corresponding to Val444 of BmGH31 is highlighted in *cyan*. N and A/B mean nucleophilic and acid/base catalytic residues, respectively. (B) The homology models of BmCBM32 (*magenta*) and EfCBM32 (*cyan*) are superimposed on the crystal structures of CpCBM32-1 (PDB 4LPL, *red*) and CpCBM32-3 in complex with GalNAc (4P5Y, *orange*). The residues predicted to bind GalNAc (*yellow*) are labeled with the same colors as the stick models.