Crystal structure of the Enterococcus faecalis α-N-acetylgalactosaminidase, a member of the glycoside hydrolase family 31.

メタデータ	言語: eng
	出版者:
	公開日: 2020-05-27
	キーワード (Ja):
	キーワード (En):
	作成者: Miyazaki, Takatsugu, Park, Enoch Y.
	メールアドレス:
	所属:
URL	http://hdl.handle.net/10297/00027486

1   **Crystal structure of the *Enterococcus faecalis* α-*N*-acetylgalactosaminidase,**

2   **a member of the glycoside hydrolase family 31**

3

4   Takatsugu Miyazaki[1,*], and Enoch Y. Park[1]

5

6   [1] Green Chemistry Research Division, Research Institute of Green Science and Technology,

7   Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan.

8

9   ***Correspondence:** Takatsugu Miyazaki, Green Chemistry Research Division, Research

10   Institute of Green Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku,

11   Shizuoka,   422-8529,   Japan;   Tel: +81-54-238-4886;   E-mail:

12   miyazaki.takatsugu@shizuoka.ac.jp

13

14   **Running Title:** Structure of GH31 α-*N*-acetylgalactosaminidase

15

16

1

18 **Abstract**

19 Glycoside hydrolase family 31 (GH31) contains α-glucosidase, α-xylosidase, α-galactosidase,

20 and α-transglycosylase. Recent work has expanded the diversity of substrate specificity of

21 GH31 enzymes, and α-*N*-acetylgalactosaminidases (αGalNAcases) belonging to GH31 have

22 been identified in human gut bacteria. In this study, we determined the crystal structure of a

23 truncated form of GH31 αGalNAcase from *Enterococcus faecalis*. The enzyme has a similar

24 fold as other reported GH31 enzymes and an additional fibronectin type 3-like domain.

25 Additionally, the structure in complex with *N*-acetylgalactosamine reveals that conformations

26 of the active site residues, including its catalytic nucleophile, change to recognize the ligand.

27 The catalytic site residues are completely conserved among GH31 αGalNAcases but vary in

28 comparison to other reported GH31 enzymes except for the catalytic residues.

29

32

35

Introduction

Glycoside hydrolases (GH, EC 3.2.1.-) are widely distributed in various organisms and enable carbohydrate utilization by degrading glycosidic bonds. These enzymes are classified into more than 160 families based on their amino acid sequences [1], and recent work has established new GH families by discovering novel GHs from mainly bacteria and fungi [2–7]. In addition, different substrate specificities in known GH families have been found [2]. The glycoside hydrolase family 31 (GH31) is a large family that contains more than 12,000 protein sequences in the CAZy database (http://www.cazy.org/, accessed on April 8, 2020). GH31 enzymes have been identified as acting on α-glycosidic bonds, including α-glucosidase [8–13], α-1,3-glucosidase [14], mannosyl-oligosaccharide α-1,3-glucosidase [15,16], α-xylosidase [17–19], sucrase-isomaltase [20], α-galactosidase [2,21], sulfoquinovosidase [22,23], dextranase [24,25], and α-glucan lyase [26]. GH31 enzymes employ a retaining mechanism, and several enzymes catalyze transglycosylation to produce α-glucosidic linkages such as α-1,2, α-1,3, α-1,4, and α-1,6 [27–30]; cycloalternan-forming enzyme and cycloalternan degrading enzyme are also GH31 members [31,32]. All GH31 members share a $(\beta/\alpha)_8$-barrel catalytic domain and some β-sandwich domains at the N- and C-termini.

Rahfeld et al. identified α-*N*-acetylgalactosaminidases (αGalNAcases) that belong to GH31 by screening metagenomic DNA libraries from the human gut microbiome [33]. GH31 αGalNAcases specifically release the initial *N*-acetylgalactosamine (GalNAc) residue from mucin-type *O*-glycans. Exo-αGalNAcases are found in GH27, GH36, GH109, and GH129 families. GH27 αGalNAcases from fungi and mammals hydrolyze mucin-type *O*-glycans and blood type A-antigens [34–36]. GH109 αGalNAcases were identified to release GalNAc from blood type A-antigen and employ an NAD^+-dependent hydrolytic mechanism [37]. *Bifidobacterium bifidum* GH129 αGalNAcase cleaves Tn-antigen (GalNAcα1-Ser) and has endo activity on core-1-type *O*-glycan (Galβ1–3GalNAcα-) [38,39]. GH101 enzymes are endo-acting αGalNAcases that release a disaccharide from the core-1 structure of *O*-glycans [40]; the catalytic domains of GH101 and also GH27, GH36, and GH129 enzymes adopt a $(\beta/\alpha)_8$-barrel fold. Together with GH31, GH27 and GH36 enzymes form the GH-D clan. In this clan, the two catalytic aspartic acid residues are structurally conserved and work as

66   nucleophile and general acid/base catalysts in the retaining mechanism (Fig. 1). An H^1-NMR

67   study on GH31 αGalNAcase hydrolysis revealed that the enzyme employs the same retaining

68   mechanism as other GH31 enzymes [33]. However, no crystal structure of GH31 αGalNAcase

69   has yet been reported.

70   In this study we characterized a GH31 αGalNAcase from a human gut bacterium,

71   *Enterococcus faecalis* (hereafter EfNag31A), and determined its crystal structure as the first

72   structure for the GH31 αGalNAcases. The structures reveal that EfNag31A has at least two

73   conformations in the active site, including its catalytic residue, to accept substrates. The

74   results provide novel insight into substrate recognition mechanism in GH31 enzymes.

75

Materials and Methods

77

Materials and strains

79   *p*-Nitrophenyl *N*-acetyl-α-D-galactosaminide (pNP-α-GalNAc) and *p*-nitrophenyl

80   α-D-galactopyranoside (pNP-α-Gal,) were purchased from Cayman Chemical (Ann Arbor,

81   MI, USA) and Tokyo Chemical Industry (Tokyo, Japan), respectively. *p*-Nitrophenyl

82   α-D-glucopyranoside (pNP-α-Glc) and *p*-nitrophenyl α-D-mannopyranoside (pNP-α-Man)

83   were obtained from Merck (Darmstadt, Germany). *p*-Nitrophenyl *N*-acetyl-α-D-glucosaminide

84   (pNP-α-GlcNAc) and *p*-nitrophenyl α-D-xylopyranoside (pNP-α-Xyl) were from Carbosynth

85   (Compton, Berkshire, UK). All other reagents were of analytical grade and purchased from

86   Wako Pure Chemical Industry (Osaka, Japan) or Merck unless otherwise stated. *E. faecalis*

87   NBRC 12964 (ATCC 10100) was obtained from NITE Biological Resource Center (Chiba,

88   Japan). *Escherichia coli* strains DH5α and BL21 (DE3) were used for DNA manipulation and

89   protein expression, respectively, unless otherwise stated.

90

Expression and purification of EfGH31 and EfGH31-CBM32

92   A signal sequence was predicted using the SignalP server [41] and domains were

93   identified according to Conserved Domain Database [42] and InterPro [43]. DNA for coding

94   amino acid residues 43–1126 (named EfGH31-CBM32) and 43–984 (named EfGH31) of

95   EfNag31A (GenBank, EOK08638.1) were amplified from bacteria by colony-directed PCR

96　　using a KOD FX Neo DNA polymerase (Toyobo, Osaka, Japan) with primers 5′-TTT TT<u>C</u>

97　　<u>ATA TG</u>C AAG AGC AAA CAG CAA AAG AAG-3′ (for EfGH31-CBM32 and EfGH31),

98　　5′- TTT T<u>CT CGA G</u>CT ATG GTT GTT TGT AAA AGA GCA TC-3′ (for

99　　EfGH31-CBM32), and 5′-TTT T<u>CT CGA G</u>CT ATT TAT AGG GAT CAT CTT GTG-3′ (for

100　　EfGH31). The PCR products was digested with NdeI and XhoI (sites underlined in primers),

101　　ligated into pET-28a vector, and their identity confirmed by DNA sequencing. *E. coli* BL21

102　　(DE3) harboring the expression plasmids was grown at 37 ºC in Luria-Bertani medium with

103　　50 μg/mL kanamycin. Protein expression was induced at an optical density of 0.6 (at 600 nm)

104　　with isopropyl-β-D-thiogalactopyranoside (final concentration, 0.1 mM) with overnight

105　　incubation at 20 ºC. Cells were harvested by centrifugation at 10,000 × *g* for 5 min and

106　　resuspended in 50 mM sodium phosphate buffer (pH 8.0) containing 20 mM imidazole and

107　　300 mM NaCl. After disruption by sonication, insoluble debris was removed by

108　　centrifugation at 20,000 × *g* for 20 min and the supernatant applied to a Ni^{2+} nitriolotriacetic

109　　acid agarose (Qiagen, Hilden, Germany) column equilibrated with the same buffer. The

110　　column was buffer washed, and recombinant proteins eluted with 50 mM sodium phosphate

111　　buffer (pH 8.0) containing 250 mM imidazole and 300 mM NaCl. Enzymes were

112　　concentrated with an Amicon Ultra ultrafiltration device (Merck) and purified by gel filtration

113　　chromatography using an ÄKTAexplorer system (GE Healthcare) equipped with a Superdex

114　　200 Increase 10/300 column (GE Healthcare, Chicago, IL, USA) in 20 mM sodium phosphate

115　　buffer (pH 7.0) containing 300 mM NaCl. Protein fractions were pooled and concentrated in

116　　10 mM HEPES-NaOH buffer (pH 7.0) as described above. Selenomethionine

117　　(SeMet)-substituted EfGH31 was obtained by *E. coli* B834 (DE3) cultured in LeMaster

118　　medium [44] and was purified in the same manner as native protein. Protein purity was

119　　confirmed by SDS-PAGE. Protein concentration was measured at 280 nm based on

120　　theoretical molar absorption coefficients, 188,510 M^{-1} cm^{-1} for EfGH31-CBM32 and 159,060

121　　M^{-1} cm^{-1} for EfGH31.

122

123　　**Enzyme assays**

124　　　　Hydrolytic activity toward various *p*-nitrophenyl glycosides was measured in 50 μL

125　　reaction mixtures containing 5.0 μg/mL enzyme, 0.5 mM of substrate, and 50 mM Britton–

126   Robinson (phosphate–acetate–borate) buffer (pH 6.0) at 37 °C. To examine the effect of pH

127   on hydrolytic activity; reaction mixtures containing 0.5 mM pNP-α-GalNAc were prepared

128   with the same buffer at pH 2.0–12.0 and were incubated for 10 min at 37 °C. The temperature

129   dependence was examined using 0.5 mM pNP-α-GalNAc and 50 mM Britton–Robinson

130   buffer (pH 6.0) from 20 °C to 70 °C. Initial velocities of the hydrolytic reaction for

131   pNP-α-GalNAc were determined using 50 mM Britton–Robinson buffer (pH 6.0) and five

132   concentrations (0.1–2.0 mM) of pNP-α-GalNAc at 37 °C. Enzyme concentrations used were

133   41 nM for EfGH31-CBM32 and 47 nM for EfGH31. All reactions were performed in

134   triplicate and quenched by adding two volumes of 1 M Na_2CO_3. Released p-nitrophenol was

135   measured at 405 nm. Kinetic parameters were calculated by fitting to the Michaelis–Menten

136   equation using non-linear regression analysis by KaleidaGraph software (Synergy Software,

137   Reading PA, USA).

138

139   **Protein crystallography**

140   Proteins (15–16 mg/mL) were crystallized at 20 °C using the hanging-drop vapor

141   diffusion method, where 1.0 μL of protein solution was mixed with an equal volume of a

142   crystallization reservoir solution. Initial crystallization screening was performed using Crystal

143   Screen, Crystal Screen 2, PEG/Ion Screen, PEG/Ion 2 Screen kits (Hampton Research, Aliso

144   Viejo, CA, USA). EfGH31 crystals appeared in several conditions whereas no crystal of

145   EfGH31-CBM32 was obtained. Well-diffracted crystals of native and SeMet-substituted

146   EfGH31 were obtained with a crystallization solution containing 25% (w/v) polyethylene

147   glycol 3,350 (Hampton Research) and 200 mM ammonium citrate tribasic. Crystals of

148   EfGH31 in complex with GalNAc were obtained by co-crystallization under the same

149   condition with 5 mM GalNAc. All crystals were cryoprotected with the reservoir solution

150   supplemented with ethylene glycol at a final concentration of 20% (v/v) and then flash-frozen

151   in liquid nitrogen.

152   Diffraction data were collected at the BL5A beamline (Photon Factory, Tsukuba, Japan).

153   Data were processed using XDS [45]. Initial phase was calculated from the single-wavelength

154   anomalous dispersion dataset of SeMet-substituted EfGH31 crystals using the AutoSol

155   program in PHENIX [46]. Sixteen selenium sites were found with a figure of merit of 0.315.

156   A rough model of SeMet-substituted EfGH31 (R_{work} = 0.241; R_{free} = 0.266) was obtained and

157   further built using PHENIX AutoBuild [47]. Structures of apo EfGH31 and its complex with

158   GalNAc were solved by the molecular replacement method using MOLREP [48] with the

159   coordinates of SeMet-EfGH31 and native EfGH31, respectively, as search models.

160   Refinement and manual model building were performed using REFMAC5 [49] and COOT

161   [50], respectively. Coordinates and structural factors were deposited in the Worldwide Protein

162   Data Bank (http://wwpdb.org/). Molecular images were prepared using PyMOL (Schrödinger

163   LLC, New York, NY, USA). Structural similarity searches were carried out using the Dali

164   server [51].

165

166

167   **Results and Discussion**

168

169   **Expression and characterization**

170   EfNag31A consists of multiple domains with 1866 amino acid residues (Fig. 2A). The

171   enzyme has an N-terminal signal peptide, several GH31 domains (including DUF4968,

172   GH31_N, GH31_CPE1046, and DUF5110 domains), a fibronectin type 3 (FN3), an F5/F8

173   type C (assigned as carbohydrate-binding module family 32, CBM32, according to the CAZy

174   database) region, a dockerin II domain, a type III cohesin domain, six FIVAR domains, and a

175   C-terminal membrane anchoring region in this order. With an exception of the C-terminal

176   domains, the domain organization (from the N-terminus to the type III cohesion domain) is

177   similar to reported GH31 α-*N*-acetylgalactosaminidases from *Bacteroides caccae* (BcGH31,

178   sequence identity is 44%), *Bacteroides plebeius* (BpGH31, 44%), and *Clostridium*

179   *perfringens* (CpGH31, 43%) [33]. To enzymatically and structurally characterize EfNag31A,

180   two truncated proteins, EfGH31-CBM32 (43–1126 residues) and EfGH31 (43–984), were

181   expressed in *E. coli* (Fig. S1); the C-terminal domains, including the dockerin and cohesin

182   domains, were not predicted to be important for catalytic activity. The purified enzymes

183   displayed hydrolytic activity toward pNP-α-GalNAc but not pNP-α-Glc, pNP-α-Gal,

184   pNP-α-GlcNAc, pNP-α-Man, and pNP-α-Xyl. The optimal pH and temperature of both

185   recombinant enzymes were identical at 6.0 and 50 ºC, respectively (Fig. S2). The kinetic

186   parameters of EfGH31-CBM32 and EfGH31, for the hydrolysis of pNP-α-GalNAc, were

187   similar to the full-length and truncated forms of the other GH31 α-N-acetylgalactosaminidases

188   (αGalNAcases) in comparison with isozymes belonging to different GH families (Table 1).

189   These results suggest that the C-terminal regions, including the CBM32 domain, do not affect

190   hydrolytic reaction.

191

192   **Overall structure**

193   The crystal structure of EfGH31 was determined at 1.4 Å resolution. The crystal belongs

194   to the space group $P2_12_12_1$ with one molecule in the asymmetric unit (Table 2). The electron

195   density ($2|F_o|-|F_c|$) map at 1 σ shows continuous electron density for almost all amino acid

196   residues from Ser57 to Gln979. EfGH31 is composed of five domains: an N-terminal

197   β-sandwich domain (N-domain, residues 57–278); a (β/α)$_8$-barrel catalytic domain

198   (A-domain, 279–614); a proximal C-terminal β-sandwich domain (615–705); a distal

199   C-terminal β-sandwich domain (706–883); and an FN3 domain (884–979) (Fig. 2B). A

200   structural homology search using the Dali server [51] revealed that the highest Z scores were

201   observed for *Cellvibrio japonicus* α-xylosidase (CjXyl31A, PDB 2XVK) [18] and

202   *Flavobacterium johnsoniae* dextranase (FjDex31A, PDB 6JR8) [25], which belong to GH31

203   but display different substrate specificity (Table S1). EfGH31 also shows structural

204   homologies to the other GH31 enzymes whose structures have been determined, though their

205   sequence identities are lower than 25%. By contrast, GH27 αGalNAcases, which are

206   classified in clan GH-D together with GH31 and GH36, and GH129 αGalNAcases have low

207   structural homology (sequence identity ≤ 10%, Z score < 16%) to GH31 enzymes.

208   The N-, A-, proximal C-, and distal C-domains of EfGH31 are conserved in many

209   GH31 proteins, and their folds resemble those of CjXyl31A and FjDex31A (Fig. 2C).

210   However, a superimposition of EfGH31 with the structures of GH31 enzymes shows different

211   structural components. The catalytic A-domain of EfGH31 is interrupted by an extra

212   subdomain (named A′-subdomain, 312–350) with an antiparallel β-sheet and a short α-helix

213   between the first β-strand and the second α-helix of the (β/α)$_8$-barrel fold. The A′-subdomain

214   is located near the putative catalytic site at the center of the (β/α)$_8$-barrel, but its function is

215   unclear as no similar component is seen in other GH31 enzyme structures. The A′-subdomain

216   is conserved in BcGH31 and BpGH31, whereas the corresponding region of CpGH31 is

217   shorter (Fig. 3). Therefore, the A′-subdomain may be unnecessary for hydrolytic activity but

218   may be involved in substrate binding.

219   EfGH31 has an FN3 domain after the conservative distal C-domain. A Dali search using

220   the FN3 domain of EfGH31 revealed that it has a small degree of structural similarity to the

221   CBM56 domains of β-1,3-glucanases (PDB 5T7A and 5H9Y, Z score = 8.7 and 7.7, rmsd =

222   2.4 and 2.4 Å, sequence identity = 12% and 8%, respectively), an FN3 domain of GH78

223   α-rhamnosidase (PDB 6GSZ, Z score = 8.6, rmsd = 2.3 Å, sequence identity = 20%), and an

224   N-terminal 'CBM-like' domain of *Trueperella pyogenes* GH31 cycloalternan degrading

225   enzyme (PDB 5F7S, Z score = 7.6, rmsd = 2.6 Å, sequence identity = 11%). FN3 domains are

226   found in many GHs and are often located between a catalytic domain and a CBM domain as a

227   linker [52–54]. Three CBM32 domains of CpGH31, which display 17%–30% sequence

228   identity to the CBM32 domain of EfGH31, have the ability to bind galactose and GalNAc

229   [55]. The FN3 domain of EfGH31 is likely to function as a stable linker allowing optimal

230   positioning and/or flexibility of the catalytic domain and the CBM32 domains.

231

Active site

233   To investigate the catalytic mechanism and substrate recognition, the crystal structure of

234   EfGH31 complexed with GalNAc was determined at 1.9 Å resolution. An electron density

235   map for β-anomer of GalNAc (GalNAc −1) was present at the center of the catalytic

236    $(β/α)_8$-barrel, which is predicted to be subsite −1 (subsite nomenclatures are according to

237   Davies et al. [56]) (Fig. 4A). The GalNAc −1 molecule interacts with seven amino acid

238   residues (side chains of Asp384, Tyr386, Trp420, Lys453, Asp455, and Asp508, and main

239   chain of Val456) directly *via* hydrogen bonds (Table S2). Asn308, Trp505, and Asp538 form

240   hydrogen bond networks with the ligand *via* water molecules. Ile542 and Met567 create a

241   hydrophobic environment to interact with the apolar face of GalNAc −1, and Trp221, Val456,

242   Leu492, and Trp505 form a hydrophobic pocket that accepts the methyl group of the

243   acetamide group of GalNAc −1. All amino acid residues interacting with GalNAc −1 are

244   conserved in GH31 αGalNAcases (Fig. 3). Asp455 and Asp508, which are located near C1

245   atom of GalNAc −1, are predicted to be the catalytic nucleophile and the general acid/base,

246   respectively (see Fig. 1), agreeing with Rahfeld et al., who observed that mutagenesis in the

247   corresponding residues of BpGH31 resulted in a remarkable decrease of hydrolytic activity

248   [33].

249   Superimposition of the unliganded and GalNAc-complex structures demonstrates

250   structural differences between the active site residues (Fig. 4B). In the GalNAc-complex,

251   three loops after the second, third, and fourth β-strands of the $(\beta/\alpha)_8$-barrel move closer to

252   GalNAc −1 compared with the unliganded enzyme. Accordingly, the nucleophile Asp455

253   moves closer to C1 of GalNAc −1 and the side chain of Tyr386 flips to interact with 1-OH of

254   GalNAc −1. The side chains of Met567 and Tyrp570 also move to interact with GalNAc −1.

255   Therefore, EfGH31 has two forms of the active site, an open form and a closed form (Fig.

256   4C). The distances between Asp455 and Asp508 catalysts are 10.0 and 6.5 Å in the open and

257   closed forms, respectively. Although it is not clear whether such changes occur when a

258   substrate (α-anomer) enters the active site, these conformational changes may occur during

259   substrate binding and the hydrolytic reaction in order to accommodate the substrate's GalNAc

260   residue and enable the nucleophile Asp455 to attack the GalNAc C1 in the proper position.

261   No such conformational change was reported in known GH31 enzyme structures.

262   Superimposition of EfGH31 onto the structures of known GH31 enzymes indicates that

263   the subsite −1 residues are different, though the two catalytic aspartic acid residues are

264   completely conserved (Fig. 5). Residues corresponding to Leu492 of EfGH31 are Arg in all

265   previously reported GH31 enzymes. The substitution provides a space to accept an acetamide

266   group from GalNAc as described above. Trp420 and Lys453, which recognize an axial 4-OH

267   atom of GalNAc, are conserved in the *Pseudopedobacter saltans* α-galactosidase (PsGal31A)

268   [21], which recognizes an axial 4-OH of galactose (Fig. 5D). In other GH31 enzymes that

269   recognize sugars with an equatorial 4-OH, such as glucose and xylose, these residues are

270   substituted to different amino acids (Fig. 5A–C). By contrast, the recognition mechanism of

271   the apolar faces of galacto-configured sugars is different between EfGH31 and PsGal31A.

272   Ile542 between the seventh β-strand and α-helix and Met567 between the eighth β-strand and

273   α-helix are substituted to Arg and Asn, respectively, in PsGal31A. Instead, Trp486 located

274   between the eighth β-strand and α-helix of PsGal31A hydrophobically interacts with the

275   nonpolar face of galactose. These differences in substrate recognition mechanism may result

276   in different substrate specificity.

277

278   **Implication for function**

279   EfNag31A contains the N-terminal signal peptide, multiple FIVAR domains. Although

280   the function of FIVAR domains are not yet clear, they are often found in cell wall associated

281   proteins (Pfam 07554). This potential function suggests that EfNag31A the enzyme is

282   secreted from bacterial cells and is associated with the cell membrane and the cell wall.

283   Dockerin and cohesin domains are used to build multi-enzyme complexes on the cell surface,

284   e.g., the cellulosome, which is an enzyme complex that degrades cellulose and hemicellulose

285   [57]. Therefore, EfNag31A may form part of a multi-enzyme complex that degrades

286   carbohydrates. No genes coding for carbohydrate-active enzymes were found associated with

287   the EfNag31A gene in the genome of *E. faecalis*, whereas its ortholog BcGH31 forms a

288   cluster together with genes for several GHs, peptidases, a sugar-binding protein, and a sugar

289   transporter, which are predicted to degrade mucin-like *O*-glycoproteins and

290   glycosaminoglycans [33]. A BLAST search using EfNag31A found putative GH31

291   αGalNAcases (> 40% sequence identity) from Firmicutes (e.g., *Enterococcus*, *Streptococcus*,

292   and *Lactobacillus* spp.), Bacteroidetes (e.g., *Bacteroides* and *Elizabethkingia* spp.), and

293   Verrucomicrobia (*Akkermansia muciniphila*) (data not shown), most of which are mammalian

294   gut bacteria and/or are associated with pathogenicity. *E. faecalis* possess a GH101

295   endo-αGalNAcase that releases di- and trisaccharides from core-1, -2, and -3-type

296   oligosaccharides and mucin glycopeptides [58]. Thus, bacteria including *E. faecalis* may have

297   acquired these enzymes to scavenge GalNAc and other monosaccharides, which are released

298   from mucin produced in the host.

299   In conclusion, we characterized EfNag31A, which exhibits a selective αGalNAcase

300   activity that is inactive with other α-glycosides, and determined the crystal structure as the

301   first GH31 αGalNAcase. The enzyme has a domain architecture that is similar to the GH31

302   enzymes reported to date; however, the active site residues vary and strictly recognize

303   GalNAc with conformational changes. These results indicate that EfNag31A is involved in

304   the degradation of α-GalNAc-linked glycoproteins, such as mucin. Further studies, such as

305   mutational analysis and examination of complex structure with a substrate, are required to

306   completely understand the mechanism of hydrolysis and substrate recognition.

307

308   **Acknowledgments**

314

315   **Conflict of interest**

316   The authors declare that they have no conflicts of interest with the contents of this article.

317

318   **Author contributions**

319   TM conceived and supervised the study; TM designed and performed experiments; TM and

320   EYP analyzed the data; TM wrote the manuscript; TM and EYP revised and approved the

321   manuscript.

322

323   **References**

324

325   1. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM and Henrissat B (2014) The

326   carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490–

327   D495.

328   2. Helbert W, Poulet L, Drouillard S, Mathieu S, Loiodice M, Couturier M, Lombard V,

329   Terrapon N, Turchetto J, Vincentelli R and Henrissat B (2019) Discovery of novel

330   carbohydrate-active enzymes through the rational exploration of the protein sequences

331   space. *Proc Natl Acad Sci U S A* **116**, 6063–6068.

332   3. Kuhaudomlarp S, Pergolizzi G, Patron NJ, Henrissat B and Field RA (2019) Unraveling

333   the subtleties of β-(1→3)-glucan phosphorylase specificity in the GH94, GH149, and

334   GH161 glycoside hydrolase families. *J Biol Chem* **294**, 6483–6493.

335   4. Tanaka N, Nakajima M, Narukawa-Nara M, Matsunaga H, Kamisuki S, Aramasa H,

336   Takahashi Y, Sugimoto N, Abe K, Terada T, Miyanaga A, Yamashita T, Sugawara F,

337   Kamakura T, Komba S, Nakai H and Taguchi H (2019) Identification, characterization,

338   and structural analyses of a fungal endo-β-1,2-glucanase reveal a new glycoside hydrolase

339   family. *J Biol Chem* **94**, 7942–7965.

340 5. Briliūtė J, Urbanowicz PA, Luis AS, Baslé A, Paterson N, Rebello O, Hendel J, Ndeh DA,

341   Lowe EC, Martens EC, Spencer DIR, Bolam DN and Crouch LI (2019) Complex

342   *N*-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus

343   encoded by multiple co-regulated genetic loci. *Nat Microbiol* **4**, 1571–1581.

344 6. Le Mauff F, Bamford NC, Alnabelseya N, Zhang Y, Baker P, Robinson H, Codée JDC,

345   Howell PL and Sheppard DC (2019) Molecular mechanism of *Aspergillus fumigatus*

346   biofilm disruption by fungal and bacterial glycoside hydrolases. *J Biol Chem* **294**, 10760–

347   10772.

348 7. Hettle AG, Hobbs JK, Pluvinage B, Vickers C, Abe KT, Salama-Alber O, McGuire BE,

349   Hehemann JH, Hui JPM, Berrue F, Banskota A, Zhang J, Bottos EM, Van Hamme J and

350   Boraston AB (2019) Insights into the κ/ι-carrageenan metabolism pathway of some

351   marine *Pseudoalteromonas* species. *Commun Biol* **2**, 474.

352 8. Ernst HA, Lo Leggio L, Willemoës M, Leonard G, Blum P and Larsen S (2006) Structure

353   of the *Sulfolobus solfataricus* α-glucosidase: implications for domain conservation and

354   substrate recognition in GH31. *J Mol Biol* **358**, 1106–1124.

355 9. Sim L, Quezada-Calvillo R, Sterchi EE, Nichols BL and Rose DR (2008) Human

356   intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and

357   basis of inhibition and substrate specificity. *J Mol Biol* **375**, 782–792.

358 10. Tan K, Tesar C, Wilton R, Keigher L, Babnigg G and Joachimiak A (2010) Novel

359   α-glucosidase from human gut microbiome: substrate specificities and their switch.

360   *FASEB J* **24**, 3939–3949.

361 11. Tagami T, Yamashita K, Okuyama M, Mori H, Yao M and Kimura A (2013) Molecular

362   basis for the recognition of long-chain substrates by plant α-glucosidases. *J Biol Chem*

363   **288**, 19296–19303.

364 12. Chaudet MM and Rose DR (2016) Suggested alternative starch utilization system from

365   the human gut bacterium *Bacteroides thetaiotaomicron*. *Biochem Cell Biol* **94**, 241–246.

366   13. Roig-Zamboni V, Cobucci-Ponzano B, Iacono R, Ferrara MC, Germany S, Bourne Y1
367   Parenti G, Moracci M and Sulzenbacher G (2017) Structure of human lysosomal acid
368   α-glucosidase-a guide for the treatment of Pompe disease. *Nat Commun* **8**, 1111.

369   14. Kang MS, Okuyama M, Mori H and Kimura A (2009) The first α-1,3-glucosidase from
370   bacterial origin belonging to glycoside hydrolase family 31. *Biochimie* **91**, 1434–1442.

371   15. Caputo AT, Alonzi DS, Marti L, Reca IB, Kiappes JL, Struwe WB, Cross A, Basu S,
372   Lowe ED, Darlot B, Santino A, Roversi P and Zitzmann N (2016) Structures of
373   mammalian ER α-glucosidase II capture the binding modes of broad-spectrum iminosugar
374   antivirals. *Proc Natl Acad Sci U S A* **113**, E4630–E4638.

375   16. Satoh T, Toshimori T, Yan G, Yamaguchi T and Kato K (2016) Structural basis for
376   two-step glucose trimming by glucosidase II involved in ER glycoprotein quality control.
377   *Sci Rep* **6**, 20575.

378   17. Lovering AL, Lee SS, Kim YW, Withers SG and Strynadka NC (2005) Mechanistic and
379   structural analysis of a family 31 α-glycosidase and its glycosyl-enzyme intermediate. *J*
380   *Biol Chem* **280**, 2105–2115.

381   18. Larsbrink J, Izumi A, Ibatullin FM, Nakhai A, Gilbert HJ, Davies GJ and Brumer H
382   (2011) Structural and enzymatic characterization of a glycoside hydrolase family 31
383   α-xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification. *Biochem*
384   *J* **436**, 567–580.

385   19. Cao H, Walton JD, Brumm P and Phillips GN Jr (2020) Crystal structure of α-xylosidase
386   from *Aspergillus niger* in complex with a hydrolyzed xyloglucan product and new insights
387   in accurately predicting substrate specificities of GH31 family glycosidases. *ACS Sustain*
388   *Chem Eng* **8**, 2540–2547.

389   20. Sim L, Willemsma C, Mohan S, Naim HY, Pinto BM and Rose DR (2011) Structural
390   basis for substrate selectivity in human maltase-glucoamylase and sucrase-isomaltase
391   N-terminal domains. *J Biol Chem* **285**, 17763–17770.

392   21. Miyazaki T, Ishizaki Y, Ichikawa M, Nishikawa A and Tonozuka T (2015) Structural and
393   biochemical characterization of novel bacterial α-galactosidases belonging to glycoside
394   hydrolase family 31. *Biochem J* **469**, 145–158.

395   22. Speciale G, Jin Y, Davies GJ, Williams SJ and Goddard-Borger ED (2016) YihQ is a

396   sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat Chem*

397   *Biol* **12**, 215–217.

398   23. Abayakoon P, Jin Y, Lingford JP, Petricevic M, John A, Ryan E, Wai-Ying Mui J, Pires

399   DEV, Ascher DB, Davies GJ, Goddard-Borger ED and Williams SJ (2018) Structural and

400   biochemical insights into the function and evolution of sulfoquinovosidases. *ACS Cent Sci*

401   **4**, 1266–1273.

402   24. Gozu Y, Ishizaki Y, Hosoyama Y, Miyazaki T, Nishikawa A and Tonozuka T (2016) A

403   glycoside hydrolase family 31 dextranase with high transglucosylation activity from

404   *Flavobacterium johnsoniae*. *Biosci Biotechnol Biochem* **80**, 1562–1567.

405   25. Tsutsumi K, Gozu Y, Nishikawa A and Tonozuka T (2020) Structural insights into

406   polysaccharide recognition by *Flavobacterium johnsoniae* dextranase, a member of

407   glycoside hydrolase family 31. *FEBS J* **287**, 1195–1207.

408   26. Rozeboom HJ, Yu S, Madrid S, Kalk KH, Zhang R and Dijkstra BW (2013) Crystal

409   structure of α-1,4-glucan lyase, a unique glycoside hydrolase family member with a novel

410   catalytic mechanism. *J Biol Chem* **288**, 26764–26774.

411   27. Larsbrink J, Izumi A, Hemsworth GR, Davies GJ and Brumer H (2012) Structural

412   enzymology of *Cellvibrio japonicus* Agd31B protein reveals α-transglucosylase activity in

413   glycoside hydrolase family 31. *J Biol Chem* **287**, 43288–43299.

414   28. Ichinose H, Suzuki R, Miyazaki T, Kimura K, Momma M, Suzuki N, Fujimoto Z, Kimura

415   A and Funane K (2017) *Paenibacillus* sp. 598K 6-α-glucosyltransferase is essential for

416   cycloisomaltooligosaccharide synthesis from α-(1 → 4)-glucan. *Appl Microbiol*

417   *Biotechnol* **101**, 4115–4128.

418   29. Fujimoto Z, Suzuki N, Kishine N, Ichinose H, Momma M, Kimura A and Funane K

419   (2017) Carbohydrate-binding architecture of the multi-modular α-1,6-glucosyltransferase

420   from *Paenibacillus* sp. 598K, which produces α-1,6-glucosyl-α-glucosaccharides from

421   starch. *Biochem J* **474**, 2763–2778.

422   30. Ma M, Okuyama M, Tagami T, Kikuchi A, Klahan P and Kimura A (2019) Novel

423   α-1,3/α-1,4-glucosidase from *Aspergillus niger* exhibits unique transglucosylation to

424   generate high levels of nigerose and kojibiose. *J Agric Food Chem* **67**, 3380–3388.

425   31. Tagami T, Miyano E, Sadahiro J, Okuyama M, Iwasaki T and Kimura A (2016) Two

426   novel glycoside hydrolases responsible for the catabolism of cyclobis-(1→6)-α-nigerosyl.

427   *J Biol Chem* **291**, 16438–16447.

428   32. Light SH, Cahoon LA, Halavaty AS, Freitag NE and Anderson WF (2016) Structure to

429   function of an α-glucan metabolic pathway that promotes *Listeria monocytogenes*

430   pathogenesis. *Nat Microbiol* **2**, 16202.

431   33. Rahfeld P, Wardman JF, Mehr K, Huff D, Morgan-Lang C, Chen HM, Hallam SJ and

432   Withers SG (2019) Prospecting for microbial α-*N*-acetylgalactosaminidases yields a new

433   class of GH31 *O*-glycanase. *J Biol Chem* **294**, 16400–16415.

434   34. Garman SC, Hannick L, Zhu A and Garboczi DN (2002) The 1.9 Å structure of

435   α-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases.

436   *Structure* **10**, 425–434.

437   35. Clark   NE   and   Garman   SC   (2009)   The   1.9   Å   structure   of   human

438   α-*N*-acetylgalactosaminidase: The molecular basis of Schindler and Kanzaki diseases. *J*

439   *Mol Biol* **393**, 435–447.

440   36. Ashida H, Tamaki H, Fujimoto T, Yamamoto K and Kumagai H (2000) Molecular

441   cloning of cDNA encoding α-*N*-acetylgalactosaminidase from *Acremonium* sp. and its

442   expression in yeast. *Arch Biochem Biophys* **384**, 305–310.

443   37. Liu QP, Sulzenbacher G, Yuan H, Bennett EP, Pietz G, Saunders K, Spence J, Nudelman

444   E, Levery SB, White T, Neveu JM, Lane WS, Bourne Y, Olsson ML, Henrissat B and

445   Clausen H (2007) Bacterial glycosidases for the production of universal red blood cells.

446   *Nat Biotechnol* **25**, 454–464.

447   38. Kiyohara M, Nakatomi T, Kurihara S, Fushinobu S, Suzuki H, Tanaka T, Shoda S,

448   Kitaoka   M,   Katayama   T,   Yamamoto   K   and   Ashida   H   (2012)

449   α-*N*-Acetylgalactosaminidase from infant-associated bifidobacteria belonging to novel

450   glycoside hydrolase family 129 is implicated in alternative mucin degradation pathway. *J*

451   *Biol Chem* **287**, 693–700.

452   39. Sato M, Liebschner D, Yamada Y, Matsugaki N, Arakawa T, Wills SS, Hattie M, Stubbs

453   KA, Ito T, Senda T, Ashida H and Fushinobu S (2017) The first crystal structure of a

454   family 129 glycoside hydrolase from a probiotic bacterium reveals critical residues and

455   metal cofactors. *J Biol Chem* **292**, 12126–12138.

456   40. Fujita K, Oura F, Nagamine N, Katayama T, Hiratake J, Sakata K, Kumagai H and

457   Yamamoto K (2005) Identification and molecular cloning of a novel glycoside hydrolase

458   family of core 1 type *O*-glycan-specific endo-α-*N*-acetylgalactosaminidase from

459   *Bifidobacterium longum*. *J Biol Chem* **280**, 37415–37422.

460   41. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S,

461   von Heijne G and Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using

462   deep neural networks. *Nat Biotechnol* **37**, 420–423.

463   42. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI,

464   Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C,

465   Lanczycki CJ and Marchler-Bauer A (2020) CDD/SPARCLE: the conserved domain

466   database in 2020. *Nucleic Acids Res* **48**, D265–D268.

467   43. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang

468   HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A,

469   Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C,

470   Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND,

471   Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA,

472   Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong SY and Finn RD (2019)

473   InterPro in 2019: improving coverage, classification and access to protein sequence

474   annotations. *Nucleic Acids Res* **47**, D351–D360.

475   44. LeMaster DM and Richards FM (1985) ^{1}H-^{15}N heteronuclear NMR studies of *Escherichia*

476   *coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263–

477   7268.

478   45. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.

479   46. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung

480   LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ,

481   Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger

482   TC, Urzhumtsev AG, Videau LL, Williams CJ and Adams PD (2019) Macromolecular

483   structure determination using X-rays, neutrons and electrons: recent developments in

484   Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877.

485   47. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW,

486   Read RJ and Adams PD. (2008) Iterative model building, structure refinement and density
487   modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr*
488   **64**, 61–69.
489   48. Vagin A and Teplyakov A (1997) MOLREP: an automated program for molecular
490   replacement. *J Appl Crystallogr* **30**, 1022–1025.
491   49. Murshudov GN, Vagin AA and Dodson EJ (1997) Refinement of macromolecular
492   structures by the maximum-likelihood method. *Acta Crystallogr Sect D Biol Crystallogr*
493   **53**, 240–255.
494   50. Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of
495   Coot. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 486–501.
496   51. Holm L (2019) Benchmarking fold detection by DaliLite v.5. *Bioinformatics* **35**, 5326–
497   5327.
498   52. Jee JG, Ikegami T, Hashimoto M, Kawabata T, Ikeguchi M, Watanabe T and Shirakawa
499   M (2002) Solution structure of the fibronectin type III domain from *Bacillus circulans*
500   WL-12 chitinase A1. *J Biol Chem* **277**, 1388–1397.
501   53. Valk V, van der Kaaij RM and Dijkhuizen L (2017) The evolutionary origin and possible
502   functional roles of FNIII domains in two *Microbacterium aurum* B8.A granular starch
503   degrading enzymes, and in other carbohydrate acting enzymes. *Amylase* **1**, 1–11.
504   54. Fujita M, Tsuchida A, Hirata A, Kobayashi N, Goto K, Osumi K, Hirose Y, Nakayama J,
505   Yamanoi T, Ashida H and Mizuno M (2011) Glycoside hydrolase family 89
506   α-*N*-acetylglucosaminidase from *Clostridium perfringens* specifically acts on GlcNAc
507   α1,4Gal β1R at the non-reducing terminus of *O*-glycans in gastric mucin. *J Biol Chem*
508   **286**, 6479–6489.
509   55. Grondin JM, Duan D, Kirlin AC, Abe KT, Chitayat S, Spencer HL, Spencer C,
510   Campigotto A, Houliston S, Arrowsmith CH, Allingham JS, Boraston AB and Smith SP
511   (2017) Diverse modes of galacto-specific carbohydrate recognition by a family 31
512   glycoside hydrolase from *Clostridium perfringens*. *PLoS One* **12**, e0171606.
513   56. Davies GJ, Wilson KS and Henrissat B (1997) Nomenclature for sugar-binding subsites in
514   glycosyl hydrolases. *Biochem J* **321**, 557–559.
515   57. Smith SP and Bayer EA (2013) Insights into cellulosome assembly and dynamics: from

516   dissection to reconstruction of the supramolecular enzyme complex. *Curr Opin Struct Biol*

517   **23**, 686–694.

518   58. Goda HM, Ushigusa K, Ito H, Okino N, Narimatsu H and Ito M (2008) Molecular

519   cloning, expression, and characterization of a novel endo-α-*N*-acetylgalactosaminidase

520   from *Enterococcus faecalis*. *Biochem Biophys Res Commun* **375**, 441–446.

521   59. Hsieh HY, Calcutt MJ, Chapman LF, Mitra M and Smith DS (2003) Purification and

522   characterization of a recombinant α-*N*-acetylgalactosaminidase from *Clostridium*

523   *perfringens*. *Protein Expr Purif* **32**, 309–316.

524   60. Diederichs K and Karplus PA (2013) Better models by discarding data? *Acta Crystallogr*

525   *D Biol Crystallogr* **69**, 1215–1222.

526   61. Robert X and Gouet P (2014) Deciphering key features in protein structures with the new

527   ENDscript server. *Nucleic Acids Res* **42**, W320–W324.

528

529 **Table 1. Kinetic parameters for the hydrolysis of pNP-α-GalNAc by EfGH31, EFGH31-CBM32,**

530 **and other α-*N*-acetylgalactosaminidases.**

Enzyme	Family	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} mM^{-1})	Reference
EfGH31-CBM32	GH31	4.45 ± 0.04	136 ± 6	33	This study
EfGH31	GH31	6.54 ± 0.12	158 ± 12	41	This study
BcGH31 [a]	GH31	3.6 ± 0.2	110 ± 10	32	[33]
tBcGH31 [b]	GH31	2.1 ± 0.1	140 ± 20	16	[33]
BpGH31 [a]	GH31	7.2 ± 0.3	270 ± 20	27	[33]
tBpGH31 [b]	GH31	2.2 ± 0.1	140 ± 20	16	[33]
Human NAGAL [c]	GH27	16.3 ± 0.1	700 ± 30	23	[35]
CpAagA [d]	GH36	N. A. [g]	1100	N. A.	[59]
EmGH109 [e]	GH109	9.84 ± 0.16	77 ± 6	128	[37]
NagBb [f]	GH129	11.0 ± 0.4	2060 ± 230	2.35	[39]

531 [a] full-length enzyme.

532 [b] truncated form containing an FN3 domain but not C-terminal domains including a CBM32 domain.

533 [c] human lysosomal αGalNAcase NAGAL.

534 [d] *C. perfringens* αGalNAcase AagA.

535 [e] *Elizabethkingia meningoseptica* αGalNAcase.

536 [f] *Bifidobacterium bifidum* αGalNAcase NagBb.

537 [g] not available.

538

539 **Table 2. Data collection and refinement statistics.**

	SeMet	Apo	GalNAc-complex
Data collection			
Beamline	PF BL5A	PF BL5A	PF BL5A
Wavelength (Å)	0.9792	1.0000	1.0000
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	82.4, 83.1, 145.2	82.4, 83.0, 149.7	82.6, 83.2, 147.8
Resolution range (Å)	50–1.80 (1.90–1.80)	50–1.40 (1.48–1.40)	50–1.90 (2.00–1.90)
Measured reflections	1,212,826	2,560,235	486,092
Unique reflections	92,458	201,712	80,809
Completeness (%)	99.4 (98.8)	100 (100)	99.8 (99.4)
Redundancy	13.1 (13.5)	12.7 (13.1)	6.0 (5.0)
Mean $I/\sigma(I)$	16.6 (4.0)	18.3 (2.7)	13.8 (3.6)
R_{meas}	0.101 (0.713)	0.071 (0.913)	0.084 (0.504)
$CC_{1/2}$ [60]	(0.924)	(0.884)	(0.844)
Wilson B factor	23.1	14.2	23.4
Refinement statistics			
R_{work} / R_{free}	0.241 / 0.266	0.171 / 0.195	0.217 / 0.261
RMSD			
Bond lengths (Å)		0.008	0.010
Bond angles (°)		1.470	1.682
Number of atoms			
Protein		7,307	7,101
Ligand/Ion		36	36
Water		1,063	309
Average B (Å^2)			
Protein		23.4	33.6
Ligands		29.4	28.8
Water		32.4	27.5
Ramachandran plot			
Favored (%)		96.4	96.0
Outliers (%)		0.1	0.2
Clashscore		3.54	4.36
PDB codes		6M76	6M77

540 The values for the highest resolution shells are given in parentheses.

541

542   **Figure Legends**

543

544   **Figure 1. Reaction mechanism of GH31 enzymes.**

545

546   **Figure 2. Structure of EfNag31A. (A)** Primary structures of full-length EfNag31A (*upper*)

547   and its truncated constructs, EfGH31-CBM32 (*middle*) and EfGH31 (*lower*). Domains of

548   EfNag31A are presented based on Conserved Domain Database and InterPro, and

549   recombinant protein domains are described based on the crystal structure with the same colors

550   as **B**. **(B)** Overall structure of EfGH31: N-domain, *blue*; A-domain, *red*; A′-subdomain,

551   *yellow*; proximal C-domain, *cyan*; distal C-domain, *orange*; FN3 domain, *pink*. **(C)** Structural

552   comparison with known GH31 enzymes: EfGH31, *red*; CjXyl31A, *cyan*; FjDex31A, *yellow*.

553   In CjXyl31A, the PA14 domain (residues 238–384) inserted into the N-domain is omitted.

554

555   **Figure 3. Sequence alignment of GH31 α-*N*-acetylgalactosaminidases.** Partial sequences

556   of EfGH31, BcGH31 (GenBank ID, ASM65008.1), BpGH31 (EDY97082.1), and CpGH31

557   (ABG84084.1) were aligned using Clustal Omega and the figure generated by ESPript 3.0

558   [61]. Secondary structures and domain architecture of EfGH31 are described above the

559   sequences: N-domain, *blue*; A-domain, *red*; A′-subdomain, *yellow*; proximal C-domain, *cyan*;

560   distal C-domain, *orange*; FN3 domain, *pink*. Identical residues are shown in white with a red

561   background and conservative changes in red with a white background. Catalytic residues and

562   residues interacting with GalNAc are in *cyan* and *green*, respectively. Active site residues

563   whose conformations are different between apo and GalNAc-complex structures are indicated

564   with *orange circles*.

565

566   **Figure 4. Active site of EfGH31. (A)** Stereo view of the active site of EfGH31 in complex

567   with GalNAc. Side chains of amino acid residues interacting with GalNAc as well as the main

568   chain of Val456 are shown in *blue stick* models, and catalytic residues and GalNAc are in

569   *cyan* and *yellow*, respectively. Hydrogen bonds are shown as *dashed lines* and water

570   molecules as *red spheres*. An F_o-F_c omit electron density map of GalNAc (contoured at 3 σ)

571   is shown as *blue mesh* in the dotted box. **(B)** Structural differences between apo (*white*) and

572 GalNAc-complex (*blue*). Conformational changes of the amino acid residues are described by

573 *black arrows*. **(C)** Molecular surfaces of the active site in apo form (*upper*) and

574 GalNAc-complex (*lower*) of EfGH31. A GalNAc molecule is superimposed to the apo form

575 (*upper*). N and A/B mean nucleophilic and acid/base catalytic residues, respectively.

576

577 **Figure 5. Structural comparison of active sites of GH31 enzymes.** Active site of EfGH31

578 (*blue*) is superimposed to CjXyl31A in complex with a covalent inhibitor (PDB 2XVK, *cyan*)

579 **(A)**, FjDex31A D312A mutant in complex with isomaltotriose (PDB 6JR8, *yellow*) **(B)**,

580 human NtMGAM in complex with acarbose (PDB 2QMJ, *magenta*) **(C)**, and PsGal31A in

581 complex with galactose (PDB 4XPP, *orange*) **(D)**. Ligands are shown in thin stick models.

582 Catalytic residues are underlined with *red lines*, and subsites according to Davies et al. [56]

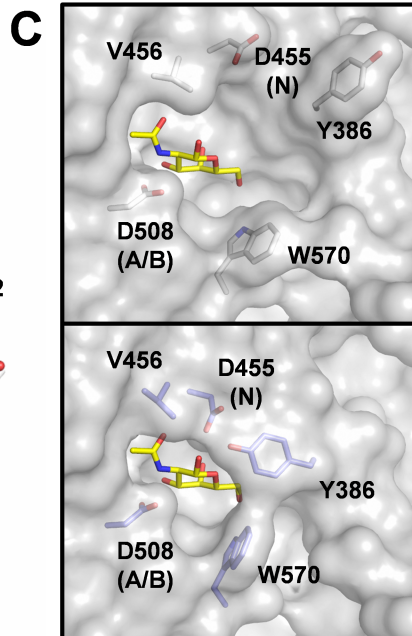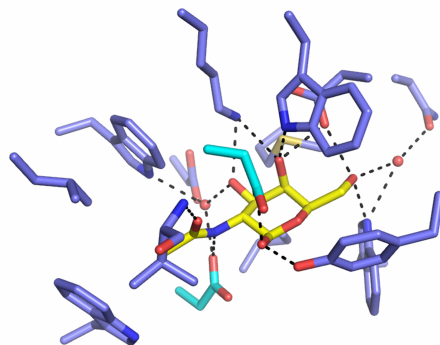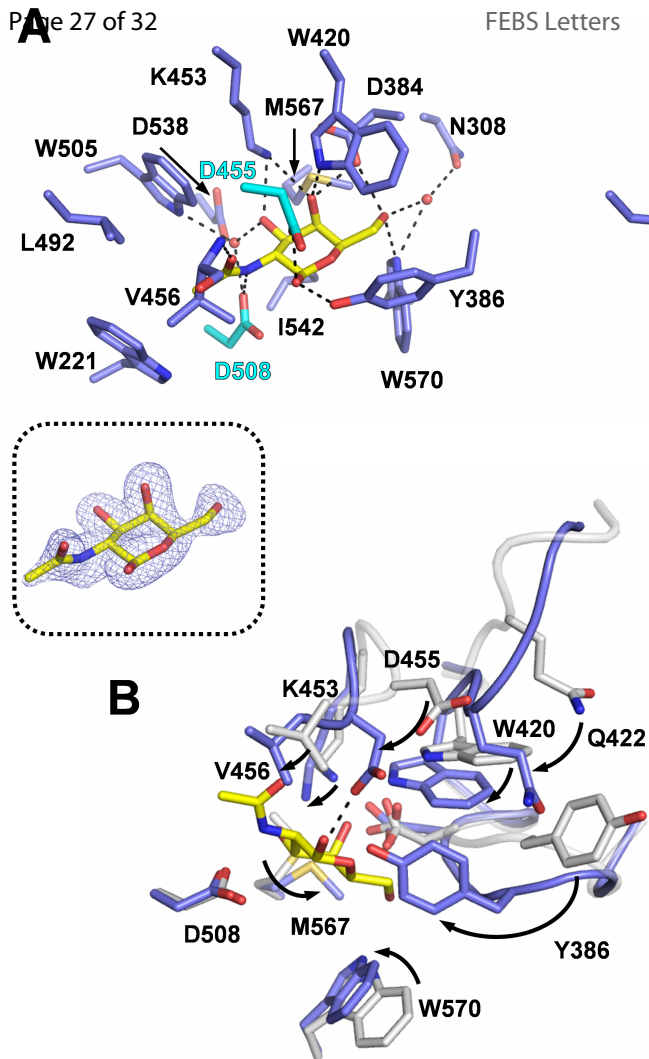583 are also indicated.

584

nucleophile
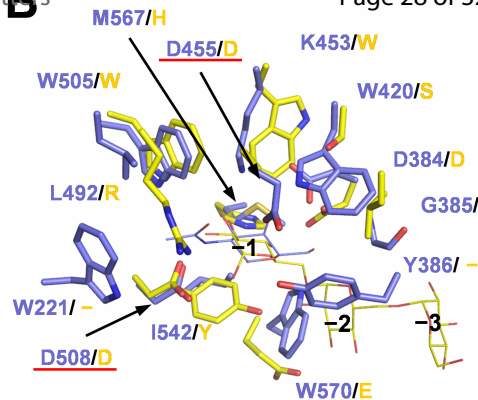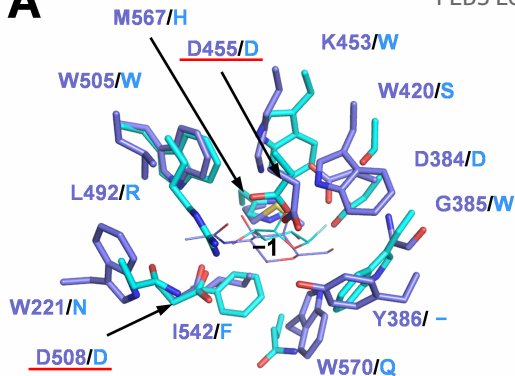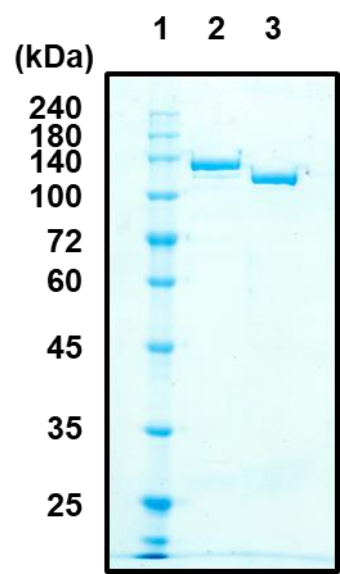
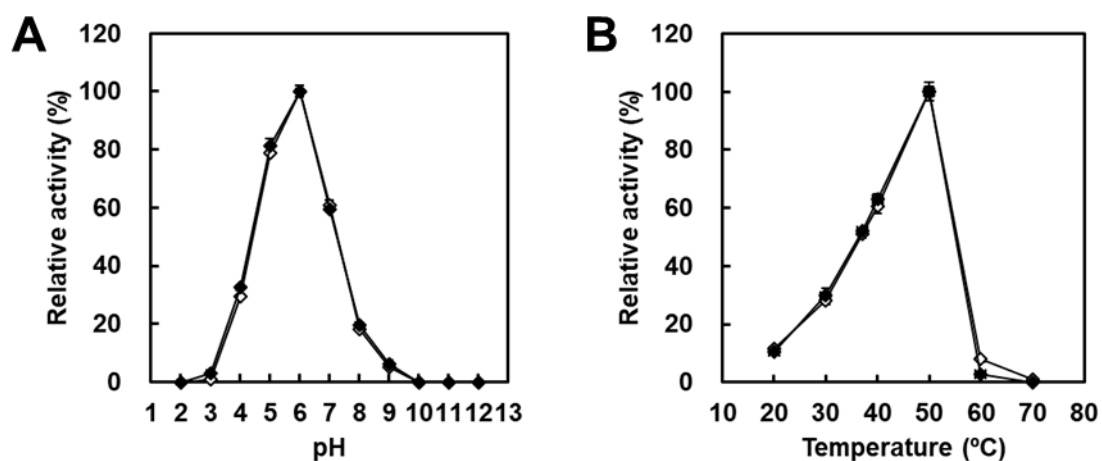**acid/base**
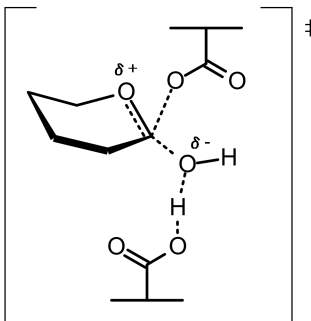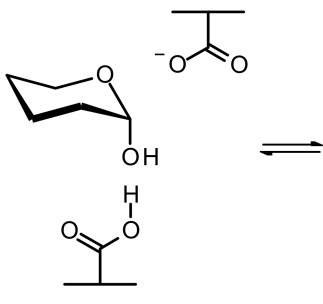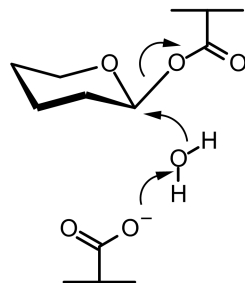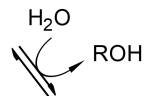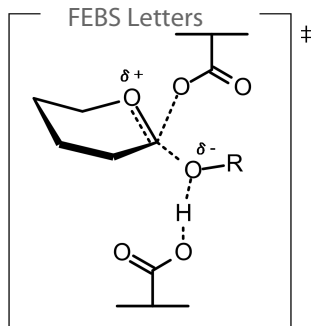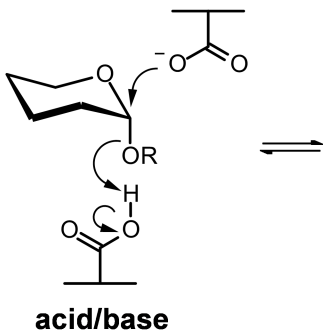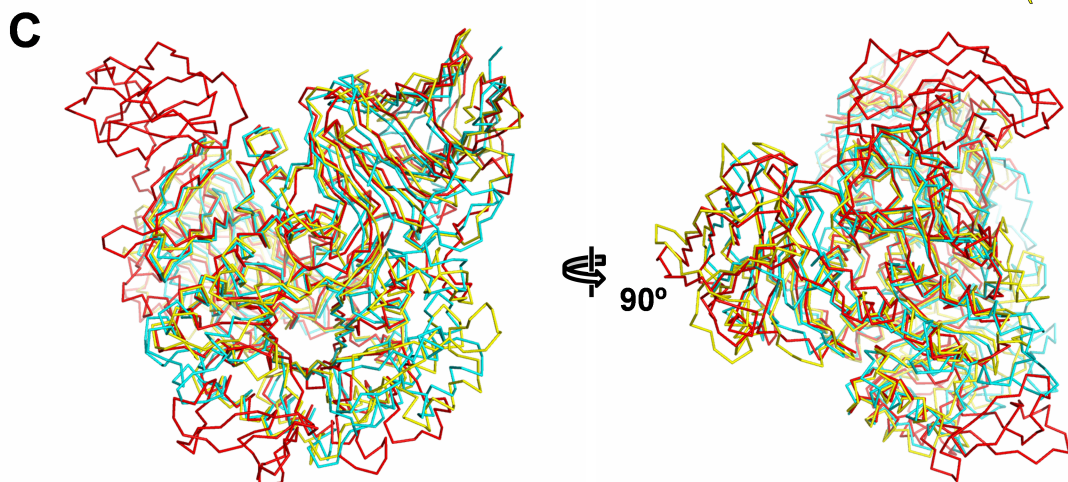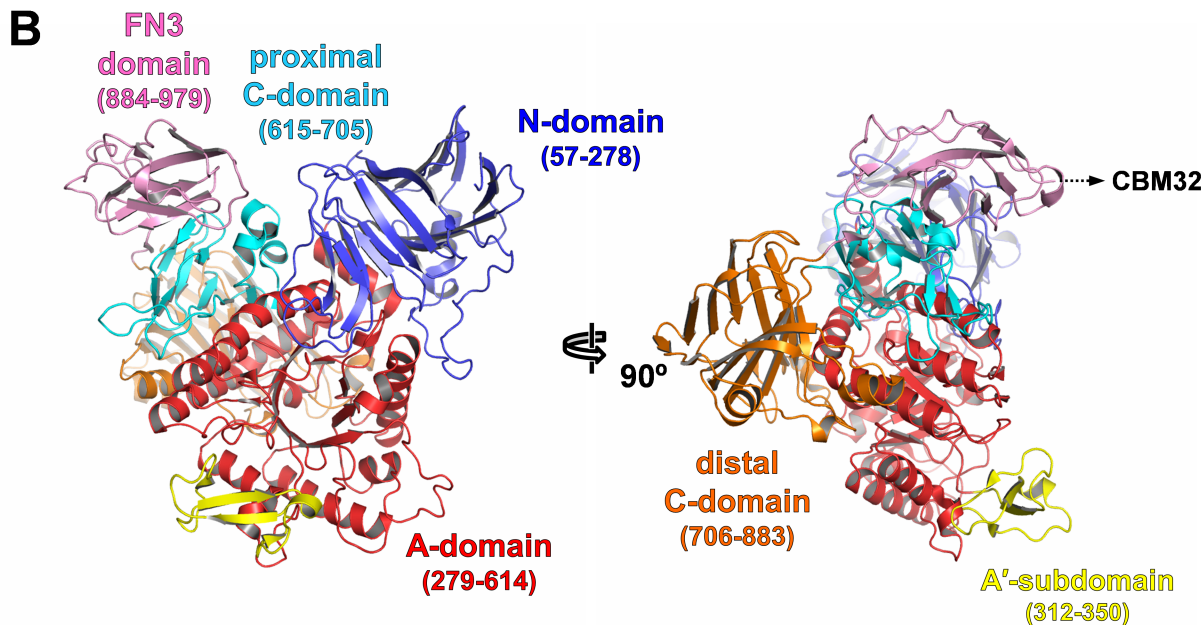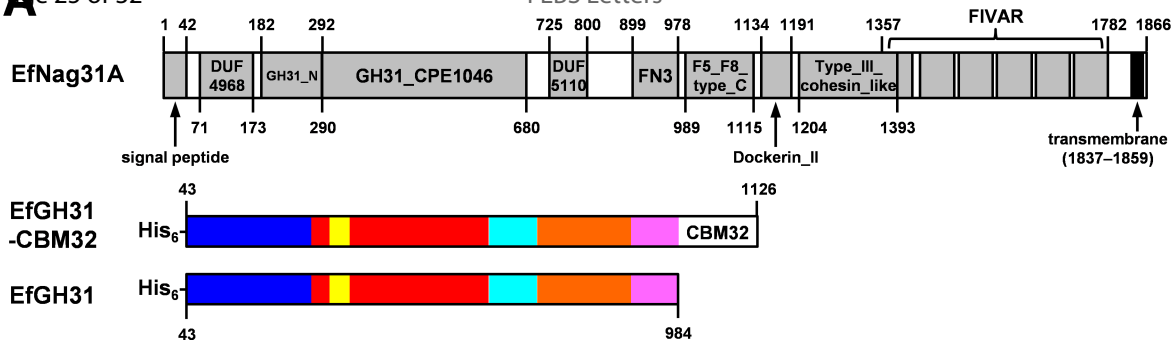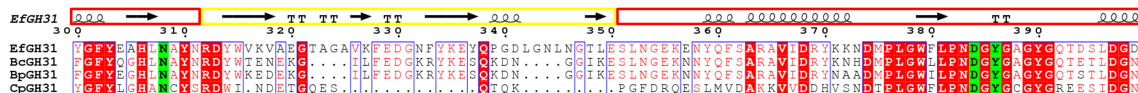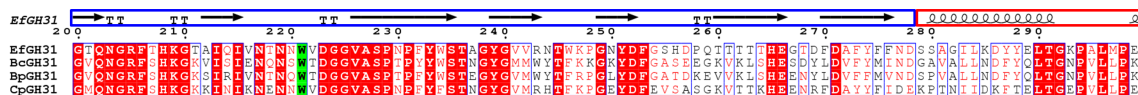
D455 (Nuc)

D508 (A/B)
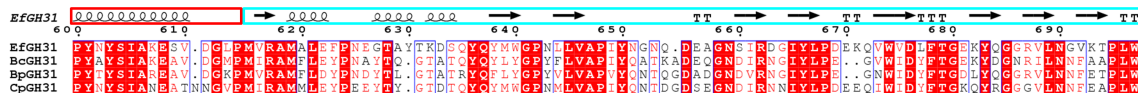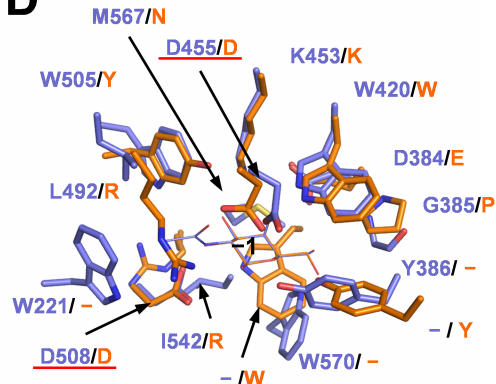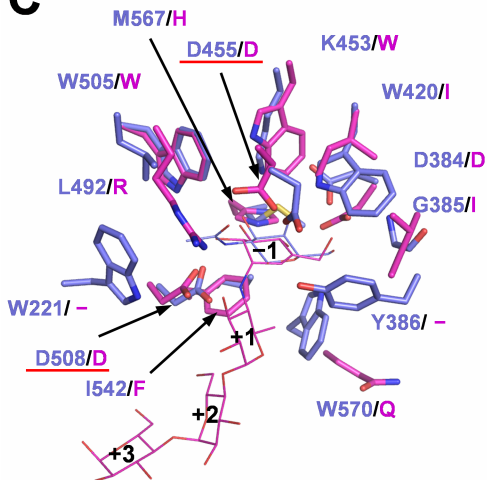
Figure S1. Purities of recombinant EfGH31-CBM32 and EfGH31 used in this study.

Proteins were analyzed by SDS-PAGE with a 10% acrylamide gel. Lane 1, ExcelBand All Blue Broad Range Plus Protein Marker (PM1700, SMOBIO Technology, Hsinchu, Taiwan); lane 2, EfGH31-CBM32; lane 3, EfGH31.

1

Figure S2. Effects of pH and temperature for hydrolytic activity of EfGH31-CBM32 and EfGH31. pH dependence **(A)** and temperature dependence **(B)** of hydrolytic activity the recombinant EfGH31-CBM32 (filled diamond) and EfGH31 (open diamond) hydrolytic activity toward pNP-α-GalNAc substrate. pH dependence was measured at 30 ºC using Britton–Robinson buffer (pH 2.0–12.0) and temperature dependence was measured at 20 ºC–70 ºC using Britton–Robinson buffer (pH 6.0).

Table S1. Glycoside hydrolases structurally homologous to EfGH31 searched using the Dali server.

Enzyme	EC number	PDB	Z score	RMSD (Å)	Sequence identity (%)
GH31 family					
Cellvibrio japonicus α-xylosidase CjXyl31A	3.2.1.177	2XVK	34.8	2.3	24
Flavobacterium johnsoniae dextranase FjDex31A	3.2.1.11 2.4.1.–	6JR8	34.7	2.1	24
Homo sapiens maltase-glucoamylase (MGAM) N-terminal subunit	3.2.1.20	2QMJ	34.0	2.8	18
α-Xylosidase from a soil metagenome MeXyl31	3.2.1.177	5ZN6	33.6	2.4	21
Mus musculus α-glucosidase II (GANAB)	3.2.1.207	5HJR	33.0	2.7	18
Kribbella flavida cycloalternan-specific α-1,3-isomaltosidase	3.2.1.204	5X3I	32.9	2.4	18
Cellvibrio japonicus oligosaccharide α-1,4-transglucosylase CjAgd31B	2.4.1.161	5I24	32.9	2.5	20
Blautia obeum α-glucosidase Roα-G1	3.2.1.20	3POC	32.9	2.8	19
Listeria monocytogenes cycloalternan-forming enzyme LmCAFE	2.4.1.–	5HPO	32.6	2.7	21
Trueperella pyogenes cycloalternan-specific α-1,3-isomaltosidase	3.2.1.204	5I0G	32.2	2.2	20
Bacteroides ovatus α-xylosidase BoGH31A	3.2.1.177	5JOU	32.0	2.5	20
Homo sapiens lysosomal α-glucosidase (GAA)	3.2.1.20	5NN3	32.0	2.7	19
Homo sapiens sucrase-isomaltase N-terminal subunit	3.2.1.10 3.2.1.48	3LPO	31.8	2.7	17
Saccharolobus solfataricus α-glucosidase	3.2.1.20	2G3M	31.0	2.7	21
Beta vulgaris α-glucosidase SBG	3.2.1.20	3W38	30.5	2.8	17
Escherichia coli α-xylosidase YicI	3.2.1.177	2F2H	30.3	2.6	19
Aspergillus niger α-xylosidase	3.2.1.177	6DRU	30.1	2.6	21
Homo sapiens maltase-glucoamylase (MGAM) C-terminal subunit	3.2.1.20	3TOP	29.3	3.2	16
Chaetomium thermophilum α-glucosidase II	3.2.1.207	5DKX	28.9	2.8	18
Paenibacillus sp. 598K 6-α-glucosyltransferase Ps6GT31A	2.4.1.–	5X7Q	28.5	17.1	22
Gracilariopsis lemaneiformis α-1,4-glucan lyase	4.2.2.13	2X2J	27.4	2.9	18
Bacteroides thetaiotaomicron α-glucosidase BT0339	3.2.1.20	5F7C	27.2	2.2	23
Bacteroides thetaiotaomicron α-glucosidase BT3299	3.2.1.20	5DJW	27.1	2.7	20
Agrobacterium tumefaciens sulfoquinovosidase	3.2.1.199	5OHS	26.5	2.6	20
Pseudopedobacter saltans α-galactosidase PsGal31A	3.2.1.22	4XPQ	26.3	3.0	22
Escherichia coli sulfoquinovosidase YihQ	3.2.1.199	5OHT	26.3	2.7	18
Other family					
Homo sapiens GH27 α-*N*-acetylgalactosaminidase	3.2.1.49	3H55	15.8	3.7	9
Gallus gallus GH27 α-*N*-acetylgalactosaminidase	3.2.1.49	1KTB	15.4	3.9	10
Bfidobacterium bifidum GH129 α-*N*-acetylgalactosaminidase	3.2.1.49	5WZP	8.7	4.8	9

3

Table S2. Distances of hydrogen bonds between EfGH31 and GalNAc.

GalNAc atom	Protein atom	Distance (Å)	
Direct hydrogen bond			
O1	Asp455 OD1	2.5	
	Tyr386 OH	2.7	
N2	Asp508 OD1	2.7	
O3	Lys453 NZ	3.0	
O4	Ly453 NZ	3.4	
	Asp384 OD1	2.7	
	Trp420 NE1	2.7	
O6	Asp384 OD1	2.7	
	Trp570 NE1	2.9	
O7 (carbonyl O of acetamide group)	Val456 N	3.0	

GalNAc atom	Protein atom	Distance (Å)	
		Sugar and water	Water and protein
Hydrogen bond via water			
O3	Trp505 NE1	2.5	3.0
	Asp508 OD1	2.5	2.7
	Asp538 OD1	2.5	2.7
O6	Asp308 OD1	3.2	2.9

4