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

Supplementary data

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

Takatsugu Miyazaki[1,2,*], Marina Ikegaya[2], Santiago Alonso-Gil[3,*]

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

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

[3] Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Vienna, Austria.

***Correspondence:** Takatsugu Miyazaki, miyazaki.takatsugu@shizuoka.ac.jp; Santiago Alonso-Gil, santiago.alonso.gil@univie.ac.at

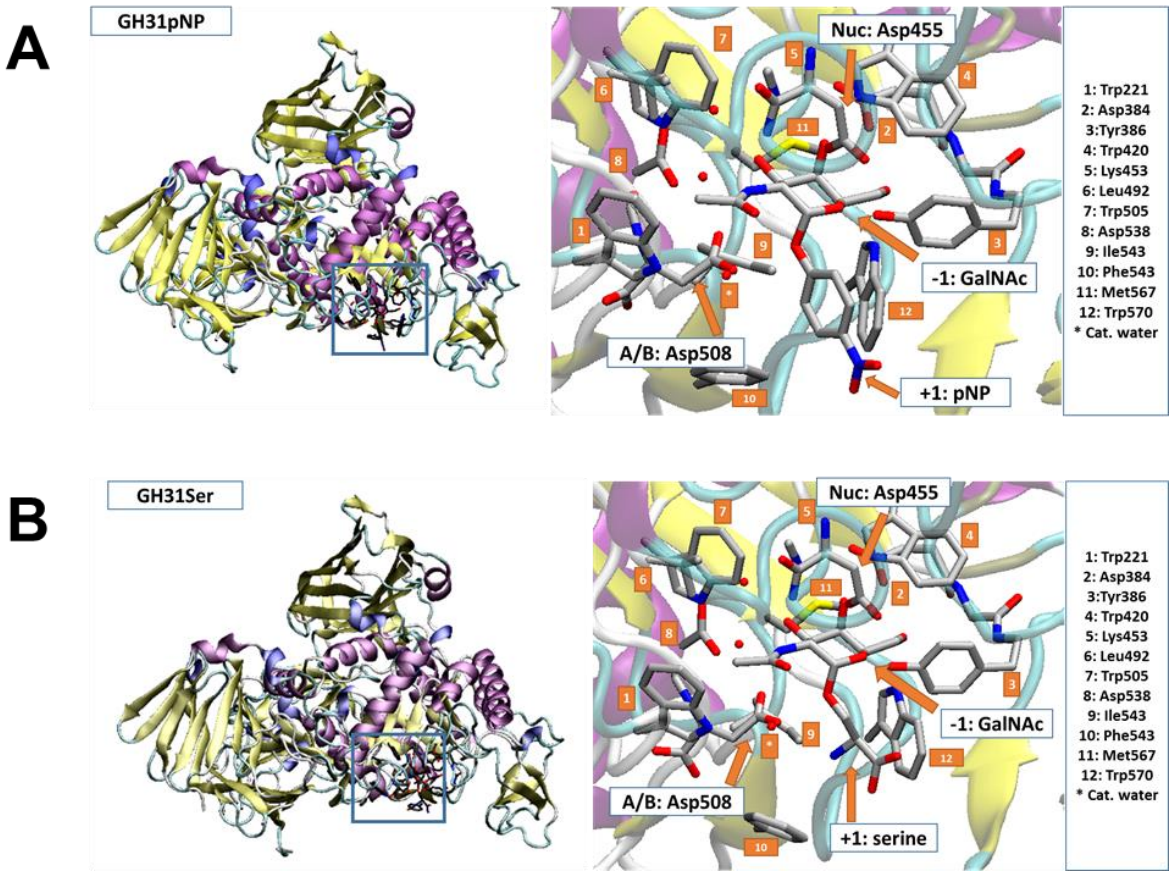

Figure S1. Three-dimensional representation of the GH31pNP (A) and GH31Ser (B) models and the topology of their active sites.
For the sake of clarity, the hydrogen atoms are not shown.

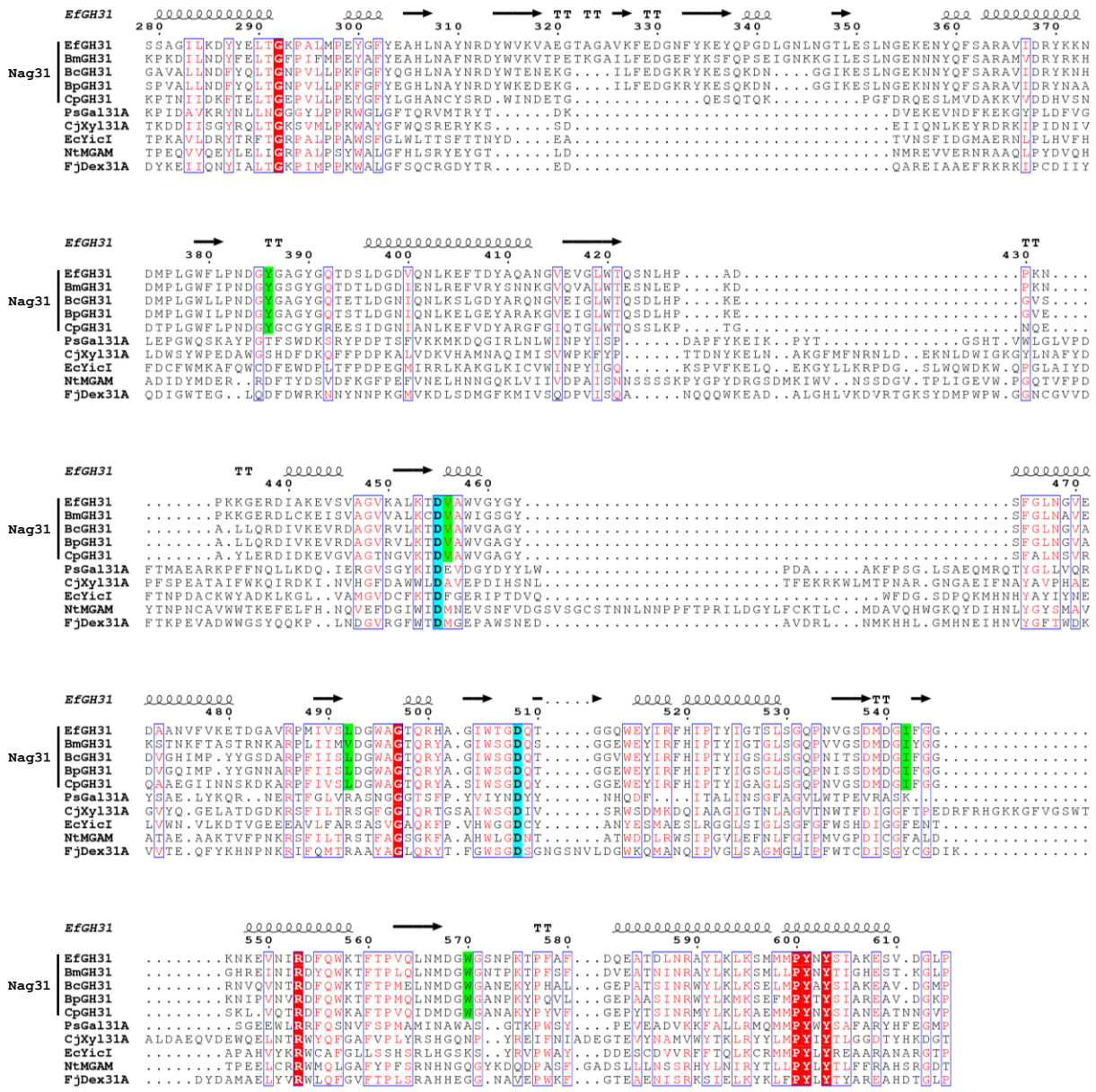

Figure S2. Structure-based alignment of the catalytic domains of Nag31s and the structure-determined GH31 enzymes.

The alignment was performed by PROMALS3D server (Pei *et al*., 2008) using the coordinates of EfGH31 (PDB 6M77), *Pseudopedobacter saltans* α-galactosidase PsGal31A (PDB 4XPO), *Cellvibrio japonicus* α-xylosidase CjXyl31A (PDB 2XVG), *Escherichia coli* α-xylosidase EcYicI (PDB 1WE5), human maltase NtMGAM (PDB 2QLY), *Flavobacterium johnsoniae* dextranase FjDex31A (PDB 6JR6), and the primary sequences of Nag31s from *Bombyx mori* (BmGH31, GenBank BCL50884.1), *Bacteroides caccae* (BcGH31, ASM65008.1), *Phocaeicola plebeius* (BpGH31, EDY97082.1), and *Clostridium perfringens* (CpGH31, ABG84084.1). The figure was generated using ESPript 3.0 (Robert and Gouet, 2014). Secondary structures and amino acid residue numbers of EfGH31 are described above the sequences. Identical residues are shown in white with a red background, while conservative changes are shown in red with a white background. The residues mutated in this study and the catalytic residues are highlighted in green and cyan, respectively.

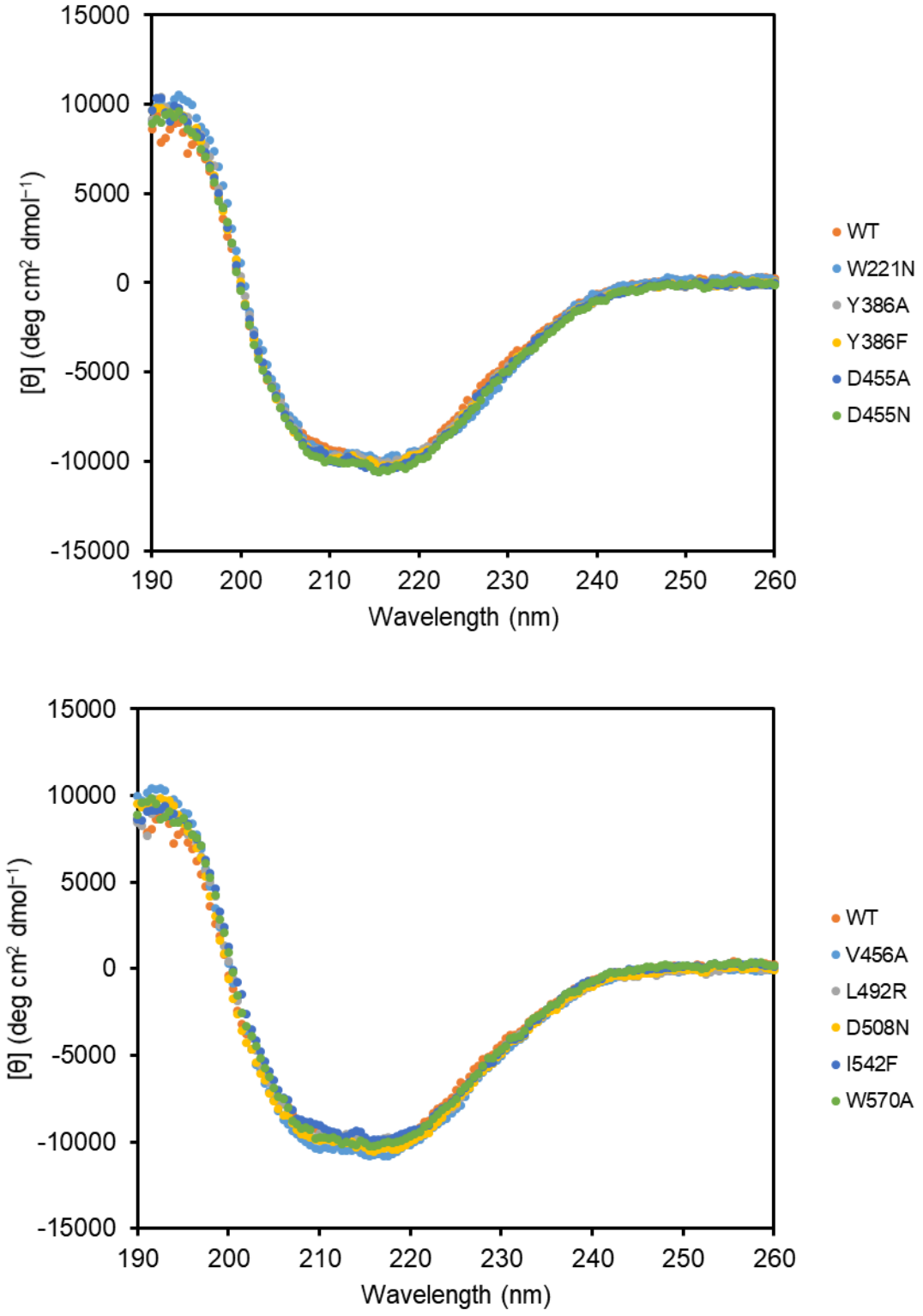

Figure S3. Circular dichroism (CD) spectra of EfGH31 WT and mutants.
The enzymes at a concentration of 0.1 mg/mL (in 10 mM sodium phosphate buffer, pH 7.0) were analyzed using a J-820 spectropolarimeter (JASCO Co., Tokyo, Japan) at room temperature. The upper panel shows the CD spectra of WT, W221N, Y386A, Y386F, D455A, and D455N. The lower panel shows the CD spectra of WT, V456A, L492R, D508N, I542F, and W570A.

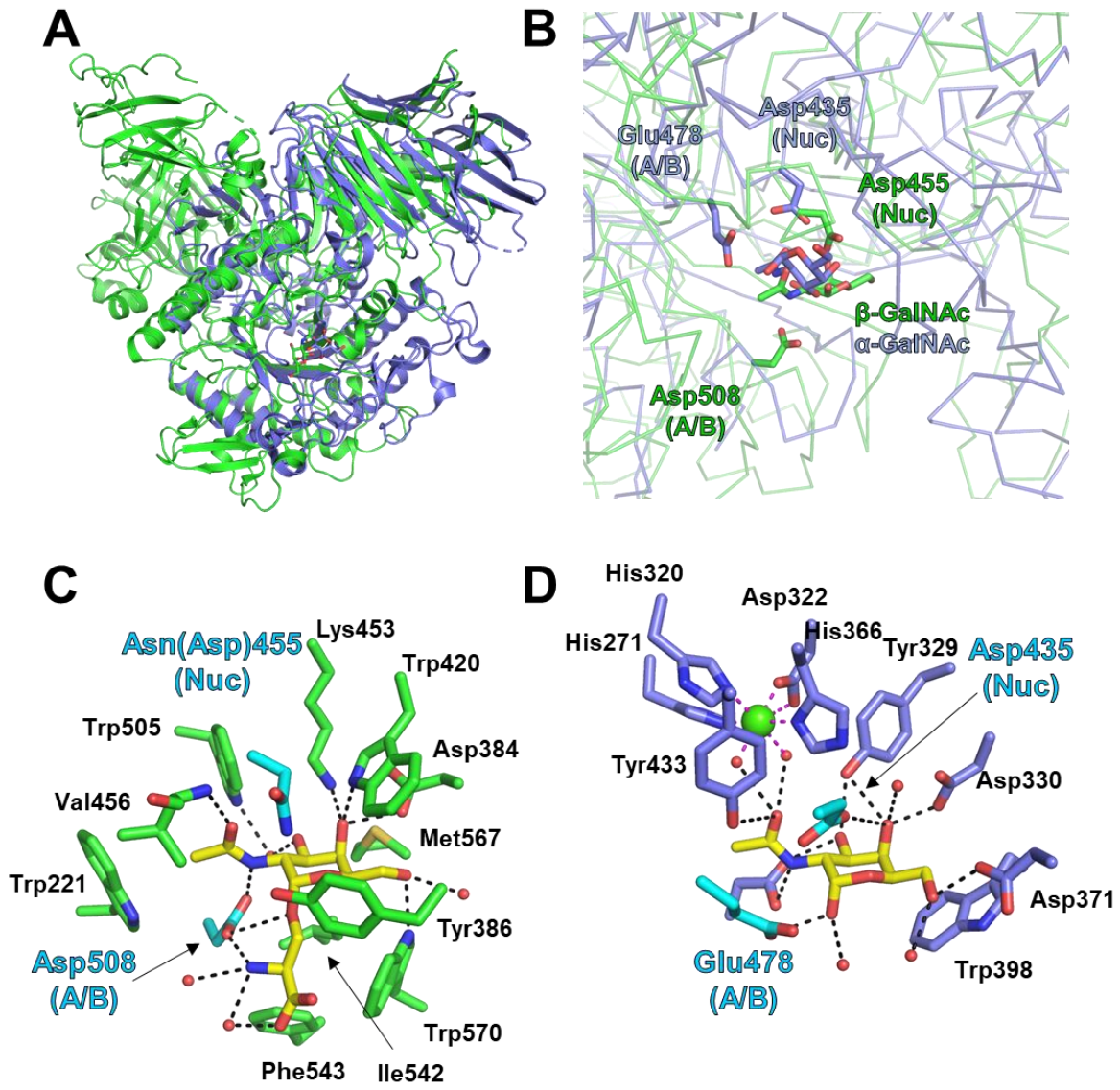

Figure S4. Structural comparison between EfGH31 and GH129 α-*N*-acetylgalactosaminidase NagBb.

(A) Superposition of the overall structures of EfGH31 in complex with β-GalNAc (PDB 6M77, *green*) and NagBb in complex with α-GalNAc (PDB 5WZN, *slate blue*). (B) Active sites comparison based on the whole structure superposition. The catalytic nucleophiles are sterically conserved, while the catalytic acid/base are non-sterically conserved. The orientations of the GalNAc molecules are also different from each other. (C and D) The active site residues of EfGH31-D455N in complex with GalNAcα-Ser (this study) as well as those of NagBb in complex with α-GalNAc (PDB 5WZN). The catalytic residues are colored in *cyan*, ligands are colored in *yellow*, and the others have the same colors as (A). Hydrogen bonds are indicated as *black dashed lines*. A calcium ion is indicated as a *green sphere* and the coordination is indicated as *magenta dashed lines*.

Table S1. Sequences of oligonucleotides for mutagenesis used in this study.

Oligonucleotide	Sequence (5′→3′)
W221N_F	ACAATAACGTCGACGGCGGAGTAGCATCGCCTA
W221N_R	TCCGCCGTCGACGTTATTGTTCGTATTGACAAT
Y386A_F	ACGGCGCCGGCGCTGGCTATGGCCAAACAG
Y386A_R	CCAGCGCCGGCGCCGTCATTTGGCAAGAAC
Y386F_F	GGCTTCGGTGCCGGCTATGGCCAAACAGATTC
Y386F_R	AGCCGGCACCGAAGCCGTCATTTGGCAAGAAC
D455A_F	CAAAACTGCCGTGGCGTGGGTCGGGTA
D455A_R	GCCACGGCAGTTTTGAGAGCTTTAACAC
D455N_F	CAAAACTAACGTGGCGTGGGTCGGGTA
D455N_R	GCCACGTTAGTTTTGAGAGCTTTAACAC
V456A_F	CTGACGCGGCGTGGGTCGGGTATGG
V456A_R	CCACGCCGCGTCAGTTTTGAGAGCTT
L492R_F	TGTTTCACGAGATGGTTGGGCCGGAAC
L492R_R	ACCATCTCGTGAAACAATCATCGGACG
D508N_F	TGGACCGGTAATCAAACGGGTGGTCAATG
D508N_R	TTGATTACCGGTCCAAATACCAGCGTGC
I542F_F	GGATGGATTTTTTGGTGGAAAAAATA
I542F_R	CCAAAAAATCCATCCATATCTGAACC
W570A_F	CGGCGCGGGATCCAATCCAAAAACGCCA
W570A_R	TTGGATCCCGCGCCGTCCATATTTAGCTG

Table S2. Cremer and Pople's puckering coordinates for the 6-membered ring in the subsite -1 of the GH31pNP and GH31Ser cluster models.

Structure	(φ, θ, Q) GH31pNP	(φ, θ, Q) GH31Ser
MC	(187.27, 11.14, 0.55)	(184.99, 10.68, 0.55)
TSgly	(211.19, 38.85, 0.51)	(218.45, 33.30, 0.49)
GEI	(226.14, 62.09, 0.55)	(233.94, 50.57, 0.52)
GEI*	(215.12, 71.62, 0.61)	(221.97, 66.72, 0.58)
TSdegly	(229.84, 42.98, 0.50)	(233.44, 42.74, 0.50)
PC	(164.20, 8.60, 0.54)	(191.88, 8.51, 0.55)

References

Pei J, Kim BH, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res*. 36: 2295–2300.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 42: W320–W324.