




Please cite the Published Version

Zhang, Z, Dong, M, Zallot, R , Blackburn, GM, Wang, N, Wang, C, Chen, L, Baumann, P, Wu, Z, Wang, Z , Fan, H , Roth, C, Jin, Y and He, Y (2023) Mechanistic and Structural Insights into the Specificity and Biological Functions of Bacterial Sulfoglycosidases. *ACS Catalysis*, 13 (1). pp. 824-836. ISSN 2155-5435

DOI: <https://doi.org/10.1021/acscatal.2c05405>

Publisher: American Chemical Society (ACS)

Version: Supplemental Material

Downloaded from: <https://e-space.mmu.ac.uk/636705/>

Usage rights:  In Copyright

Additional Information: This is an author accepted manuscript of an article published in *ACS Catalysis*, by American Chemical Society.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Supplementary Information for

Mechanistic and Structural Insights into the Specificity and Biological Function of Bacterial Sulfoglycosidases

Zhen Zhang[¶] + [∇], Mochen Dong^{+∇}, Rémi Zallot^{^&∇}, George Michael Blackburn[‡], Nini Wang[¶],

Chengjian Wang[∇], Long Chen⁺, Patrick Baumann⁺, Zuyan Wu⁺, Zhongfu Wang[∇], Haiming Fan[¶],

Christian Roth[√], Yi Jin^{+&}, Yuan He^{¶*}*

[¶] Key Laboratory of Synthetic and Natural Functional Molecule, College of Chemistry and Materials Science, Northwest University, Xi'an, 710127, P. R. China

[#] College of Life Science and Technology, Northwest University, Xi'an, 710069, P. R. China

[^] Institute of Life Sciences, Swansea University Medical School, Swansea, SA2 8PP, United Kingdom

[‡] School of Biosciences, University of Sheffield, Firth Court, Western Bank, Sheffield, S10 2TN, United Kingdom

[√] Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Arnimallee 22, 14195 Berlin, German

[∇] Glycobiology and Glycotechnology Research Center, College of Food Science and Technology, Northwest University, Xi'an 710069, P. R. China.

⁺ School of Chemistry, Cardiff University, Cardiff, CF10 3AT, United Kingdom

[&] Present address: Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

[▽]Z.Z, M.D., and R.Z. contributed equally to this work.

*Email: yi.jin@manchester.ac.uk; yuanhe@nwu.edu.cn.

Supplementary Table 1 Oligonucleotides used for sequencing and mutagenesis.	12
Supplementary Table 2 Data collection and refinement statistics.	14
Supplementary Table 3 Complex structures containing 6SO3 (CAzy Carb_ligand: 6SO3).16	
Supplementary Figure 1. Identification of cellular localizations of BbhII in <i>Bifidobacterium bifidum</i> JCM1254.	18
Supplementary Figure 2: Detection of released 6S-GlcNAc from natural and synthetic substrates using HPLC-MS.	19
Supplementary Figure 3. The standard curves of 4MU in the universal buffer at various pHs and pH dependence of the hydrolysis of 4MU-6S-GlcNAc by BbhII _{WT} , Bt4394 _{WT} , SGL _{WT} , and Niako3494 _{WT}	21
Supplementary Figure 4 Michaelis-Menten plots for illustrating substrate specificity of non-sulfated 4MU-GlcNAc with BbhII _{WT} , Bt4394 _{WT} , SGL _{WT} and Niako3494 _{WT}	22
Supplementary Figure 5 Michaelis-Menten plots for hydrolysis of 4MU-6S-GlcNAc by wild-type BbhII and its variants at pH 6.0.	23
Supplementary Figure 6 Structures of the 6S-NAG-oxazoline intermediate complexes of BbhII in comparison with the SpHex and other GH20 structures.	25
Supplementary Figure 7. Identification of cellular localizations of Bt4394 in <i>Bacteroides thetaiotaomicron</i>	26
Supplementary Figure 8 Inhibition of recombinant Bt4394 by Na ₂ SO ₃ , Na ₂ SO ₄ , and NAG-thiazoline.	27
Supplementary Figure 9 Michaelis-Menten plots for 4MU-6S-GlcNAc hydrolysis by Bt4394 _{WT} and Bt4394 variants at pH 5.5.	28
Supplementary Figure 10 Structures of the oxazoline intermediate complexes of Bt4394 in comparison with ScHex structure and other GH20 hexosaminidases structures with catalytic conformation.	29
Supplementary Figure 11 The structure comparison of BbhII with Bt4394 indicates the size difference of the leaving group in the substrates.	30
Supplementary Figure 12 Multiple sequence alignment (MSA) by T-coffee with default setting (https://www.ebi.ac.uk/Tools/msa/tcoffee/) for the known 6S-GlcNAcases.	31
Supplementary Figure 13 Michaelis-Menten plot for the hydrolysis of 4MU-6S-GlcNAc by SGL _{WT} at pH 6.0.	32

Supplementary Figure 14 Michaelis-Menten plots for the hydrolysis of 4MU-6S-GlcNAc by wild-type Niako3494 and its variants.	33
Supplementary Figure 15 Active site of Niako3494 and its sequence aligned with “sulfate-binding site” in other 6S-GlcNAcases.	35
Supplementary Figure 16 Michaelis-Menten plots for illustrating substrate specificity of pNP-6P-GlcNAc with BbhII _{WT} , Bt4394 _{WT} , and SGL _{WT}	36
Supplementary Figure 17 Sequences with 6S-GlcNAcase activity and the reference GH20 SpHex (all highlighted in •) on the UniRef90 SSN for the protein family PF00728, GH20, at AST 83 for the full length (a) and AST 57 for the GH20 domain only (b).	38
Supplementary Figure 18 6S-GlcNAcases, Bt4394, BbhII, SGL, human HexA, Niako3494, F3-ORF26, F10-ORF19, and the reference GH20 SpHex (all highlighted in •) mapped on the UniRef90 SSN for the protein family PF00728 representing GH20.	41
Supplementary Figure 19 The GH20 domain sequences with the sulfate-binding feature as in Bt4394.	44
Supplementary Figure 20 The GH20 domain sequences with the sulfate-binding feature as in BbhII.	46
Supplementary Figure 21 Comparison of the flexibility of the catalytic loop with the D-E diad relative to the conserved histidine (H467 in BbhII, H270 in Bt4394), which has been proposed to modulate the p <i>K</i> _a s of catalytic diad by oscillating between the two in a GH20 family, in 6S-NAG-oxazoline intermediate structures of BbhII and Bt4394. The whole catalytic loop has been colored in yellow.	47
Supplementary Figure 22 The observed NAG thiazoline-SO ₂ ⁻ species found in the Bt4394 active site.	48
Supplementary Figure 23 MS spectrum of 1 (ESI-MS).	49
Supplementary Figure 24 ¹ H NMR spectrum of 1 in D ₂ O (500 MHz).	50
Supplementary Figure 25 ¹³ C NMR spectrum of 1 in D ₂ O (125 MHz).	50
Supplementary Figure 26 ³¹ P NMR spectrum of 1 in D ₂ O (202 MHz).	51

1. Experimental procedures

Molecular Cloning

BbhII: dsDNA encoding BbhII (GenBank: AB504522.1 from *Bifidobacterium bifidum* JCM1254, Uniprot ID: D4QAP5) from residue Q31 (without the signal peptide sequence in bold) was codon optimized for *E. coli*, synthesized (Genscript, USA) and cloned into the pET23(a) expression vector using the *NdeI/XhoI* restriction sites. The resulting plasmid encoded a C-terminal His₆-tagged protein.

>BAI94823.1

MNIKRRGLARFMSLICASAMLLVPASSALAQHNAEAAAAAESSTSTVSNLATMATVTASGREVVSSGFGPEL
AADNQDLPDNPDKSVHNASGASRWSADRGSGPWLAYEFPGEATISSVNIWGNNTYATNYSIQTSDDGSN
WTDVKTGLKATAQAQVWKTTFDTPIKTRHIRMIATTKSQSWSLSVWEMRTMGTISAVATDPLSRLTPRPLY
AQSDAGEAFELKKNCTCVSVSDGSLLPADVMDRDELGTSYGLKLAEGTNCPIFTFLDENLDVTGHVGSQAQSI
TADAEYTIIVSDADSVTVKARSATAGIWAQAQTLQLLQIGPWTNSTVKLADVAFIPAVNIADAPRYQWRGVLVD
PARSFYPLDEMKGQ MIDVMSAYKMNTLHLHLSEDEGFRVEITNDGRADGDTTDTYQLAIKSGAISYQSAWTS
NWSPAQDGRGTGYWTQSEFIELVAYAADHGIAIVPEIDGPGHSFSLHGLAELNTGNSNPKPAAGEDTPAFI
QSAQGRSSLATDADITYTVLGHIMDQLDGMIDKGIKASTMPASELKRMYFHLGGDELFLSGGAGNKTERLQ
EYLGSRGALVKERDKTTIVWVDGLDAVDQIPEGSVVQHWGNAANNASIQKLLNQRNGKIIMSPAGNTYFP
QRPGTETTGVTWACGACTTSNFYQWNPTSSAGTTEDKVLGVEDALWSEHLRSLNDAEFLMYTRMMATAEVG
WTQQNRKDYDNWNKRVGDIAIDLNRGANFHKATEVTSWKGSYAAVDAEQKVTGDKVLVGRYAEPGLTGT
DGLSFTATYTAEGGTAVNLPVTPDMKQTYSSQQQLKNGRLVNGAHMNSIVDVVYVTLPSDVLAAADSEAVGRL
DVSVSSTYPIPSDSSMSIAIKDGKVTQTWTGDERPTDPDPEPEVTVVSIKASTSQSDVKVGDTFDPSKV
KVVATKSDKTTAVLAAADYTIAVTDKDGNAIDVTKPFEEAGDLTVTVALKDDGSIKDSFTMTVTDKGAVDP
DPDPTPKPNPDPQKPSGDNKPKQIKPEGGKPGDVVAETGASVSGAALAAMICAAGAIVMLAVRRQRR

Bt4394: dsDNA encoding Bt4394 (GenBank: AAO79499.1 from *Bacteroides thetaiotaomicron* VPI-5482, Uniprot ID: Q89Z13) from residue Q22 (without the signal peptide sequence in bold) was codon optimized for *E. coli*, synthesized (Sangon Biotech, Shanghai) and cloned into the pET23(a) expression vector using the *NdeI/XhoI* restriction sites. Primers are shown in Supplementary Table 1. The resulting construct encoded a C-terminal His₆-tagged recombinant protein.

>WP_011109236.1

MKIKHFLPLLLLLGSNEMLTAQEIALTPQPAHLTVKDGREFEFGNQLKAKVTPYQGDSIRMVFESEFKKELQE
ATGIKVSSTQKEAKARIILDLNPQLPAEAYKLVNSKKQVRIEASRPAGFYALQTLKQLMPRNV MAGVATS
DHSQWSLPSVEIEDAPRFEWRGFMLDEGRHFFGKDEIKRVIDMMAIYKMNRHFHWHLTEDQGWRIEIKKYPK
LTETGAWRNSKVLAYGDVVKPDGERYGGFYTKDIKEIVAYAKKFFIEIPEIDI PGHSQAAVAAYPEFLAC
DPRDKHEVWLQQGISTDVINVANPKAMQFAKEVIDELTELEFPFNYIHLGGDECPTRKWQKNDECKKLLSEI
GSSNFRDLQIYFYKQLKDYIATKPADQQRQLIFWNEVLHGNTSILGNDITIMAWIGANAAAKQAAKQGMNT
ILSPQIPYYINRKQSKLPTEPMSQGHGTETVEAVYNYQPLKDVAALQPYKGVQANFWTEWVTEPSVLEY
LMLPRLAAVAEAGWTPQEKRNVEDFKERIRKDAELYDLKGWNYGKHIMK

SGL: dsDNA encoding SGL (GenBank: AAW30398.1 from *Prevotella* strain RS2, Uniprot ID: Q5MAH5) from residue Q20 was codon optimized for *E. coli*, synthesized and cloned into the pET22(a) expression vector using the *NcoI/XhoI* restriction sites (Genscript, USA). The resulting plasmid encoded a C-terminal His₆-tagged protein with a pelB signal peptide for translocation to the periplasm.

>AAW30398.1

MKKLCFALVMLLSCVCSYAQTITTYSLSEAPITKEELLKGNLKVLRVWTTSKDLYLSNANNSNAFDANGS
TVFAVEAVKGGVKLVKNSNEYFGGEGAALARTADA AKAKVFSVVKIGSKVAN THADADETRSFWITCQAG
GTKYLNTNTSN NATVQYAGGNWSTFYIYKVSEEQI AVPYPTCIIPAPRGAELN GEGRLALSAMDNISFT
TDPALAEAYVLNITADGISVASSTEKGFYALQSLAQLAEGNAEGLPLVRIADKPRFGYRGFMLDVSRLF
FSVAEVKMMIDIMARYKMNVFHWHLTDDQGWRAEIKRYPKLTTVGATRSNDYDTPITKIEENGQVYWTGNG
AKTGKPYGPYFYTQDEMREVVAYAKERHIEVLPEVDMPGHFVAAMAAYPEYSCNPSRAPQVWTGGGISSDV

LNVANPQAVEFAKNILDELCDIFPYPYIHVGGDECPTTQWEHNDLCQQKYKELGLTSYRQLQAHFIKDLAD
FVATKNKHLVCWNEAITAGGADLDLMKQTQSTIMSWNPCQEGVAVKVKLGLPAIVTEYHKDGGYYICRK
QSN DYGEPSGAGYGN DGVGECYNYVPVQGM YTQEQMALVKG VQGTFWTEHVGTNEYLEYLALPR LICVAEA
GWTPQVFNWDFRTRLANQTQWLDDHGYVYARHWMPGYVPRTPMPENEKVAATPELSSLKSPKWYRIKF
TAGGTYLQMN GSNLSTNALFQNDKSQY YAI IPTSTKKPSLSSVKLYSANGYVWTTKEGTATNGQSGTFVCG
TESEAKALYLN LQKHTADTQK WVI TKIADKTTGFNTWGGDNP GANIGFWNVNSNSDKVEFTTDDYNAGVDT
AIGTIKADDAHNNLTA AVYNAQGMRLNGMQQLNVVLFADGTAQKVFMK

Niako3494: dsDNA encoding Niako_3494 (GenBank: AEV99795.1 from *Niastella koreensis* GR20-10, Uniprot ID: G8TKW6) from residue 28 was codon optimized for *E. coli*, synthesized (Fishier Scientific, UK) and cloned into the pET28 expression vector using the *NcoI* / *XhoI* restriction sites. The resulting plasmid encoded an N-terminal His₆-tagged recombinant protein.

>AEV99795.1

MKYPRTCCFFSALLLFYSMIAAINFASAQTSEQFEWNLKLPVKAMLLTVPHPEDVPEFCRFI KEVLPKEGVNTLVLR
IRYNYKFKSHPELAGERAISEQQLKQIVQTCKEAKIRFIPKMNLLGHQSDRDHIDPLLAKYPQFDESPDYNPPVP
WKDAGPFDFYCKSLCPSHPDLLKTI FPLMDELIDVCGADAFHVGLDEVWILGYEKPCRCGGRDKAALFAEYATKL
HDHLKEKKCKMWMWSDR LIDGKTTNLLGWQASMNATFRAIDL IPTDIMICDWKYESAPPTPGYFAIKGFNVLPSS
CSNSEVALAQLAQVRLARKDGT RAPWAVTLAERMQGVFVTMWEDSKEFIDAYYGRNGKKLPSAETFKAVFAQIRK
EVMN

Site directed mutagenesis

The oligonucleotide primers listed in Supplementary Table 2 were used to carry out site-directed mutagenesis. For Bt4394 and BbhII, two separate PCRs were set up in 25 μ L reaction mixture containing either 0.04 μ M forward or reverse primers each, 500 ng template DNA, 0.2 mM dNTPs, and 0.5 U of Hieff® II Pfu DNA Polymerase (Yeasen, Shanghai) for 28 cycles. The products of these two reactions were purified by agarose gel electrophoresis and mixed in equimolar amounts and cooled gradually to allow reannealing of the complementary strands. The temperature gradient was set as follows: 95 °C for 5 min, 90 °C for 1 min, 80 °C for 1 min, 70 °C for 0.5 min, 60 °C for 0.5 min, 50 °C for 0.5 min, 40 °C for 0.5 min and holding at 37 °C. PCR products were immediately digested by *DpnI* for 18 h at 37 °C and transformed into *E. coli* TOP10 competent cells. The plasmids were purified using a miniprep kit (CoWin Biosciences, Beijing), and subsequently verified by sequencing to ensure the mutations were successful.

For Niako3494, 25 μ L PCR reactions were setup in a mixture containing both 300 ng forward and reverse primers, 500 ng Niako3494 template DNA, 12.5 μ L of PrimeSTAR® Max DNA Polymerase Premix (Takara-bio INC). Three steps were used for all the reactions as follows: 98 °C for 10 s, 55 °C for 5 s, 72 °C for 5 s/kb for 35 cycles and final at 4 °C. PCR products were immediately digested by FastDigest *DpnI* (Thermo Fisher Scientific) for 15 min at 37 °C and transformed into *E. coli* XL1-Blue competent cells. The plasmids were purified using a miniprep kit (QIAGEN, Germany), and subsequently verified by sequencing to ensure the mutations were successful.

Gene expression and protein purification

Bt4394: All Bt4394 wild-type and variants were expressed in *E. coli* Rosetta™ (DE3) pLysS strain (Novagen). The transformed cells were selected with 100 μ g/mL ampicillin (Sangon Biotech, Shanghai) and 15 μ g/mL chloramphenicol (Sangon Biotech, Shanghai) on Luria-Bertani (LB) agar medium (10 g/L tryptone, 5 g/L yeast extract and 10 g/L of NaCl, pH 7.4) by overnight incubation at 37 °C. Transformed cells were grown in LB media containing 100 μ g/mL ampicillin and 15 μ g/mL chloramphenicol at 37 °C until the OD₆₀₀ reached 0.6. The culture was cooled to 18 °C before a final concentration of 0.3 mM isopropyl- β -D-thiogalactopyranoside (IPTG, Sigma-Aldrich) was added, and then further incubated for 20 h at 180 rpm. Cells were subsequently harvested by centrifugation at 6000 rpm for 20 min at 4 °C, resuspended in 20 mL buffer A (25 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 mM imidazole) with lysozyme (Sangon Biotech, Shanghai, 0.5 mg/mL), and incubated at 4 °C for 30 min. The cells were lysed, and the lysate was centrifuged at 15000 rpm, 4 °C for 45 min. The supernatant was filtered through a 0.45 μ m syringe filter (Jinteng, Tianjin) before being loaded onto a 5 mL HisTrap™ column (GE Healthcare) which had been pre-equilibrated with 50 mL

buffer A. Subsequently, the column was washed with 60 mL buffer A and the protein was eluted with 20 mL buffer B (25 mM Tris-HCl pH 8.0, 300 mM NaCl, 250 mM imidazole). The fractions containing the eluted protein, confirmed by SDS-PAGE, were concentrated and further purified by size exclusion chromatography (SEC) on a HiLoad 16/600 Superdex 200pg column (GE Healthcare) with buffer C (25 mM Tris-HCl pH 8.0, 300 mM NaCl). The purity of the fractions was assessed by SDS-PAGE and all of those that are >95% pure were combined and concentrated prior to the crystallization screening and kinetics measurements.

BbhII: Wild-type BbhII and its variants proteins were prepared in a similar way to those of the Bt4394. The differences are that the plasmids were transformed into *E. coli* BL21(DE3), 100 µg/mL ampicillin was used, and the final concentration of IPTG for induction was 0.1 mM. In addition, the concentration for all purification buffers was 50 mM HEPES, pH 8.0.

SGL: Wild-type *sgl* gene was expressed in *E. coli* Shuffle T7 Express (NEB). The transformed cells were incubated on LB agar with 100 µg/mL ampicillin overnight at 30 °C. Transformed cells were grown in LB media containing 100 µg/mL ampicillin at 30 °C until the OD₆₀₀ reached 0.6. The culture was cooled to 16 °C before a final concentration of 0.4 mM IPTG was added, and further incubated for 20 h at 230 rpm. Cells were subsequently harvested by centrifugation at 6000 rpm for 20 min at 4 °C, resuspended in 20 mL Lysis Buffer (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 40 mM imidazole), and incubated at 4 °C for 30 min. The cells were lysed by sonication and the lysate was centrifuged at 18000 rpm, 4 °C for 45 min. The supernatant was filtered through a 0.45 µm syringe filter before being loaded onto a 20 mL Ni-NTA batch purification column (GE Healthcare) which had been pre-equilibrated with 100 mL Lysis Buffer. Subsequently, the column was washed with 60 mL Lysis Buffer and the protein was eluted with 100 mL Elution Buffer (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole). The fractions containing the eluted protein, confirmed by SDS-PAGE, were pooled and concentrated, buffer-exchanged into Buffer C (25 mM Tris-HCl pH = 8.0). This protein solution was further purified by being loaded onto a 20 mL pre-equilibrated Q-Sepharose column in Buffer C, and eluted by applying a 0-60% gradient of Buffer D (25mM Tris-HCl pH = 8.0, 1 M NaCl) over 15 CV. The purity of the fractions was assessed by SDS-PAGE and all of those that are >95% pure were combined and concentrated prior to the crystallization screening and kinetics measurements.

Niako3494: The genes of wild-type Niako3494 and its variants were expressed in BL21(DE3) Star strain (NEB). The transformed cells were incubated on LB agar with 50 µg/mL Kanamycin overnight at 37 °C. Transformed cells were grown in LB media containing 50 µg/mL kanamycin at 37 °C until the OD₆₀₀ reached 0.6. The culture was cooled to 25 °C before a final concentration of 0.5 mM IPTG, and further incubated for 16 h at 200 rpm. Cells were subsequently harvested by centrifugation at 4500 rpm for 25 min at 4°C, resuspended in 40 mL Buffer A (Tris-HCl 20 mM, imidazole 40 mM, NaCl 500 mM, pH 7.5, PMSF 1 mM), and incubated at 4 °C for 30 min. The cells were lysed by sonication and the lysate was centrifuged at 20000 rpm, 4 °C for 25 min. The supernatant was filtered through a 0.45 µm syringe filter before being loaded onto a pre-equilibrated 5 mL HisTrap column (GE Healthcare). Subsequently, the column was washed with buffer A until UV absorbance decreased to baseline and Niako3494 was eluted with 20%~80% gradient buffer B (Tris-HCl 20 mM, imidazole 500 mM, NaCl 500 mM, pH 7.5). The purity of the fractions was assessed by SDS-PAGE and the fractions containing Niako3494 protein were combined and concentrated before being loaded onto the SEC (GE Superdex75 26/600). The protein was eluted by buffer C (Tris-HCl 20 mM, NaCl 500 mM, pH 7.5). The purity of the fractions was assessed by SDS-PAGE and all those that are >95% pure were combined and concentrated prior to the crystallization screening and kinetics measurements.

pH profile of activity for Bt4394, BbhII and SGL

All the initial rates of reactions were performed by monitoring the fluorescence change for one minute at 25 °C. 15 µM 4-methylumbelliferyl 6-sulfo-2-acetamido-2-deoxy-β-D-glucopyranoside (4MU-6S-GlcNAc, Merck) dissolved in deionized water was used as substrate. Both wild-type Bt4394 and BbhII enzymes were assayed by measuring the

fluorescence of 4-methylumbelliferone (4MU) at $\lambda_{\text{ex}} = 360 \pm 10$ nm and $\lambda_{\text{em}} = 450 \pm 10$ nm. The buffer used contained 25 mM Bis-tris propane, 25 mM citrate, 300 mM NaCl and was titrated with HCl to the final full range of pHs. All experiments were performed in triplicate.

Michaelis-Menten Kinetic

Michaelis-Menten kinetics for wild-type and variants of BbhII, Bt4394, Niako3494, and sgl were measured for the enzyme-catalyzed hydrolysis of 4MU-6S-GlcNAc (Merck) and 2-acetamido-2-deoxy- β -D-glucopyranoside (4MU-GlcNAc, Merck). The initial release rates of fluorescent 4MU in 100 μ L reactions were monitored continuously at $\lambda_{\text{ex}} = 360$ nm and $\lambda_{\text{em}} = 450$ nm using a BMG Fluostar fluorescence microplate reader. All reactions were performed at 25 °C in the buffer containing 25 mM Bis-tris propane, 25 mM citrate, 300 mM NaCl and titrated with HCl to optimal pHs. The concentration of the 4MU formed was assessed using a 4MU standard curve in the same buffer as the kinetics assays. Kinetic parameters (k_{cat} , K_{M} , $k_{\text{cat}}/K_{\text{M}}$) were calculated using the Michaelis-Menten equation $y = E_{\text{t}} * k_{\text{cat}} * x / (K_{\text{M}} + x)$, which is in the GraphPad Prism 6.01 Software. All experiments were performed in triplicate.

Michaelis-Menten kinetics for wild-type BbhII, Bt4394, and SGL were measured for the enzyme-catalyzed hydrolysis of pNP-6P-GlcNAc synthesized in-house. Because the last step of the purification after the synthetic step was reverse phase chromatography (C18), the concentration of the pNP-6P-GlcNAc stock was calibrated by the extinction coefficient of 10800 $\text{cm}^{-1}\text{M}^{-1}$ at 300 nm in 0.1 M NaOH buffer, pH 13. The initial release rates of pNP in 100 μ L reactions were monitored continuously at $A_{405\text{nm}}$ using a UV spectrophotometer (Shimadzu UV-2600). All reactions were performed at 25 °C in the buffer containing Bis-tris propane-HCl 50 mM, NaCl 100 mM, pH 7.0. The concentration of the pNP released was assessed using a pNP standard curve in the same buffer as the kinetics assays. Kinetic parameters (k_{cat} , K_{M} , $k_{\text{cat}}/K_{\text{M}}$) were calculated using the Michaelis-Menten equation $y = E_{\text{t}} * k_{\text{cat}} * x / (K_{\text{M}} + x)$, which is in the GraphPad Prism 6.01 Software. All experiments were performed in triplicate.

The inhibition effect of Na_2SO_3 , Na_2SO_4 and NAG-thiazoline on Bt4394

The activity inhibition for the wild-type Bt4394 on 4MU-6S-GlcNAc was measured in the presence of the potential inhibitors, including Na_2SO_3 , Na_2SO_4 and NAG-thiazoline (Merck). A dilution series of NAG-thiazoline were pre-dissolved in DMSO (final concentration 10 μ M to 2 mM). All the reaction rates were measured in the buffer containing 100 mM Tris-HCl pH 8.0, 300 mM NaCl and 0.4% v/v DMSO. The release of the fluorescent 4MU was monitored continuously for 60 s at 25 °C using a fluorescence spectrophotometer ($\lambda_{\text{em}} = 445$ nm and $\lambda_{\text{ex}} = 365$ nm).

Crystallization

Bt4394: All Bt4394 proteins for crystallography were in the SEC buffer. To form the Bt4394_{WT}-6S-GlcNAc product complex, 500 mM 4MU-6S-GlcNAc (dissolved in deionized water) was mixed with Bt4394_{WT} protein stock (10 mg/mL) at 1:49 prior to crystallization. Co-crystallization was initially screened by sitting drop vapor diffusion method at 9 °C, by mixing 0.2 μ L premixed protein solution and 0.2 μ L precipitant against 75 μ L of reservoir solutions from 5 different commercial screens, including Crystal Screen HT, Salt Rx HT, PEG Rx HT, PEG/Ion HT (HR2-139) and Index HT (HR2-134) from Hampton Research. After a month, well-diffracting lamellar crystals appeared in one drop containing 0.1 M bicine, pH = 8.5, 20% (w/v) PEG10000 from the commercial screen PEG Rx HT. Crystals were cryoprotected by the mother liquor containing 15% (v/v) glycerol before being flash-frozen in liquid nitrogen. Apo Bt4394_{WT} and Bt4394_{D335N} variant crystals were grown at 9 °C using the hanging drop vapor diffusion method by mixing 1 μ L of the protein stock (10 mg/mL) and an equal volume of precipitant containing 23 – 25% (w/v) PEG10000, 0.1 M MES-NaOH, pH 5.5 – 6.5 or 0.1 M HEPES-NaOH, pH 7.0 – 8.5. Some needle crystals appeared in two weeks, and after optimizing conditions, well-diffracting crystals were obtained by mixing 1 μ L precipitant (23% (w/v) PEG10000, 0.1 M MES-NaOH, pH 5.5) and 1 μ L protein stock (10 mg/mL), 25% (v/v) glycerol was used as

cryoprotectant for crystals before flash freezing in liquid nitrogen. To trap the Bt4394_{D335N}-6S-GlcNAc-oxazoline intermediate complex, well-diffracting crystals were doped directly with 4MU-6S-GlcNAc powder for 11 min at 20 °C. To avoid glycerol occupying the active site, a solution of 34% (v/v) PEG10000, 0.1 M MES, pH 5.5 was used for cryoprotection before the crystals were flash-frozen in liquid nitrogen. Bt4394_{WT}-NAG-thiazoline-SO₂ complex was formed by premixing 10 mg/mL Bt4394_{WT} stock, 500 mM Na₂SO₃ and 1.5 M NAG (dissolved in deionized water) at 1:1:48 prior to crystallization. The complex was initially screened by sitting drop vapor diffusion method at 9 °C, by mixing 0.2 µL premixed protein solution and 0.2 µL precipitant against 50 µL reservoir solutions containing 23 – 25% (w/v) PEG10000, 0.1 M MES-NaOH, pH 5.5 – 6.5 or 0.1 M HEPES-NaOH, pH 7.0 – 8.5. After two weeks, only one well-diffracting crystal appeared in one drop containing 0.1 M MES, pH = 5.5, 23% (w/v) PEG10000. Crystal was cryoprotected by the mother liquor containing 15% (v/v) glycerol before being flash-frozen in liquid nitrogen.

BbhlI: Initial crystallization screening of native BbhlI_{WT} was conducted by sitting drop vapor diffusion method at 20 °C, by mixing 0.2 µL of the protein solutions (23 mg mL⁻¹ and 46 mg mL⁻¹ in buffer Tris-HCl 25 mM, pH 8.0, NaCl 200 mM) and an equal volume of precipitant against 54 µL of reservoir solution from 4 different commercial screens, including PACT and JCSG from Molecular Dimensions, and Salt Rx and Index HT from Hampton Research. Some preliminary hits were obtained in several conditions and after refinement of the crystallization conditions, well-diffracting crystals for both BbhlI_{WT} and BbhlI_{E553Q} were obtained by mixing 0.5 µL 21.6 mg/mL protein stock with 0.5 µL precipitant composed of PEG3350 21% (w/v), 0.1 M Bis-tris propane, pH 6.5, and 0.2 M NaNO₃ at 20 °C, using the sitting drop vapor diffusion method. To trap the BbhlI_{E553Q}-6S-GlcNAc-oxazoline intermediate complex, apo BbhlI_{E553Q} crystals in the drop were doped directly with 4MU-6S-GlcNAc powder for 30 min at 20 °C, before cryoprotection by the mother liquor containing 20% (v/v) PEG400 and frozen in liquid nitrogen.

SGL: The crystallization was initially screened by sitting drop vapor diffusion method at 20 °C, by mixing 0.2 µL protein solutions (16.5 mg/mL and 31.9 mg/mL) and 0.2 µL precipitant against 50 µL of reservoir solutions from 4 different commercial screens, including PACT and JCSG from Molecular Dimensions, and Salt Rx and Index HT from Hampton Research. After two months, protein crystals appeared in the drops D10 from Index HT and A3 from Salt RX. The condition from Index D10 (0.1 M Bis-tris, pH 6.5, 18 – 21.5% (w/v) PEG5000 MME) was successfully reproduced by optimization with 23.7 mg/mL SGL protein in the buffer containing HCl 25 mM, pH 8.0, and NaCl 200mM. Crystals were cryoprotected by the mother liquor containing 20% (v/v) PEG400 and frozen in liquid nitrogen.

Niako3494: The crystallization was initially screened by sitting drop vapor diffusion method at 4 °C, by mixing 0.2 µL protein solutions (16.5 mg/mL and 31.9 mg/mL) and 0.2 µL precipitant against 50 µL of reservoir solutions from 4 different commercial screens. The optimization of the hit condition was conducted by mixing 0.5 uL 18mg/mL protein stock with 0.5 uL precipitant: 0.1 M MMT buffer (DL-malic acid, MES monohydrate, tris base at 1:2:2 ratio), pH 6.0, 25% w/v PEG1500 with seeding from the initial hit drop. Crystals were cryoprotected by the mother liquor containing 20% (v/v) PEG400 and frozen in liquid nitrogen

Structure solution and refinement

Bt4394: Diffraction data from Bt4394 crystals were collected at the Shanghai Synchrotron Radiation Facility (SSRF, Shanghai) on beamlines 17U1 and 18U1. Data were indexed and integrated using XDSKabsch¹ and HKL-3000², scaled with CCP4i2 suite³ of programs. The apo structures (PDB: 7DUP) were solved by using an initial model (pdb 3RCN) identified by MrBUMP⁴ with a bioinformatic alignment, followed by molecular replacement with Phaser⁵. Full-length GH20 catalytic domain in Bt4394 has well defined density, with two conserved C284 and C348 at the zinc/copper binding site seen in some other GH20s, including Bt0459

(PDB 6Q63) and AmGH20 (PDB 6JE8). All the complex structures (PDB: 7DVA and 7DVB) were solved by direct molecular replacement with MOLREP⁶, followed by alternating rounds of manual model building using Coot⁷ and REFMAC5⁸, respectively, with NCS parameters at later stages.

BbhII, SGL and Niako3494: Diffraction data were collected at 100 K on beamlines i03 and i04 of the Diamond Light Source UK, respectively. Reflections were automatically processed with the *xia2* pipeline⁹ of the CCP4 software suite. BbhII structures were solved by molecular replacement with the PDB:1HP5 model using MOLREP. SGL and Niako3494 structures were both solved by molecular replacement with the highest-scored model calculated by AlphaFold using the sequences provided¹⁰. All structures were refined by alternating rounds of manual model building by Coot and refinement using REFMAC5, respectively. However, density is missing in BbhII for the predicted C-terminal trans-membrane anchoring region and the N-terminal coagulation factor 5/8 type C domain (predicted to be residues 52-190 as carbohydrate-binding motif 32 CBM32).

Growth of *Bifidobacterium bifidum* and *Bacteroides thetaiotaomicron*

Bifidobacterium bifidum JCM1254 and *Bacteroides thetaiotaomicron* JCM5827 were incubated on Gifu Anaerobic Medium (GAM, Haibo, Qingdao) agar plates under an anaerobic condition at 37 °C for 48 h. Bacteria colonies were scraped off from the GAM plates, resuspended with 500 µL basic anaerobic medium (0.5% peptone, 1% yeast extract, 0.008% NaCl, 0.3% L-cysteine, 0.0008% MgSO₄, 0.5% polypeptone, 0.004% KH₂PO₄ and 0.004% K₂HPO₄, pH 7.2), centrifuged at 6000 rpm, and subsequently washed three times in basic anaerobic medium. To compare the effect of different carbon sources on gene induction, bacterial suspensions of *Bifidobacterium bifidum* and *Bacteroides thetaiotaomicron* were plated on the basic anaerobic agar supplemented with 1% (w/v) glucose (Damao, Tianjin) or 1% (w/v) PGM (Type II, Sigma-Aldrich), and cultivated anaerobically at 37 °C for 48 h.

Identification of cellular localization for BbhII

Colonies of *Bifidobacterium bifidum* grown on 1% glucose or 1% PGM were scraped off from the above-mentioned basic anaerobic medium plates, carefully washed three times in 1 mL of assay buffer (25 mM tris-HCl, pH 7.5), resuspended and normalized to the same OD before being split into the following four assays using 125 µM 4MU-6S-GlcNAc as the substrate in 25 mM tris-HCl, pH 7.5 reaction buffer. The glucose-grown and PGM-grown cells were treated in exactly the same way. For assay with whole cells, 100 µL bacteria suspension and 100 µL 4MU-6S-GlcNAc were added to 800 µL buffer. For assay with lysed cells, 100 µL bacterial suspension was centrifuged at 10000 rpm for 10 min, and the cell pellet was subsequently suspended and incubated in 100 µL lysis buffer (B-PER[®] II Bacterial Protein Extraction, Thermo Fisher) for 30 min, before 100 µL 4MU-6S-GlcNAc and 800 µL of buffer were added. For assay using the soluble fraction of lysed cells, the supernatant of 100 µL cells lysate was clarified by centrifugation and was subsequently added to the reaction solution containing 100 µL 4MU-6S-GlcNAc and 800 µL buffer. For assay using the insoluble fraction of the lysed cells, the pellet of 100 µL lysate was resuspended in 100 µL reaction buffer, before 100 µL 4MU-6S-GlcNAc and 800 µL reaction buffer were added. All above reactions were incubated at 37 °C for 2 h, and the release of the 4MU group was monitored as above. EDTA inhibition was not carried out because BbhII is predicted to be extracellular and no sulfatases in *Bifidobacterium bifidum* have been detected¹¹.

Identification of cellular localization for Bt4394

Colonies of *Bacteroides thetaiotaomicron* cells grown on 1% glucose or 1% PGM were treated in a similar way as described above for the *Bifidobacterium bifidum* cells. But these assays do not distinguish the localization being periplasmic or cytoplasmic. To further pinpoint the localization of Bt4394, an established procedure was used¹². The colonies grown on the

above-mentioned basic anaerobic agar supplemented with 1% (w/v) PGM were scraped off, washed, and gently suspended in 0.5 mL TSE buffer (200 mM Tris-HCl, pH 8.0, 500 mM sucrose, 1 mM EDTA) and incubated on ice for 15 min. 0.5 mL of ice-cold deionized water was added subsequently, and the suspension was further incubated on ice for an additional 15 min. The supernatant which contains the soluble periplasmic proteins was clarified at 10000 rpm, 4 °C for 1 h. The pellet was suspended in 1 mL lysis buffer (B-PER® II Bacterial Protein Extraction, Thermo Fisher) and incubated for 30 min, before the supernatant containing the cytoplasmic protein was collected by centrifugation.

Following the extraction of periplasmic and cytoplasmic proteins separately, the following assays were carried out using 125 µM 4MU-6S-GlcNAc as the substrate in the reaction buffer of 200 mM Tris-HCl, pH 8.0, 500 mM sucrose, and 20 mM EDTA. For assay A, 100 µL periplasmic protein fraction, 100 µL 4MU-6S-GlcNAc and 450 µL two-fold assay buffer stock (400 mM Tris-HCl, pH 8.0, 1 mM sucrose, and 40 mM EDTA) were added to 350 µL deionized water. For assay B, 100 µL cytoplasmic protein fraction was added to the reaction solution containing 100 µL 4MU-6S-GlcNAc, 500 µL two-fold assay buffer stock and 300 µL deionized water. The above reactions were incubated at 37 °C for 2.5 h, and the release of 4MU group was monitored as above.

LC-MS analysis of released 6S-GlcNAc from PGM

To remove any trace of free 6S-GlcNAc in the commercial Type II PGM (Sigma), the PGM was dissolved in deionized water, and extensively dialyzed 5 times against deionized water at 4 °C within 24 h before being lyophilized by freeze-drying. Subsequently, 0.3 g of this PGM powder was dissolved in 5 mL PBS buffer (Sigma-Aldrich) and reacted with 1 mg/mL recombinantly purified wild-type Bt4394 or BbhII at 37 °C for 48 h.

0.6 g of the above freeze-dried PGM powder was dissolved in 5 mL PBS buffer and was incubated with *Bifidobacterium bifidum* cells scraped from the basic anaerobic agar supplemented with 1% (w/v) PGM for 48 h at 37 °C. This reaction was first incubated and centrifuged at 15000 rpm, 4 °C for 45 min to remove the cells and insoluble substrates. The supernatant was loaded onto 10 kDa centrifugal filters and the flow-through was collected and lyophilized. To remove proteins in the PGM samples, the lyophilized powder was subsequently dissolved in deionized water and loaded onto a Sep-Pak C₁₈ cartridge column (Waters, USA) preequilibrated with 20% – 100% acetonitrile and deionized water. The flow-through fractions were collected, lyophilized, and redissolved in deionized water.

Online LC-MS analysis was performed using a TSK-GEL AMIDE-80 (2.0 mm 150 mm, 3 µm) (TOSOH, Japan) on an HPLC system (Thermo Fisher) equipped with a PDA detector and an LTQ XL linear ion trap ESI-MS system. 10 µL of the sample was injected, and the mobile phase flow rate was 200 µL/min. With solvent A (acetonitrile) and solvent B (10 mM ammonium acetate pH 4.5), the elution gradient was set as follows: from 98% C, 2% D to 70% C, 30% D in 30 min, then from 70% C, 30% D to 50% C, 50% D in 15 min.

SSN and EFI-CGFP analysis

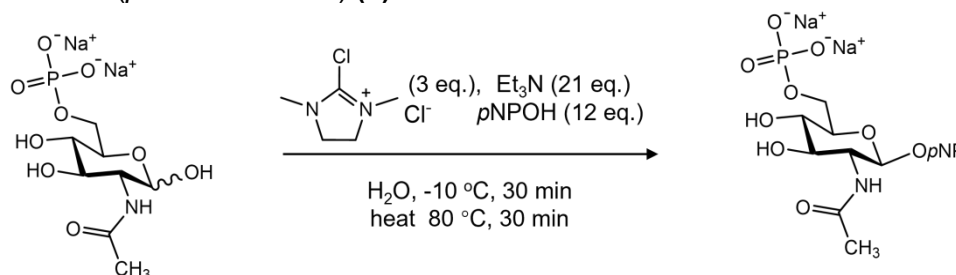
The generation the SSN and final list of GH20 proteins in the two clusters of BbhII and Bt4394. We excluded the sequences that do not have a Q at the equivalent position of 431 in Bt4394 which has been shown to provide strong binding for 6S-GlcNAc. Thus, further refinement finally identified 46 sequences, so that only the sequences possess N at the equivalent position of 437 before R, and Q at the equivalent position of 437 in Bt4394 are included (Figure S19). Because of this, F3-ORF26 will not be selected by this criterion. The orthologous group of Bt4394 is composed of species in the mammalian gut including *Prevotella*, *bacteroid faecis*, *bacteroid caccae*, *bacteroid acidifaciens*, *muribaculaceae bacterium*, *Paludibacter*, *Paramuribaculum intestinale*. The similar approach was used for identifying 40 sequences that have a Q and a W at the equivalent positions of Q640 and W651 in BbhII (Figure S20).

Among those includes the previously unidentified BbHII orthologues, *Bifidobacterium samirii* and *Bifidobacterium jacchi*, as well as *Arthrobacter*, *Trueperella*, *Actinomyces*, *Agromyces*, *Nonomuraea*, *Tessaracoccus*, *Cellulosimicrobium*, *Actinomadura*, *Streptomyces*, *Georgenia*, *Micromonospora*, *Buchananella*, and *Pseudactinotalea* etc.

The two refined SSN clusters were then used for Chemically Guided Functional Profiling CGFP¹³. CGFP identifies SSN clusters that are abundant in metagenome datasets to prioritize targets for functional characterization. After the "Identify" step and "quantify" steps, the organisms and their identified marker locations were analyzed. Heatmaps representing the quantification of sequences from SSN clusters per metagenome are available. The color scale displays the Average genome size (AGS) normalized abundance of the number of gene copies for the "hit" per microbial genome in the metagenome sample. The metagenomes are grouped according to body site so that trends/consensus across the six body sites can be easily discerned. The default heat map is calculated using the median method to report abundances.

Chemical synthesis

Synthesis of *p*-Nitrophenyl-6-*O*-phospho-2-acetamido-2-deoxy- β -D-glucopyranoside diammonium salt (*p*NP-6P-GlcNAc) (**1**).



*p*NP-GlcNAc-6P synthesis was followed the method provided by Qiu. *et al*^{14, 15}. *p*-Nitrophenol (*p*NP, 12 equiv, 667.7 mg) was added to a stirred solution of 6-*O*-phospho-*N*-acetyl-D-glucosamine (GlcNAc-6P, 350 mg, 1 equiv.), and triethylamine (1115 μ L, 20 equiv.) in water (2.2 mL) in a 50 mL sealed round bottom flask at room temperature. The mixture was stirred for 20 min until GlcNAc-6P dissolved and the reaction became less turbid. The reaction was then cooled to -10 °C and stirred for 10 min. 2-chloro-1,2-dimethylimidazolium chloride (DMC, 202.86 mg, 3 equiv.) was then added, and the mixture was stirred for a further 30 min. Consequently, the reaction was refluxed at 80 °C for 50 min. After the reaction was cooled to room temperature, triethylamine (223 μ L, 4 equiv.) was added, and the reaction was cooled down to -10 °C and stirred for 10 min. DMC (135.24 μ L, 2 equiv.) was then added, and the mixture was stirred for a further 30 min, before being heated and refluxed at 80 °C for 50 min. After the reaction was cooled to room temperature, triethylamine (223 μ L, 4 equiv.) was added, and the reaction was cooled down to -10 °C and stirred for 10 min. DMC (135.24 μ L, 2 equiv.) was then added, and the mixture was stirred for a further 30 min, before being heated and refluxed at 80 °C for 50 min. After the third iteration, TLC plates (solvent system CHCl_3 : MeOH, 2:1) were used to indicate the consumption of the starting material and the formation of a less polar product. The reaction was concentrated and diluted by the addition of 35% w/w aqueous ammonium hydroxide (~20 mL). The mixture was then evaporated under a vacuum to remove triethylamine. The residue was re-dissolved in 35% w/w aqueous ammonium hydroxide (~20 mL), and again concentrated under a vacuum to further remove any remaining triethylamine and ammonium hydroxide. All *p*NP-6P-GlcNAc diammonium salt was dissolved in deionized water and diethyl ether was used to remove the excessive *p*-nitrophenyl in the reaction three time. The water phase was loaded onto a Shimadzu reverse phase prep HPLC (column: Shimadzu column 5 μ M C18, 20 x 150 mm) for final purification. Eluent A (H_2O + 0.1% TFA) and B

(MeCN + 0.1% TFA), sample was run at 18 mL min⁻¹, with 100% A for 10 min, followed by gradient 0-30% B over 30 min; UV detection: 280 nm. HPLC: *t*_R = 24.21 min, product yield: 10 mg as a white powder. δ_{H} (500 MHz, D₂O), ¹H NMR (500 MHz, D₂O) δ 8.20 (d, *J* = 9.2 Hz, 2H), 7.15 (d, *J* = 9.0 Hz, 2H), 5.27 (d, *J* = 8.1 Hz, 1H), 4.20 – 3.94 (m, 3H), 3.81 – 3.57 (m, 3H), 1.95 (s, 3H). ¹³C NMR (125 MHz, D₂O) δ 174.98, 161.66, 142.67, 126.14, 116.54, 98.67, 75.19, 75.12, 73.28, 69.25, 66.16, 58.01, 55.39, 22.16. ³¹P NMR (202 MHz, D₂O) δ 4.40 (t, *J* = 6.4 Hz). MS⁻ (ESI-MS) calcd. *m/z* for C₁₄H₁₉N₂O₁₁P⁻ [M]⁻ 422.07, found 422.07.

Supplementary Table 1 Oligonucleotides used for sequencing and mutagenesis.

Primer	Sequence
T7	TAATACGACTCACTATAGGG
T7-term	GCTAGTTATTGCTCAGCGG
Bt4394-seq-f	GCAGCACATATGCAGGAAATCGCGCTGAC
Bt4394-seq-r	GCAGCACTCGAGTTTCATGATGTGTTTACCGTAGTTCC
Bt4394-D335N-f	CATTCACCTGGGCGGCA AAT GAATGTCCGACC
Bt4394-D335N-r	GGTCGGACATTC ATT GCCGCCAGGTGAATG
Bt4394-E336Q-f	CTGGGCGGCGAT CAAT GTCCGACCC
Bt4394-E336Q-r	GGGTCGGAC ATT GATCGCCGCCAG
Bt4394-Q431E-f	CACCATCCTGTCTCCG GAGAT CCCGTACTACATC
Bt4394-Q431E-r	GATGTAGTACGGGAT CTCC GGAGACAGGATGGTG
Bt4394-N437D-f	CGCAGATCCCGTACTACATC GACC GTAACAGTCTAAACTG
Bt4394-N437D-r	CAGTTTAGACTGTTTAC GGT CGATGTAGTACGGGATCTGCG
Bt4394-R438A-f	GATCCCGTACTACATCAAC GCA AAACAGTCTAAACTGCCGAC
Bt4394-R438A-r	GTCGGCAGTTTAGACTGTTTT TGC GTTGATGTAGTACGGGATC
Bt4394-R438K-f	GATCCCGTACTACATCAAC AAAAA ACAGTCTAAACTGCCGAC
Bt4394-R438K-r	GTCGGCAGTTTAGACTGTTTT TTT GTTGATGTAGTACGGGATC
Bt4394-D266N-f	CCGGAAAT TAA CATCCCAGGTCAC
Bt4394-D266N-r	CCTGGGAT GTTA ATTTCCGGGATG
Bt4394-Y435F-f	GATCCCGTACT TTT ATCAACCGTAAAC
Bt4394-Y435F-r	GGTTGAT AAAG TACGGGATCTGC
Bt4394-H270F-f	GATATCCCAGG TTTCT CTCAGGC
Bt4394-H270F-r	CCTGAGAG GAA ACCTGGGATATC
BbhII-seq-f	GGATGAAGGCTTCCGTGTTGAG

BbhII-seq-r	TGTACATCAGGAACTCCGCGTC
BbhII-D552N-f	CCACCTGGGTGGCA AAT GAACTGTTTCTGAGC
BbhII-D552N-r	GGTGGACCCACCG T TACTTGACAAAGACTCG
BbhII-E553Q-f	CCACCTGGGTGGCGAT CAG CTGTTTCTGAGC
BbhII-E553Q-r	GGTGGACCCACCGCTA G T C GACAAAGACTCG
BbhII-Q610D-f	GTAACACCTACTTTCCG GAT CGTCCGGGTACCGAAAC
BbhII-Q610D-r	GTTTCGGTACCCGGACG A T C CGGAAAGTAGGTGTTAC
BbhII-Q610K-f	CACCTACTTTCCG AAAC GTCCGGGTACC
BbhII-Q610K-r	GGTACCCGGACG T TTCCGAAAGTAGGT
BbhII-W621F-f	CACCGGTGTGACCT T TGCGTGCGGTGCGTG
BbhII-W621F-r	CACGCACCGCACGC AAAG GTACACCGGTG
BbhII-C626A-f	CGTGCGGTGCG GCC ACCACCAGCAACTTTTATC
BbhII-C626A-r	GATAAAAGTTGCTGGTGGT GGCC GCACCGCACG
BbhII-C626S-f	CGTGCGGTGCG TCC ACCACCAGCAACTTTTATC
BbhII-C626S-r	GATAAAAGTTGCTGGTGGT GGAC GCACCGCACG
BbhII-D552N-f	ACCTGGGTGGCA AAT GAACTGTTTCTG
BbhII-D552N-r	CAGAAACAGTTC A TTGCCACCCAGGTG
BbhII-Y637F-f	TAACAC T TTCCGCAACGTC
BbhII-Y637F-r	GTTGCG GAA AGAAGGTGTTACC
BbhII-H467F-f	ATCGCGATTGTTCCGGAGATTGACGGTCCGGG C TTTAGCTTTAGCC TGCTGCATGG
BbhII-H467F-r	ATCTCCGGAACAATCGCGATG
Niako3494-D196N-f	GTTGGTCTG AAC GAAGTTTGGATTC
Niako3494-D196N-r	CCAAACTTC G TT C AGACCAACATG
Niako3494-E197Q-f	GGTCTGGAT CAG GTTTGGATTCTGG
Niako3494-E197Q-r	CCAAAC C TGATCCAGACCAACATG

Supplementary Table 2 Data collection and refinement statistics.

	Bt4394 _{WT} apo	Bt4394 _{D335N} -6S-GlcNAc-oxazoline	Bt4394 _{WT} -6S-GlcNAc	BbhII apo	BbhII _{E553Q} -6S-GlcNAc-oxazoline
PDB	7DUP	7DVB	7DVA	6YXZ	6Z14
Data collection					
Space group	P 1 2 1 1	P 1 2 1 1	P 1 2 1 1	P 2 1 2 1 2 1	P 2 1 2 1 2 1
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	66.58, 63.00, 78.69	82.03, 67.00, 192.80	49.50, 125.60, 84.64	56.57, 126.75, 152.71	54.42, 124.88, 151.68
α , β , γ (°)	90.00, 111.34, 90.00	90.00, 94.67, 90.00	90.00, 90.63, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	73.29 (1.62)	19.71 (2.05)	70.19 (1.55)	63.38 (1.75)	64.82 (1.67)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.091(0.594)	0.107 (1.024)	0.097 (1.005)	0.123 (1.654)	0.116 (0.732)
<i>I</i> / σ <i>I</i>	12.6 (2.0)	9.1 (1.6)	11.5 (2.0)	9.0 (1.1)	7.4 (1.0)
Completeness (%)	99.0 (93.0)	99.2 (98.2)	97.4 (96.4)	100.0 (100.0)	100.0 (100.0)
Redundancy	5.9 (3.0)	6.6 (6.7)	6.8 (7.1)	7.4 (7.4)	7.0 (6.9)
Refinement					
Resolution (Å)	1.62	2.05	1.55	1.75	1.67
No. reflections	76333	130225	145227	111416	120604
<i>R</i> _{work} / <i>R</i> _{free}	0.145, 0.189	0.228, 0.259	0.142, 0.184	0.160, 0.220	0.160, 0.205
No. atoms					
Protein	4301	16710	8563	5301	5281
Ligand/ion	32	91	68	10	43
Water	488	454	567	644	655
<i>B</i> -factors					
Protein	23.36	37.54	22.29	28.99	28.51
Ligand/ion	29.69	32.98	25.03	48.00	35.01
Water	39.23	33.86	33.17	39.82	38.44
R.m.s. deviations					
Bond lengths (Å)	0.0057	0.0033	0.0045	0.0090	0.0080
Bond angles (°)	1.232	0.944	1.217	1.511	1.440

*Values in parentheses are for highest-resolution shell.

	Bt4394 _{WT} -NAGthiazoline-SO ₂	SGL apo	Niako3494 apo
PDB	8BDP	8BBL	8BAL
Data collection			
Space group	P 1 21 1	P 1	C 2 2 21
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	66.83, 60.78, 74.22	132.95, 133.42, 225.37	146.99, 254.59, 139.58
α , β , γ (°)	90.00, 107.91, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	70.63 (1.47)	94.18 (2.71)	73.49 (2.27)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.033 (0.542)	0.154 (0.516)	0.329 (2.524)
<i>I</i> / σ <i>I</i>	1.80 (1.47)	2.02 (2.73)	1.66 (2.27)
Completeness (%)	95.4 (65.0)	97.6 (97.8)	99.9 (99.7)
Redundancy	1.9 (1.9)	1.9 (1.8)	13.9 (13.6)
Refinement			
Resolution (Å)	1.47	2.71	2.27
No. reflections	91859	406022	120442
<i>R</i> _{work} / <i>R</i> _{free}	0.130, 0.182	0.273, 0.321	0.283, 0.356
No. atoms			
Protein	8518	31754	13716
Ligand/ion	2	0	5
Water	458	18	28
<i>B</i> -factors			
Protein	24.86	24.71	32.47
Ligand/ion	23.14	-	42.18
Water	37.37	16.3	27.59
R.m.s. deviations			
Bond lengths (Å)	0.0094	0.0114	0.0066
Bond angles (°)	1.518	2.310	1.613

*Values in parentheses are for highest-resolution shell.

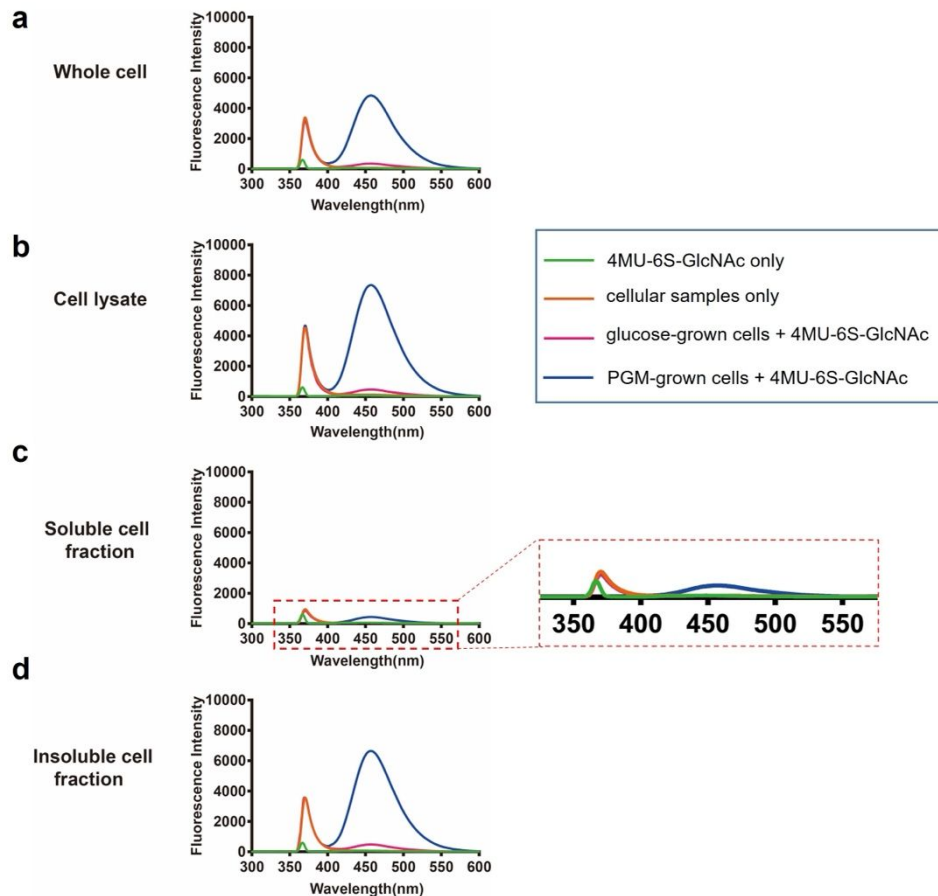
Supplementary Table 3 Complex structures containing 6SO3 (CAzy Carb_ligand: 6SO3).

(From <http://www.cazy.org/search?page=recherche&recherche=6SO3&tag=12>, by 30th Jan 2022)

PDB ID	ligand	Protein name	Resolution (Å)	6SO3 H-bond donor residue (H-bond ≥ 2.5 Å)	Family	EC
6Z14	6SO ₃ -NAG-oxazoline	BbhII	1.67	Q640 _{NH2} W651 _{NH} H ₂ O	GH20	3.2.1.-
7DVB	6SO ₃ -NAG-oxazoline	Bt4394	2.05	Q431 _{NH2} N437 _{NH2} R438 _{NH2} H ₂ O	GH20	3.2.1.-
7DVA	6SO ₃ -GlcNAc	Bt4394	1.55	Q431 _{NH2} N437 _{NH2} R438 _{NH2} H ₂ O	GH20	3.2.1.-
1HMW	β-D-4-deoxy-GlcpA-(1-3)-β-D-GalpNAc4SO ₃ -(1-4)-β-D-GlcpA-(1-3)-β-D-GalpNAc6SO ₃	chondroitin AC lyase	1.91	none	PL8	4.2.2.5
2FUT	α-L-4-deoxy-IdopA2SO ₃ -(1-4)-α-D-GlcpNSO ₃ 6SO ₃ β-D-4-deoxy-GlcpA2SO ₃ -(1-4)-α-D-GlcpNSO ₃ 6SO ₃ (sits on surface)	heparin lyase II / heparinase II	1.85	Y429 _{OH} Y436 _{NH} N437 _{NH2}	PL21	4.2.2.7 4.2.2.8
3ANK	α-L-4-deoxy-IdopA-(1-3)-β-D-Galp6SO ₃	unsaturated glucuronyl hydrolase useless – surface location	1.80	S368 _{OH}	GH88	3.2.1.-
3CU0	β-D-Galp-(1-3)-β-D-Galp6SO ₃	UDP-GlcA: β-1,4-galactosyl- xylosylprotein β-1,3-glucuron- osyl-transferase 3	1.92	Q318 _{NH2}	GT43	2.4.1.135
3ILF	β-L-Galp6SO ₃ -(1-3)-β-D-Galp-(1-4)-α-L-Galp6SO ₃ -(1-3)- α-D-Galp	β-porphyrinase A	2.65	H53 _{imi} V134 _{NH} R133 _{NH2+}	GH16	3.2.1.178

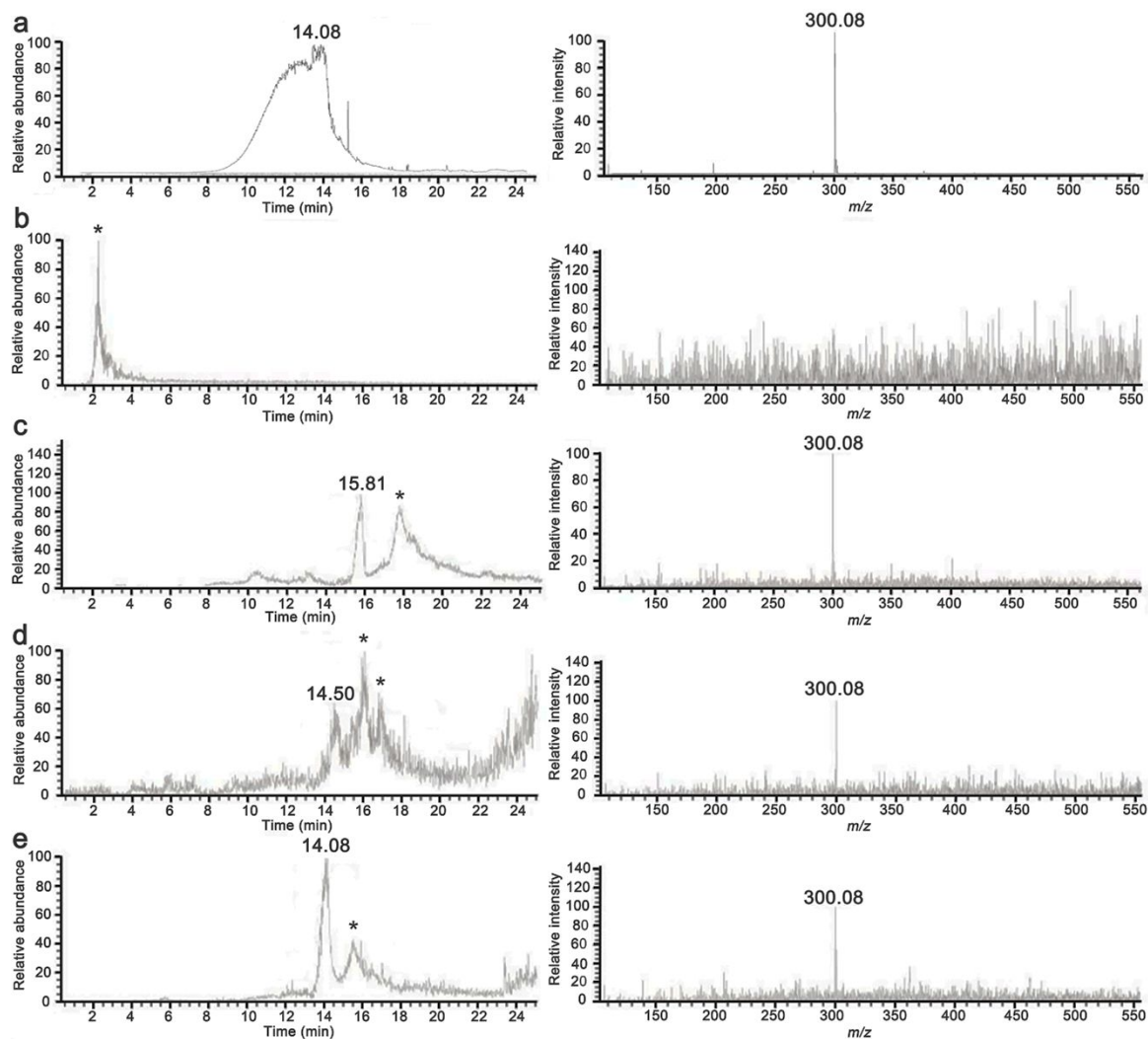
3ILR	β -D-4-deoxy-AllpA2SO ₃ -(1-4)- α -D-GlcpNSO ₃ 6SO ₃ (Poor Density)	lyase cleaving heparin and heparan sulfate	1.45	Y351 _{OH} K353 _{NH3+}	PL13	4.2.2.7
3IN9	α -L-4-deoxy-IdopA2SO ₃ -(1-4)- α -D-GlcpNSO ₃ 6SO ₃	lyase cleaving heparin and heparan sulfate	1.90	K241 _{NH3+} N243 _{NH2} N20 _{NH2}	PL13	4.2.2.7
3OGX	β -D-4-deoxy-GlcpA2SO ₃ -(1-4)- α -D-GlcpNSO ₃ 6SO ₃ (No density)	peptidoglycan recognition protein	1.70	K90 _{NH3} E142 _{COOH}	CBM0	NA
4AW7	α -L-Galp6SO ₃ -(1-3)- α -D-Galp-(1-4)- α -L-3,6-anhydro-Galp-(1-3)- β -D-Galp-(1-4)- α -L-Galp6SO ₃ -(1-3)- α -D-Galp α -L-Galp6SO ₃ -(1-3)- β -D-Galp-(1-1)-<non_carb>	β -porphyranase	1.25	W78 _{indoNH} K87 _{NH3} W331 _{indoNH} His53 _{imi}	GH86	3.2.1.178
4KRV	β -D-GlcpNAc6SO ₃	β -1,4-galactosyltransferase T1	2.20	none	GT7	2.4.1.22 2.4.1.38 2.4.1.90
4X7R	α -D-GlcpNAc-(1-1)-<non_carb> α -D-GlcpNSO ₃ 6SO ₃ -(1-4)- β -D-GlcpA-(1-4)- α -D-GlcpNSO ₃ 3SO ₃ 6SO ₃ -(1-4)- α -L-IdopA2SO ₃ -(1-4)- α -D-GlcpNSO ₃ 6SO ₃ -(1-1)-methyl	UDP-GlcNAc: teichoic acid α -N-acetylglucosaminyl-transferase	2.90	G16 _{NH} T19 _{OH} M18 _{NH}	GT4	2.4.1.-
5BO9	α -D-Neup5Ac-(2-3)- β -D-Galp-(1-4)- β -D-GlcpNAc6SO ₃	α -2,8-sialyltransferase III	1.85	none	GT29	2.4.99.-
5E9C	β -D-4-deoxy-GlcpA2SO ₃ -(1-4)- α -D-GlcpNSO ₃ 6SO ₃ -(1-4)- α -L-IdopA-(1-4)- β -D-GlcpNSO ₃ 6SO ₃	heparanase (endo- β -gluc-uronidase)	1.64	Q270 _{NH2} Q270 _{NH} R303 _{NH2+}	GH79	3.2.1.166
5WD7	α -D-GlcpNSO ₃ 6SO ₃ -(1-4)- β -D-GlcpA-(1-4)- α -D-GlcpNSO ₃ 3SO ₃ 6SO ₃ -(1-4)- α -L-IdopA2SO ₃ -(1-4)- α -D-GlcpNSO ₃ 6SO ₃ -(1-1)-methyl	CMP-Neu5Ac: α -2,8-poly-sialyltransferase	2.10	N76 _{NH2}	GT38	2.4.99.-
6LJA	α -L-4-deoxy-IdopA2SO ₃ -(1-4)- α -D-GlcpNSO ₃ 6SO ₃	heparin lyase I	2.00	G71 _{NH}	PL15	4.2.2.7
6LJL	β -D-4-deoxy-GlcpA-(1-4)- α -D-GlcpNSO ₃ 6SO ₃ -(1-4)- α -L-IdopA2SO ₃ -(1-4)- α -D-GlcpNSO ₃ 6SO ₃	heparin lyase I	1.99	R233 _{NH2+} G71 _{NH}	PL15	4.2.2.7

All 19 entries for the 6SO3 available in the PDB database to January 2022 are listed in the table (resolution better than 3.0 Å).



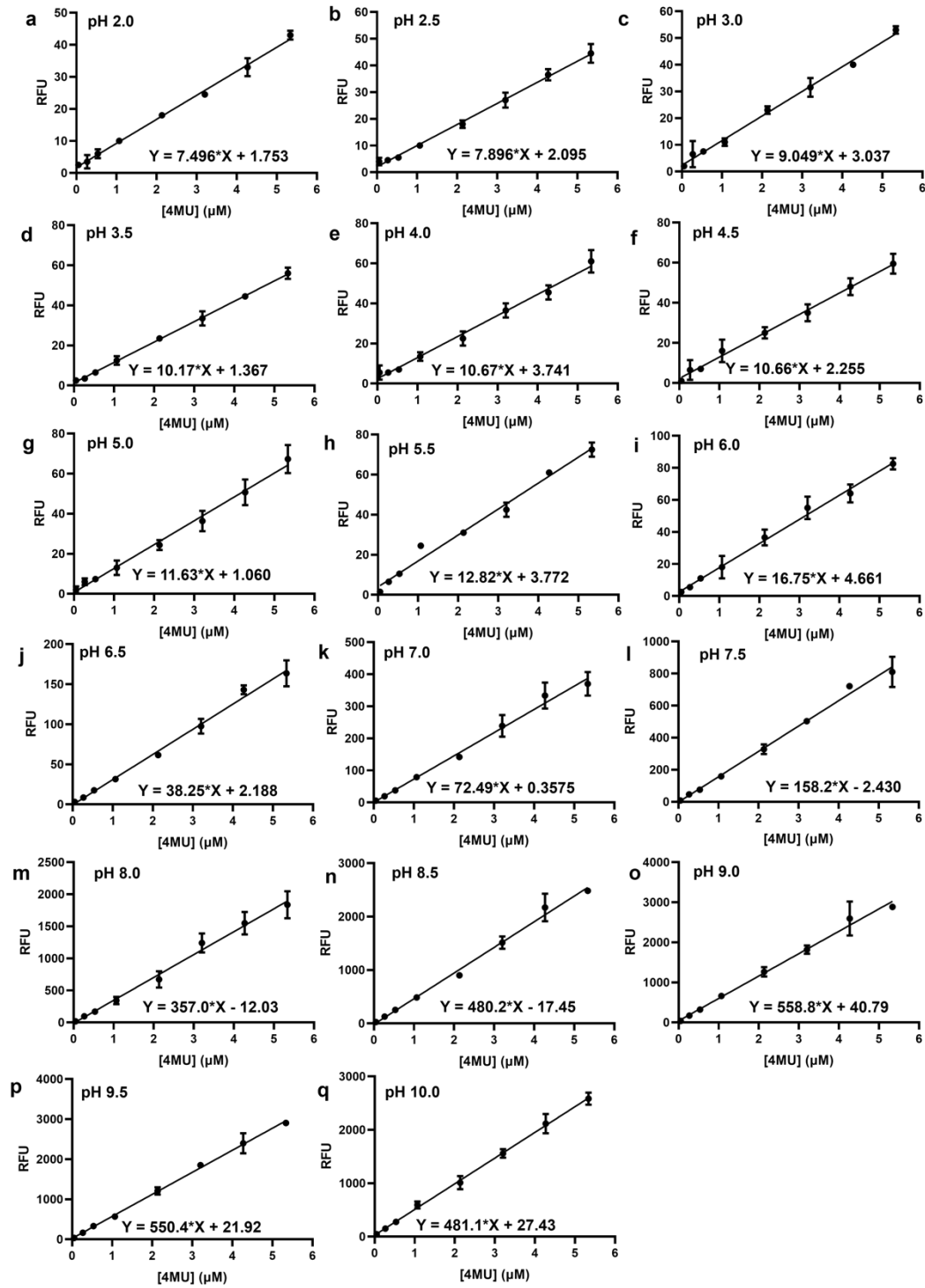
Supplementary Figure 1. Identification of cellular localizations of BbhII in *Bifidobacterium bifidum* JCM1254.

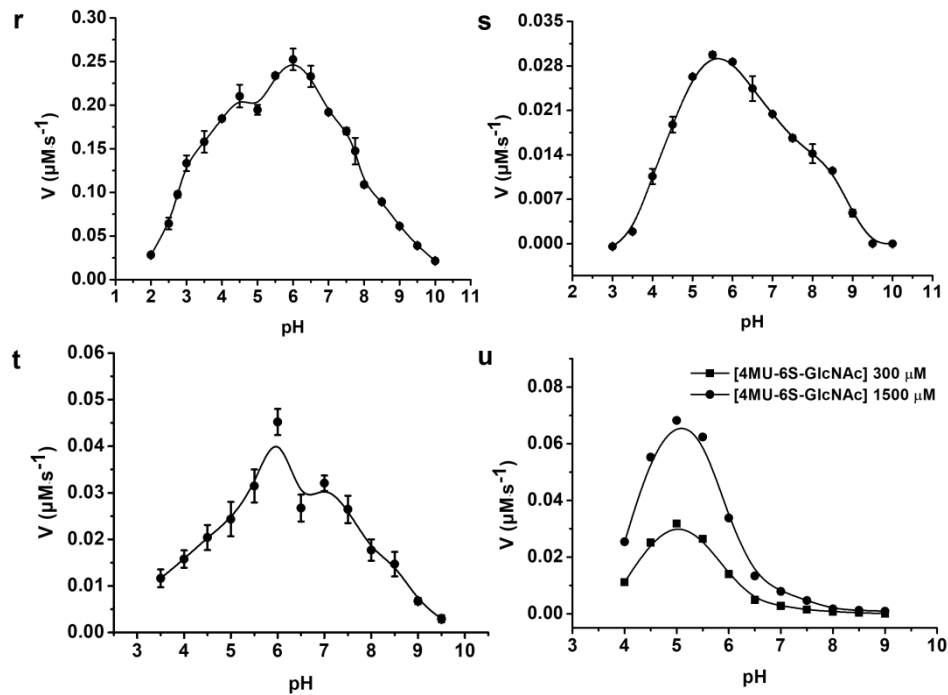
Cells grown on 1% glucose or 1% PGM were scraped off from the agar plate, washed three times and split into the following assays: 125 μ M 4MU-6S-GlcNAc was incubated with whole cells, lysed cells, soluble components in the supernatant fraction from lysed cells, and insoluble fraction in the cell debris pellet from lysed cells. All assays were performed in 200 mM NaCl and 25 mM Tris-HCl, pH 7.5 at 37 $^{\circ}$ C for 2 h. The 4MU group liberated was monitored from 300 - 600 nm using a fluorescence spectrophotometer showing the peak at $\lambda_{em} = 445$ nm: 4MU-6S-GlcNAc alone as control (green curve); 4MU-6S-GlcNAc and growth medium only as control (orange curve); 4MU-6S-GlcNAc with samples from glucose-grown cells (pink curve); 4MU-6S-GlcNAc and samples from PGM-grown cells (blue curve).



Supplementary Figure 2: Detection of released 6S-GlcNAc from natural and synthetic substrates using HPLC-MS.

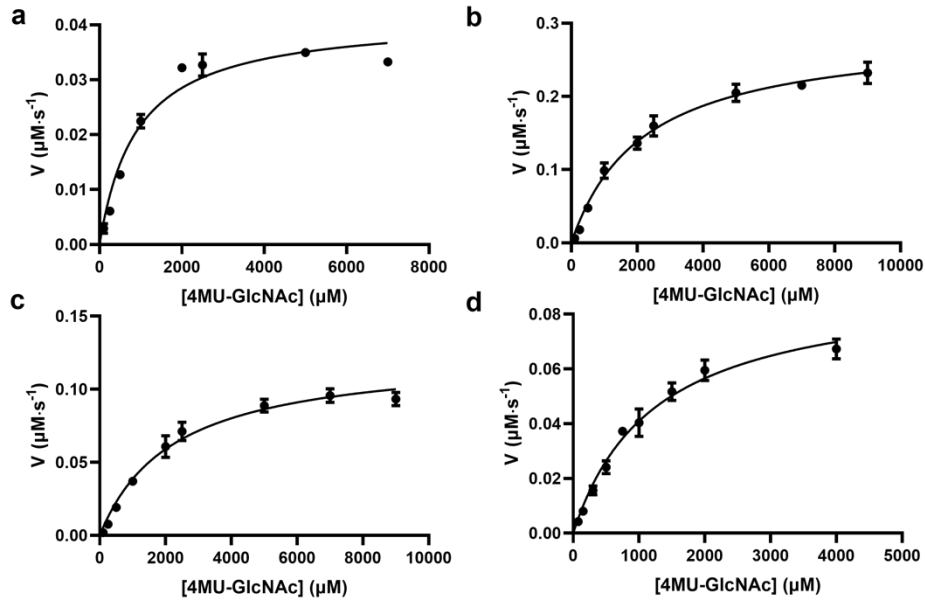
Commercial Type II PGM were digested with the GH20 BbhII and Bt4394 and the liberated 6S-GlcNAc was analyzed by LC-ESI-MS. The insets are the EICs and MS profiles of (a) commercial 6S-GlcNAc as the standard, (b) PGM without being incubated with any enzyme, (c) saccharides released from PGM treated with living *Bifidobacterium bifidum* cells. (d) saccharides released from PGM treated with BbhII, (e) saccharides released from PGM treated with Bt4394, and the asterisk denotes unknown signals.





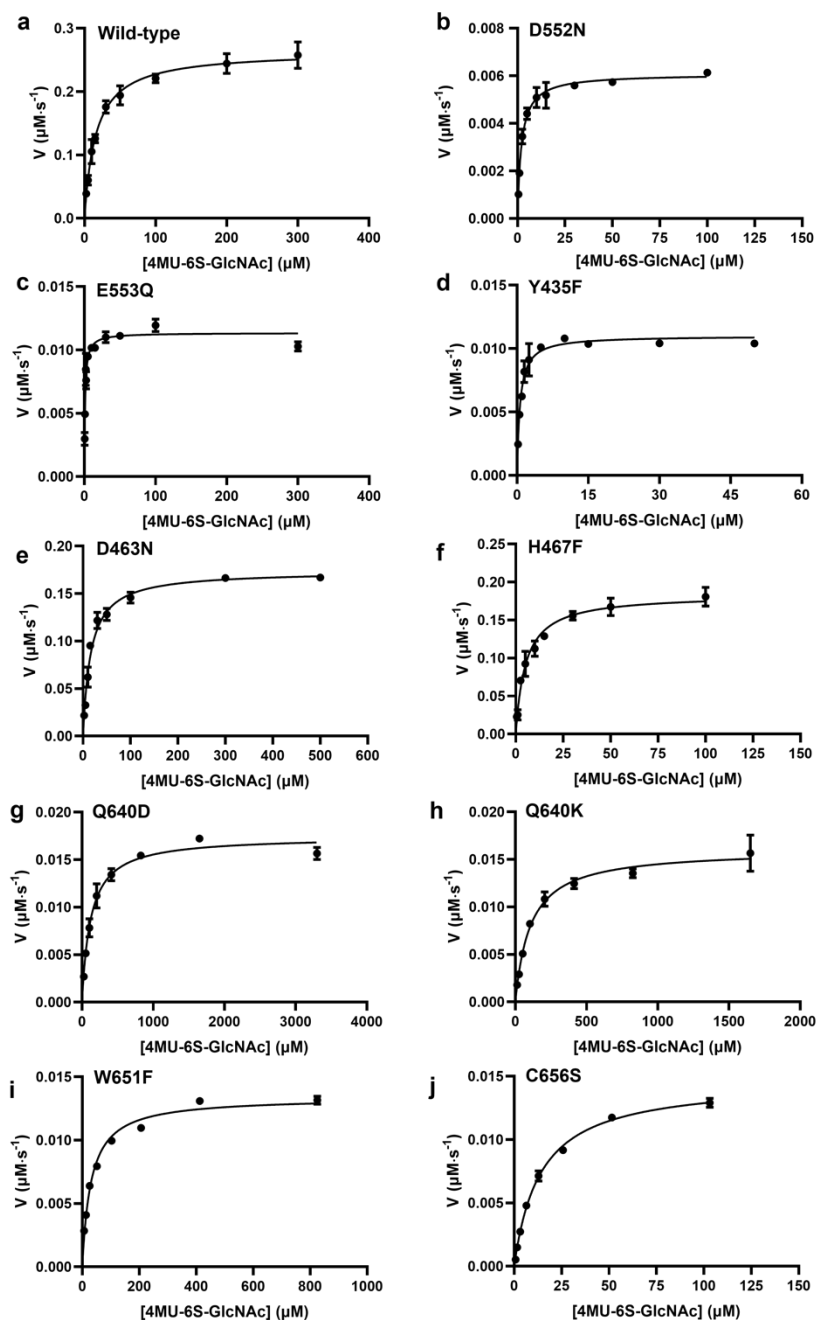
Supplementary Figure 3. The standard curves of 4MU in the universal buffer at various pHs and pH dependence of the hydrolysis of 4MU-6S-GlcNAc by BbhII_{WT}, Bt4394_{WT}, SGL_{WT}, and Niako3494_{WT}. Curves **a** to **q** fitted by linear regression; smooth curves **r** to **u** fitted in Excel

(**a-q**) All the standard curves were plotted by monitoring the fluorescence from compound 4MU at $\lambda_{\text{ex}} = 360 \text{ nm}$ and $\lambda_{\text{em}} = 450 \text{ nm}$ in a BMG Fluostar fluorescence microplate reader with gain 1200, in the buffer of 25 mM Bis-tris propane, 25 mM citrate, and 300 mM NaCl, pH 2.0 - 10.0. The pH profiles of 4MU-6S-GlcNAc hydrolysis by wild-type enzymes were performed in the same buffer as the standard curves. (**r**) pH profile for 6 nM BbhII_{WT} hydrolyzes 300 μM substrate with the pH range of 2.0 - 10.0, with an optimal pH of 6.0. (**s**) pH profile for 2 nM Bt4394_{WT} hydrolyzes 500 μM substrate with the pH range of 3.0 - 10.0, with an optimal pH of 5.5. (**t**) pH profile for 1.2 nM SGL_{WT} hydrolyzes 300 μM substrate with the pH range of 3.5 - 9.5, with an optimal pH of 6.0. (**u**) pH profile for 2 μM Niako3494_{WT} hydrolyzes 300 μM and 1500 μM substrate with the pH range of 4.0 - 9.0, with an optimal pH of 5.0.



Supplementary Figure 4 Michaelis-Menten plots for illustrating substrate specificity of non-sulfated 4MU-GlcNAc with BbhII_{WT}, Bt4394_{WT}, SGL_{WT} and Niako3494_{WT}.

Initial hydrolytic rates of substrate 4MU-GlcNAc catalyzed by four wild-type enzymes were performed in the buffer of 25 mM Bis-tris propane, 25 mM citrate, and 300 mM NaCl at 25°C (pH 5.5 for Bt4394_{WT}, pH 6.0 for BbhII_{WT} and SGL_{WT}, pH 5.0 for Niako3494_{WT}). All reactions were monitored continuously by the release of the fluorescent 4MU ($\lambda_{\text{ex}} = 360$ nm and $\lambda_{\text{em}} = 450$ nm, BMG Fluostar fluorescence microplate reader). (a) 40 nM BbhII_{WT} hydrolyzes 100 μM to 7000 μM substrate. (b) 100 nM Bt4394_{WT} hydrolyzes 100 μM to 9000 μM substrate. (c) 50 nM SGL_{WT} hydrolyzes 100 μM to 9000 μM substrate. (d) 2 μM Niako3494_{WT} hydrolyzes 75 μM to 4000 μM substrate.

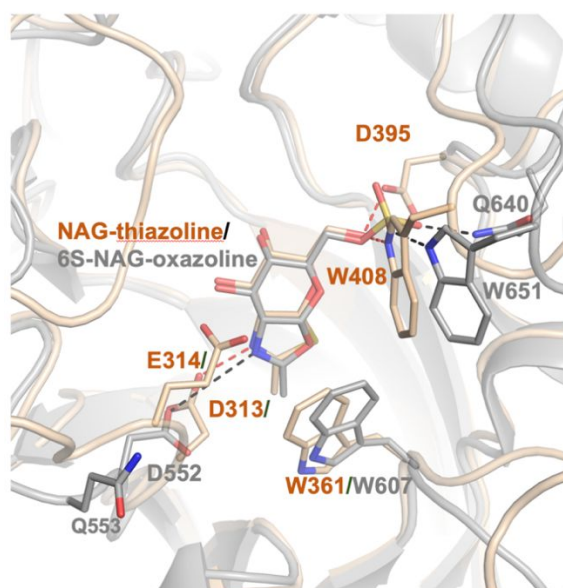


Supplementary Figure 5 Michaelis-Menten plots for hydrolysis of 4MU-6S-GlcNAc by wild-type BbhII and its variants at pH 6.0.

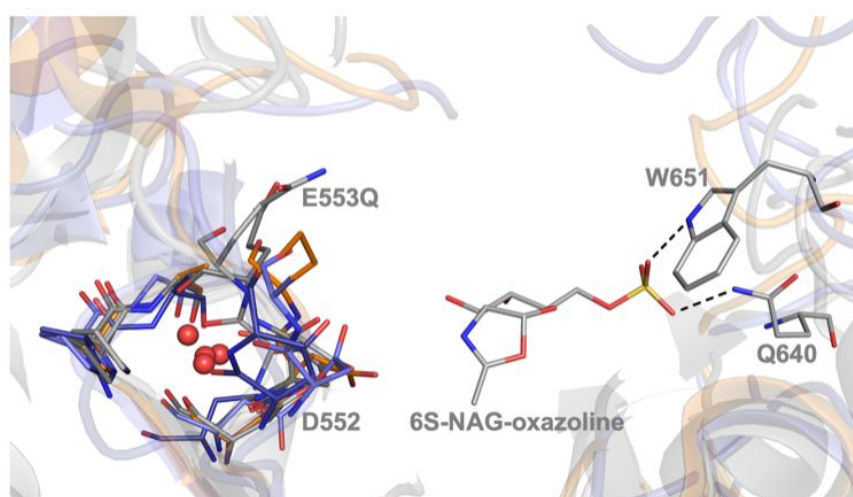
Initial hydrolytic rates of substrate 4MU-6S-GlcNAc catalyzed by BbhII_{WT} and its variants were performed in the buffer of 25 mM Bis-tris propane-HCl, 25 mM citrate, and 300 mM NaCl at pH 6.0, 25°C. All reactions were monitored continuously by the release of the fluorescent 4MU ($\lambda_{\text{ex}} = 360 \text{ nm}$ and $\lambda_{\text{em}} = 450 \text{ nm}$). (a) 6 nM BbhII_{WT} hydrolyzes 2.5 μM to 300 μM substrate.

(b) 100 nM BbhII_{D552N} hydrolyzes 2.5 μ M to 100 μ M substrate. (c) 100 nM BbhII_{E553Q} hydrolyzes 0.05 μ M to 300 μ M substrate. (d) 150 nM BbhII_{Y435F} hydrolyzes 0.5 μ M to 50 μ M substrate. (e) 6 nM BbhII_{D463N} hydrolyzes 2.5 μ M to 500 μ M substrate. (f) 0.5 μ M BbhII_{H467F} hydrolyzes 0.5 μ M to 100 μ M substrate. (g) 2.2 nM BbhII_{Q640D} hydrolyzes 25.8 μ M to 3300 μ M substrate. (h) 1.4 nM BbhII_{Q640K} hydrolyzes 12.9 μ M to 1650 μ M substrate. (i) 0.6 nM BbhII_{W651F} hydrolyzes 6.4 μ M to 825 μ M substrate. (j) 1.6 nM BbhII_{C656S} hydrolyzes 0.8 μ M to 103.1 μ M substrate.

a

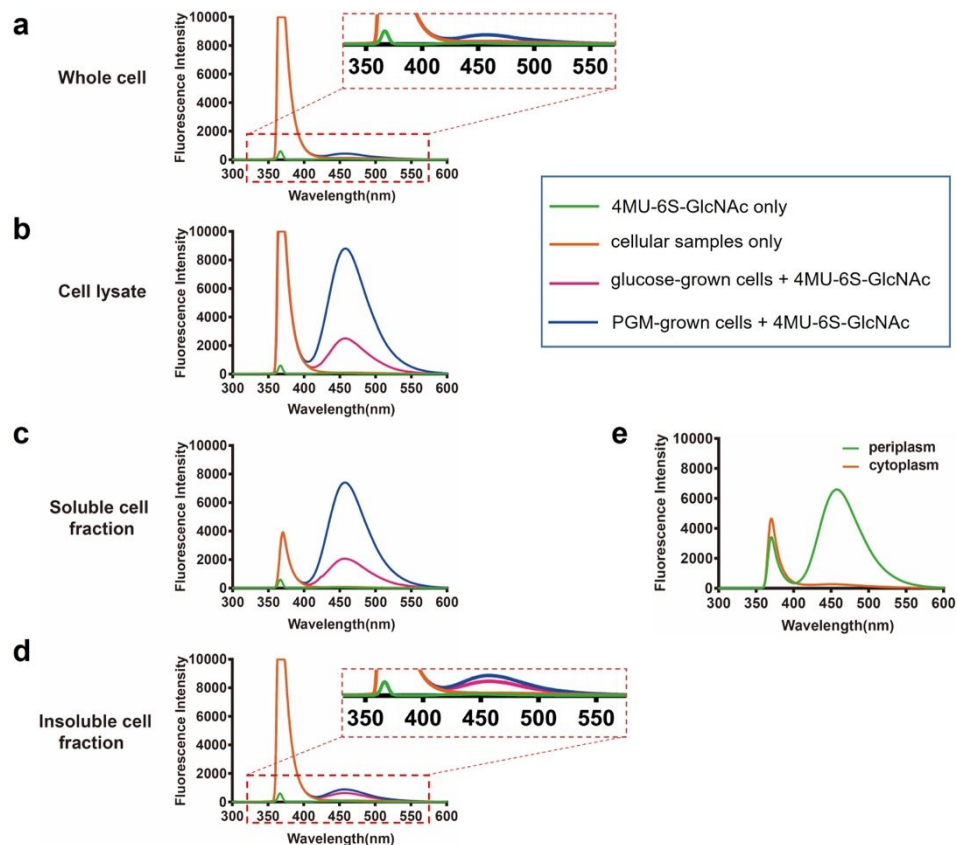


b



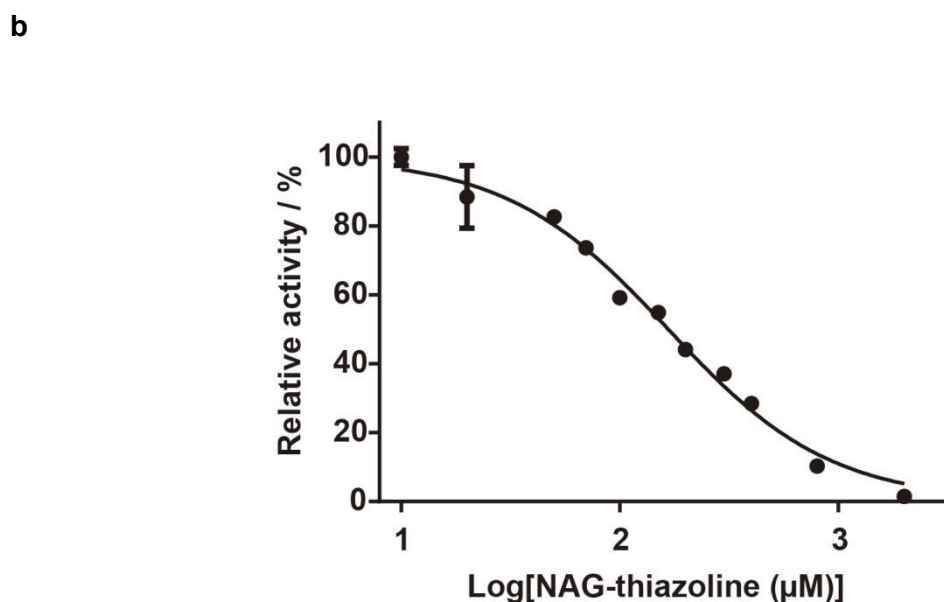
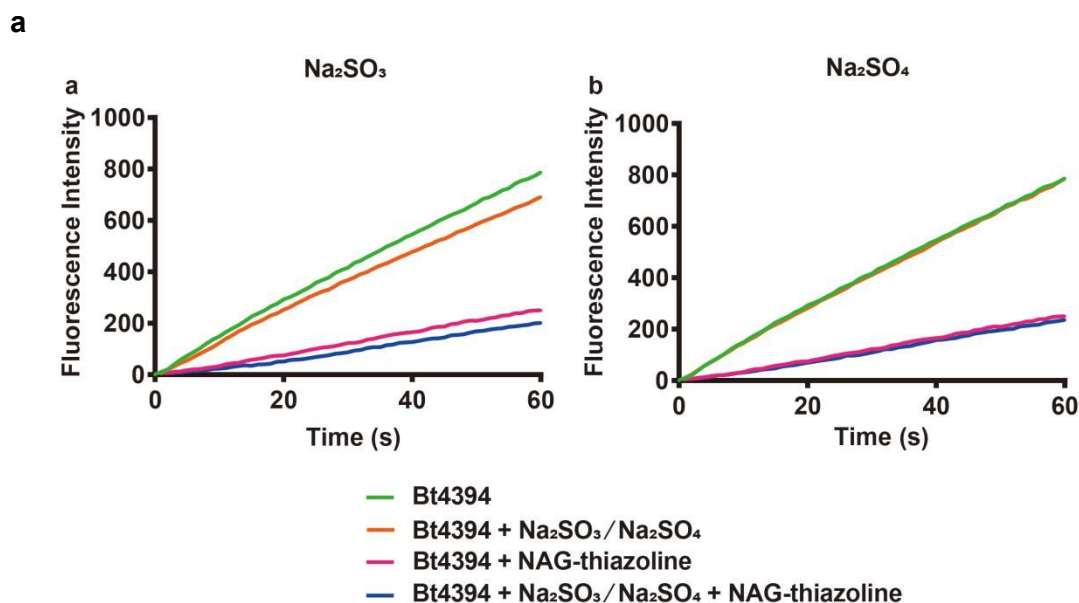
Supplementary Figure 6 Structures of the 6S-NAG-oxazoline intermediate complexes of BbhII in comparison with the SpHex and other GH20 structures.

(a) Overlay of the structures of BbhII (in gray) in its oxazoline intermediate complex with the loop containing D552/E553 in a catalytically incompetent conformation, and SpHex (in salmon, PDB: 1hp5) in its thiazoline intermediate complex with the loop containing D313/E314 in the catalytic conformation. **(b)** Overlay of the structures of BbhII (PDB: 6Z14 in gray) with other GH20 structures that have the loop containing the D/E pair equivalent to D552/E553Q in a catalytically incompetent conformation. The GH20 structures used are (PDB codes): 3RPM, 5A6J, 3NSM, and 6JE8.



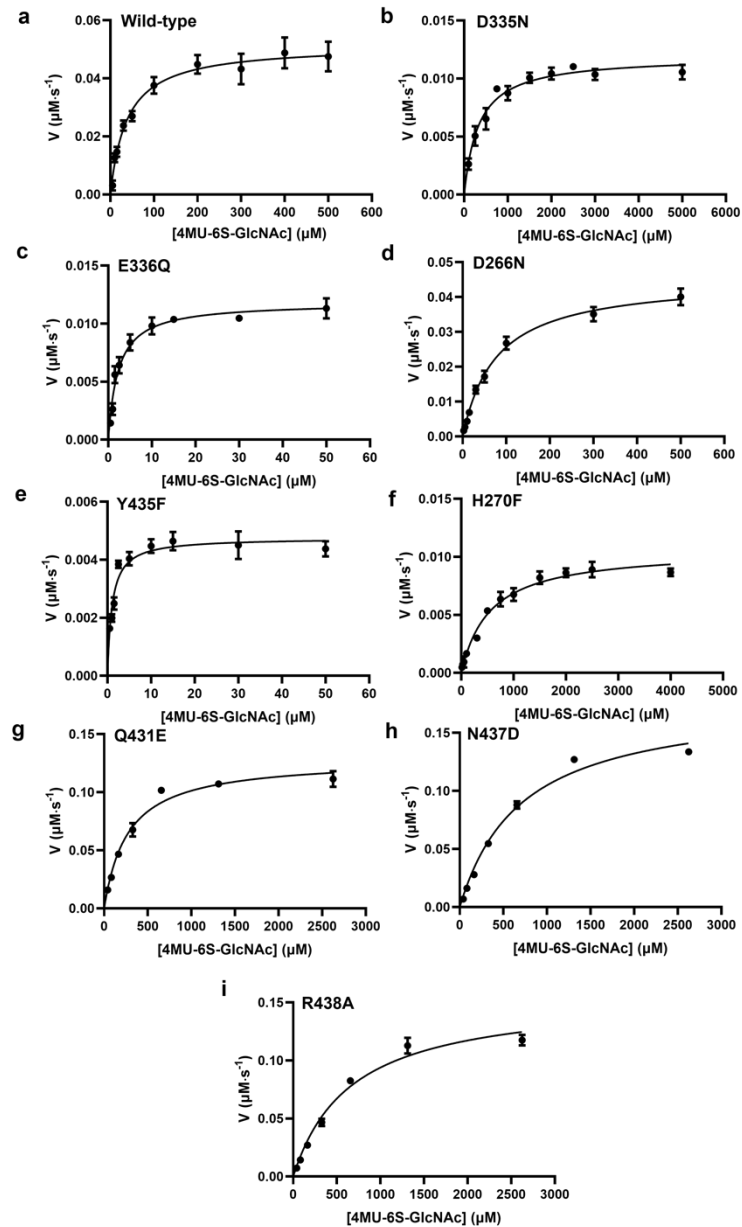
Supplementary Figure 7. Identification of cellular localizations of Bt4394 in *Bacteroides thetaiotaomicron*.

(a-d): 125 μ M 4MU-6S-GlcNAc was treated with whole cells, lysed cells, soluble components in the supernatant fraction from lysed cells, and insoluble fraction in the cell debris pellet from lysed cells. 4MU-6S-GlcNAc only as control (green); corresponding cellular samples only as control (orange); samples from glucose-grown cells (pink); samples from PGM-grown cells (blue). **(e)**: 125 μ M 4MU-6S-GlcNAc treated with periplasmic fraction (green); 125 μ M 4MU-6S-GlcNAc treated with cytoplasmic fraction (orange). Detected activities in the cell lysate, soluble fraction, and periplasmic fraction are attributed primarily due to periplasmic Bt4394 but might possibly result from a collective effect of sulfate removal by potential intracellular 6S-O-sulfatases (BT1628, BT3177, BT4656)¹⁶, followed by glycosidic bond cleavage by the action of other GlcNAcases.



Supplementary Figure 8 Inhibition of recombinant Bt4394 by Na_2SO_3 , Na_2SO_4 , and NAG-thiazoline.

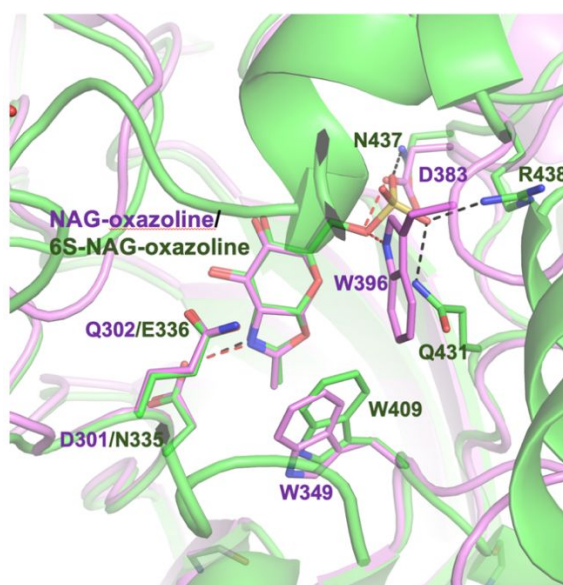
(a) Rates of fluorescence intensity change at 445 nm were measured under steady-state conditions using 100 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 μM 4MU-6S-GlcNAc, and 5 nM Bt4394_{WT}. Final concentration of each component used in the reaction: 200 mM $\text{Na}_2\text{SO}_3\text{-HCl}$ (pH 8.0), 200 mM $\text{Na}_2\text{SO}_4\text{-NaOH}$ (pH 8.0), or 200 μM NAG-thiazoline. (b) IC_{50} plot of NAG-thiazoline inhibition of the hydrolysis of 4MU-6S-GlcNAc by Bt4394_{WT} (curve fitted by nonlinear regression analysis). The determined IC_{50} value for NAG-thiazoline is $166.6 \pm 10 \mu\text{M}$.



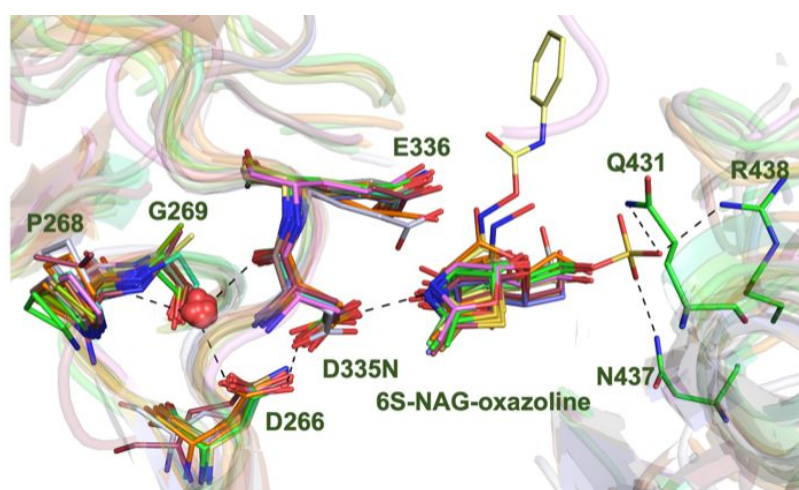
Supplementary Figure 9 Michaelis-Menten plots for 4MU-6S-GlcNAc hydrolysis by Bt4394_{WT} and Bt4394 variants at pH 5.5.

Initial rates of the substrate 4MU-6S-GlcNAc 5 μM hydrolysis catalyzed by Bt4394_{WT} and its variants were measured in the buffer of 25 mM Bis-tris propane, 25 mM citrate, and 300 mM NaCl, pH 5.5, at 25°C by monitoring the release of fluorescence from 4MU using $\lambda_{\text{ex}} = 360 \pm 10$ nm and $\lambda_{\text{em}} = 450 \pm 10$ nm. (a) 2 nM Bt4394_{WT} hydrolyses 5 μM to 500 μM substrate. (b) 5 nM Bt4394_{D335N} hydrolyses 100 μM to 5 mM substrate. (c) 100 nM Bt4394_{E336Q} hydrolyses 0.5 μM to 50 μM substrate. (d) 2 nM Bt4394_{D266N} hydrolyses 2.5 μM to 500 μM substrate. (e) 155 nM Bt4394_{Y435F} hydrolyses 0.5 μM to 50 μM substrate. (f) 1 μM Bt4394_{H270F} hydrolyses 0.5 μM to 50 μM substrate. (g) 99.4 nM Bt4394_{Q431E} hydrolyses 41 μM to 2625 μM substrate. (h) 14.2 nM Bt4394_{N437D} hydrolyses 41 μM to 2625 μM substrate. (i) 16.6 nM Bt4394_{R438A} hydrolyses 41 μM to 2625 μM substrate.

a

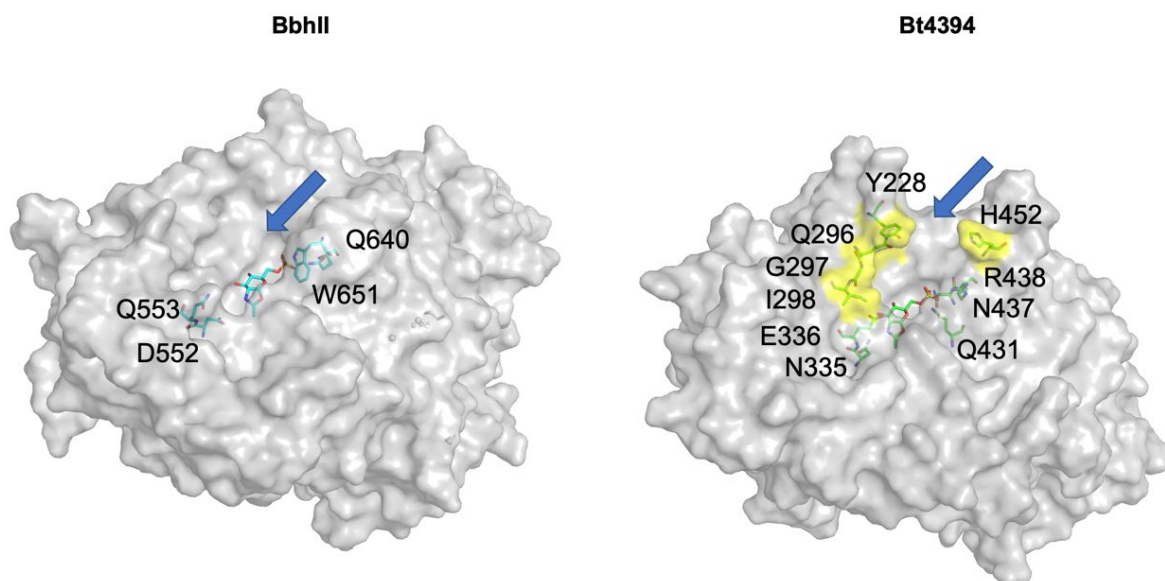


b



Supplementary Figure 10 Structures of the oxazoline intermediate complexes of Bt4394 in comparison with ScHex structure and other GH20 hexosaminidases structures with catalytic conformation.

(a) Overlay of the oxazoline intermediate complex structures of Bt4394 (in green) with the loop containing N335/E336 in a catalytic conformation, and ScHex (in magenta, PDB:4c7g) with the loop containing D301/Q302 also in the catalytic conformation. The sulfate binding sites in BbhII and Bt4394 are not available in ScHex or SpHex, because it is occupied by D383/D395 and W396/W408 coordinating the 6'-OH of the NAG-oxazoline. **(b)** Overlay of the structures of Bt4394 (PDB: 7DVB green sticks) with GH20 structures that have the loop containing the D/E pair equivalent to D335/E336 in a catalytically competent conformation. The GH20 structures used are (PDB codes): 1HP5, 4JAW, 5BXP, 5OAR, 4AZC, 6K35, 2EPN, 4C7G, 3OZP, 6Q63, 3GH4, 1YHT, 1QBA, 2YL6, 3RCN, 7CBN, and 6JQF.



Supplementary Figure 11 The structure comparison of BbhII with Bt4394 indicates the size difference of the leaving group in the substrates.

The narrower opening near the leaving group of the substrate formed by Y228, Q296, G2967, I298 and H452 (in yellow surface) in Bt4394 is narrower than a wider opening in BbhII. GAGB residues and sulfate-binding residues are shown in sticks.

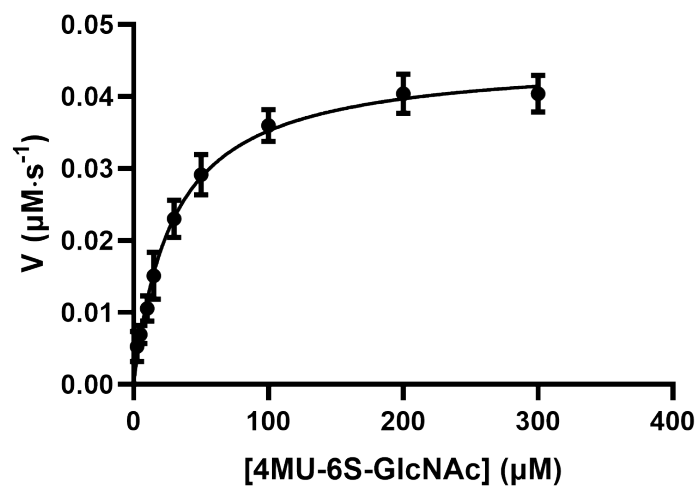

```

Bt4394      G-GFYTQDKDIKEIVAYAKKKFIEIPEIDIPGHSQAAVAAYPEFLA-CDP
F10-ORF19  G-GFYTQNDIKEIVAYAKTKFVEIPEIDIPGHSQAAVAAYPEFLA-CDP
F3-ORF26   S-GFYTQEEAKEIVKYAADRFITVPEVDLPGHMMGALVSYPEL-G-CTG
HexAa      --HIYTAQDVKEVIEYARLRGIRVLAEFDTPGHTLSWGPPIGLLTPCYS
sgl        GPYFYTQDEMREVVAYAKERHIEVLPEVDMPGHFVAAMAAYPEY-S-CNP
           : ** . : : * : : * * : : : : . * * * * . . * *
Bt4394      RDKHEVWLQQGISTDVINVANPKAMQFAKEVIDELTELFPPFYIHLGGDE
F10-ORF19  ENMHEVWLRQGISTDVINVANPQAMQFARDVIDELIDLFPFRYIHLGGDE
F3-ORF26   -GPYEIPCKWGVFPDVLCGGNDRALQFAKDLVNEIMDIFPSPYIHLGGDE
HexAa      GSEP-----SGTF-GPVNPSLNNTYEFMSTFFLEVSSVFPDFYLHLGGDE
sgl        SRAPQVWTGGGISSDVLNVANPQAVEFAKNILDELCDIFPYPIYHVGDE
           * . : . . : : * . : * : . : * * * : * : * * * *
Bt4394      CPTRKWQKNDECKLLSEIGS-SN-----FRDLQIYFYKQLKDYIATKP
F10-ORF19  CPTNKWQKNEECQSLLKEMGS-TN-----FRDLQIYFYKQLKDYMATKP
F3-ORF26   CPKVRWEKCSACQAKIRELGL-KDTPKHSKENQLQTYFMSEVGVKINDR-
HexAa      VDFTCWKSNPEIQDFMRKKGFGED-----FKQLESFYIQTLLDIVSSY-
sgl        CPTTQWEHNDLCQQYKELGL-TS-----YRQLQAHFIKDLADFVATK-
           * : : : * . . : * : . : . : . :
Bt4394      ADQQRQLIFWNEVLHGNTS----ILGNDITIMAWIGAN-----AA-AKQA
F10-ORF19  ANQQRRLVFWNEVLHGNTA----LLGNDITIMAWIGAD-----AA-AQNA
F3-ORF26   ---GRKMLGWDEMLEGGL-----APGATVMSWTGVK-----GG-IE-A
HexAa      ---GKGYVWVQEVFDNKVK-----IQPDTIIQVWREDIPVNYMKE-LELV
sgl        ---NKHLVCWNEAITAGGADLDLMKQTQSTIIMSWNPCQ-----EGVAKAV
           : : * * : : * : : * : :
Bt4394      AKQGMNTILSP----QIPYYINRRKQSKLPTEPMSQGHGTETVEAVYNYQP
F10-ORF19  AKQGMSTILSP----QIPYYINRRQSDLPTPEPMSQGHGTETVEAVYNYQP
F3-ORF26   ARLHHDAIMTPI----QFLYFSNPT-----YNRIKGTKSLERVYTFEP
HexAa      TKAGFRALLSA-----PWYLNRI-----SYGP-DWKDFYIYVEP
sgl        KKLGLPAIVTEYHKDGGYICRKQSNQDYGEPGAGYGNQDVGVEGCYNYVP
           : : : : : : : : : * : * *
Bt4394      LKDVDAAL-QPYKGVQANFWTEWVTEPSVLEYLMLPRLAAVAEAGWTPQ
F10-ORF19  MKGVEADL-QPYYSGVQANFWTEWVVDSSVLEYLMLPRLAAVAEAGWTPQ
F3-ORF26   VSNELAEDERKYIIGTQGCWTEWTRDSLKMEWQILPRMAALSEIQWTEP
HexAa      LAFEGTPEQKALVIGGEACMWGEYVDNTN-LVPRLWPRAGAVARLWLNK
sgl        VQGMYTQEQMALVKGVGQTFWTEHVTGNEYLEYLALPRLICVAEAGWTPQ
           : : * . : * * . : * * * . : * *
Bt4394      EKRNIEDFKERIRKDAELYDLKGWNYGKHIM-----
F10-ORF19  EKRDYVSFKERIRKDAVLYDLKGWNYGKHIM-----
F3-ORF26   LHKNFDSFLKRLPALLAIYRDRGYDFRQDIYDVNI-DIVPAPDEGKAKI-
HexAa      LTSDLTFAYERLSHFRCCELLRRGVQA-Q-----P-----
sgl        VFKNWDNFRTRLANQTQWLDDHGYYARHWMPGYVPRTQMPENEKVAAT
           : * : : * :

```

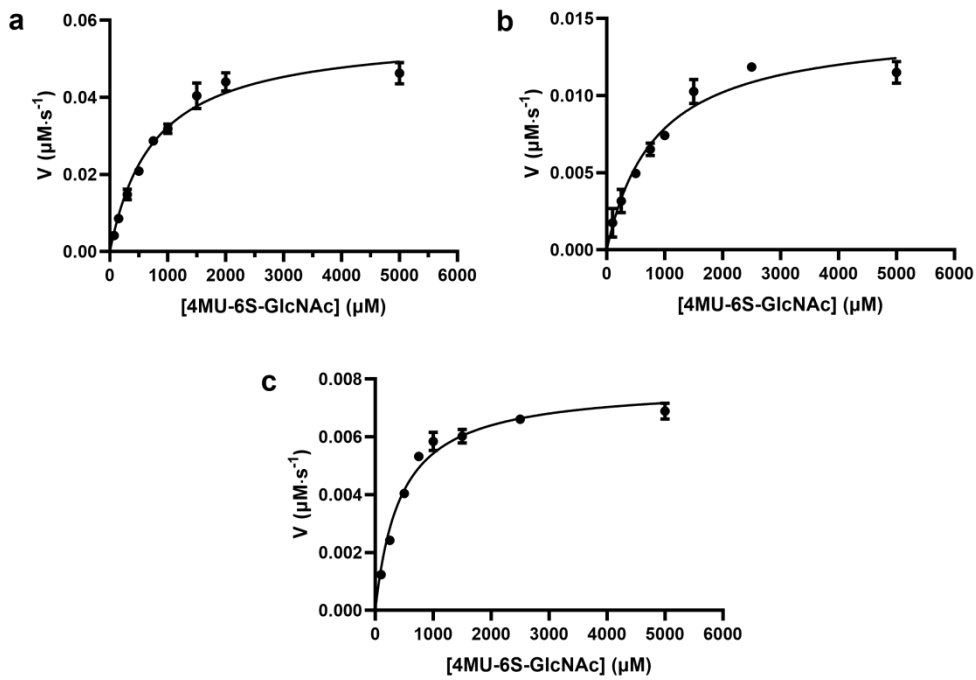
Supplementary Figure 12 Multiple sequence alignment (MSA) by T-coffee with default setting (<https://www.ebi.ac.uk/Tools/msa/tcoffee/>) for the known 6S-GlcNAcases.

MSA of Bt4394, human HexA α subunit, F10-ORF19 and F3-ORF26 and SGL show the predicted sulfate binding loop (highlighted in yellow) and sulfate bind residues in bold. The catalytic tyrosine that coordinated with the oxazoline ring-oxygen is highlighted in green.



Supplementary Figure 13 Michaelis-Menten plot for the hydrolysis of 4MU-6S-GlcNAc by SGL_{WT} at pH 6.0.

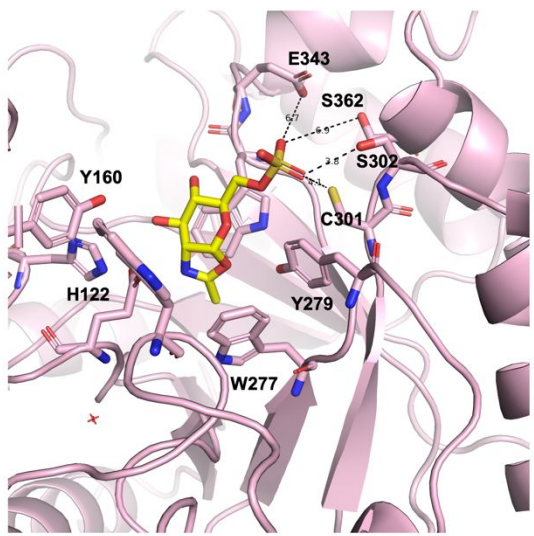
Initial hydrolytic rate of the substrate 4MU-6S-GlcNAc (2.5 μM to 300 μM) by 1.2 nM SGL_{WT} was performed in 25 mM Bis-tris propane-HCl, 25 mM citrate and 300 mM NaCl, pH 6.0 at 25°C. The release of the fluorescent 4MU was monitored using $\lambda_{\text{ex}} = 360 \text{ nm}$ and $\lambda_{\text{em}} = 450 \text{ nm}$).



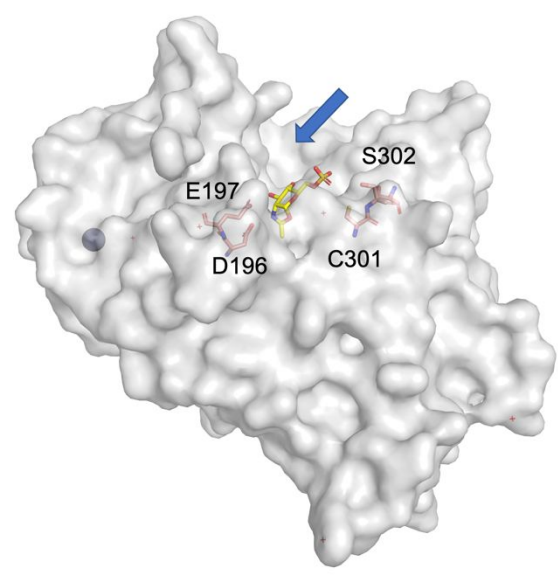
Supplementary Figure 14 Michaelis-Menten plots for the hydrolysis of 4MU-6S-GlcNAc by wild-type Niako3494 and its variants.

The initial rates of 4MU-6S-GlcNAc hydrolysis by Niako3494_{WT} and its variants were measured in the buffer of 25 mM Bis-tris propane, 25 mM citrate, and 300 mM NaCl, pH 5.0. All the reactions were measured at 25°C and monitored by release of fluorescent 4MU ($\lambda_{\text{ex}} = 360$ nm and $\lambda_{\text{em}} = 450$ nm). (a) 1 μM Niako3494_{WT} hydrolyses 75 μM to 5 mM substrate. (b) 393.6 μM Niako3494_{D196N} hydrolyses 100 μM to 5 mM substrate. (c) 80 μM Niako3494_{E197Q} hydrolyses 100 μM to 5 mM substrate.

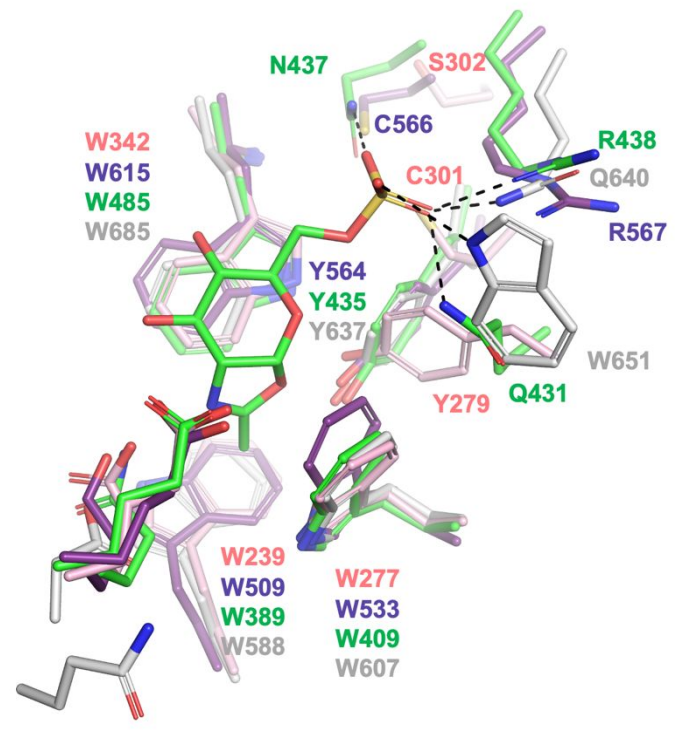
a



b



c



d

```

Bt4394      FIEIPEIDIPGHSQ-----AAVAAYPEFLACDPR-DKHEVWLQGGIST
HexAa      GIRVLAEFDTPGHTL-----SWGPGIPGLL---TPCYSGSEPSGTFG---
Niako3494  KIRFIPKMNLLGHQSDRDHIDPLLAKYPQFD-ESPDYNPPVPWKDAGPFD
sgl        HIEVLPEVDMPGHFV-----AAMAAYPEYS-CNPS-RAPQVWTGGGISS
          *..:..: **      . . *      *      *

Bt4394      ---DVINVANPKAMQFAKEVIDELTELFPFNYIHLGGDECPTRKWQKNDE
HexAa      ----PVNPSLNNTYEFMSTFFLEVSSVFPDFYLHLGGDEVDFTCWKSNPE
Niako3494  FYCKSLCPSHPDLLKTIFFPLMDELIDVCGADAFHVGLDEVWILGYEKCPR
sgl        ---DVLNVANPQAVEFAKNILDELCDIFPYPIYIHVGGDECPTTQWEHNDL
          : : . : .: *: .: :*: * * : :

Bt4394      CKKLLSEIGS-SNFRDLQIYFYKQLKDYIATKPADQQRQLIFWNEVLHGN
HexAa      IQDFMRKKGFGEDFKQLESFYIQTLLDIVSSY----GKGYVWQEVDFDNK
Niako3494  CGGRD-----KAALFAEYATKLHDHLKEK----KCQMWMWSDRLIDG
sgl        CQQKYKELGL-TSYRQLQAHFIKDLADFVATK----NKHLVCHNEAITAG
          * : * * :      *.: :

Bt4394      TS-----ILGNDITIMAWIGANA-----AAK--QAA
HexAa      VK-----IQPDTIIQVWREDIPVNYMKELE--LVT
Niako3494  KTTNLLGWQASMNATFRAID-LIPTDIMICDWKYESA-----PPTPGYFA
sgl        GA-----DLDLMKQTQSTIMSNPCQE-----GVAK--AVK
          : * *

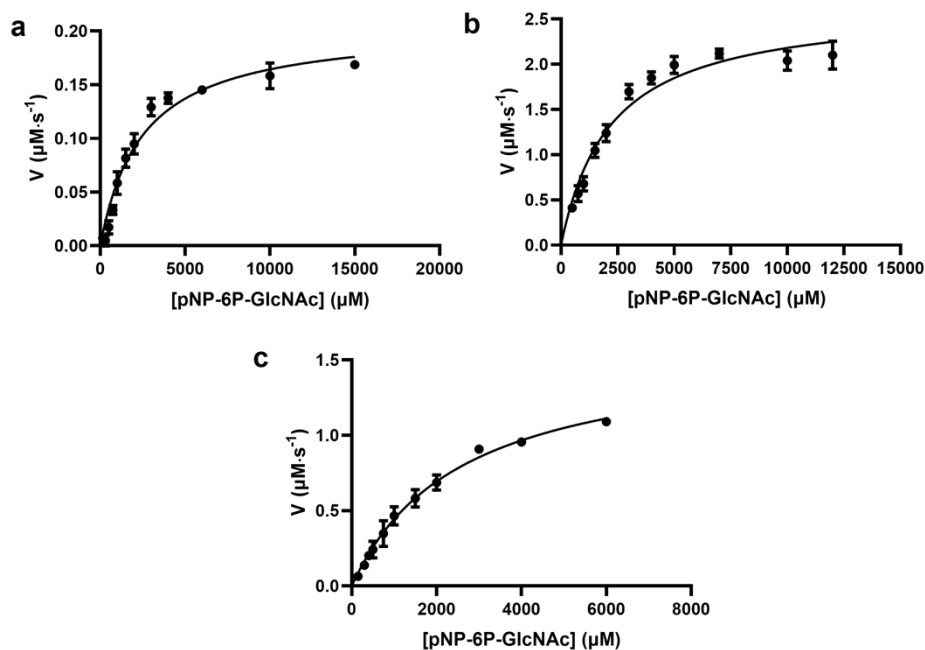
Bt4394      KQGMNTILSP----QIPYYINRKQSKLPTEPMSQGHGTETVEAVYNYQPL
HexAa      KAGFRALLS-----APWYLNRI-----SYGPDW-KDFYIVEPL
Niako3494  IKGFNVLPS-----CSNSEVALAQLAQVRLARKDGTTRAPWAVTL----
sgl        KLGLPAIVTEYHKDGGYYICRKQSN DYGEPSGAGY GNDGVEGCYNYVVP
          *: .: :      *

Bt4394      KDVDAALQ-PYKGVQANFTEWVTEPSVLEYLMLPRLAAVAEAGWTPQE
HexAa      AFEGTPEQKALVIGGEACMNGEYVDNTNLVPRL-WPRAGAVAERLWSNKL
Niako3494  -----AERMQGVFVTMWEDSKFIDAYY-----
sgl        QGMYTQEQMALVKGVQGTFTTEHVG TNEYLEYLALPRLICVAEAGWTPQV
          *: : * : : .

```

Supplementary Figure 15 Active site of Niako3494 and its sequence aligned with “sulfate-binding site” in other 6S-GlcNAcases.

(a) Active site of apo Niako3494 overlaid with the Bt4394-6S-oxazoline intermediate complex (only 6S-oxazoline is shown for clarity) to show the residues in Niako3494’s “predicted sulfate binding sites” are far away from the 6-sulfate. (b) Surface of apo Niako3494 overlaid with the Bt4394-6S-oxazoline intermediate complex (only 6S-oxazoline is shown for clarity) to show the open active site with a potential sugar binding site beyond 3'-OH and 4'-OH for bigger substrates and why Niako3494 could be an *endo*-glycosidase. (c) Structure of Niako3494 aligned with other 6S-GlcNAcase structures to show the Y279 enters the active site from a different direction. (d) Multiple sequence alignment (MSA) for Bt4394, human HexA α -subunit, Niako3494 and SGL by T-coffee with default setting (<https://www.ebi.ac.uk/Tools/msa/tcoffee/>) show the residues predicted to participate the sulfate binding (highlighted in yellow). The “back up” residues (N118 for Niako3494, D463 for BbhII, D266 for Bt4394, D391 for SGL) are highlighted in cyan. The catalytic diad and other conserved residues are highlighted in magenta. Our Niako3494 structure and kinetics studies have shown that its aligned region (in yellow and green) in the sequence cannot interact with the 6-sulfate moiety effectively.

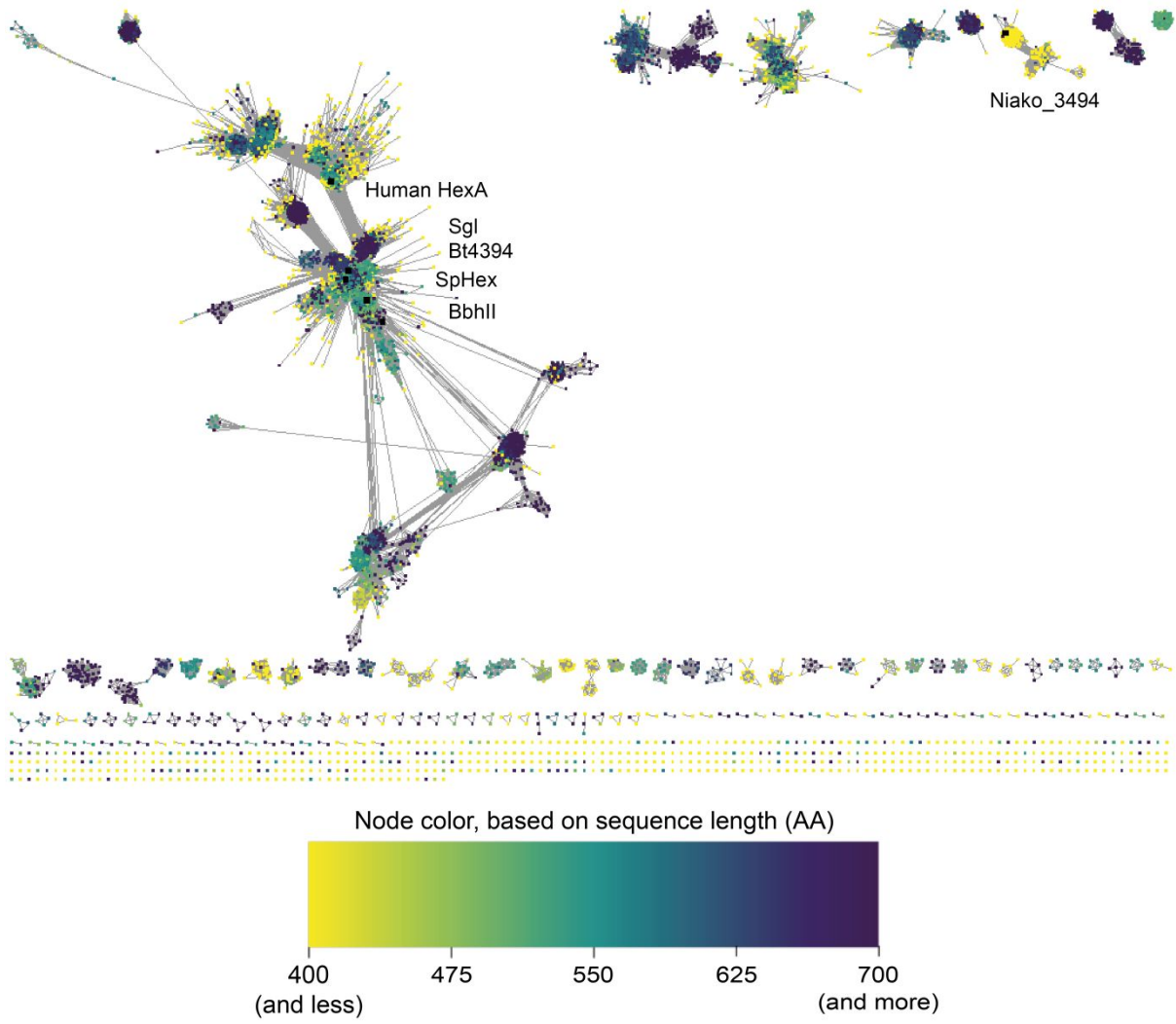


Supplementary Figure 16 Michaelis-Menten plots for illustrating substrate specificity of pNP-6P-GlcNAc with BbhII_{WT}, Bt4394_{WT}, and SGL_{WT}.

The initial rates of hydrolysis were measured in 50 mM Bis-tris propane, 100 mM NaCl (Bt4394_{WT}, BbhII_{WT}, and SGL_{WT} at pH 7.0) at 25°C. The release of the pNP was monitored at 405 nm. (a) 50 nM BbhII_{WT} hydrolyses 150 μM to 15 mM substrate. (b) 1 μM Bt4394_{WT} hydrolyses 500 μM to 12 mM substrate. (c) 100 nM Sgl_{WT} hydrolyses 150 μM to 6 mM substrate.

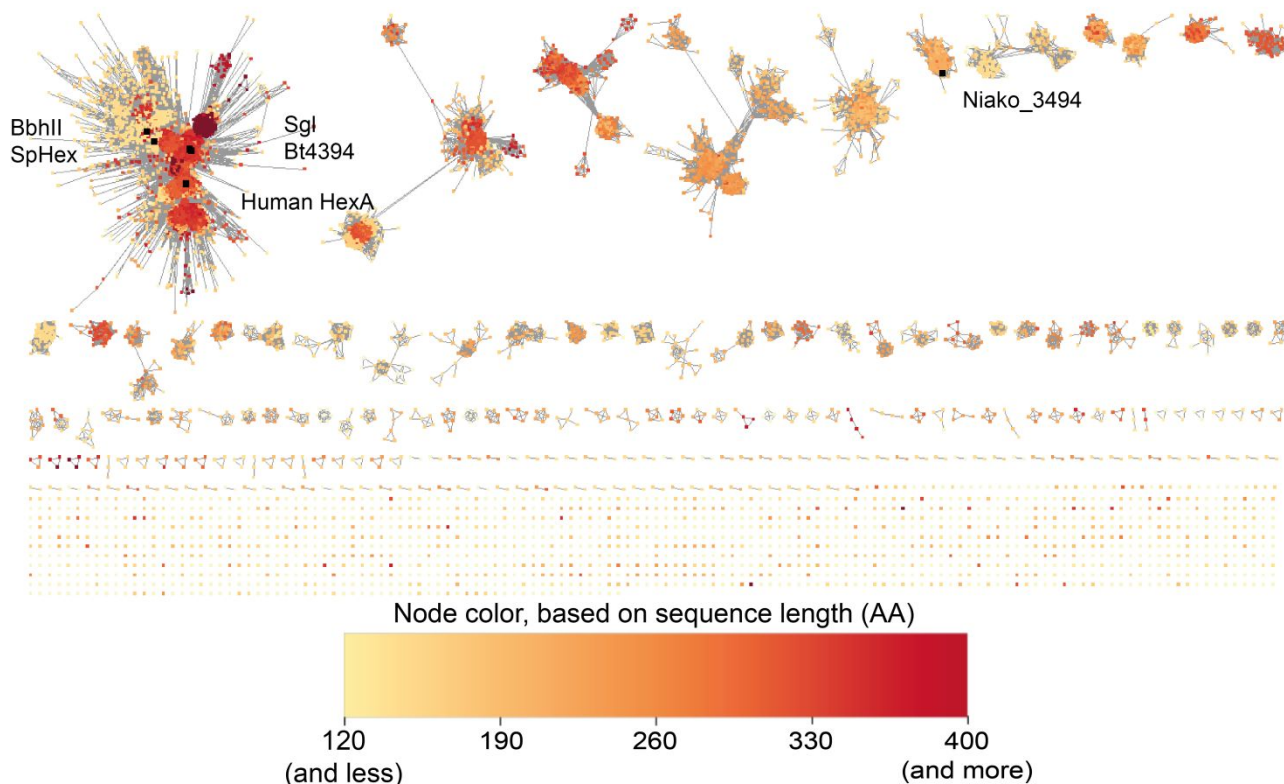
a

GH20 (PF00728), UniRef90, AST83 ($\pm 35\%$ identity), No Fragments



b

GH20 (PF00728) Domain only, UniRef90, AST57 ($\pm 35\%$ identity), No Fragments



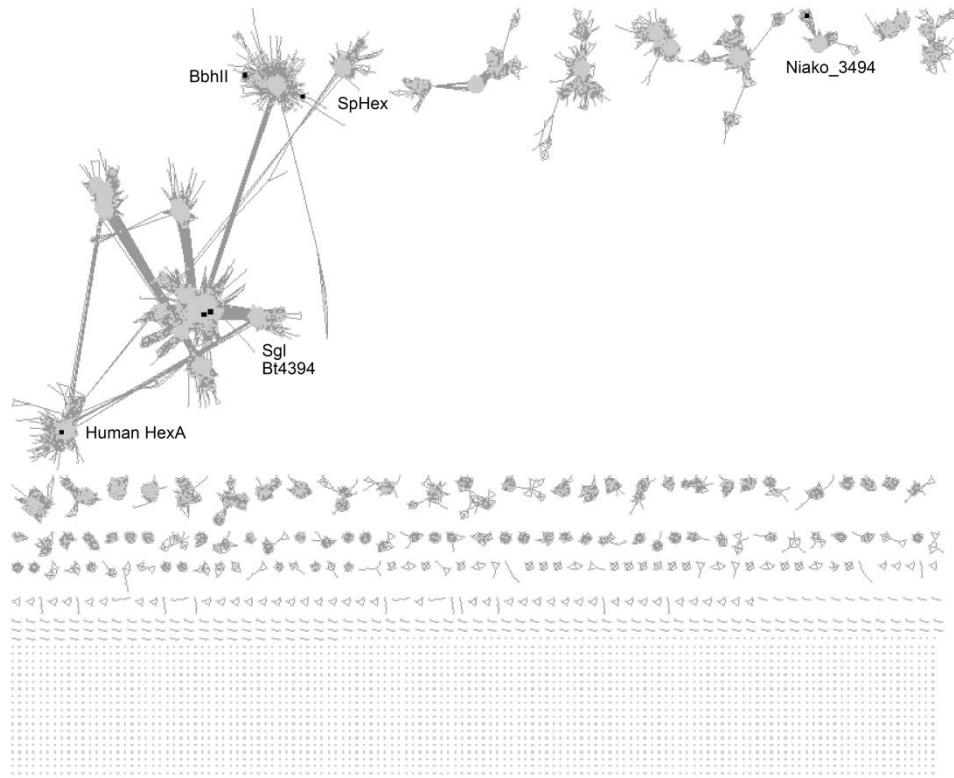
Supplementary Figure 17 Sequences with 6S-GlcNAcase activity and the reference GH20 SpHex (all highlighted in •) on the UniRef90 SSN for the protein family PF00728, GH20, at AST 83 for the full length (a) and AST 57 for the GH20 domain only (b).

Number of Nodes: 16,977 (representing 35,970 sequences); Number of Edges: 11,727,606. (a) When plotting with the full length, at AST 83, only sequences that share around 35% identity or more are connected. Coloring scheme is based on the length of the representative sequences. The variability in color shows that GH20 members are often found fused to other domains that are not necessarily equipped with beta-*N*-acetylhexosaminidase activities. The domain fusions can influence networks. (b) When plotting with GH20 domain only, at AST 57, only sequences that share around 35% identity or more are connected. The coloring scheme is based on the lengths of the GH20 domain in the representative sequences. Variability in sequence length is much smaller than in A, and the sequences with similar lengths are clustered together.

Niako_3494 is only characterized for its 6S-GlcNAcase activity (GH20_distant) by using pNP-6S-GlcNAc¹⁷ and it is very different from the others at the sequence level. BbhII, Bt4394, SGL as well as SpHex all fall into the same cluster at low similarity threshold (35%).

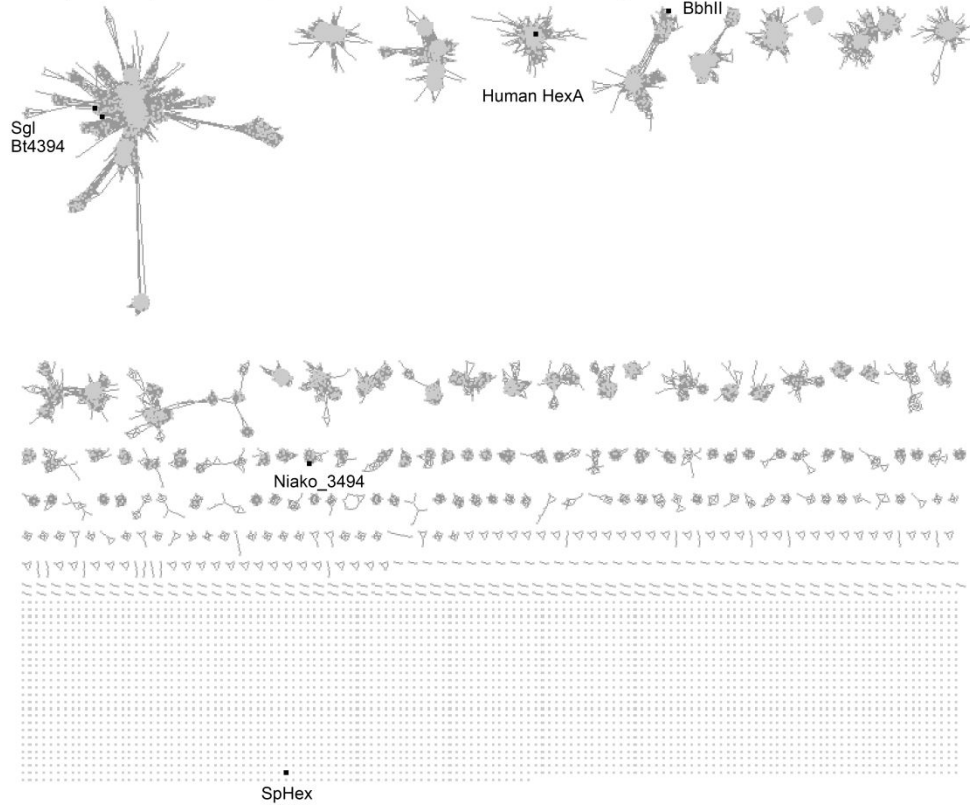
a

GH20 (PF00728) Domain only, UniRef90, AST75 ($\pm 40\%$ identity), No Fragments



b

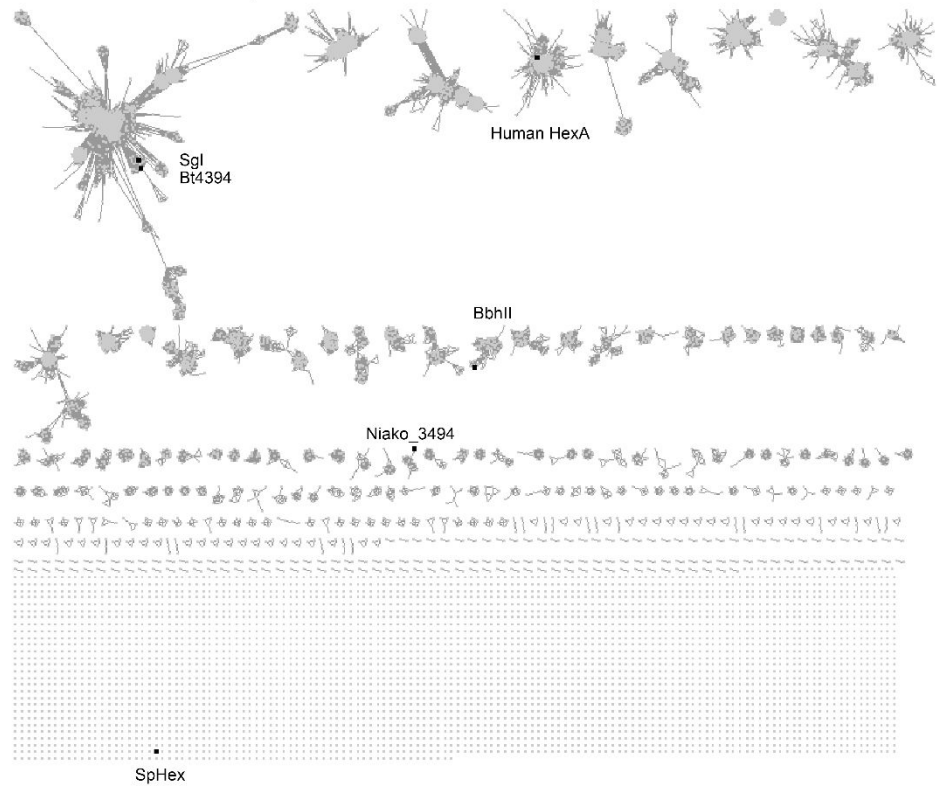
GH20 (PF00728) Domain only, UniRef90, filtered for AST84, No Fragments



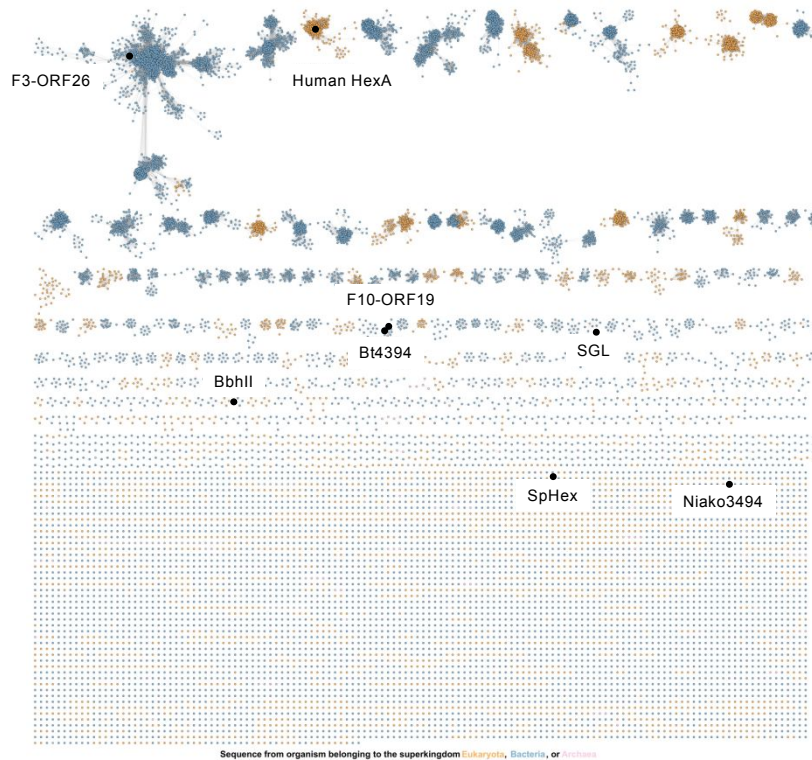
c

d

GH20 (PF00728) Domain only, UniRef90, filtered for AST88, No Fragments



GH20 (PF00728) Domain only, UniRef90, AST130, No Fragments



Supplementary Figure 18 6S-GlcNAcases, Bt4394, BbhII, SGL, human HexA, Niako3494, F3-ORF26, F10-ORF19, and the reference GH20 SpHex (all highlighted in •) mapped on the UniRef90 SSN for the protein family PF00728 representing GH20.

The pair-wise alignment was using the sequences from just the GH20 catalytic domain. AST75 is not yet enough to separate the subgroups. Increasing the AST thresholds from (a) 75, (b) 84, (c) 88 to (d) 130 leads to the separation of subclusters that corresponds to differences in the GH20 domain sequences. At the sequence level, the other sequences are closer together than they are to SpHex, one of the best studied GH20. BbhII is different enough that it goes in its own cluster at a medium threshold. SGL and Bt4394 are close at the sequence level, so that they are only separated when the AST is much higher. The minimum alignment score threshold (AST) to separate SGL from Bt4394 is 130.

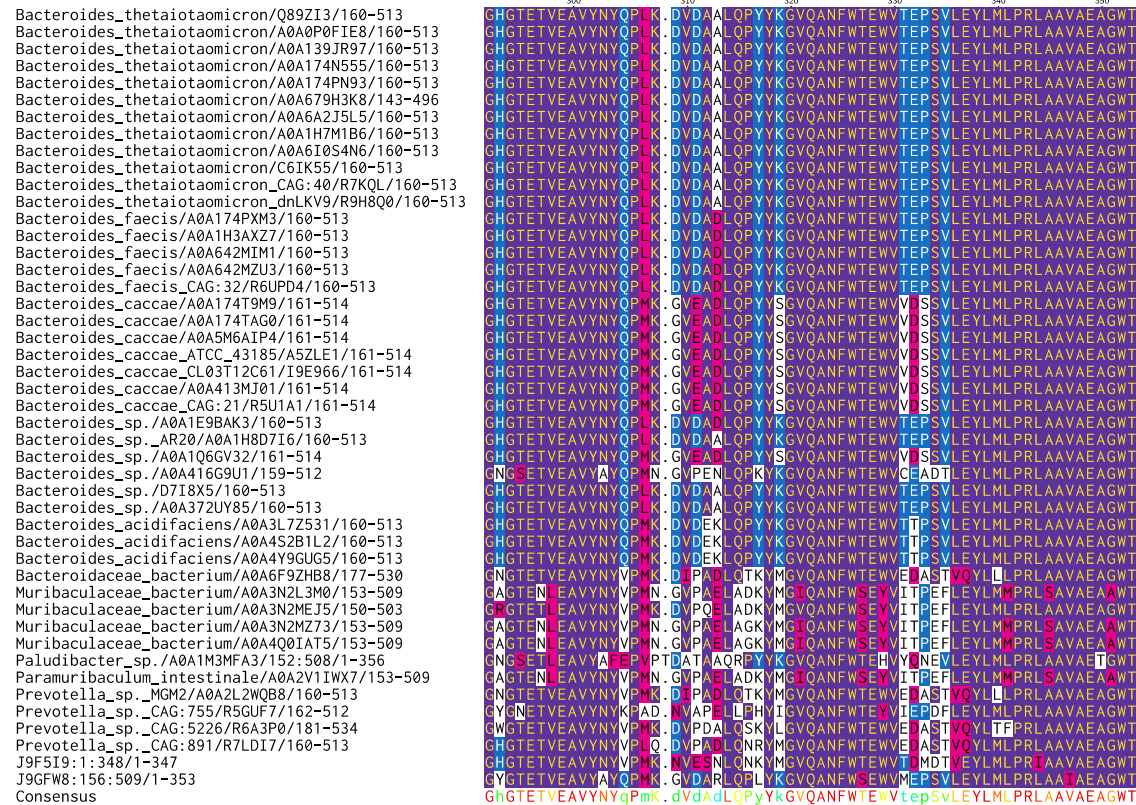
GH20
GH20
GH20
GH20

	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290
Bacteroides_thetaiotaomicron/O89Z13/160-513	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/A0A0P0FIE8/160-513	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/A0A139J97/160-513	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/A0A174N955/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/A0A174PN93/160-513	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/A0A679H3K8/143-496	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/A0A6A2J5L5/160-513	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/A0A1H7M1B6/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/A0A6I0S4N6/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron/C6IK55/160-513	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron_CAG:40/R7KQL/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_thetaiotaomicron_dnlKV9/R9H8Q0/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_faecis/A0A174PXM3/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_faecis/A0A1H3AXZ7/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_faecis/A0A642MIM1/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_faecis/A0A642MZU3/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_caccae/A0A174T9M9/161-514	VANPKAMQFARDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ADAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_caccae/A0A174TAG0/161-514	VANPKAMQFARDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ADAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_caccae/A0A5M6IP4/161-514	VANPKAMQFARDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ADAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_caccae_ATCC_43185/A5ZLE1/161-514	VANPKAMQFARDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ADAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_caccae_CL03T12C61/I9E966/161-514	VANPKAMQFARDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ADAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_caccae_CAG:21/R5U1A1/161-514	VANPKAMQFARDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ADAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_sp./A0A1E9BAK3/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_sp._AR20/A0A1H8D7I6/160-513	VANPKAMQFARDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ADAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_sp./A0A1Q6V32/161-514	VAKPEAVQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_sp./A0A416G9U1/159-512	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_sp./D7I8X5/160-513	VANPKAMQFAKEVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_sp./A0A372UY85/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_acidifaciens/A0A3L7Z531/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_acidifaciens/A0A4S2B1L2/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroides_acidifaciens/A0A4Y9GUG5/160-513	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Bacteroidaceae_bacterium/A0A6F9ZH88/177-530	VSNPRAVQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Muribaculaceae_bacterium/A0A3N2L3M0/153-509	VSNPAAVKMSTDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Muribaculaceae_bacterium/A0A3N2MEJ5/150-503	VANPAAVRMSTDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Muribaculaceae_bacterium/A0A3N2M73/153-509	VSNPAAVKMSTDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Muribaculaceae_bacterium/A0A4Q0IAT5/153-509	VSNPAAVKMSTDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Paludibacter_sp./A0A1M3MFA3/152:508/1-356	VANPAAVQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Paramuribaculum_intestinale/A0A2V1IWX7/153-509	VSNPAAVKMSTDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Prevotella_sp._MGM2/A0A2L2WQB8/160-513	VSNPRAVQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Prevotella_sp._CAG:755/R5GUF7/162-512	VSNPKSIOFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Prevotella_sp._CAG:5226/R6A3P0/181-534	VSNPRAVQFANDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Prevotella_sp._CAG:891/R7LD17/160-513	VANPKAVQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
J9F519:1:348/1-347	VANPKARELARNIDVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
J9GFW8:156:509/1-353	VANPKALQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									
Consensus	VANPKAMQFAKVIDELTELFFPNI	IHLGGDECPTRKWQKNECKKLLS	IGSSNFRDLQIYFYKOLKDYIATKPA	DQORQLIFWNEVLHGNTSL	LGNDDITIMAWIG	ANAAAKQAAGKGMNTILSPOIPYYINRROSKLPTPEMSQ									

↑
GAGB

↑
SO3 binding

GH20



- non-conserved
- similar
- ≥ 50% conserved
- ≥ 80% conserved

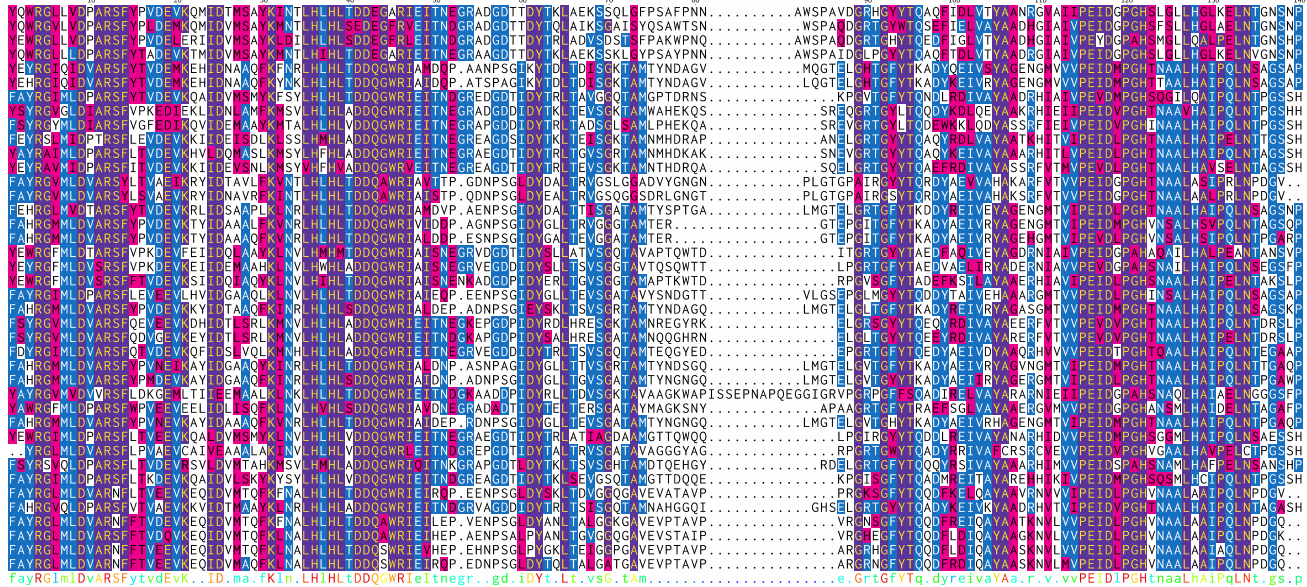
Supplementary Figure 19 The GH20 domain sequences with the sulfate-binding feature as in Bt4394.

The key conserved residues known to be involved in classical GH20 activity are labelled at the top of the sequences. The sulfate recognizing residues and catalytic D-E amino acid pair are labelled at the bottom of the sequences.

GH20

GH20

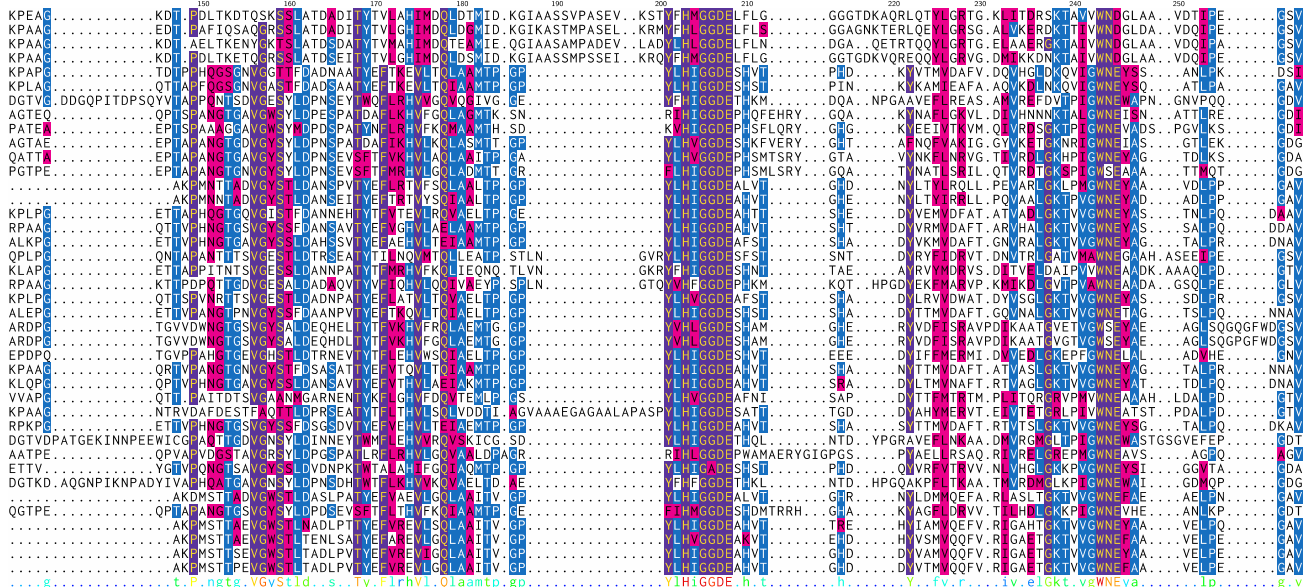
Bifidobacterium_sp._LMG31471/A0A61IGJ90/346-711
 Bifidobacterium_bifidum_PRL2010/A0A0H3E56/347-712
 Bifidobacterium_samirlii/A0A430FN00/340-705
 Bifidobacterium_jacchi/A0A5N5R154/349-716
 Arthro bacter_alpinus/A0A052M086/305-649
 Arthro bacter_glacialis/A0A232W48/326-671
 Tru eperella_bernardiae/A0A0W1KJ3/317-672
 Actinomyces_tangfeifanii/A0A1D9MLR8/337-684
 Actinomyces_tangfeifanii/A0A1D9MLR8/337-684
 Actinomyces_israelii/A044481T18/336-683
 Actinomyces_graeviniitzi_C83/G9PHV6/343-690
 Actinomyces_sp._oral_taxon_897/A0A250LQB6/221-568
 Actinokineospora_terrerae/A0A1H9XKJ2/173-506
 Actinokineospora_auranticolor/A0A256GKZ8/173-506
 Agromyces_sp._OV568/A0A3E0HXL3/326-671
 Nonomuraea_wenchangensis/A0A1I0B0W8/309-648
 Nonomuraea_sp._C10/A0A5C8J0A4/309-648
 Tessaracoccus_bendigoensis_DSM12906/A0A1M6K157/315-662
 Tessaracoccus_sp._Z501/A0A1R1H99/323-672
 Tessaracoccus_sp._OH4464/A0A3P1WP49/320-669
 Cellulosimicrobium_cellulans_J1/A0A1X7DV88/326-670
 Actinoadura_meyerae/A0A239F4F/316-661
 Streptomyces_sp._CMB-S1M0423/A0A2K9FN11/190-538
 Streptomyces_sp._WAC06738/A0A38VVT9/190-538
 Georgenia_satyanarayana/A0A2Y9AB68/201-554
 Micromonospora_arida/A0A3N9XZ3/324-669
 Micromonospora_sp._MP36/A0A5D0UQ6/188-533
 Buchananella_hordeovulneris/A0A3P1W775/328-686
 Pseudactinotalea_sp._HY158/A0A5Q2W410/339-700
 Streptosporangium_roseum/D2B927/335-680
 Olseniella.sp._An290/A0A1Y4B5P5/515-872
 Actinomyces_howellii/A0A448H632/163-509
 Cutibacterium_acnes/A0A5R8PR66/131-1655
 Collinsella_stercoris_DSM13279/B6G8V9/346-698
 Paenarthrobacter_nicotinovorans/A0A318WB3/142-471
 Bowdeniella_nasicola/A0A1Q5Q1W4/315-661
 Paenarthrobacter_aurescens/A0A4Y3NPR5/137-466
 Arthro bacter_sp._TS-15/A0A545BM0/45-374
 Arthro bacter_sp._Myb23/A0A259D4H6/138-467
 Arthro bacter_sp._EprR571/A0A117N8T2/148-477
 Consensus



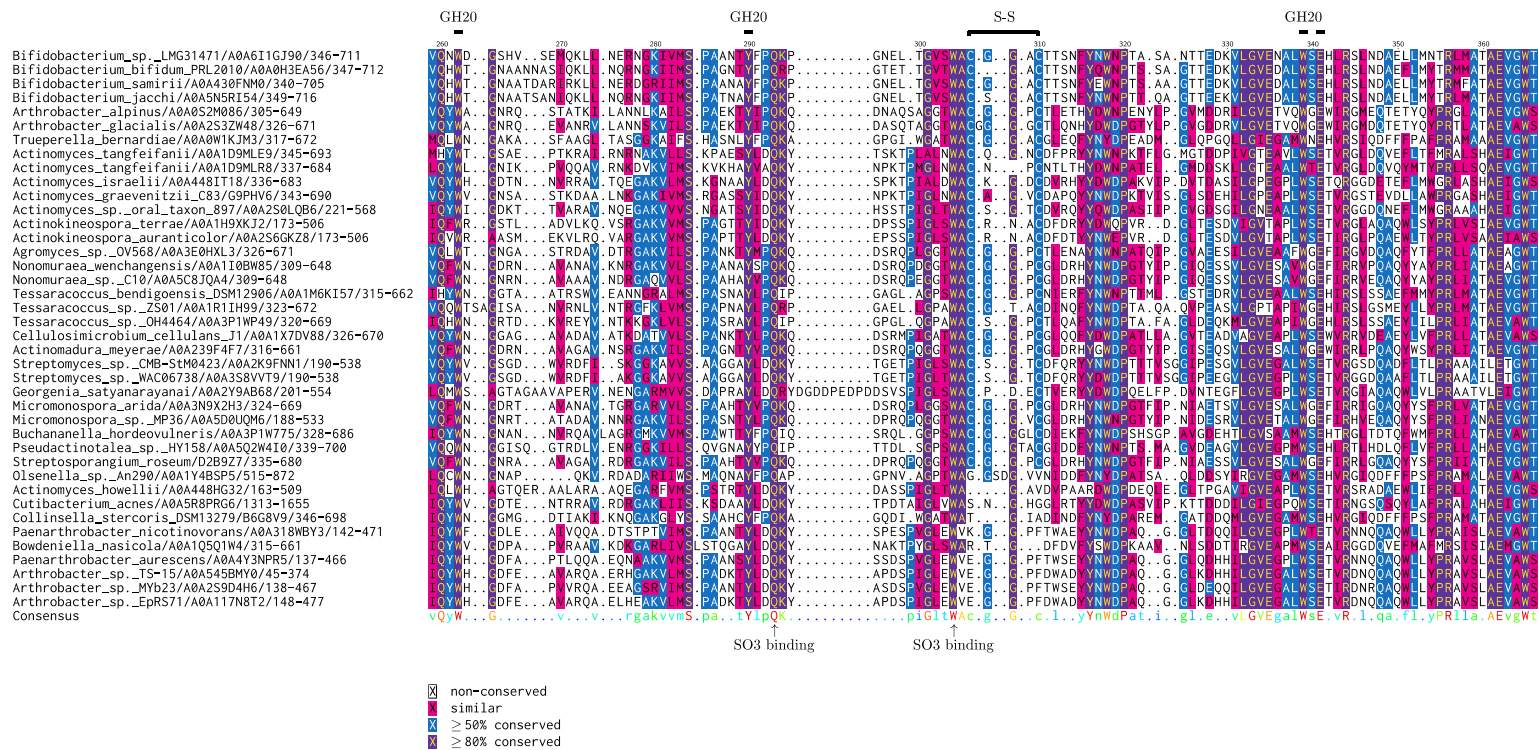
GH20

GH20

Bifidobacterium_sp._LMG31471/A0A61IGJ90/346-711
 Bifidobacterium_bifidum_PRL2010/A0A0H3E56/347-712
 Bifidobacterium_samirlii/A0A430FN00/340-705
 Bifidobacterium_jacchi/A0A5N5R154/349-716
 Arthro bacter_alpinus/A0A052M086/305-649
 Arthro bacter_glacialis/A0A232W48/326-671
 Tru eperella_bernardiae/A0A0W1KJ3/317-672
 Actinomyces_tangfeifanii/A0A1D9MLR8/337-684
 Actinomyces_tangfeifanii/A0A1D9MLR8/337-684
 Actinomyces_israelii/A044481T18/336-683
 Actinomyces_graeviniitzi_C83/G9PHV6/343-690
 Actinomyces_sp._oral_taxon_897/A0A250LQB6/221-568
 Actinokineospora_terrerae/A0A1H9XKJ2/173-506
 Actinokineospora_auranticolor/A0A256GKZ8/173-506
 Agromyces.sp._OV568/A0A3E0HXL3/326-671
 Nonomuraea_wenchangensis/A0A1I0B0W8/309-648
 Nonomuraea_sp._C10/A0A5C8J0A4/309-648
 Tessaracoccus_bendigoensis_DSM12906/A0A1M6K157/315-662
 Tessaracoccus_sp._Z501/A0A1R1H99/323-672
 Tessaracoccus.sp._OH4464/A0A3P1WP49/320-669
 Cellulosimicrobium_cellulans_J1/A0A1X7DV88/326-670
 Actinoadura_meyerae/A0A239F4F/316-661
 Streptomyces_sp._CMB-S1M0423/A0A2K9FN11/190-538
 Streptomyces.sp._WAC06738/A0A38VVT9/190-538
 Georgenia_satyanarayana/A0A2Y9AB68/201-554
 Micromonospora_arida/A0A3N9XZ3/324-669
 Micromonospora.sp._MP36/A0A5D0UQ6/188-533
 Buchananella_hordeovulneris/A0A3P1W775/328-686
 Pseudactinotalea.sp._HY158/A0A5Q2W410/339-700
 Streptosporangium_roseum/D2B927/335-680
 Olseniella.sp._An290/A0A1Y4B5P5/515-872
 Actinomyces_howellii/A0A448H632/163-509
 Cutibacterium_acnes/A0A5R8PR66/131-1655
 Collinsella_stercoris_DSM13279/B6G8V9/346-698
 Paenarthrobacter_nicotinovorans/A0A318WB3/142-471
 Bowdeniella_nasicola/A0A1Q5Q1W4/315-661
 Paenarthrobacter_aurescens/A0A4Y3NPR5/137-466
 Arthro bacter_sp._TS-15/A0A545BM0/45-374
 Arthro bacter_sp._Myb23/A0A259D4H6/138-467
 Arthro bacter_sp._EprR571/A0A117N8T2/148-477
 Consensus

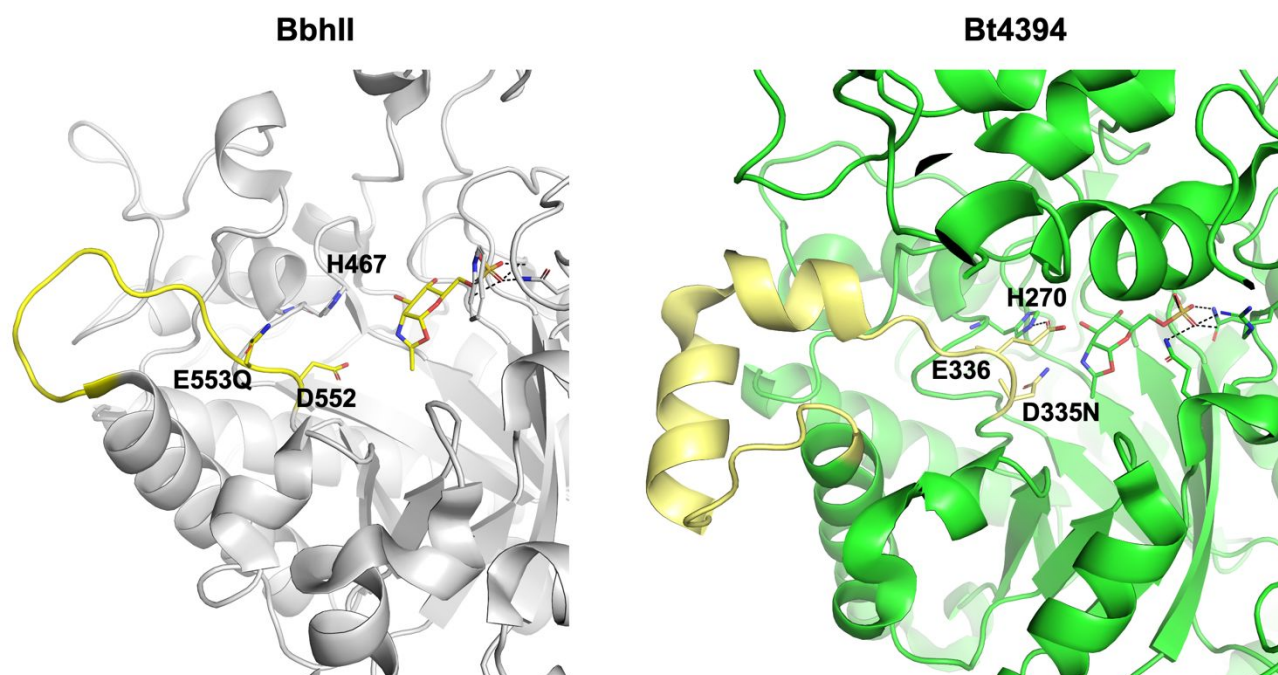


GAGB



Supplementary Figure 20 The GH20 domain sequences with the sulfate-binding feature as in Bbhl.

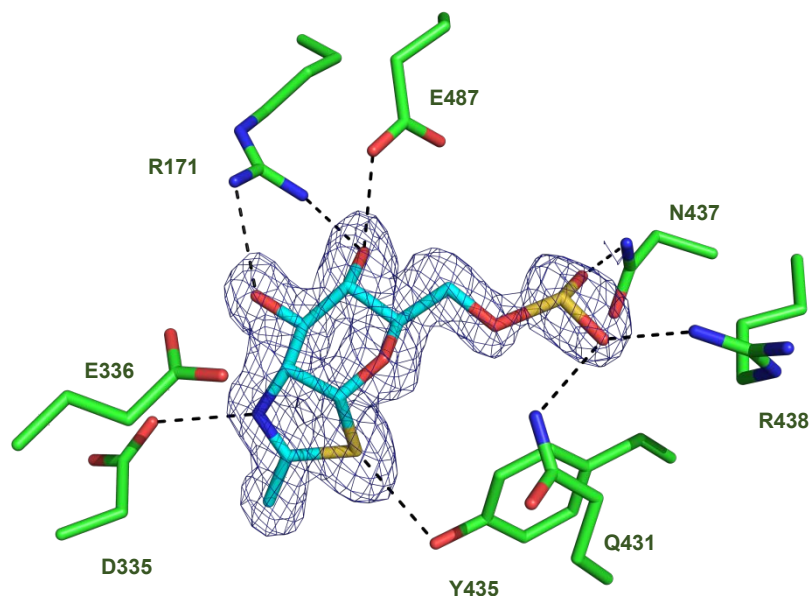
The key conserved residues known to be involved in classical GH20 activity are labelled at the top of the sequences. The sulfate recognizing residues and catalytic D-E amino acid pair are labelled at the bottom of the sequences.



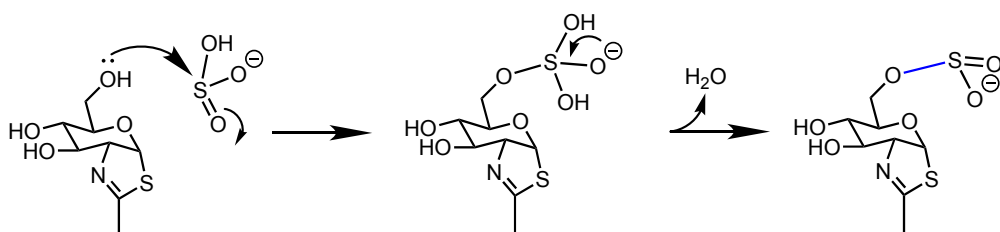
Supplementary Figure 21 Comparison of the flexibility of the catalytic loop with the D-E diad relative to the conserved histidine (H467 in BbhII, H270 in Bt4394).

This histidine has been proposed to modulate the pK_a s of catalytic diad by oscillating between the two in a GH20 family, in 6S-NAG-oxazoline intermediate structures of BbhII and Bt4394. The whole catalytic loop has been colored in yellow.

a



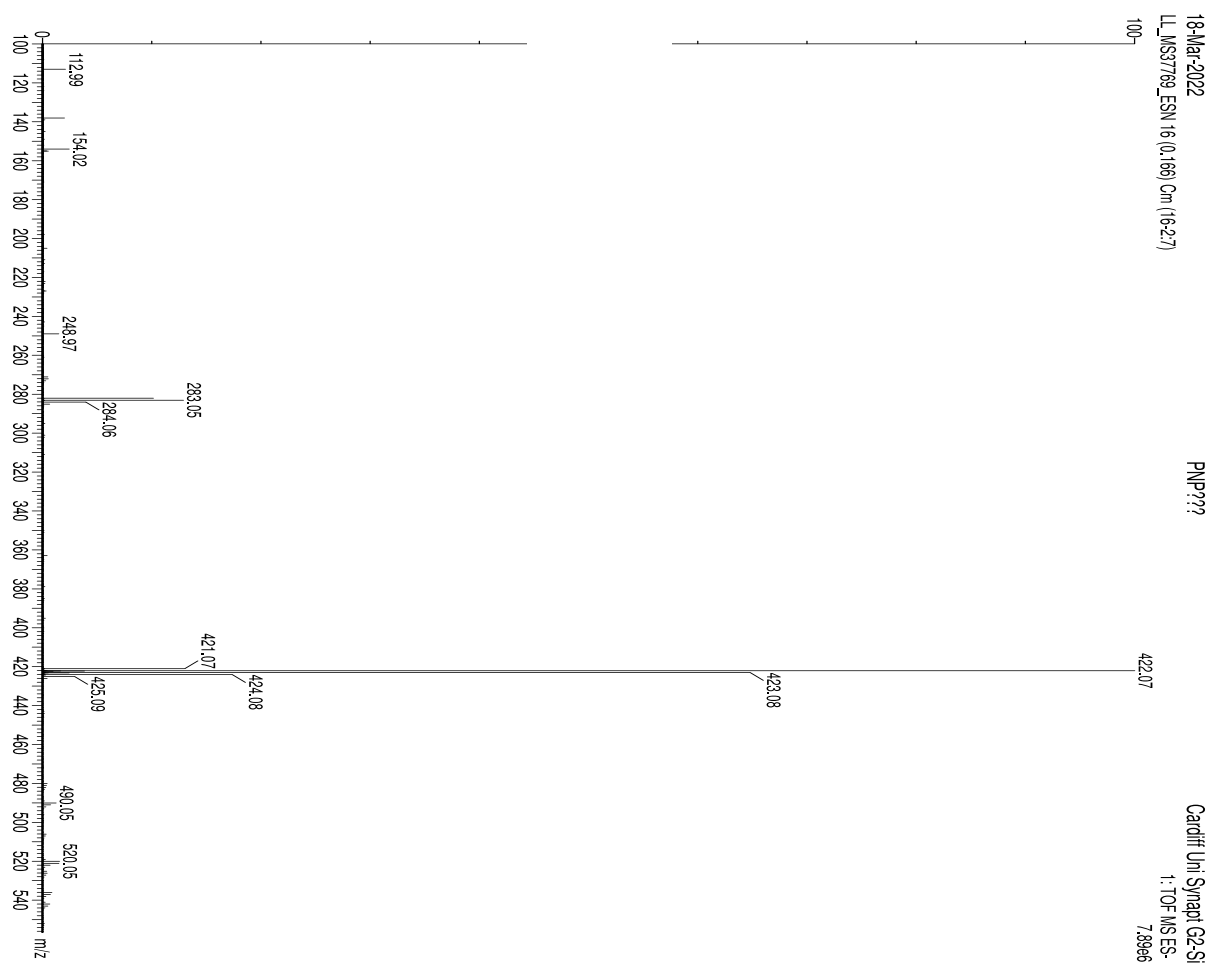
b



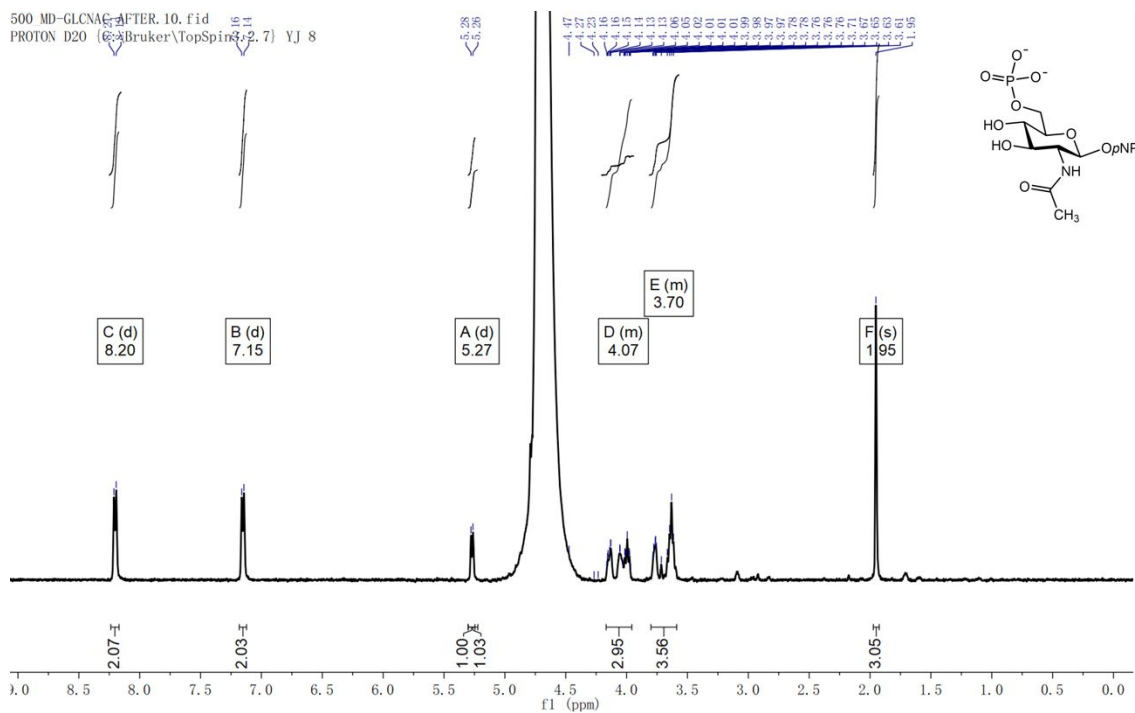
Supplementary Figure 22 The observed NAG thiazoline-SO₂⁻ species found in the Bt4394 active site.

(a) Active site of Bt4396 with NAG thiazoline-SO₂⁻ bound. $2F_o-F_c$ map contoured at 1.0 σ ($0.333 \text{ e}/\text{\AA}^3$) for NAG-thiazoline and the partially covalently linked SO₂ species. (b) The proposed mechanism of the condensation reaction between NAG-thiazoline and sulfite through dehydration during the co-crystallization of Bt4394, NAG-thiazoline and sodium sulfite in the 1.47 Å resolution structure.

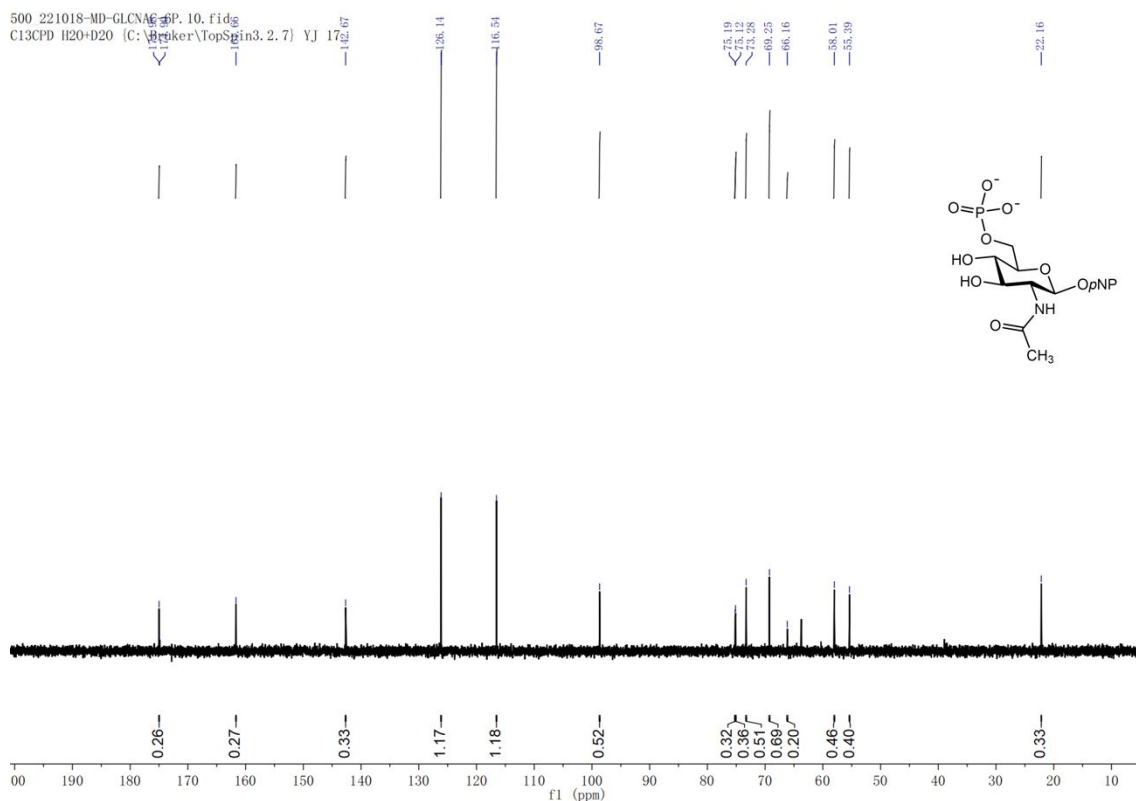
4. HRMS, ¹H-NMR and ¹³C-NMR spectra



Supplementary Figure 23 MS spectrum of **1** (ESI-MS). calcd. m/z for $C_{14}H_{19}N_2O_{11}P^-$ [M]⁻ 422.07, found 422.07.

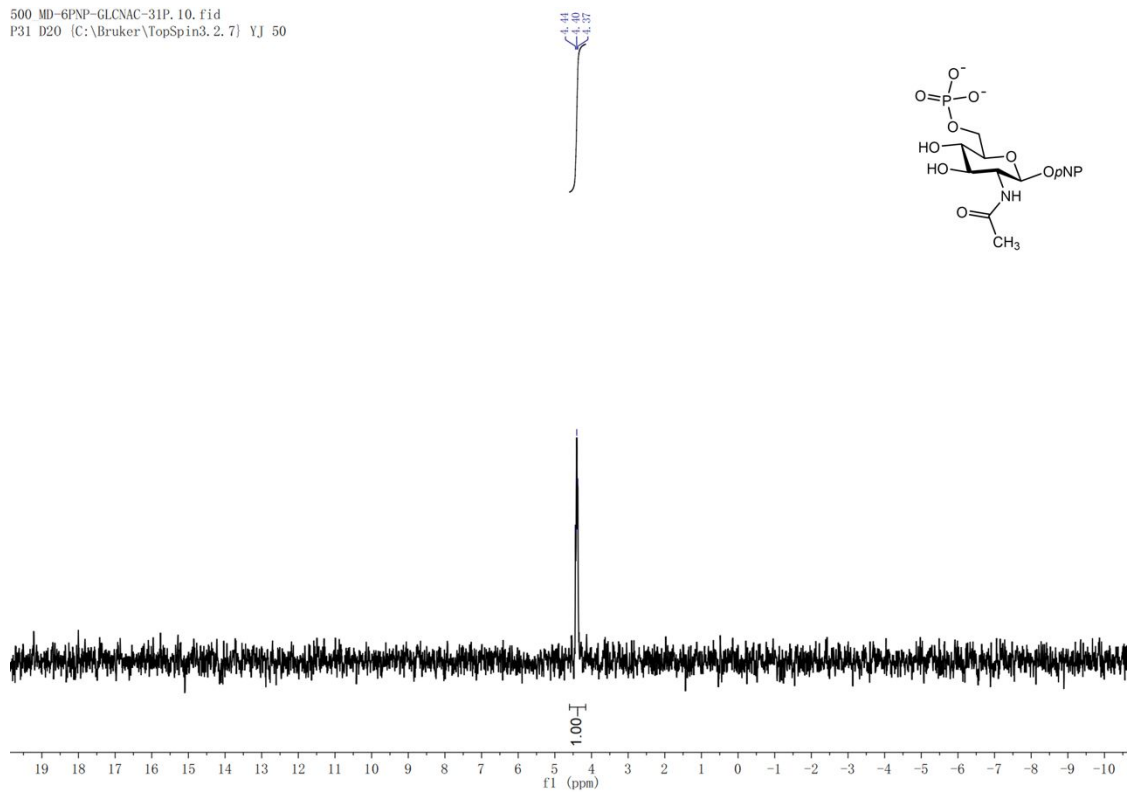


Supplementary Figure 24 ^1H NMR spectrum of **1** in D_2O (500 MHz).



Supplementary Figure 25 ^{13}C NMR spectrum of **1** in D_2O (125 MHz).

500 MD-6PNP-GLCNAC-31P_10.fid
P31 D20 [C:\Bruker\TopSpin3.2.7] YJ 50



Supplementary Figure 26 ^{31}P NMR spectrum of **1** in D_2O (202 MHz).

5. SI References :

- (1) Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125-132.
- (2) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 859-866.
- (3) Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K.; Lebedev, A.; McNicholas, S. J.; Nicholls, R. A.; Noble, M.; Pannu, N. S.; Roth, C.; Sheldrick, G.; Skubak, P.; Turkenburg, J.; Uski, V.; von Delft, F.; Waterman, D.; Wilson, K.; Winn, M.; Wojdyr, M. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **2018**, *74*, 68-84.
- (4) Keegan, R. M.; Winn, M. D. MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64*, 119-124.
- (5) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658-674.
- (6) Vagin, A.; Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 22-25.
- (7) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126-2132.
- (8) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355-367.
- (9) Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **2010**, *43*, 186-190.
- (10) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583-589.
- (11) Katoh, T.; Maeshibu, T.; Kikkawa, K.-i.; Gotoh, A.; Tomabechei, Y.; Nakamura, M.; Liao, W.-H.; Yamaguchi, M.; Ashida, H.; Yamamoto, K.; Katayama, T. Identification and characterization of a sulfoglycosidase from *Bifidobacterium bifidum* implicated in mucin glycan utilization. *Biosci. Biotech. Biochem.* **2017**, *81*, 2018-2027.
- (12) Quan, S.; Hiniker, A.; Collet, J.-F.; Bardwell, J. C. A. Isolation of bacteria envelope proteins. In *Bacterial Cell Surfaces, Methods Mol. Biol.* **2013**. *966*, 359-366.
- (13) Zallot, R.; Oberg, N. O.; Gerlt, J. A. 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77-85.

(14) Qiu, X.; Fairbanks, A. J. Scope of the DMC mediated glycosylation of unprotected sugars with phenols in aqueous solution. *Org. Biomol. Chem.* **2020**, *18*, 7355-7365.

(15) Qiu, X.; Fairbanks, A. J. Direct synthesis of para-nitrophenyl glycosides from reducing sugars in water. *Org. Lett.* **2020**, *22*, 2490-2493.

(16) Luis, A. S.; Jin, C.; Pereira, G. V.; Glowacki, R. W. P.; Gugel, S. R.; Singh, S.; Byrne, D. P.; Pudlo, N. A.; London, J. A.; Basle, A.; Reihill, M.; Oscarson, S.; Evers, P. A.; Czjzek, M.; Michel, G.; Barbeyron, T.; Yates, E. A.; Hansson, G. C.; Karlsson, N. G.; Cartmell, A.; Martens, E. C. A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* **2021**, *598*, 332-337.

(17) Helbert, W.; Poulet, L.; Drouillard, S.; Mathieu, S.; Liodice, M.; Couturier, M.; Lombard, V.; Terrapon, N.; Turchetto, J.; Vincentelli, R.; Henrissat, B. Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. *Proc. Natl. Acad. Sci. U S A* **2019**, *116*, 10184-10185.