Electronic Supplementary Material (ESI) for ChemComm.

SUPPORTING INFORMATION

Structural and mechanistic insights into a *Bacteroides vulgatus* retaining *N*-acetyl-β-galactosaminidase that uses neighboring group participation

C. Roth,[a] M. Petricevic,[b] A. John,[c,d] E.D. Goddard-Borger,[c,d,*] G.J. Davies[a,*] and S. J. Williams[b,*]

Contents

Supplementary figures

(a)

>ENA|ABR39859|ABR39859.1 Bacteroides vulgatus ATCC 8482 conserved hypothetical protein

ATGAAAAAACTCCTATTTCTAGGCGCTCTCCTGTTGTCAACAGTATGTATGAACGCCCAAACTTCCGAATATTATCAGGA
AGCAGCCAATCCTATTGCCACCAACCCCGCTTTATGGGCGAAAGTTACAGCTCCACAAATCAGCTGGGGAAGCACAGACA
TCCGCTATAAGAAAGAGGAACCGGCTCCCATCCATAGTGCGCAGAAAAGTATGAATCTTACTGCATGGAAAGGTGAGAAA
ATTTCAGCCCAACTGGTAGTATGGACTCCCAAAGTGCTAAATGACTTGACTTTCATGGTCAGCGATTTAACCTCAGGCAG
TGCAACCATCAGTAAAGAGAATATCCGAACAGGCTTTGTCCGTTACGTTATAACCGATGAACTGAACAAAGATGGTTTGG
GCGCATGCGGCTATCGAAACAGTGCTGATTTTGACTCAACTTTAGTAGCAGATGTAATAGACCATATCACTCCTACTCTA
ACCCTTCCCGCCAACTCCACCCAAGGGGGATGGATCAGCGTAAACATCCCTCAGGGCACTAAAGCCGGGAAATACACAGG
AACCGTCACAGTGAAAGCCGACGGTATCACCTTGTCTGAATTAAAACTGAACCTCCAAGTGAAGAACCGTACTCTGCCTC
CTCCTTCCGAATGGGCTTTCCATCTGGATTTATGGCAGAACCCTTATGCAGTATCCCGTTACTACAATGTGGAACCGTTC
AGCAAAAAACATTTCGATTTGATGCGCCCATTGATGAAACTGTATGCCGATGCAGGTGGTAAGGTAATCACAGCCTCCAT
TATGCACAAGCCTTGGAACGGACAGACCTATGATGCTTTTGAAAGCATGGTCACTTGGTTGAAAAAAGCGGATGGAACCT
GGTATTTTGACTATACCGTATTCGACAAGTGGGTGGAATTTATGATGGATCTTGGTGTCAAAAAACAAATCAGTTGCTAT
TCTATGGTTCCCTGGCGCCTTTCTTTCCAATATTTTGATCAGGCCAGCAACTCTTTCAAATTTTTGGACGCCAAACCGGG
TGAAGTTGCTTATGAAGAATTTTGGATGAATATGCTGCAAGATTTCTCAAAGCATCTGAAAGCAAAAGGCTGGTTCGATA
TCACTCACATTGCGATGGACGAACGCCCGATGAAGGACATGCAGGAAACACTGAAAGTGATCCGTAAGGCTGATAAAGAC
TTTAAAGTTTCTTTGGCAGGAACTTATCACAAAGAACTATTGGATGATCTGAATGATTATTGTATCACCATTGCCGAGAA
ATTTACTCCCGAAGAGATTGAAGCGCGCCGGAAGGCAGGCAAAGTAACTACTTATTATACCTGCTGCACAGAGCCCCGCC
CCAACACTTTCACTTTCAGCGAACCTGCCGAAGCTGAATGGCTGGCATGGCATAGTGCGAAAGAAAATCTGGACGGATAT
CTTCGTTGGGCTTTAAACAGCTGGGTGAAAAATCCCCTACAAGACAGCCGTTTCACAGCTTGGGCTGCCGGAGACACGTA
TATGATTTATCCGGGCGCCCGTTCATCCATCCGGCTGGAACGCCTGACAGAAGGAATACAATTTTTTGAGAAAGTACGCA
TTCTGAAAGAAGAATTTGAAGAAAAAGGCAATAAAGGAGCTATTAAGAATATAGACAAAACCTTGAAAATGTTTGATGAA
TCAAGCATGGATAAGATTTCTCCTACCACTGCCGTAAACAAAGCAAAAAAAGTTATCAACCGATACTAG

Predicted translation:

MKKLLFLGALLLSTVCMNAQTSEYYQEAANPIATNPALWAKVTAPQISWGSTDIRYKKEE
PAPIHSAQKSMNLTAWKGEKISAQLVVWTPKVLNDLTFMVSDLTSGSATISKENIRTGFV
RYVITDELNKDGLGACGYRNSADFDSTLVADVIDHITPTLTLPANSTQGGWISVNIPQGT
KAGKYTGTVTVKADGITLSELKLNLQVKNRTLPPPSEWAFHLDLWQNPYAVSRYYNVEPF
SKKHFDLMRPLMKLYADAGGKVITASIMHKPWNGQTYDAFESMVTWLKKADGTWYFDYTV
FDKWVEFMMDLGVKKQISCYSMVPWRLSFQYFDQASNSFKFLDAKPGEVAYEEFWMNMLQ
DFSKHLKAKGWFDITHIAM**DE**RPMKDMQETLKVIRKADKDFKVSLAGTYHKELLDDLNDY
CITIAEKFTPEEIEARRKAGKVTTYYTCCTEPRPNTFTFSEPAEAEWLAWHSAKENLDGY
LRWALNSWVKNPLQDSRFTAWAAGDTYMIYPGARSSIRLERLTEGIQFFEKVRILKEEFE
EKGNKGAIKNIDKTLKMFDESSMDKISPTTAVNKAKKVINRY*

underlined = signal peptide

(b)

>His6-BvGH123:

MHHHHHHLEVLFQGPQTSEYYQEAANPIATNPALWAKVTAPQISWGSTDIRYKKEE
PAPIHSAQKSMNLTAWKGEKISAQLVVWTPKVLNDLTFMVSDLTSGSATISKENIRTGFV
RYVITDELNKDGLGACGYRNSADFDSTLVADVIDHITPTLTLPANSTQGGWISVNIPQGT
KAGKYTGTVTVKADGITLSELKLNLQVKNRTLPPPSEWAFHLDLWQNPYAVSRYYNVEPF
SKKHFDLMRPLMKLYADAGGKVITASIMHKPWNGQTYDAFESMVTWLKKADGTWYFDYTV
FDKWVEFMMDLGVKKQISCYSMVPWRLSFQYFDQASNSFKFLDAKPGEVAYEEFWMNMLQ
DFSKHLKAKGWFDITHIAM**DE**RPMKDMQETLKVIRKADKDFKVSLAGTYHKELLDDLNDY
CITIAEKFTPEEIEARRKAGKVTTYYTCCTEPRPNTFTFSEPAEAEWLAWHSAKENLDGY
LRWALNSWVKNPLQDSRFTAWAAGDTYMIYPGARSSIRLERLTEGIQFFEKVRILKEEFE
EKGNKGAIKNIDKTLKMFDESSMDKISPTTAVNKAKKVINRY

Underlined = hexahistidine tag

Figure S1: Sequence of *Bv*GH123.

(a) Predicted protein sequence of *Bv*GH123. (b) Recombinant *Bv*GH123 protein sequence, consisting of an N-terminal hexahistidine tag fused to the catalytic domain without signal peptide. The general acid/base and stabilizer residues are in bold.

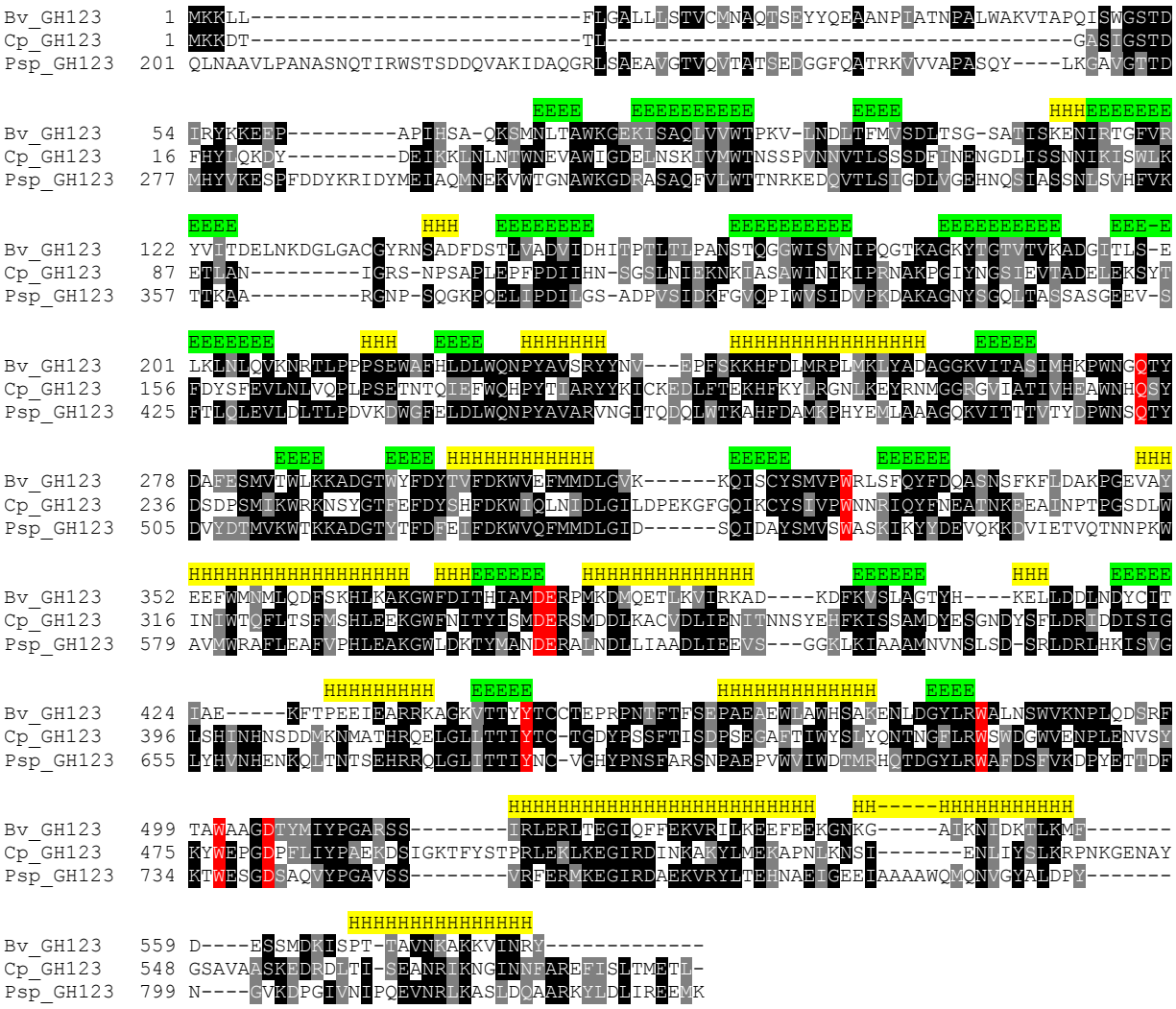
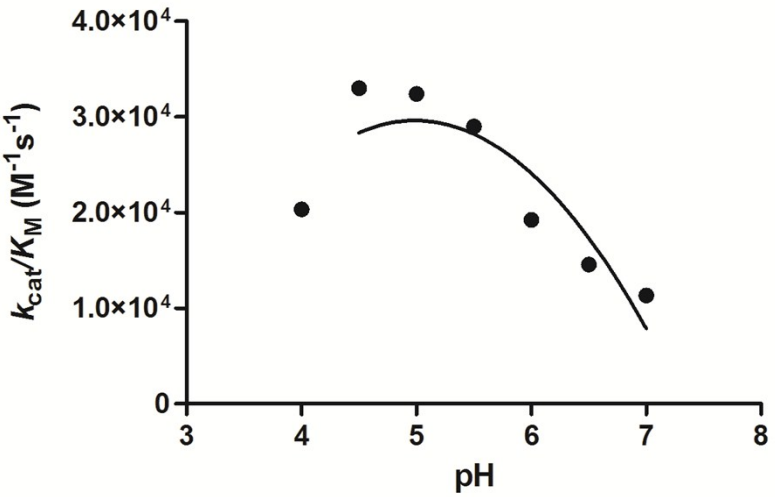


Figure S2: Sequence alignment of biochemically-characterized family GH123 sequences.

The characterized GH123s from *Bacteroides vulgatus* ATCC 8482 (ABR39859.1), *Clostridium perfringens* ATCC 13124 (ABG82546.1) and *Paenibacillus* sp. TS12 (BAJ83606.1) where aligned using ClustalW. The secondary structure of the *B. vulgatus* enzyme, as determined by the structure reported in this work, is shown above the alignment. Residues interacting with the substrate are highlighted in red.

(a)



(b)

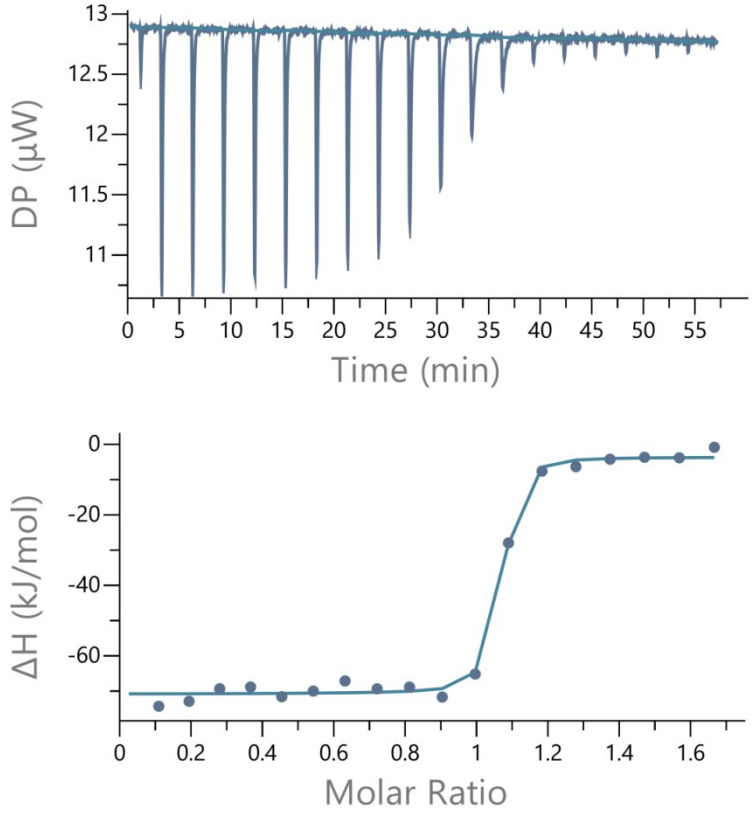


Figure S3: pH activity and inhibitor binding to *Bv*GH123.

(a) pH dependence of activity for *Bv*GH123 using pNPGalNAc as substrate. k_{cat}/K_M values were determined by the substrate depletion method. See the Experimental section for further details. (b) Isothermal titration calorimetry of Gal-thiazoline binding to *Bv*GH123.

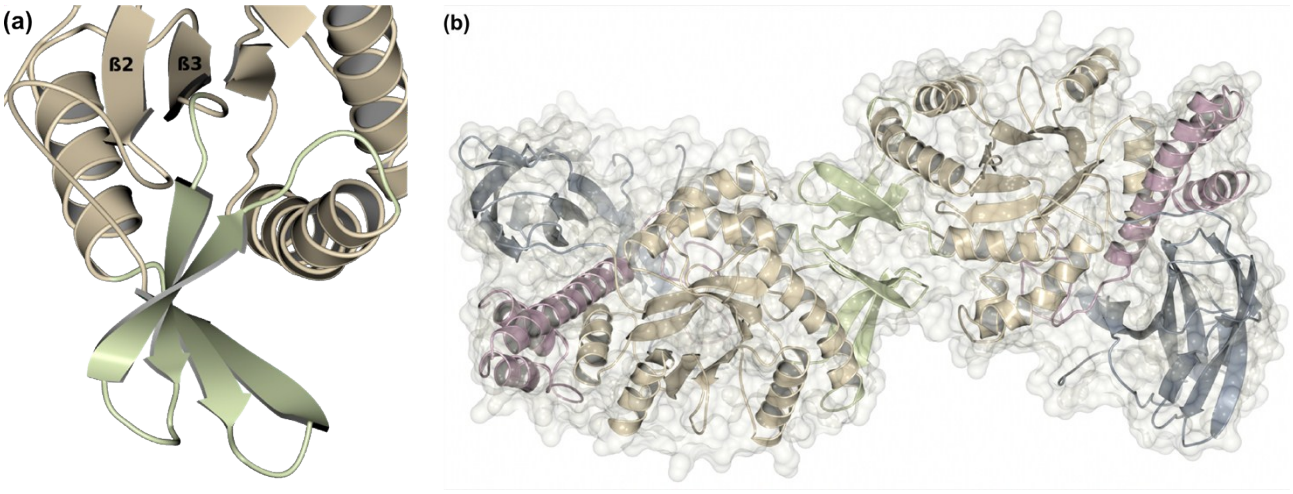


Fig. S4 Dimer of *Bv*GH123 as observed in the crystal structure.

(a) β-wing domain formed by insertion of a 4 stranded β-sheet between the β3 and 4 of the $(\beta\alpha)_8$-barrel. (b) Crystallographic dimer observed for *Bv*GH123

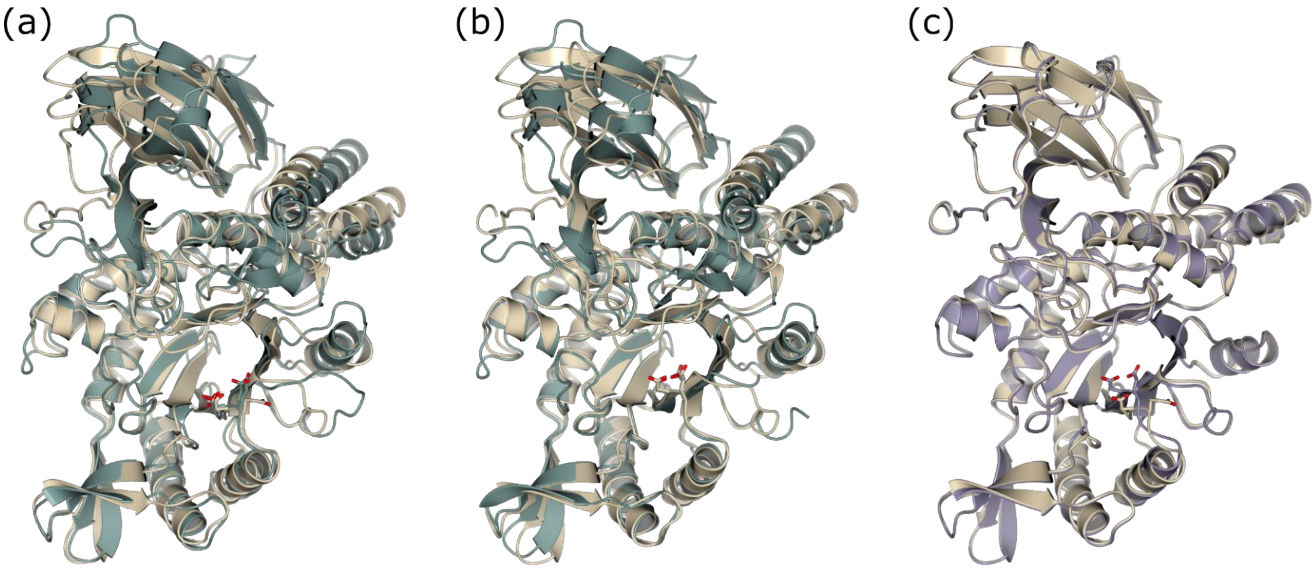


Fig. S5 Overlays of *Bv*GH123 in open and closed conformations, and with *Cp*Nga123.

Overlay of open (a) and closed (b) conformations.of *Bv*GH123 (wheat) and *Cp*Nga123 (slate gray). (c) Overlay of open (wheat) and closed (plum) form of *Bv*GH123. The active site residues D361 and E362 are shown in stick representation. The overlay was based on the TIM-barrel domains.

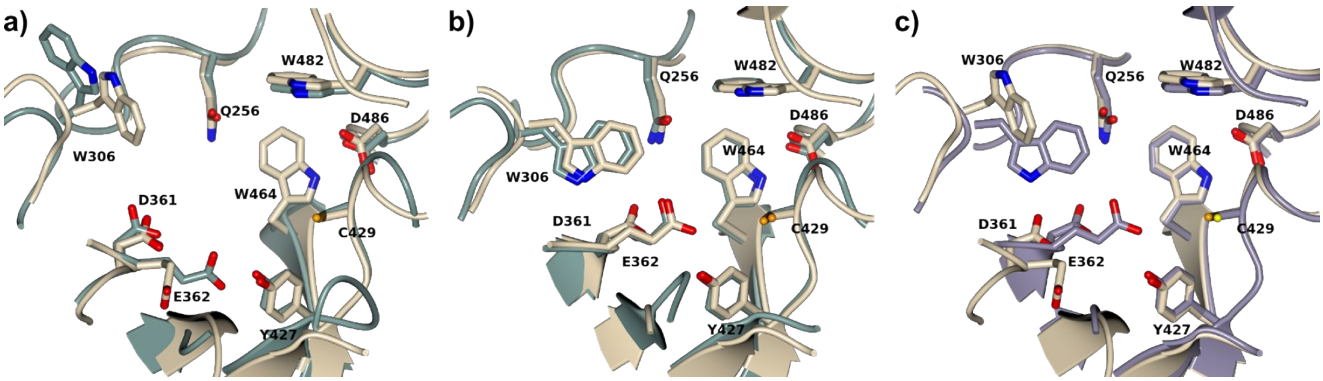


Fig. S6 Zoom of overlay of *Bv*GH123 and *Cp*Nga123 in open and closed conformations.

Overlay of open (a) and closed (b) conformations.of *Bv*GH123 (wheat) and *Cp*Nga123 (slate gray). (c) Overlay of *Bv*GH123 in open(wheat) and closed (slate gray).

Experimental

Cloning and expression

The gene of *Bv*GH123, encoding the mature protein sequence, was amplified from genomic DNA of *B. vulgatus* DSM 1447 and cloned in YSBL3CLIC creating a construct with an N-terminal cleavable His-Tag[1]. Protein expression was carried out in TB-medium. For the production of selenomethionine containing protein the gene was transformed in *E. coli* B384 and the cells were grown in minimal medium, using autoinduction to induce protein synthesis[2,3]. *Bv*GH123 was purified in a two-step procedure using IMAC and then gel filtration. Fractions containing *Bv*GH123, eluting as a dimer, were combined and concentrated to approx 25 mg/ml and stored at –80 °C.

Protein crystallography

Initial crystallisation conditions were established using commercial sparse matrix screens from Hampton Research and Molecular Dimensions. Subsequently crystals were optimised in 48 well screen sitting drop with a protein concentration of 7.5 mg/ml.

Data were collected at the Diamond Light source (Didcot UK). Data were indexed and integrated using XDS[4] as part of Xia2[5] and subsequently scaled using AIMLESS[6]. The experimental phasing was done using the Crank2 pipeline within i2 of the CCP4 software package[7,8,9]. The partial structure was then further refined against a higher resolution native dataset. The model was further improved by alternating cycles of manual model building in real space with Coot and reciprocal space refinement with Refmac5[10,11]. The quality of the final models was judged using Molprobity[12].

Determination of stereochemistry of catalysis by ^{1}H NMR analysis

*Bv*GH123 catalysed hydrolysis of pNPGalNAc was monitored by ^{1}H NMR spectroscopy using a 500 MHz instrument. A solution of *Bv*GH123 in buffered D_2O (0.15 ml, 0.095 mM in 10 mM sodium phosphate, pD 7.4) was added to a solution of pNPGalNAc (4.0 mg, 19.3 mmol) in buffered D_2O (0.6 ml, 10 mM sodium phosphate, pD 7.4) at 25 °C. ^{1}H NMR spectra were acquired at time points (t = 0, 3, 22 and 38 min, 20 h).

Kinetic analysis of *Bv*GH123

pH dependence of activity. k_{cat}/K_M values for *Bv*GH123 were measured for pNPGalNAc hydrolysis using the substrate depletion method in a stopped assay. Reactions (500 μl) were performed in 50 mM citrate/phosphate buffer, 150 mM NaCl at a range of pH values (4.0, 4.5, 5.0, 5.5, 6.0, 7.0) at 37 °C. Reactions were initiated by the addition of 10 μl of 0.106 μM *Bv*GH123 to pNPGalNAc (0.4 mM) in buffer, and aliquots (25 μl) were quenched at different time points into 75 μl glycine buffer (1 M, pH 10.0) in a 96-well plate. Absorbances were measured using a UV/visible plate reader (λ = 405 nm). Data (absorbance at 400 nm against time) were fitted to a first order rate equation ($A_t = A_0$ (1-e^{-kt})+offset, where A_t is the absorbance at time t, A_0 is the absorbance at time 0 and k is the rate constant) using the Prism 6 software package (Graphpad Scientific Software), to give values of V_{max}/K_M, which was adjusted for the enzyme concentration to give k_{cat}/K_M. The k_{cat}/K_M values at different pH values were fitted to a bell-shaped ionisation curve (k_{cat}/K_M = (limit x $10^{(pH-pKa1)}$/($10^{(2xpH-pKa-pKa2)}$ + $10^{(pH-pKa1)}$+1) using Prism 6.

Michaelis-Menten kinetics. Kinetic parameters for *Bv*GH123 hydrolysis of pNPGalNAc and pNPGlcNAc, were measured using a Varian Cary50 UV/visible spectrophotometer to measure the release of 4-nitrophenol at the isosbestic point (λ = 348 nm). Reactions were performed in 50 mM sodium phosphate, 150 mM NaCl, pH 5 at 37 °C using a final concentration of 17.8 nM *Bv*GH123

at substrate concentrations ranging from 0.25 mM to 1.25 mM. Above this concentration the signal-to-noise deteriorated, suggesting the presence of incompletely dissolved substrate. The extinction coefficient for 4-nitrophenol under the assay conditions was determined to be 6172 M^{-1} cm^{-1}. Kinetic parameters were calculated using the Prism 6 software package (Graphpad Scientific Software).

Size exclusion chromatography-multi-angle laser light scattering

Multi angle light scattering was used to assess the oligomeric state of *Bv*GH123. A sample of *Bv*GH123 was applied on a Superdex 200 10/300 (GE Healthcare), pre-equilibrated with 10 mM HEPES pH 7.5, 250 mM NaCl, 1mM DTT, with a flow rate of 0.5 ml/min, using a Shimadzu HPLC system (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler). This was linked to a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector. The column was calibrated using BSA as standard

Isothermal titration calorimetry

The determination of the binding constant of Gal-thiazoline was carried out with an Auto-ITC 200 system from Malvern. A final concentration of 17.5 µM *Bv*GH123 in 50 mM MES pH 6, 150 mM NaCl was added to the calorimeter. Gal-thiazoline was dissolved at 147 µM and used to titrate the protein. Injections of 2 µl were used with a spacing of 180s. The K_D-value was calculated using the Malvern software. The measurements were carried out at 25 °C. The K_d-value was calculated using the Malvern analysis software. The Free Gibbs energy was determined to -46.1 kJ/mol with an enthalpic contribution of 67.8 kJ/mol and an slightly unfavourable entropic contribution of -21.7 kJ/mol

Table S1: *Bv*GH123 X-ray data collection, processing and refinement statistics.

	BvGH123SeMet	*Bv*GH123 apo	*Bv*GH123 GalNAc	*Bv*GH123 Gal-thiazoline
PDB-ID		5l7r	5l7u	5l7v
X-ray source	Diamond-IO3	Diamond-IO4	Diamond-IO2	Diamond-IO2
Wavelength [Å]	0.979420	0.979490	0.91841	0.979500
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Resolution limit [Å]	2.7	1.85	2.1	2.3
Unit cell parameters				
a [Å]	55.0	55.3	55.4	55.5
b [Å]	144.7	147.8	142.2	148.4
c [Å]	145.7	150.5	145.8	147.7
Solvent content [%]	51.2	53.9	50.7	53.5
Monomers per asu	2	2	2	2
Total reflections	213617(28161)	859298(42494)	445720(28436)	389448(17480)
Unique reflections	32604(4215)	106036(5148)	68163(4539	51335(2915)
I/σ(I)*	11.6(1.4)	9.9(1.2)	12.0 (1.2)	15.6(1.1)
completeness [%]	99.5(98.2)	99.9(100)	100(100)	93.4(63.6)
R_{merge}	0.105(1.421)	0.141(1.930)	0.104(1.530)	0.063(1.552)
R_{meas}	0.124(1.674)	0.151(2.056)	0.114(1.670)	0.068(1.692)
R_{pim}	0.066(0.877)	0.053(0.703)	0.044(0.664)	0.024(0.651)
CC(1/2)	0.997(0.561)	0.998(0.331)		0.999(0.434)
Refinement				
R_{cryst}		0.181	0.193	0.187
R_{free}		0.219	0.227	0.221
Number of				
Protein residues		1122	1111	1121
Ligand		0	2	2
Water		644	179	27
r.m.s.d.				
Bond (Å)		0.016	0.014	0.013
Angle (°)		1.701	1.592	1.626
Average B-factor (Å^2)				
Protein		35.5	52.4	81.1
Ligand			71.8	61.1
Water		37.3	41.5	60.4
Ramachandran plot (%)				
favored/ allowed/disallowed		98/1.9/0.1	97.6/2.2/0.2	97.8/2.2/0

Values in parentheses refer to the highest resolution shell

References

1. M. J. Fogg, A. J. Wilkinson, *Biochem. Soc. Trans.*, 2008, **36**, 771–775.

2. H. K. Sreenath, C. A. Bingman, B. W. Buchan, K. D. Seder, B. T. Burns, H. V. Geetha, W. B. Jeon, F. C. Vojtik, D. J. Aceti, R. O. Frederick, G. N. Phillips, Jr, B. G. Fox, *Protein Expr. Purif.*, 2005, **40**, 256–267.

3. F. W. Studier, *Methods Mol. Biol.*, 2014, **1091**, 17–32.

4. W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 125–132.

5. G. Winter, C. M. C. Lobley, S. M. Prince, *Acta Crystallogr. D Biol. Crystallogr.*, 2013, **69**, 1260–1273.

6. P. R. Evans, G. N. Murshudov, *Acta Crystallogr. D Biol. Crystallogr.*, 2013, **69**, 1204–1214.

7. N. . Collaborative Computational Project, *Acta Crystallogr. D Biol. Crystallogr.*, 1994, **50**, 760–763.

8. P. Skubák, N. S. Pannu, *Nat. Commun.*, 2013, **4**, 2777.

9. G. M. Sheldrick, *Acta Crystallogr. A*, 2008, **64**, 112–122.

10. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 486–501.

11. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 355–367.

12. V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 12–21.