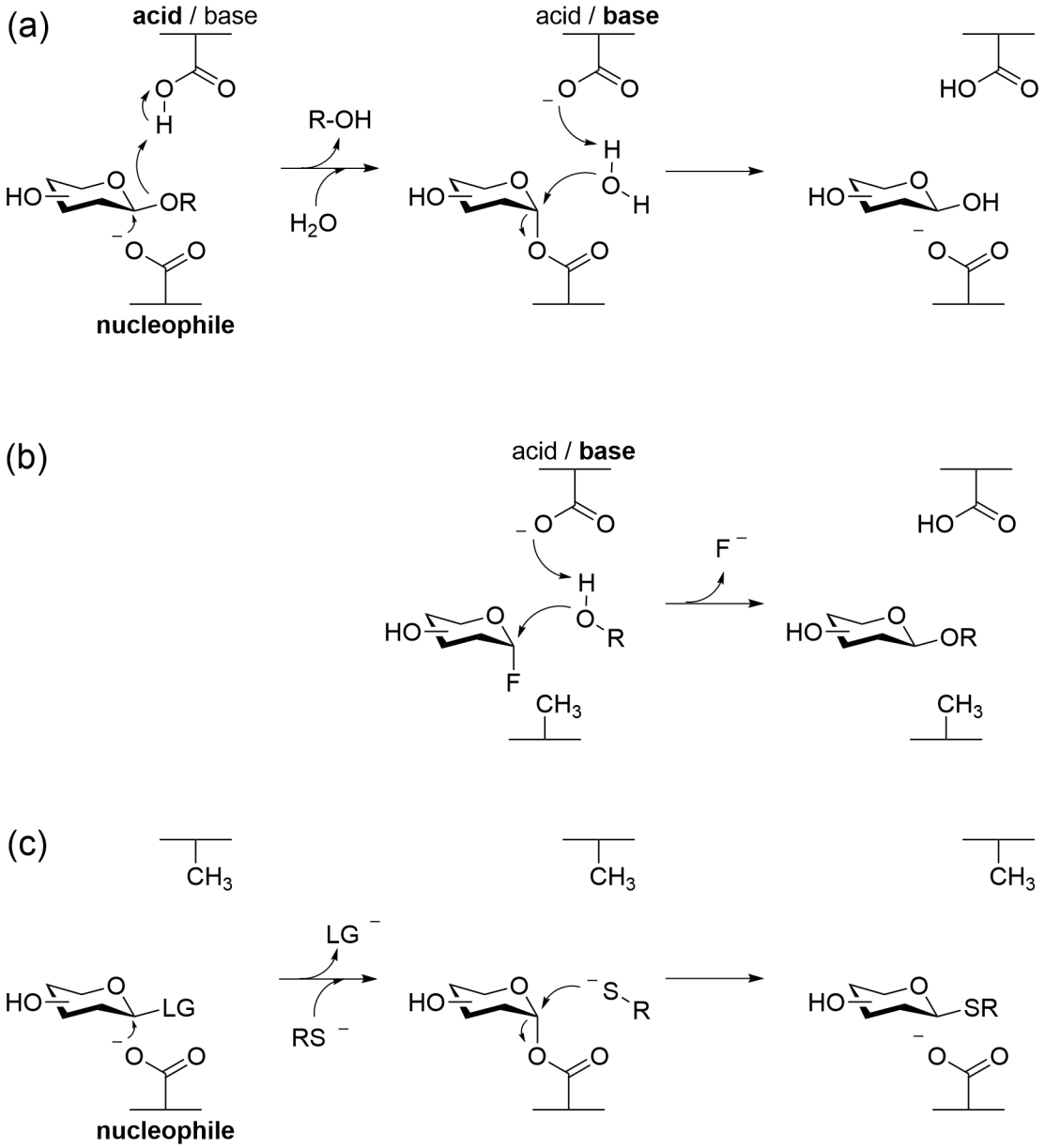


Figure S1: (a) Catalytic mechanism of a retaining glycosidase following a double-displacement mechanism, (b) nucleophile mutant of a retaining glycosidase acting as a glycosynthase using an activated donor and (c) an acid/base mutant of a retaining glycosidase acting as a thioglycoligase in the presence of an activated donor and a thiol acceptor.

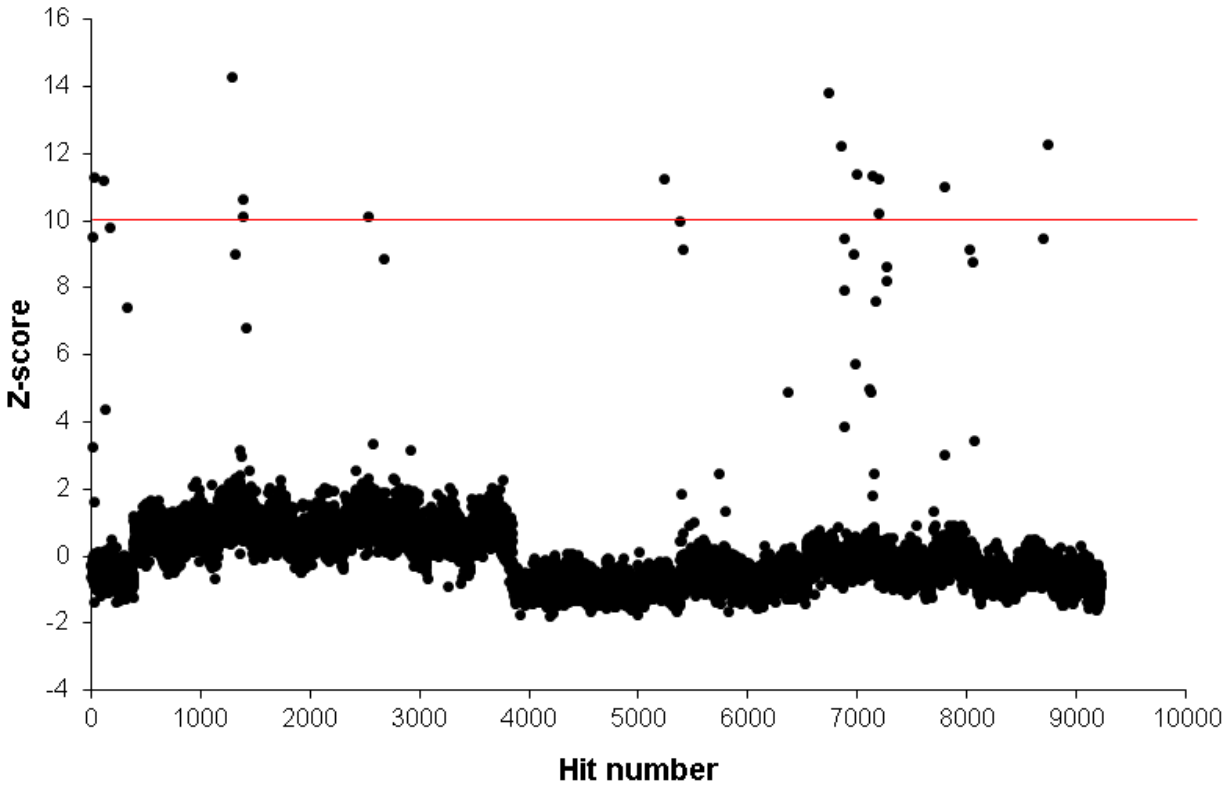


Figure S2: Z-score of the metagenomic library screen using β-GlcNAc-MU. The red line illustrates the chosen cut off at Z-score > 10; Z-score = (fluorescence – mean value) / standard deviation

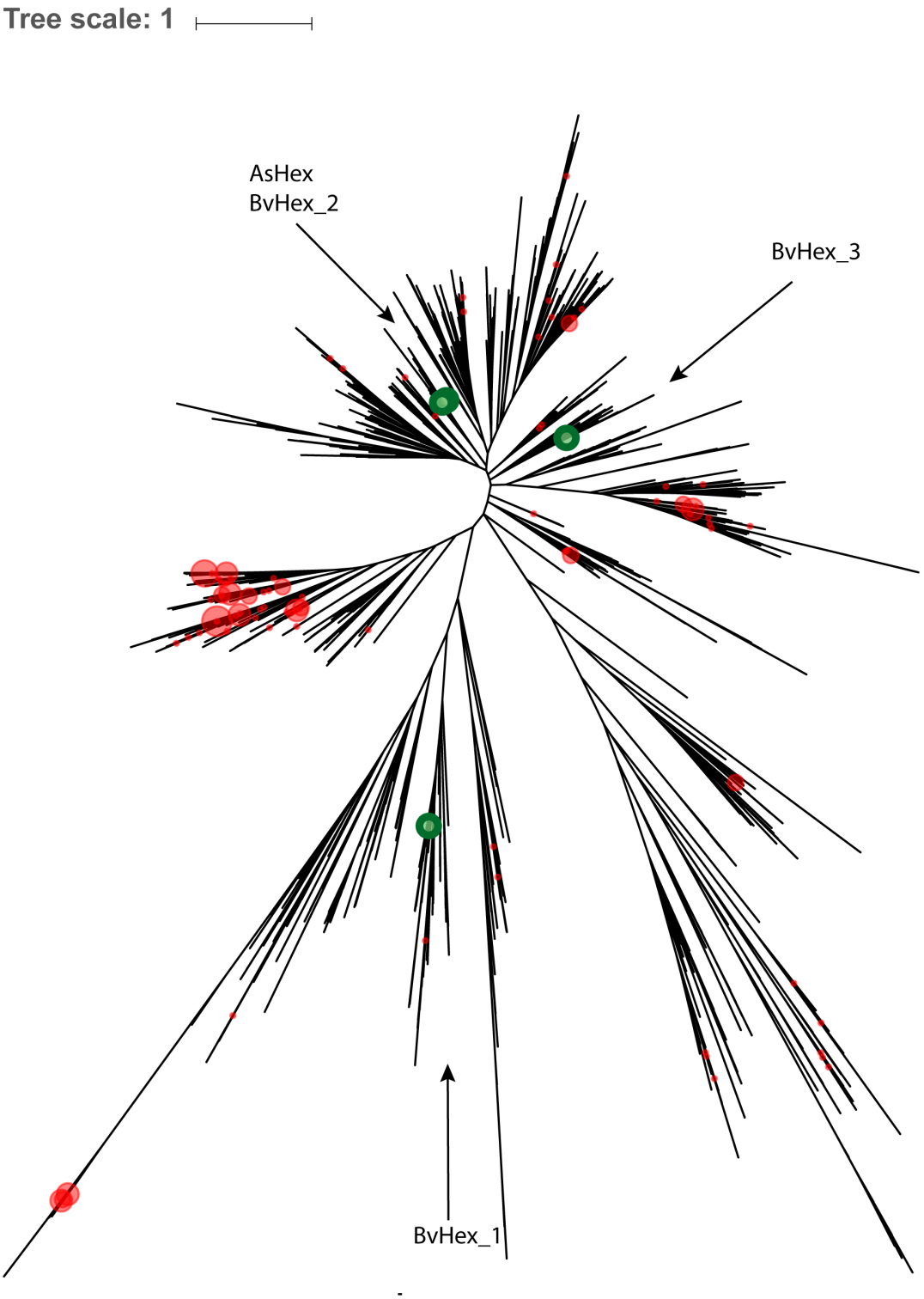


Figure S3: Phylogenetic analysis of the GH20 family. Red circles indicate characterized candidates of this family, according to data from the CAZY database. The HexNAcases identified within this study are marked by green circles.

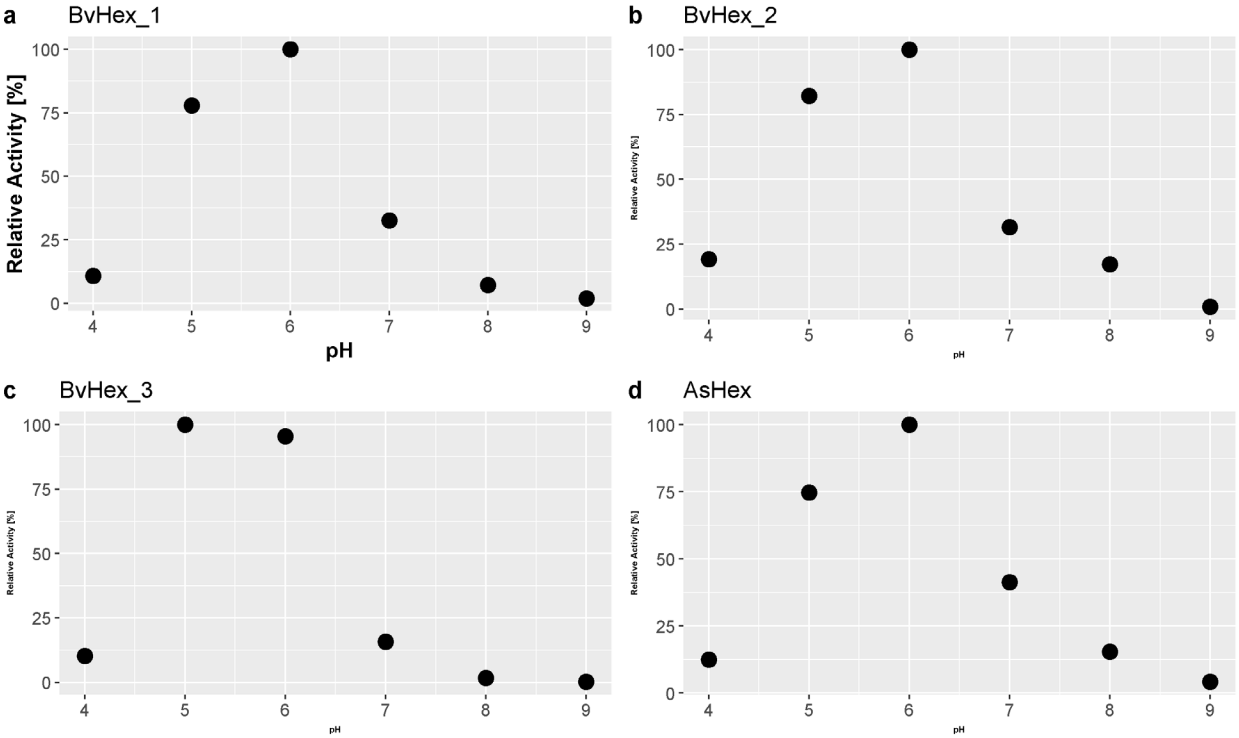
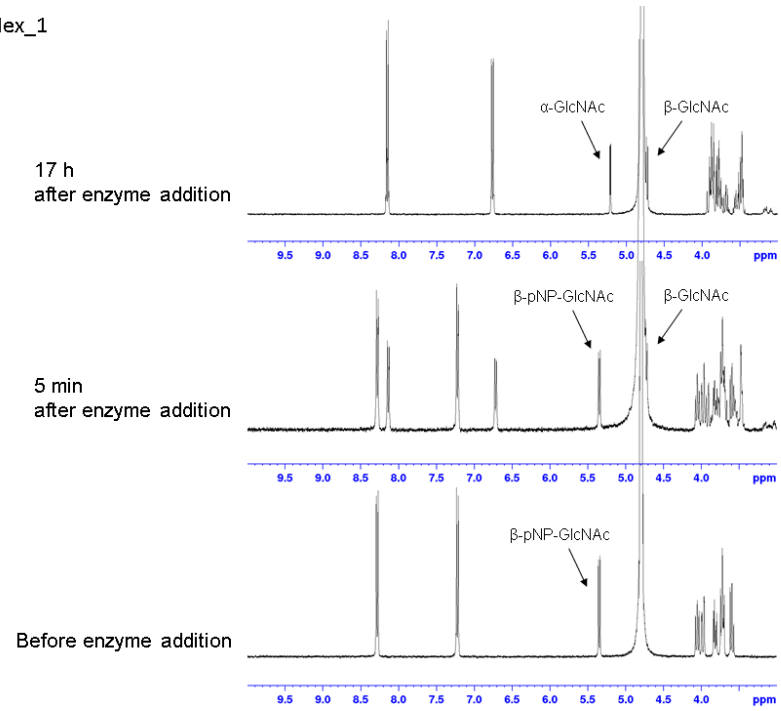
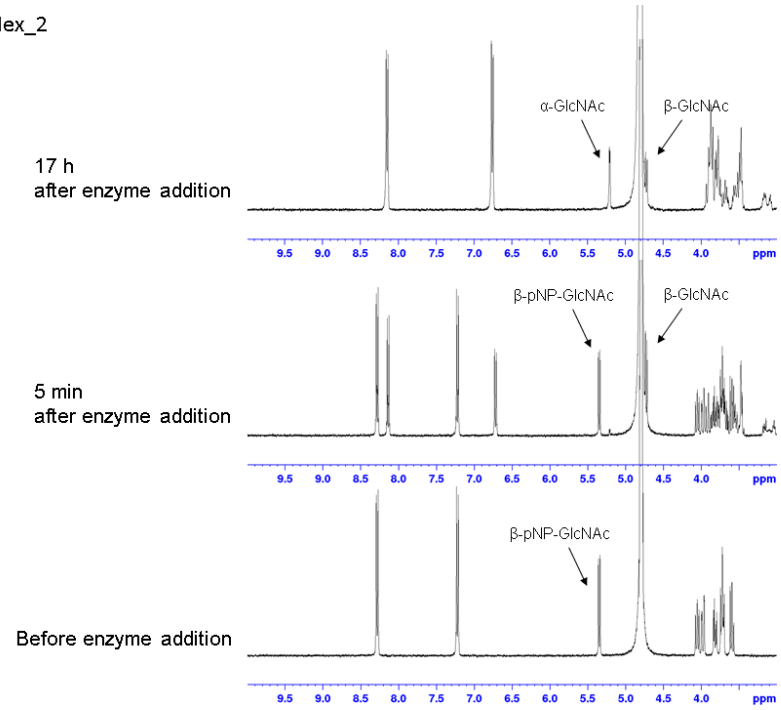


Figure S4: pH profiles of the HexNAcases

*Bv*Hex_1



*Bv*Hex_2



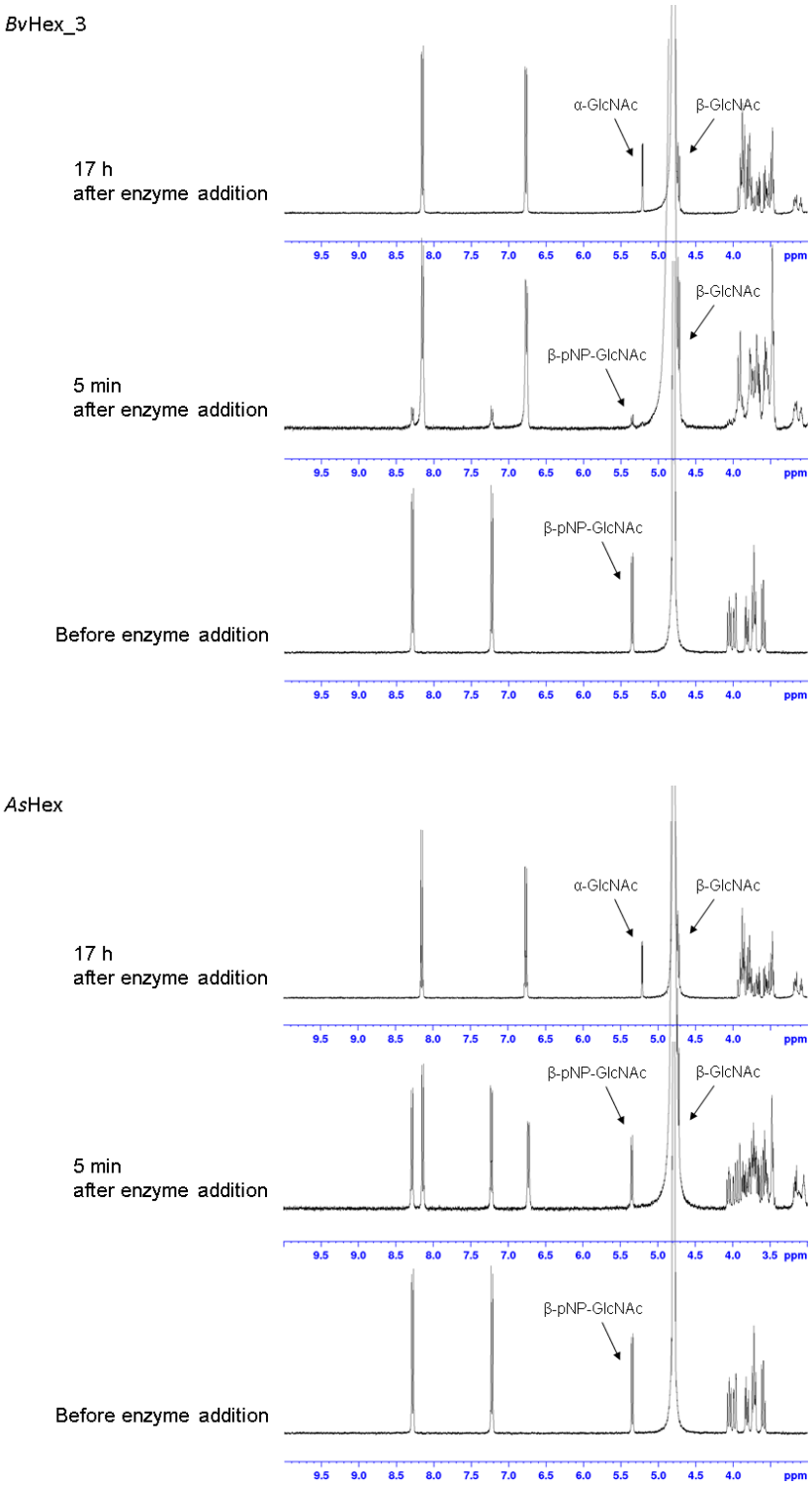


Figure S5: *In-situ* NMR analysis of the reaction mechanism of the HexNAcases, confirming a retaining mechanism for all four enzymes. After 5 min reaction time, a reducing β-GlcNAc peak appears proofing the formation of the respective β-anomer upon hydrolysis. After 17 h incubation, the α-anomer arises due to mutarotation.

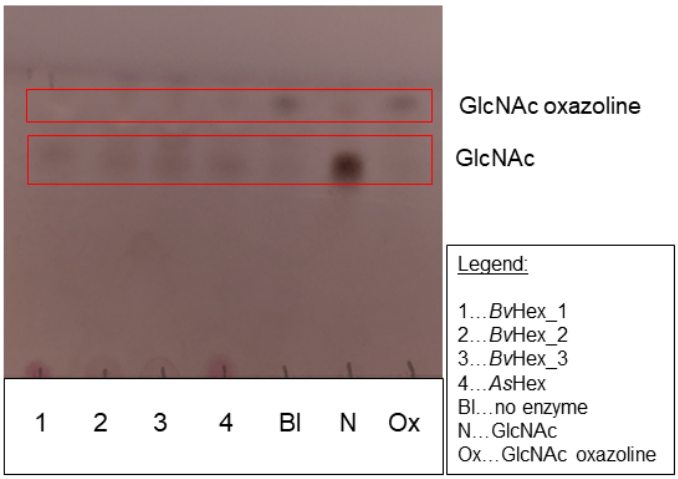


Figure S6: Enzymatic hydrolysis of GlcNAc oxazoline by the HexNAcases, analyzed by TLC (1 h reaction time)

Table S1: Kinetic parameters of the hexosaminidase wild type hexosaminidases on 6-S-β-GlcNAc-pNP and 6-N_3-β-GlcNAc-pNP (pH 7.0)

Wildtype		*Bv*Hex_1	*Bv*Hex_2	*Bv*Hex_3	*As*Hex
6-S-β-GlcNAc-pNP	K_m [mM]	-	0.94	0.68	1.3
	k_{cat} [s^{-1}]	-	13.7	0.09	1.4
	k_{cat}/K_m [$s^{-1}*mM^{-1}$]	-	14.5	0.13	1.07
6-N_3-β-GlcNAc-MU	K_m [mM]	0.86	2.1	0.97	0.70
	k_{cat} [s^{-1}]	0.006	0.01	0.0002	0.006
	k_{cat}/K_m [$s^{-1}*mM^{-1}$]	0.007	0.006	0.0002	0.009

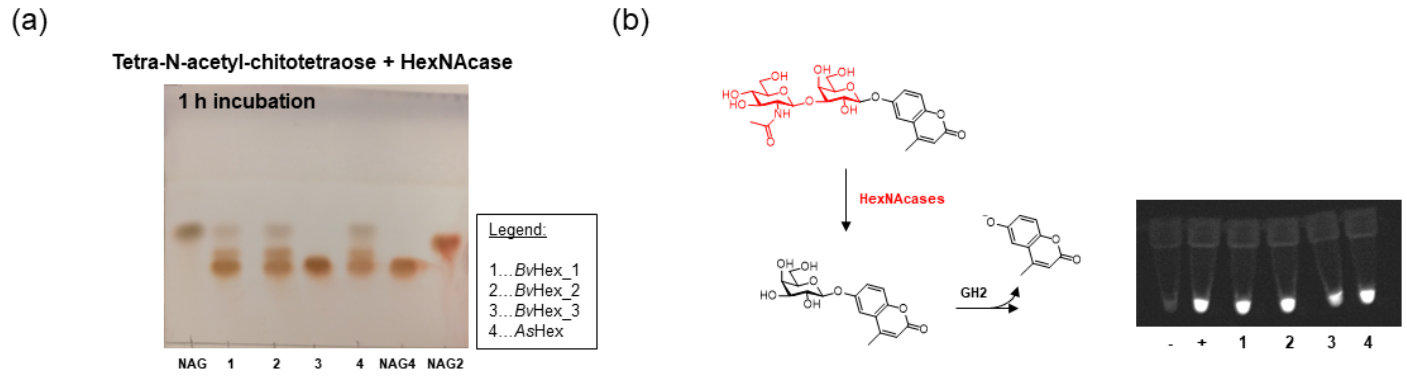


Figure S7: (a) Hydrolysis of a chitin tetramer (harboring a β-1,4 linkage) by the wild type HexNAcases, analyzed by TLC and (b) a coupled hydrolysis experiment, qualitatively analyzed detecting the released fluorophore, proving the hydrolysis activity towards β-1,3 linked GlcNAc; NAG = N-Acetylglucosamine, NAG2 = Chitobiose, NAG4 = Chitotetraose.

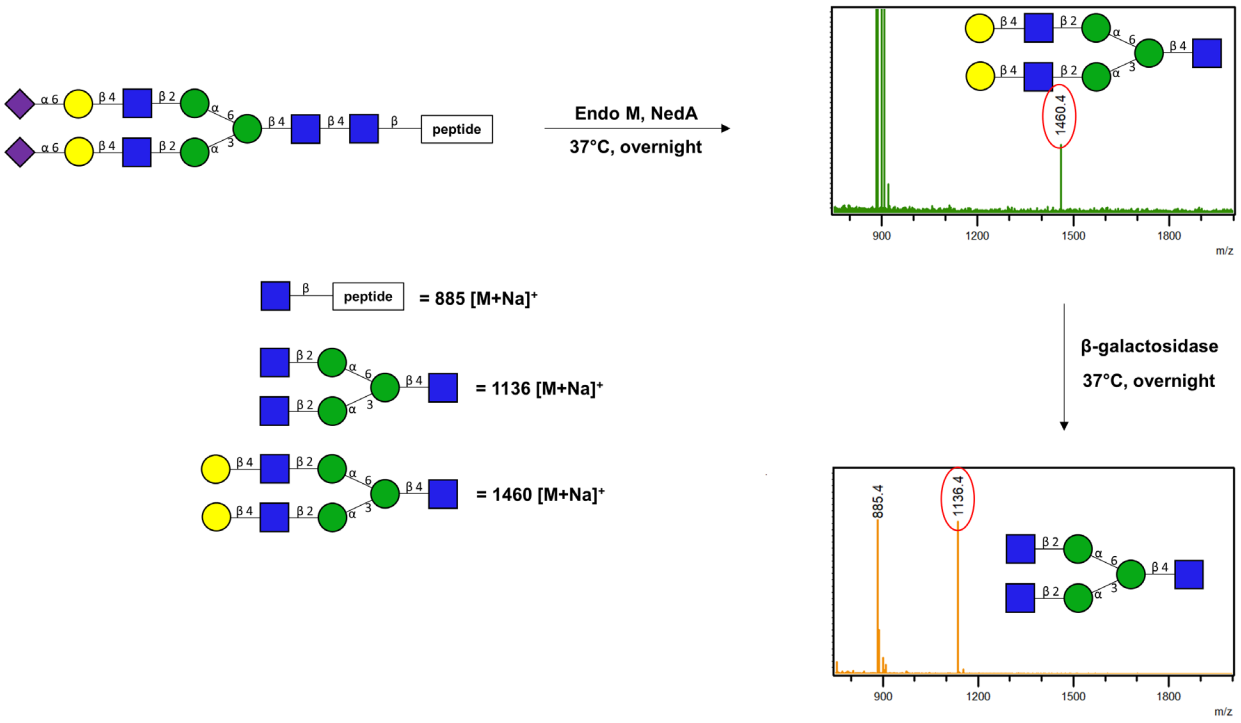


Figure S8: Trimming of sialoglycopeptides (SGPs) previously isolated from egg yolk. Cleavage from the peptide backbone was achieved by incubation with EndoM, terminal sialic acids were removed using NedA and the galactoses were cleaved off by an α-galactosidase from *A. niger*. Blue squares = GlcNAc; green circles = Man; yellow circles = Gal; purple squares = sialic acid

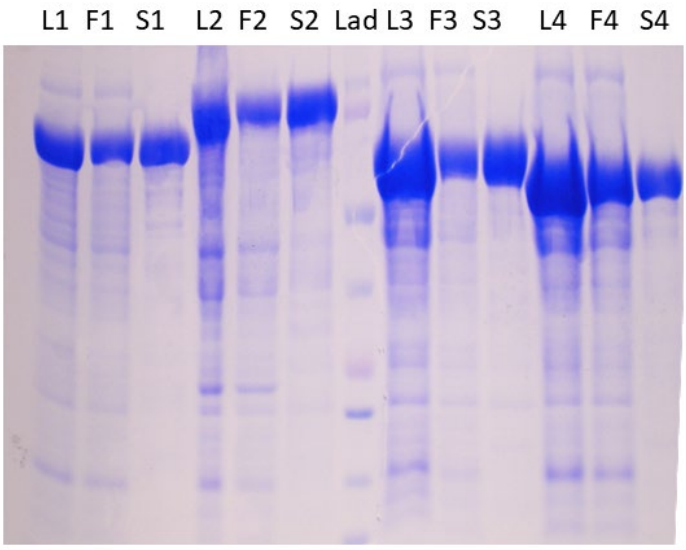


Figure S9: SDS PAGE of the HexNAcase expression and purification by affinity chromatography. L1-S1…BvHex_1; L2-S2…BvHex_2; L3-S3…BvHex_3; L4-S4…AsHex (L…lysate, F...flow through upon column loading, S…elution fraction). High expression levels resulted in an overloading of the nickel NTA column which resulted in substantial unbound protein eluting in the flow through fraction upon column loading.

Methods

Materials and reagents

All reagents and solvents were obtained from commercial suppliers and used without further purification. Analytical thin layer chromatography (TLC) was performed using Merck TLC silica gel 60-F254 plates and visualized by UV and a p-anisaldehyde stain. A antigen type 1_{penta} was a gift from D. Kwan.

Human feces metagenomic library screening

For the generation of the human metagenomics fosmid library fresh human fecal samples were collected from a healthy male with blood group AB+. Informed consent was obtained from the voluntary participant before his donation, based on the guidelines of the Clinical Research Ethics Board of the University of British Columbia. The direct DNA extraction and fosmid library creation was performed as described in Rahfeld et al..[3]

For the functional screening, 25 plates from the 51 x 384-well AB+Blood Fosmid library were randomly chosen, thawed at room temperature and replicated into 384-well plates containing 50 µl screening LB-media (12.5 µg/mL chloramphenicol, 25 µg/mL kanamycin, 100 ug/mL arabinose, 10 mM MgSO4). Plates were incubated at 37°C for 18 hours in a sealed container containing a reservoir of water to prevent excessive evaporation. Onto the grown screening plates, 45 µl of the reaction mixture (100 mM NaH_2PO_4, pH 7.4, 2%(v/v) Triton-X 100, 100 µM GlcNAc-α-MU) were added using the QFill instrument [Genetix]. The plates were then incubated at 37°C (in a sealed container) for 24 h, and the fluorescence (Ex: 365 nm Em: 435 nm, sweep-mode, gain 80) of each plate was measured at hours 1, 2, 4, 8 and 24 via a Synergy H1 plate reader [BioTek]. For all wells a Z-score was calculated, which is given by the formula: Z-score = (Fluorescence-median value)／Standard Deviation 31. All positive hits above a certain threshold (Fig. S2), were streaked out on an LB-Agar (12.5 µg/mL chloramphenicol) plate to obtain single colonies and re-screened in triplicate. All verified hits were stored separately at -70°C in LB-media (12.5 µg/mL chloramphenicol, 25 µg/mL kanamycin, 15%(v/v) glycerol, 10 mM MgSO4).

Fosmid hit sequencing, ORF prediction

To isolate the fosmid DNA for Sanger sequencing, the positive hit fosmid glycerol stocks were used to inoculate 5 mL of TB media (12.5 µg/mL chloramphenicol, 25 µg/mL kanamycin, 100 ug/mL arabinose, 10 mM $MgSO_4$), incubated overnight at 37 °C, 220 rpm. Fosmid isolation was performed using the GeneJet plasmid miniprep kit [Thermo Fisher]. Sanger sequencing was performed using the pCOS1FOS sequencing primers. The obtained sequences were annotated against the Refseq prokaryote representative genomes (Update date:2019/12/05) with Microbial Nucleotide BLAST. Only hits with forward and reverse sequences on the same bacterial genome were processed further, by detecting ORFs present on the nucleotide sequence, flanked by the primers, with ORF_finder (https://www.ncbi.nlm.nih.gov/orffinder/). All predicted ORFs with annotations to members of a GH family (with known or suspected β-N-acetylglucosaminidase activities) were cloned into pET28a plasmid using the Golden Gate cloning strategy 41, the primer sequences are noted in Supplementary Table S2.

GH20 reference package and phylogenetic mapping

GH20 protein sequences were downloaded from the CAZy database using SACCHARIS' cazy_extract.pl script.[4] TreeSAPP (available at https://github.com/hallamlab/TreeSAPP), was used to both build the reference trees and map the sequences to these trees. Briefly, treesapp create used the GH20 HMM from dbCAN version 7 to extract protein family domains from all full-length sequences downloaded from CAZy.[5,6] These sequences were then clustered at 75% sequence similarity using UCLUST to remove redundant sequences.[7] RAxML version 8.2.12 was used to build the reference trees with the '--autoMRE' flag and PROTGAMMAAUTO to select the optimal amino acid replacement model.[8,9]

Treesapp assign was then used to map the query sequences onto the reference tree and taxonomically classify them. Protein sequences were aligned to HMMs using hmmsearch and the aligned regions were extracted.[10] hmmalign was used to include the new query sequences in the reference multiple alignment before BMGE removed the unconserved positions from the alignment file.[11] RAxML's evolutionary placement algorithm was used to classify the query sequences in the reference tree through insertions.[12] Placements of each query sequence were filtered and concatenated into a single. Jplace file before being visualized in iTOL.[13,14]

Protein expression and purification

All HexNAcases were cloned into pET28a vector via Golden Gate cloning.[15] The primer sequences used are summarized in Table S2. The proteins were expressed in *E. coli* BL21(DE3) shaking (220 rpm) for 16 h at 30°C, using 50 mL autoinduction medium in the presence of 50 µg/mL kanamycin. The expression cultures were inoculated with 500 µL of overnight cultures. Cells were harvested by centrifugation at 6000 rpm for 15 min at 4 °C in a Beckman Coulter Avanti® J-E floor centrifuge (JA-10 rotor). The cell pellets were resuspended in 5 mL of lysis buffer (25 mM HEPES, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM 2-mercaptoethanol, 0.625 U/mL benzonase, 10 mg/mL $MgCl_2$, and 1 Pierce Protease Inhibitor mini tablet). Afterwards the cells were lysed by sonication and the cell debris was removed by centrifugation (15000 rpm, 30 min, 4°C, JA-20 rotor). The soluble fraction was purified by immobilized nickel affinity chromatography (1 mL). The target proteins were eluted using 3 mL HEPES buffer (25 mM HEPES, pH 8.0, 300 mM NaCl, 1 mM 2-mercaptoethanol) containing 500 mM imidazole. Fractions were analyzed by SDS- PAGE; protein containing fractions were concentrated and exchanged to the same HEPES buffer lacking imidazole, using an Amicon® Ultra-4 MWCO10-kDa centrifugal filter (Merck).
The C-terminal half of tau (244-441) was expressed as previously described.[16]
Active site mutants were generated using the QuikChange protocol using the primers noted in Table S2 below[17], expressed and purified as described above.

Table S2: Primer sequences.

Name	Sequence
Primer for isolating target genes from fosmids	
*Bv*Hex_1_fw	ATGGTCTCGCCATAAGGAAACATGTCCGAAAGTAATTC
*Bv*Hex_1_rv	ATGGTCTCGATTCTTATTGCGTGCAATTTATATTCCGC
*Bv*Hex_2_fw	ATGGTCTCGCCATGAAGTGAAAGAAGCAAATTATCAG
*Bv*Hex_2_rv	ATGGTCTCGATTCTCAATTTAATGAAATTTCATCAACAAAC
*Bv*Hex_3_fw	ATGGTCTCGCCATACGAAAGAAGAGCAACGCACAC
*Bv*Hex_3_rv	ATGGTCTCGATTCTTATTTCACTACAATTTCATCTGTAAAC
*As*Hex_fw	ATGGTCTCGCCATGCCGGTCCCGAGGCG
*Bv*Hex_rv	ATGGTCTCGATTCTTATTCGTCTTTGCCGTCGAAAAAG
Primer of QuikChange mutagenesis	
*Bv*Hex_1 D293A_fw	CATATCGGTACAGCGGAGTACAATGCCAAG
*Bv*Hex_1 D293A_rv	CTTGGCATTGTACTCCGCTGTACCGATATG
*Bv*Hex_2 D310A_fw	CATGTAGGTGGTGCGGAATGTCCTAAGG
*Bv*Hex_2 D310A_rv	CCTTAGGACATTCCGCACCACCTACATG
*Bv*Hex_3 D301A_fw	CATATAGGAGGGGCGGAGGCAGGTAAAG
*Bv*Hex_3 D301A_rv	CTTTACCTGCCTCCGCCCCTCCTATATG
*As*Hex D300A_fw	CACATCGGCGGCGCGGAATGTCCGAAAG
*As*Hex D300A _rv	CTTTCGGACATTCCGCGCCGCCGATGTG
*Bv*Hex_1 E294A_fw	CATATCGGTACAGACGCCTACAATGCCAAGG
*Bv*Hex_1 E294A_rv	CCTTGGCATTGTAGGCGTCTGTACCGATATG
*Bv*Hex_2 E311A_fw	GTAGGTGGTGACGCGTGTCCTAAGGTAC
*Bv*Hex_2 E311A_rv	GTACCTTAGGACACGCGTCACCACCTAC
*Bv*Hex_3 E302A_fw	CATATAGGAGGGGATGCCGCAGGTAAAGCC
*Bv*Hex_3 E302A_rv	GGCTTTACCTGCGGCATCCCCTCCTATATG

*As*Hex E301A_fw	CATCGGCGGCGACGCGGTGTCCGAAAGAG
*As*Hex E301A _rv	CTCTTTCGGACACCGCGTCGCCGCCGATG
*Bv*Hex_1 Y372F_fw	GTGATGCTTATCTTTTCATTGTTCCGGCTGCCGG
*Bv*Hex_1 Y372F_rv	CAGCCGGAACAATGAAAAGATAAGCATCACACGTATTG
*Bv*Hex_1 Y372N_fw	GTGATGCTTATCTTAACATTGTTCCGGCTGCCGG
*Bv*Hex_1 Y372N_rv	CAGCCGGAACAATGTTAAGATAAGCATCACACGTATTG
*Bv*Hex_2 Y408F_fw	CTAATACTTATTTGTTCTTCGACTATTACCAAACC
*Bv*Hex_2 Y408F_rv	GTAATAGTCGAAGAACAAATAAGTATTAGGAG
*As*Hex Y392N_fw	GAACGATTTCTGCAACCTGGACTACTATCAGACG
*As*Hex Y392N_rv	GATAGTAGTCCAGGTTGCAGAAATCGTTCGGCG

NMR analysis of the stereochemical course of β-GlcNAc-pNP hydrolysis

β-GlcNAc-pNP was dissolved in DMSO-d6 and diluted with a deuterated 1x PBS buffer to give 4 mM concentration of β-GlcNAc-pNP and 5% DMSO-d6. Aliquots of the concentrated HexNAcases (~20 mg/mL) were dried on a SpeedVac concentrator and afterwards dissolved in deuterated 1xPBS buffer, using maximum one half of the initial volume to further concentrate the protein and account for eventual activity losses. Initial spectra were recorded with 600 µL of the buffered β-GlcNAc-pNP before initiating the reaction by the addition of 10 µL enzyme solution. ^{1}H-NMR spectra were recorded at regular intervals. Each spectrum was acquired with 16 scans on a Bruker AV-400 MHz spectrometer.

Kinetic analysis

Kinetic studies were performed at 37°C in citrate/phosphate/glycine buffer (25/25/25 mM, 100 mM NaCl, pH 7.0). Initial rates (<20% reaction) of enzyme-catalyzed hydrolysis of β-GlcNAc-pNP were determined on a Synergy H1 plate reader (BioTek) at 405 nm in a 96 well plate (half area) with 50 µL of solution per assay. Absorbance values were converted to concentrations using the respective extinction coefficients of para-nitrophenol in the buffer. Michaelis-Menten parameters (V_{max} and K_m) were extracted from these data sets by non-linear fitting to the Michaelis-Menten equation.
The pH profile of the HexNAcases was determined using β-GlcNAc-pNP, recording initial rates at 1.0*Km at 37°C. The same buffer system as mentioned above was used for these experiments, covering a pH range from pH 4 to 9 and the release of pNP was detected at 405 nm.
The inhibition constant (Ki) was determined by incubating the respective enzyme with eight different inhibitor concentrations, bracketing the Ki value, for 5 min before the hydrolysis reaction was initiated adding the substrate. Two different substrate concentrations (<Km and >Km) were used and the initial rates were measured monitoring the increase in absorbance at 405 nm at pH 7.0. Dixon plots (1/v vs [I]) were constructed to validate the use of competitive inhibition model and to assess the fit of the rest of the data. A competitive inhibition model was fit to the data by using non-linear regression analysis with Grafit 7.0.

Exploration of the substrate specificity

Hydrolysis experiments on GlcNAc oxazoline and tetra-N-acetyl chitotetraose:
GlcNAc oxazoline was synthesized as previously reported.[18] 20 mM GlcNAc oxazoline or 15 mM tetra-N-acetyl chitotetraose: was incubated with 0.1 mg/mL of the respective HexNAcase in a HEPES buffer (50 mM HEPES, 100 mM NaCl, pH 7.0) at room temperature. Substrate hydrolysis was followed by TLC (5 BuOH: 4 MeOH: 4NH_4OH: 1 H_2O; p-anisaldehyde stain).

Hydrolysis of GlcNAc-β-1,2-Man bond in sialoglycopeptides (SGPs):
SGPs were isolated and purified from egg yolk powder following a procedure from Liu et al[19]. Pure SGP were dissolved in 1xPBS buffer pH 7.0 (7 mg/mL) and enzymatically trimmed in order to access GlcNAc-β-1,2-Man which which initiates the antennae of complex N-glycans. The SGP solution was incubated with 0.4 mg/mL EndoM at 37° overnight, cleaving of the glycan from the peptide by hydrolyzing the chitobiose core of the N-

glycan part. After MALDI-TOF MS analysis (Bruker Autoflex) indicated complete substrate hydrolysis, enzyme was removed using an Amicon Ultra centrifugal filters with a MWCO of 10 kDa (Merck). The filtrate was freeze dried, taken up in a sodium acetate buffer (50 mM, 5 mM $CaCl_2$, pH 5.5) and the pH adjusted to pH 5.5, if necessary.

NedA was added (75 µg) to the glycan solution and incubated for 48 h at 37°C to remove the terminal sialic acids. After completion of the reaction, the enzyme was removed by ultra-centrifugation. The pH was adjusted to 4.5, BSA was added (final concentration 1 mg/mL) a beta-galactosidase from *A. niger* (120 U, Sigma) was added. The mixture was incubated at 37°C overnight until complete removal of galactose was observed, and the enzyme was removed by ultra-centrifugation. The filtrate, containing the trimmed N-glycans with terminal GlcNAc β-1,6 linked to Man, was adjusted to pH 6.0 and incubated individually with 0.5 mg/mL each of the HexNAcases. After incubation at 37°C for 2 h, the reactions were monitored by MALDI-TOF MS analysis.

Hydrolysis of GlcNAc-β-1,3-Gal found in the methylumbelliferyl type 1 A blood group pentasaccharide MU-Type1A_{penta} (A antigen type 1_{penta}):

The activity of the HexNAcases on the internal GlcNAc-β-1,3-Gal linkage was determined using a modified version of a coupled enzyme assay that was described by Kwan *et al*.[20] A deacetylase (*Fp*GalNAcDeAc) (50 µg/mL),an α-galactosaminidase *(Fp*GalNase), an α-fucosidase (AfcA) and a β-galactosidase (BgaC) will (0.05 mg/ml each) trim the A antigen type 1_{penta}, resulting in a terminal GlcNAc β-1,3-linked to Gal-MU. The reaction mixture contained 0.5 mg/mL of the respective HexNAcase and MU release was initiated by BgaC only after cleaving the terminal GlcNAc from the remaining glycan. Release of MU was qualitatively monitored.

Thioglycoligation reactions

Thioglycoligase reactions were performed at pH 7.0 in HEPES buffer (50 mM HEPES, 10 mM DTT) and 37°C. Glycosylation donors and acceptors were dissolved in DMSO and final DMSO concentrations in the assay did not exceed 10%. The final assay concentrations were 10 mM donor, 5 mM thiol acceptor (thio sugars and cysteine ethyl ester) and 0.2 mg/mL of the respective mutant enzyme. The model peptides Tab1 (VPYSSAQS) and the cysteine mutant thereof, Tab1-Cys (VPYCSAQS) were synthesized and purified following a previously published procedure.[16] Thioglycoligations using the peptides as acceptors were conducted at final assay concentrations of 8 mM donor, 0.5 mg/mL of the respective peptide and 0.2 mg/mL of the respective mutant enzyme. A previously reported triple mutant of Tau was used as model protein for S-GlcNAcylation and produced as published.[16] Final assay concentrations were 8 mm donor, 12 µM of the Tau triple mutant and 0.2 mg/mL of the respective mutant enzyme. Reactions using thio sugar acceptors were monitored by TLC ((5 EtOAc: 2 MeOH: 1 H_2O), assays involving peptides were monitored by MALDI-TOF MS using a Super DHB matrix (dissolved in 0.1% TFA), whereby 1 µL of a 1:1 mixture of matrix and reaction solution was spotted and dried on a MALDI sample plate. MALDI-TOF-MS analysis allowed a semi-quantitative, comparative assessment of the product yield. Provided reactions yields refer to estimates based on the given MS data and must be considered semi-quantitative. S-GlcNAcylation of Tau was analyzed on the intact protein by ESI-TOF MS on a Waters Xevo G2-SqTOF MS. Samples were diluted in 0.1% formic acid and subjected to a Waters nanoACQUITY UPLC, equipped with a Zorbax 300SB-C8 column (Agilent). Data were analyzed with a Waters MassLynx V4.1 mass spectrometer and a MaxEnt1 algorithm was used for deconvolution of the protein mass from the multiply charged species. Controls were performed for all reactions under the same reaction conditions but lacking of the respective enzyme.

References

1 C. Val-Cid, X. Biarnés, M. Faijes and A. Planas, *PLoS One*, 2015, **10**, 1–17.

2 H. Nakayama-imaohji, M. Ichimura, T. Iwasa, N. Okada, Y. Ohnishi and T. Kuwahara, *J. Med. Investig.*, 2012, **59**, 79–94.

3 P. Rahfeld, L. Sim, H. Moon, I. Constantinescu, C. Morgan-Lang, S. J. Hallam, J. N. Kizhakkedathu and S. G. Withers, *Nat. Microbiol.*, 2019, **4**, 1475–1485.

4 D. R. Jones, D. Thomas, N. Alger, A. Ghavidel, G. Douglas Inglis and D. Wade Abbott, *Biotechnol. Biofuels*, 2018, **11**, 1–15.

5 Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao and Y. Xu, *Nucleic Acids Res.*, 2012, **40**, 445–451.

6 B. I. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat, *Nucleic Acids Res.*, 2009, **37**, 233–238.

7 R. C. Edgar, *Bioinformatics*, 2010, **26**, 2460–2461.

8 A. Stamatakis, *Bioinformatics*, 2006, **22**, 2688–2690.

9 A. Stamatakis, P. Hoover and J. Rougemont, *Syst. Biol.*, 2008, **57**, 758–771.

10 S. R. Eddy, *Bioinformatics*, 1998, **14**, 755–763.

11 A. Criscuolo and S. Gribaldo, *BMC Evol. Biol.*, , DOI:10.1186/1471-2148-10-210.

12 S. A. Berger, D. Krompass and A. Stamatakis, *Syst. Biol.*, 2011, **60**, 291–302.

13 F. A. Matsen, N. G. Hoffman, A. Gallagher and A. Stamatakis, *PLoS One*, 2012, **7**, 1–4.

14 I. Letunic and P. Bork, *Nucleic Acids Res.*, 2016, **44**, W242–W245.

15 C. Engler, R. Kandzia and S. Marillonnet, *PLoS One*, , DOI:10.1371/journal.pone.0003647.

16 G. Tegl, J. Hanson, H.-M. Chen, D. Kwan, A. G. Santana and S. Withers, *Angew. Chemie Int. Ed.*, 2019, **58**, 1–7.

17 L. Zheng, U. Baumann and J. L. Reymond, *Nucleic Acids Res.*, , DOI:10.1093/nar/gnh110.

18 C. André-Miral, F. M. Koné, C. Solleux, C. Grandjean, M. Dion, V. Tran and C. Tellier, *Glycobiology*, 2015, **25**, 394–402.

19 L. Liu, A. R. Prudden, G. P. Bosman and G.-J. Boons, *Carbohydr. Res.*, 2017, **452**, 122–128.

20 D. H. Kwan, S. Ernst, M. P. Koetzler and S. G. Withers, *Glycobiology*, 2015, **25**, 806–811.