Supporting information

Evolution of phosphorylases from *N*-acetylglucosaminide hydrolases in family GH3

Jorick Franceus[1], Jolien Lormans[1], Lore Cools[2], Matthias D'hooghe[2], Tom Desmet[1*]

[1] Centre for Synthetic Biology (CSB), Department of Biotechnology, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium.

[2] SynBioC Research Group, Department of Green Chemistry and Technology, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium

E-mail: tom.desmet@ugent.be

FIGURES

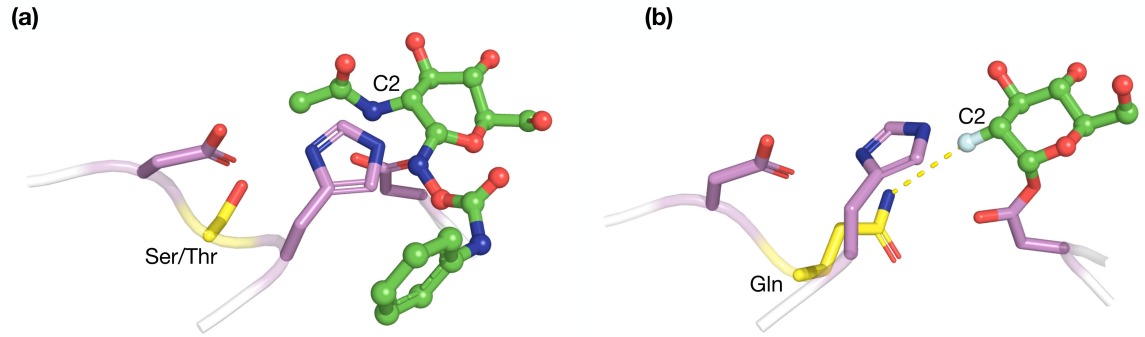

Figure S1. In *N*-acetylglucosaminidases and glycoside phosphorylases from family GH3, a different residue is flanked by the catalytic dyad. (a) *N*-acetylglucosaminidase from *Bacillus subtilis* in complex with an *N*-acetylglucosaminide analogue (green) (PDB code 3NVD). A small serine or threonine residue (yellow) ensures that there is enough space to accomodate the *N*-acetyl group at C2 of the substrate. (b) Glycoside phosphorylase BglP bound to a glucose analogue (green) (PDB code 5VQE). A glutamine residue (yellow) is predicted to form a hydrogen bond with the hydroxyl group at C2 of the substrate. Catalytic residues shown in violet.

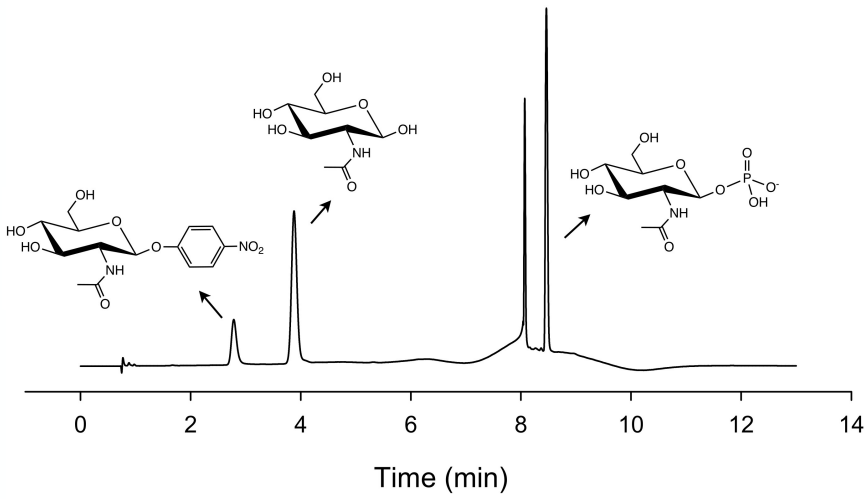

Figure S2. Profile from high-performance anion-exchange chromatography with pulsed amperometric detection, obtained after incubating *Pa*GH F33R/D62N/R70A with 5 mM *p*NP-GlcNAc and 50 mM phosphate. The peak at 8 min is caused by acetate in the eluent.

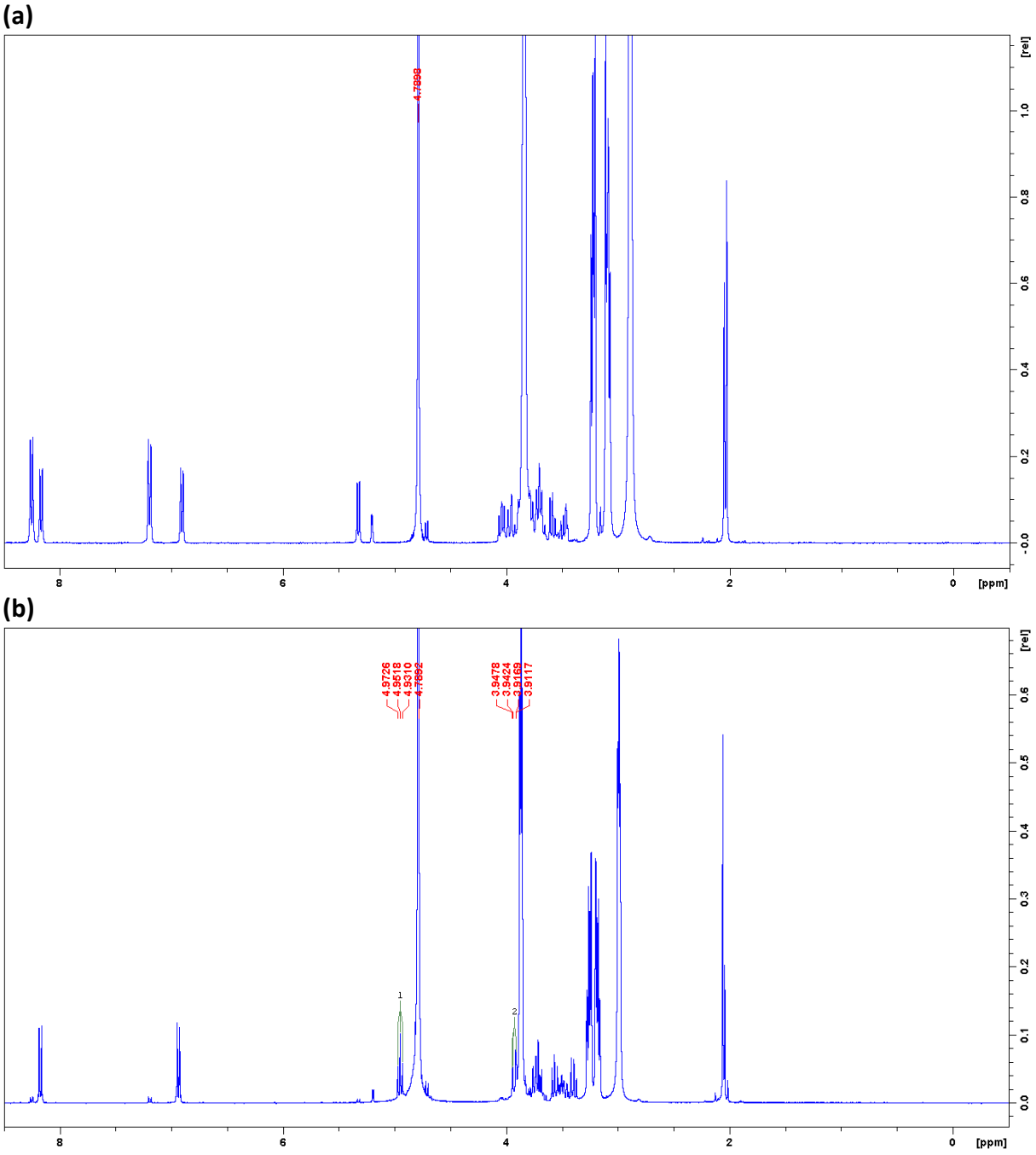

Figure S3. ^{1}H NMR spectra of reaction mixtures obtained after incubating *Pa*GH variant F33R/D62N/R70A with 5 mM *p*NP-GlcNAc with (a) 0 mM phosphate, or (b) 50 mM phosphate.

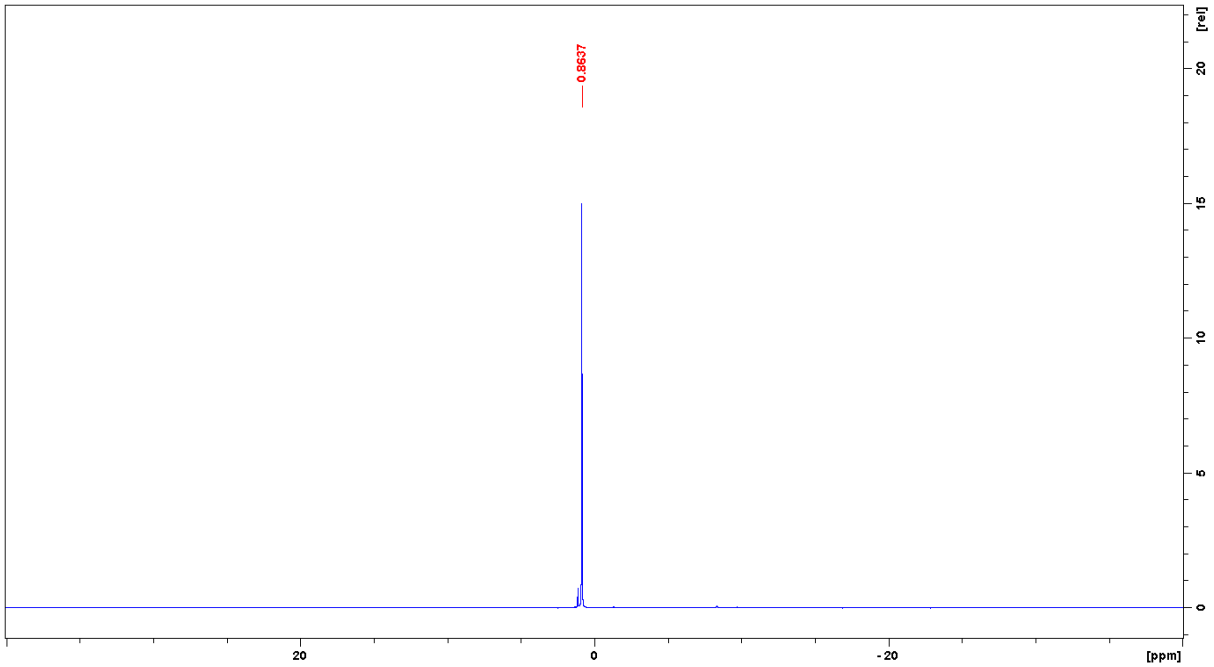

Figure S4. ^{31}P NMR spectrum of reaction mixture obtained after incubating *Pa*GH variant F33R/D62N/R70A with 5 mM *p*NP-GlcNAc and 50 mM phosphate.

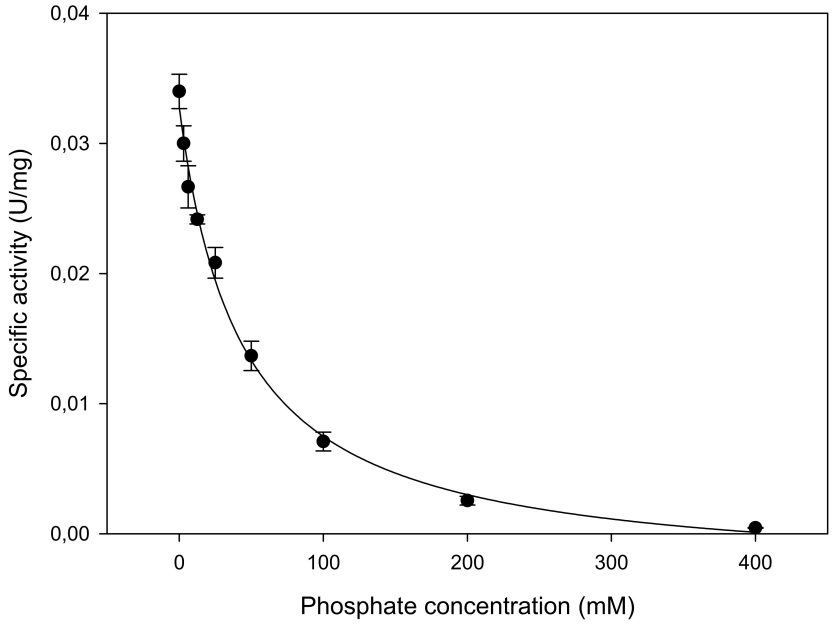

Figure S5. The hydrolytic activity of *Pa*GH F33R/D62N/R70A is suppressed by the presence of inorganic phosphate. Specific activities were measured using 5 mM *p*NP-GlcNAc as substrate at 30°C and pH 6.5.

TABLES

Table S1. Hits from the screening of library F33X/D62N/R70X (ratio between activities with and without phosphate > 1.5).

	Position	
#	**33**	**70**
1	Tyr	Ser
2	Arg	Glu
3	Thr	Lys
4	Ser	Gly
5	Arg	Gln
6	Ser	Gln
7	Gly	Arg
8	Thr	Thr
9	Gln	Ser
10	Ser	Thr
11	Thr	Gln
12	Lys	Gln
13	Tyr	Gln
14	Gly	Thr

Table S2. Amino acid sequences for *Pa*GH, *Bs*GH and *Cf*GP. His_6-tag and linker are underlined.

Enzyme	UniProt code	Sequences
*Pa*GH	Q9HZK0	MQGSLMLDIGGTWLTAEDRQILRHPEVGGLIIFARNIEHPAQVRELCAAIRA IRPDLLLAVDQEGGRVQRLRQGFVRLPAMRAIADNPNAEELAEHCGWLMATE VQAVGLDLSFAPVLDLDHQRSAVVGSRAFEGDPERAALLAGAFIRGMHAAGM AATGKHFPGHGWAEADSHVAIPEDARSLEEIRRSDLVPFARLAGQLDALMPA HVIYPQVDPQPAGFSRRWLQEILRGELKFDGVIFSDDLSMAGAHVVGDAASR IEAALAAGCDMGLVCNDRASAELALAALQRLKVTPPSRLQRMRGKGYANTDY RQQPRWLEALSALRAAQLIDLEHHHHHH
*Bs*GH	P40406	MEASASKRAIDANQIVNRMSLDEKLGQMLMPDFRNWQKEGESSPQALTKMND EVASLVKKYQFGGIILFAENVKTTKQTVQLTDDYQKASPKIPLMLSIDQEGG IVTRLGEGTNFPGNMALGAARSRINAYQTGSIIGKELSALGINTDFSPVVDI NNNPDNPVIGVRSFSSNRELTSRLGLYTMKGLQRQDIASALKHFPGHGDTDV DSHYGLPLVSHGQERLREVELYPFQKAIDAGADMVMTAHVQFPAFDDTTYKS KLDGSDILVPATLSKKVMTGLLRQEMGFNGVIVTDALNMKAIADHFGQEEAV VMAVKAGVDIALMPASVTSLKEEQKFARVIQALKEAVKNGDIPEQQINNSVE RIISLKIKRGMYPARNSDSTKEKIAKAKKIVGSKQHLKAEKKLAEKAVTVLK NEQHTLPFKPKKGSRILIVAPYEEQTASIEQTIHDLIKRKKIKPVSLSKMNF ASQVFKTEHEKQVKEADYIITGSYVVKNDPVVNDGVIDDTISDSSKWATVFP RAVMKAALQHNKPFVLMSLRNPYDAANFEEAKALIAVYGFKGYANGRYLQPN IPAGVMAIFGQAKPKGTLPVDIPSVTKPGNTLYPLGYGLNIKTGRPLLEHHH HHH
*Cf*GP	Q7WUL3	MIDLTAAPFSLDDDGIAWVRTTLAEMGEDEKLGQLFCLITYTSDPEYLGYLT RGLHVGGVMLRTMTAADAAATVTTLQSTATVPLLISANLEGGASQTVQEATH VGSNMALAATGSTDHVRRAATVIGREARALGINWAFTPVVDIDLNFRNPITN TRTFGADAATVAAMGAEYVEAIQAQGLAASAKHFPGDGVDERDQHLLASVNT MSVEEWDDSFGVVYRAAIAAGVKTVMVGHIMLPAYSRALRPGVADRDILPGV VAEELLNDLLRDRLGFNGLVVSDSTTMAGLASVLPRSQAVPRVIAAGCDMFL FTKNLDEDFGYMRAGIRDGVITPERLDEAVTRILALKASLGLHRGTNLPAQG AAGVLADPDHSATAREVAASSITLVKEEPGVLPITRERYPRVLVYDLQNGGS PIGQGARAGAVEQFVDALVEAGHDVTRFEPGGGWEGMAAPTTDVTERHDLVL YLANLSTRSNQTVVRIEWAEPMGANVPAYVHSVPTVFVSFENPYHLFDVPRV RTLINTYGSSPVVLETLLAALQGKAPFAGSSPVDAFCGQWDTHLLEHHHHHH

Table S3. Primers used in this study. Mutations are underlined.

Construct	Template	Sequences (5'-3')
*Pa*GH F33R	*Pa*GH	GTGGTCTGATTATCCGTGCACGTAATATTG CAATATTACGTGCACGGATAATCAGACCAC
*Pa*GH D62N	*Pa*GH	CTGCTGGCAGTTAATCAAGAAGGTGG CCACCTTCTTGATTAACTGCCAGCAG
*Pa*GH R70A	*Pa*GH	GTCGCGTTCAGGCACTGCGTCAGGG CCCTGACGCAGTGCCTGAACGCGAC
*Pa*GH 71-32	*Pa*GH	CTGCGTCAGGGTTTTG GATAATCAGACCACCAACTTCC
*Cf*GP R63F	*Cf*GP	GTGGTGTTATGCTGTTTACCATGACAGCAGC GCTGCTGTCATGGTAAACAGCATAACACCAC
*Cf*GP N70D	*Cf*GP	GCTGATTAGCGCAGATCTGGAAGGTGGTG CACCACCTTCCAGATCTGCGCTAATCAGC
*Cf*GP T98R	*Cf*GP	GGTGCAAGCCAGCGCGTTCAAGAAGC GCTTCTTGAACGCGCTGGCTTGCACC

Table S4. Gene fragment used in this study. Mutations are underlined.

Fragment	Sequence (5'-3')
*Pa*GH 33X-70X	GGAAGTTGGTGGTCTGATTATCNNKGCACGTAATATTGAACATCCGGCA CAGGTTCGTGAACTGTGTGCAGCAATTCGTGCCATTCGTCCGGATCTGC TGCTGGCAGTTAATCAAGAAGGTGGTCGCGTTCAGNNKCTGCGTCAGGG TTTTG