Structural and mechanistic analysis of a β-glycoside phosphorylase identified by screening a metagenomic library

Spencer S. Macdonald\textsuperscript{1,4,6}; Ankoor Patel\textsuperscript{2}; Veronica L.C. Larmour\textsuperscript{2}; Connor Morgan-Lang\textsuperscript{5}; Steven J. Hallam\textsuperscript{3,4,5,6,7}; Brian L. Mark\textsuperscript{2}; Stephen G. Withers\textsuperscript{1,4,6,8}

\textsuperscript{1}Departments of Chemistry and Biochemistry, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z1

\textsuperscript{2}Department of Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2

\textsuperscript{3}Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z3

\textsuperscript{4}Genome Science and Technology Program, University of British Columbia, 2329 West Mall, Vancouver, BC V6T 1Z4, Canada

\textsuperscript{5}Graduate Program in Bioinformatics, University of British Columbia, Vancouver, BC, Canada, Canada V6T 1Z4

\textsuperscript{6}ECOSCOPE Training Program, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3

\textsuperscript{7}Peter Wall Institute for Advanced Studies, University of British Columbia, Canada, V6T 1Z1

\textsuperscript{8}Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia, Canada, V6T 1Z4

*Running title: GH3 β-glycoside phosphorylase identified from a metagenomic library

To whom correspondence should be addressed: Stephen G. Withers, Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia, V6T 1Z1, Canada, Tel.: +1 (604) 822-3402; Fax: +1 (604) 822-2487; Email: withers@chem.ubc.ca

Keywords: Phosphorylases, cellulose, β-glucose-1-phosphate, glycosidases, biomass, biofuel, mechanism, CAZy

The abbreviations used are: DNP\textsubscript{Glc}, 2,4-dinitrophenyl β-D-glucopyranoside; pNP\textsubscript{Glc}, p-nitrophenyl β-D-glucopyranoside; pNP\textsubscript{Glc}N\textsubscript{Ac}, p-nitrophenyl N-acetyl-β-D-glucosaminide; DNP\textsubscript{2FGlc}, 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside; DFMU\textsubscript{Glc}, 6,8-difluoromethylumbelliferyl β-D-glucopyranoside; DFMU\textsubscript{Glc}N\textsubscript{Ac}, 6,8-difluoromethylumbelliferyl N-acetyl-β-D-glucosaminide, DNP, 2,4-dinitrophenyl; pNP, p-nitrophenyl; DFMU, 6,8-difluoromethylumbelliferyl.

ABSTRACT

Glycoside phosphorylases have considerable potential as catalysts for the assembly of useful glycans for products ranging from functional foods and prebiotics to novel materials. However, the substrate diversity of currently identified phosphorylases is relatively small, limiting their practical applications. To address this limitation, we developed a high-throughput screening approach using the activated substrate 2,4-dinitrophenyl β-D-glucoside (DNP\textsubscript{Glc}) and inorganic phosphate for identifying glycoside phosphorylase activity and used it to screen a large insert metagenomic library. The initial screen, based on release of 2,4-dinitrophenol from DNP\textsubscript{Glc} in the presence of phosphate, identified the gene $bglP$, encoding a retaining β-glycoside phosphorylase from the CAZy GH3 family. Kinetic and mechanistic analysis of the gene product, BglP, confirmed a double displacement ping-pong mechanism involving a covalent glycosyl–enzyme intermediate. X-ray crystallographic analysis provided insights into the phosphate-binding mode and identified a key glutamine residue in the active site important for substrate recognition. Substituting this glutamine for a serine swapped the substrate specificity from glucoside to N-acetylglucosaminide. In summary, we present a high-throughput screening approach for identifying β-glycoside phosphorylases, which was robust, simple to implement, and useful in identifying active clones within a metagenomic library.
Implementation of this screen enabled discovery of a new glycoside phosphorylase class and has paved the way to devising simple ways in which enzyme specificity can be encoded and swapped, which has implications for biotechnological applications.

Carbohydrate active enzymes (CAZymes) are the biocatalysts responsible for the assembly, degradation and modification of glycans in biological systems. They are also widely employed enzymes in industry, being used in brewing and food processing, animal feed preparation, industrial pulp and paper applications and increasingly in biofuel and bioproduct development. While the use of CAZymes is cost-effective in glycans degradation, glycan assembly generally requires the use of expensive starting materials, such as nucleotide phosphosugars. The high-cost of these materials makes de novo industrial-scale glycan synthesis difficult and usually non-viable.

One class of CAZyme that offers a potential solution to the high costs typically associated with enzymatic glycan synthesis is that of the glycoside phosphorylases (GPs), which are increasingly being recognized and used for the biocatalysis and biotransformation of glycans. These enzymes ordinarily carry out phosphorolysis by transferring a glycosyl moiety from the non-reducing end of a di- or polysaccharide substrate onto an inorganic phosphate, thereby generating a sugar-1-phosphate. GPs distinguish themselves from most CAZymes in that the hydrolytic free energy associated with the glycosidic ester linkage of the sugar-1-phosphate product is roughly equivalent to that of the glycosidic linkage in the glycan substrate. Accordingly, the equilibrium constants for the reactions catalyzed by most GPs is close to 1, thus the equilibrium position can be tipped in favor of glycoside synthesis by manipulation of reaction conditions.

Since the discovery of the first GP, glycogen phosphorylase, in 1938 only 29 distinct new GP activities have been identified; the majority in the past 15 years. However, given the vast number of GHs known, it seems likely that 30 is a significant underestimate of the actual number of GP activities present in nature, especially since the use of phosphorolysis to metabolize glycans is inherently more energetically favorable for a cell than hydrolysis since the released sugar-1-phosphates feed directly into the glycolysis pathway without the need for expenditure of ATP. Given this metabolic efficiency advantage, and considering the wide range of diverse glycans that are metabolized by microbes, it is probable that glycoside phosphorylases are more widespread than previously thought. Indeed, it seems likely that numerous forms of carbohydrates are metabolized through phosphorolysis, suggesting an abundance of GPs to discover in nature.

Functional metagenomics screening offers a means to search for novel biocatalysts from genetic material drawn directly from the environment. Metagenomic libraries provide access to the vast reservoir of uncultivated genetic diversity encoded in microbial communities inhabiting natural and engineered ecosystems. Emerging sophisticated sequence- and function-based screening technologies are being deployed to identify novel enzymes, including CAZymes, within these libraries. However, to our knowledge no function-based metagenomics screening approach has yet been developed that targets GP-encoded genes. A central challenge in designing the needed screen is that of distinguishing phosphorylase activity from the related hydrolyase activity in a high-throughput manner.

We report here the development and implementation of such a high-throughput screening approach to identify GP activity. By screening of a library derived from a passive mine tailings biochemical reactor system (BCR) fed with lignocellulosic biomass, we identified a previously unknown β-retainning GP from CAZy family GH3, and report on its structural and mechanistic characterization. Our screening approach is based on the use of activated aryl glycosides as easily monitored substrates for glycoside phosphorylases when deployed in the presence of phosphate. The BCR library is a large insert (fosmid) library with an average insert size of 40 kilobases first described by Mewis and colleagues. The authors used the BCR library to search for novel cellulose-degrading enzymes. Therefore, in testing the parallel screening approach to identify GPs, we focused in this case on identifying fosmid clones encoding enzymes with the capacity of phosphorolysing the β 1,4-
glucosidic linkages of cellulose and cello-oligosaccharides.

RESULTS

Development and testing of the metagenomic screen – The design of the screen was based upon the notion that, since many GPs are members of CAZy GH families, they might accept an activated aryl glycoside substrate, transferring the glycosyl moiety to added phosphate. This concept has not been tested previously, to our knowledge, apart from one demonstration with a nucleotide phosphorylase\(^26\), though the concept has parallels with previous work on nucleotide-dependent glycosyltransferases\(^27\). 2,4-Dinitrophenyl glucoside (DNPGlc) was used to test this idea since it is highly activated (pKa of 2,4-dinitrophenol = 4.0)\(^28\) and has proved to be a near universal substrate for glucosidases. Further, the low pKa of the phenol allows direct, continuous assays to be performed at pH values down to below 4 without the need to add base in a stopped manner. The choice of a glucoside was based on the cellulolytic origin of the library to be screened, which thus might contain cello-oligosaccharide-degrading phosphorylases, especially since a number of cellulases and β-glucosidases had already been identified within this library\(^25\).

In order to test whether glycoside phosphorylases can accept DNPGlc as a substrate in the presence of phosphate, and catalyze glucosyl transfer to generate glucose-1-phosphate (G1P), three known GPs that cleave cellobiose or cellooliters were cloned and expressed. Two of these GPs, cellobiose phosphorylase\(^29\) (RtCBP) and celloolitase phosphorylase\(^30\) (RtCDP) from Ruminiclostridium thermocellum (formerly known as Clostridium thermocellum) belong to CAZy family GH94. These enzymes use a β-inverting mechanism in which phosphate attacks directly on the sugar anomeric center, forming α-glucose-1-phosphate (αG1P) (Figure 1B). The third enzyme, the β-glucosidase/N-acetylglucosaminidase Nag3 from Cellulomonas fimi\(^31,32\), belongs to CAZy family GH3 and uses a β-retaining mechanism involving a covalent α-glycosyl-enzyme intermediate that is then attacked by phosphate, forming a β-glucose-1-phosphate (βG1P) (or βGlcNAc-1-phosphate) product (Figure 1C).

As shown in Figure 2A all three enzymes cleave DNPGlc in the presence of phosphate, as can be observed by monitoring the increase in absorbance at 400 nm. Importantly, all three enzymes are only minimally active in the absence of phosphate. Confirmation that the rate stimulation is due to phosphorylase action was provided by TLC analysis of reaction mixtures, which revealed that all three catalysed the transfer of the glucosyl moiety of DNPGlc to inorganic phosphate thereby forming G1P (Figure 2B). These findings therefore validate the concept of the screen: (1) Lysed extracts from metagenomic clones are initially assayed using DNPGlc in the presence of phosphate to detect both GHs and GPs. (2) Those clones that show activity are then rescreened in the absence of phosphate to weed out the glycosidases. (3) Since it is also possible that phosphate could stimulate a GH activity through a specific binding effect of some sort, or simply activation at higher ionic strength, leading to false positives, TLC analysis of this much smaller number of reaction mixtures is used to confirm the presence of the sugar phosphate product. (4) Fosmid clones capable of producing G1P are sequenced and analysed to identify the open reading frames (ORFs) responsible for the observed GP activity. An overview of the screening process is provided in Figure 3.

Functional screen – The BCR fosmid library was constructed using E. coli EPI300 as an expression host and contained 18048 clones in 47 x 384-well plates\(^25\). Following replication of the BCR master library, a total of 880 clones failed to grow, leaving 17168 clones to be screened. Functional screening using DNPGlc in the presence of phosphate yielded 54 active clones with activity greater than the mean + 4 SD, a hit rate of 0.31% (Figure 4B). To distinguish clones with GP activity from those that are GHs the 54 hits were re-arrayed into a 96-well master plate and rescreened both in the presence and absence of phosphate (Figure 4C). Each assay condition was run in triplicate and positive phosphate dependency was determined by performing a t-test between a clone’s A\(_{400}\) values in 0 and 50 mM phosphate. A confidence level of 95 % (p-value < 0.05) was set as a threshold to determine whether to further validate the clone. Of the 54 clones from the master library, 12 displayed activity in the presence of phosphate exceeding the set threshold. The reaction products of these 12 clones were then examined by TLC analysis (Figure 4D). Of the 12 clones so analyzed, two
were found to produce G1P in the presence of phosphate (29K06 and 31P01).

**Identification of the active glycoside phosphorilases** – Complete sequencing of the two fosmid clones revealed they were contiguous with one another, overlapping with 23247 bp at 100% identity (Supporting Figure 1A). Thirty-one open reading frames (ORFs) were predicted from the 29K06 clone and 32 from 31P01. BLASTX query of these against the CAZY database revealed a novel GH3 ORF (hereafter referred to as bglP) located within the overlapping region of the two fosmids. Based on amino acid sequence analysis (Figure 5), bglP was predicted to encode a β-glucosidase/N-acetylglucosaminidase (NagZ), from a hexosaminidase subgroup of GH3 characterized by the sequence motif\(^3\) K-H-(FI)-P-G-(HL)-G-X_4-D-(ST)-H. However, the motif found in bglP had substitutions at two residues (underlined): K-H-(FI)-P-G-D-G-X_4-D-Q-H. This modified motif is also present in the only other known glycoside phosphorylase in the GH3 family, Nag3\(^31,32\). In order to confirm that the activity detected was indeed associated with this gene it was subcloned into a pET expression vector and heterologously expressed and purified. Indeed, activity assays of purified enzyme confirmed that BglP catalyzes the phosphate-dependent cleavage of DNPGlc seen from the 29K06 and 31P01 source clones (Supporting Figure 1B).

**Kinetic and mechanistic characterization of BglP** – The other known GP in the GH3 family, C. fimi Nag3, has been subjected to mechanistic characterization to probe its proposed two-step double-displacement mechanism and confirmed to produce βG1P and βGlcNAc1P, though no 3-D structure has been obtained. In fact, Nag3 was the first retaining β-glycoside phosphorylase described and, like BglP, was shown to also catalyze hydrolysis, but at a reduced rate relative to phosphorylisis\(^2\). Comparison of their amino acid sequences (Figure 5) reveals a high degree of sequence similarity between BglP and Nag3 (55% similarity and 39% identity), making it probable that the two enzymes follow the same mechanism involving a covalent α-glucosyl enzyme intermediate (Figure 1C). In order to confirm this and to allow comparison of the two enzymes, a kinetic and mechanistic analysis of BglP was performed.

Kinetic parameters were determined for cleavage of three different substrates, paranitrophenyl β-glucoside (pNPGlc), 2’, 4’-dinitrophenyl β-glucoside (DNPGlc) and paranitrophenyl β-N-acetylglucosaminide (pNPglcNAc), both in the absence of phosphate, and in the presence of increasing concentrations (Figure 6A-C). Based upon both \(k_{cat}/K_m\) and \(k_{cat}\) values (Table 1) it is apparent that BglP prefers glucoside substrates over N-acetyl-glucosaminides, \(k_{cat}/K_m\) values for pNPglc being approx. 10-fold higher and \(k_{cat}\) values approx. 100-fold higher than those of the hexosaminide. A second distinction of the two substrate classes is that \(k_{cat}\) for the two glucoside substrates increases over 10-fold as phosphate concentrations are raised, whereas increases for pNPglcNAc are much more modest at 1.5 - 2-fold. \(K_m\) values also increase with phosphate, by up to 20 - 30-fold for the glucosides, with the net consequence being that \(k_{cat}/K_m\) values remain approximately constant. These observations for the glucosides reflect classic kinetic behavior for a ping-pong mechanism with substrates for which the second step (cleavage of the glycosyl enzyme) is rate-limiting (Figure 1C).\(^{34}\)

Introduction of an alternate, better, nucleophile (phosphate) into the reaction accelerates the decomposition of the glycosyl enzyme intermediate through provision of a second pathway \(k_{ip}\). The net effect is to not only raise \(k_{cat}\) values, but also to decrease the accumulation of the glycosyl-enzyme intermediate, thereby raising the \(K_m\) value. By contrast, the \(k_{cat}/K_m\) value, which can be expressed in terms shown in Equation 1, reflects the first irreversible step and is thus unaffected by steps occurring later in the pathway.

\[
\frac{k_{cat}}{K_m} = \frac{k_1k_2}{k_{-1} + k_2} \tag{Equation 1}
\]

These kinetic studies therefore strongly support the presumed ping-pong mechanism for BglP. They also suggest that, for the glucosides tested, and certainly for DNPGlc, the rate-limiting step is the turnover of the glycosyl-enzyme intermediate. However, the smaller effect of phosphate on \(k_{cat}\) and \(K_m\) for pNPglcNAc might suggest that the formation of the glycosyl-enzyme intermediate remains at least partially rate-limiting for this substrate. This implies that the presence of the acetamide moiety lowers reaction rates,
possibly due to different interactions with the signature loop. If this is the case then measuring the pre-steady state phase of BglP-catalyzed hydrolysis of the glucoside and N-acetylglucosaminide substrates should reveal substantial differences between the substrates. The use of 6,8-difluoromethylumbelliferyl β-glucoside (DFMUGlGlc) and 6,8-difluoromethylumbelliferyl β-N-acetyl glucosaminide (DFMUGlGlcNAc) improved signal sensitivity and substrate reactivity (phenol pKₐ = 4.7) over their pNP counterparts. Indeed, a clear burst phase was observed for cleavage of DFMUGlGlc, but not for DFMUGlGlcNAc, supporting the above assignment of rate-determining steps (Figure 6D). Direct evidence for a two-step covalent mechanism was then sought by using electrospray ionization mass spectrometry to detect the covalent glycosyl-enzyme intermediate. As seen in Figure 7A, the BglP mass increased by 163 Da, equivalent to a glucosyl moiety, upon incubation with DNPGlc consistent with formation of the covalent intermediate. Notably, the presence of phosphate, which induces rapid turnover of the intermediate, resulted in disappearance of the corresponding glycosyl-enzyme mass.

Having confirmed a two-step mechanism involving a glycosyl-enzyme intermediate, we sought to lengthen the lifetime of the intermediate to allow further mechanistic and structural studies. This was achieved by use of 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (DNP2FGlc) as a slow substrate for which the deglycosylation step (k₃) is much slower than the glycosylation step (k₂). This arises as a consequence of inductive destabilization of the oxocarbenium ion-like transition states of the C2 fluorine, which slows both steps (k₂ and k₃), while the incorporation of an excellent (DNP) leaving group ensures that k₂ > k₃, thus that the intermediate accumulates. Indeed, incubation of BglP with DNP2FGlc in the absence of phosphate resulted in time-dependent, pseudo-first order inactivation of the enzyme, as shown in Figure 7B. The rate of inactivation varied, in a saturable manner, with the concentration of the 2-fluoroosugar, allowing extraction of kinetic parameters for inactivation of BglP of kᵢ = 0.17 min⁻¹ and Kᵢ = 32 mM (Supporting Figure 2A). The 2-fluoro-glycosyl-BglP (2FGlc-BglP) covalent intermediate species was shown to be mechanistically relevant by removal of excess inactivator and then measuring rates of reactivation by assaying aliquots of the enzyme as a function of time in buffer containing increasing concentrations of phosphate. As can be seen in Figure 7C phosphate did indeed stimulate reactivation of the enzyme in a time-dependent fashion. The plot of kᵢ versus [phosphate] showed no saturation behavior, revealing that binding of inorganic phosphate is weak. However, the slope of the line yielded a value for the second order rate constant for reactivation of kᵢ/Kᵢ = 3.2 x 10⁻⁵ min⁻¹ mM⁻¹ (Supporting Figure 2B). This absence of saturation at phosphate concentrations of 100 mM might seem to be a concern, given the apparent saturation binding behavior of phosphate seen in Figure 6. However, it is clear from the difference in maximal rates observed between DNPGlc and pNPGlc in Figure 6A and B that the curvature seen in these plots has its origin in changes in rate-limiting step as phosphate concentrations increase: the glycosylation rate constant (k₂) for DNPGlc is greater than that for pNPGlc.

X-ray crystal structure of BglP – Crystals of unliganded BglP were obtained through hanging drop vapor diffusion in a mother liquor containing 5 mg/mL BglP and 27% polyethylene glycol 1000, and the structure was solved to a resolution of 2.1 Å (Table 3). BglP is a 567 a.a. protein that adopts a two-domain fold (Figure 8A) and is monomeric in solution according to gel filtration (data not shown). The N-terminal domain (a.a. 1-392) forms a (β/α)₈ TIM barrel structure typical of many glycoside hydrolase catalytic domains, with an active site composed of residues within loops that radiate out from the C-terminal ends of the β-strands of the barrel. The C-terminal domain (a.a. 393-562) adopts an α/β sandwich that does not participate directly in catalysis, with the exception of amino acids 476-487, which pack along the edge of the TIM barrel active site to stabilize a loop that emanates from β-strand 5 of the barrel and bears the sequence K-H-(F1)-P-G-D-G-Xᵣ-D-Q-H. This loop contains the sequence that defines the GH3 subgroup to which BglP belongs (albeit with the two alterations noted previously) and accommodates the general acid/base His206 (Figure 8B).

Substrate recognition – To gain structural insights into the mechanism and specificity of BglP, crystals of the enzyme were soaked with 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (DNP2FGlc) to form the long-lived glycosyl-enzyme intermediate that had been
characterized kinetically. The structure so determined revealed clear electron density for a 2-deoxy-2-fluoro-α-D-glucosyl moiety in a \(^{4}\text{C}_1\) chair conformation covalently bound via C1 to Asp288 (Figure 8B). This residue is conserved in GH3 enzymes and is indeed suitably positioned within the BglP active site to serve as the catalytic nucleophile forming the key reaction intermediate. The active site contains most of the same residues identified in other GH3 enzymes as being involved in substrate recognition, such as Lys193 and His194 of the signature sequence. Also consistent with other members of the GH3 subfamily to which BglP belongs, the imidazole of a histidine residue (His206), is positioned ~5.5 Å from the carboxylate of Asp288 on the opposite face of the hexose ring of the bound intermediate and is thus well positioned to assist in catalysis by acting as a general acid/base. It first protonates the oxygen of the scissile bond to assist leaving group departure, followed, in this case, by activation of an incoming phosphate that reacts with the covalent glycosyl-enzyme intermediate to generate a phosphorylated product with net retained anomeric stereochemistry (Figure 8B). The ~5.5 Å distance between Asp288 and His206 of BglP is consistent with glycosidases that employ a configuration-retaining mechanism. One key difference, however, is the identity of the residue interacting with the C2-position of the sugar, this being the amino acid flanked by His206 and Asp204 (the so-called His-Asp catalytic dyad) in the signature sequence. In the GH3 GlcNAc-cleaving enzymes, such as the NagZs, the residue at that position is a serine, while in BglP it is a glutamine, as seen in the structural overlay of Figure 8C wherein the 2Fglc-BglP structure is superimposed on the structures of two other GH3 β-N-acetylglucosaminidases (NagZ from Burkholderia cenocepacia in complex with GlcNAc [PDB: 4GNV] and NagZ from Bacillus subtilis in complex with O-(2-acetamido-2-deoxy-D-glucopyranosylidenoamino) \(N\)-phenylcarbamate (PUGNAc) [PDB: 3NVD]). The overlay reveals that Gln205 of BglP is likely to clash with the NAc of GlcNAc based substrates, as the amide of the Gln side chain sits ~2 Å from the methyl group of the NAc from GlcNAc and PUGNAc. In the natural glucosyl intermediate, where a hydroxyl group is present at C2, we predict the Gln205 amide would be 3.0-3.5 Å away from the hydroxyl group, suggesting that Gln205 could form a hydrogen bond with the C2 hydroxyl group of glucoside substrates. In both cases the interaction at that position is between an amide and a hydroxyl, but the directionality of the interaction is inverted; a beautiful example of specificity swapping. Indeed, it had previously been speculated that the broadened substrate specificity seen for Nag3, which also prefers glucoside to \(N\)-acyethylhexosaminide substrates, was due to the substitutions seen in the modified sequence motif: K-H-(F1)-P-G-D-G-Xa-D-Q-H (underlined residues correspond to Asp193 and Gln200 in Nag3). The other differing residue, Asp198 is not close to the active site thus is unlikely to be involved in the broadened substrate specificity of the enzyme.

To test whether Gln205 indeed plays a role in discriminating between glucoside and GlcNAc substrates it was replaced by serine in BglP using site-directed mutagenesis to form the mutant, BglP-Q205S. Determination of the kinetic parameters for the cleavage of pNPGlc, pNPglcNAc and DNPGlc by BglP-Q205S (Figure 9A-C) revealed that the Gln to Ser mutation caused the preferred substrate to switch from pNPglc to pNPglcNAc as shown graphically in (Figure 9D). Kinetic parameters determined both in the absence and presence of 50 mM phosphate are shown in (Table 2). The \(k_{\text{cat}}\) and \(k_{\text{cat}}/K_m\) values of the Q205S mutant for pNPglcNAc are both ~10-fold higher than those of the wild type, with no obvious stimulation from phosphate. However, TLC analysis confirms that GlcNAc1P is the predominant product in the presence of phosphate, thus the rate-limiting step for the mutant with the GlcNAc substrate remains the glycosylation step (Supporting Figure 3). The opposite situation is seen for pNPglc, for which only \(k_{\text{cat}}/K_m\) values could be obtained since saturation was never reached, even at 50 mM substrate. The \(k_{\text{cat}}/K_m\) of the Q205S mutant is ~10-fold lower than that of wild type, and again no significant effect of phosphate on rates was observed, pointing to rate-limiting formation of the covalent intermediate. For DNPGlc, a clear stimulation of activity by phosphate could be seen in the mutant, and once again \(k_{\text{cat}}\) (0 mM phosphate) and \(k_{\text{cat}}/K_m\) (50 mM phosphate) are ~10-fold lower than their counterparts for the wild type enzyme. The presence of the better leaving group (i.e. DNP) accelerates the glycosylation step sufficiently that
deglycosylation \((k_3)\) becomes rate-limiting, thus stimulation by phosphate is observed.

**Phosphate recognition** — Hoping to gain experimental insights into the structure of the product complex, we tried to obtain a crystal structure of BgIP bound to 2-deoxy-2-fluoro-\(\beta\)-glucose-1-phosphate. Soaking BgIP crystals with this analogue (times ranging from 5 min to 1 h) either generated the same 2-fluoroglucosyl-enzyme intermediate species described above, or nothing was found bound in the active site. We therefore resorted to manual docking studies to predict how BgIP accommodates a phosphoryl group in its active site. The product \(\beta\)Glc1P was modeled into the active site by superimposing the hexose ring of Glc onto the experimentally determined hexose ring of the covalently bound 2-fluoroglucosyl moiety (**Figure 10**). When modeled in the energetically favoured \(4C_1\) chair conformation, the phosphate of \(\beta\)Glc1P sterically clashed with the side chain of Met 292. This was not too surprising, however, considering that retaining glucosidases are known to assist bond cleavage by distorting the substrate to position the leaving group in a pseudooaxial position, which brings the conformation of the substrate closer to that of the reaction transition state\(^{39}\). Thus, the repulsive interactions between Met292 and the phosphate of \(\beta\)Glc1P, in conjunction with the favourable binding interactions between the substrate glucosyl and phosphate moieties likely force the phosphate of bound \(\beta\)-Glc-1-phosphate product into a pseudooaxial position. This is also likely to be the case for the bound oligosaccharide substrate. Interestingly, when this distortion is accounted for in our model by distorting the Glc ring toward a \(1S\) skew-boat conformation (based on a crystal structure of a GH3 NagZ from *Bacillus subtilis*\(^{48}\)), which places the phosphate group pseudooaxial, the clash with Met292 is relieved with no other steric interferences arising. The pseudooaxial position also places the oxygen atom of the phosphoester linkage of the product within \(~3.2\) Å of the imidazole of His206 (**Figure 10**), which is predicted to be where the oxygen of the glycosidic bond of the substrate would reside when it is protonated during bond cleavage. If the enzyme were to employ a glutamic acid residue as acid/base catalyst, as is the case for the majority of GH3 glycosidases, there would likely be significant Coulombic repulsion with the substrate phosphate. This Glu to His substitution was proposed earlier as a modification that gave members of the subgroup of GH3 enzymes the ability to carry out phosphorylasis rather than hydrolysis\(^{32}\). Recently, Ducatti et al. demonstrated that while the Glu to His substitution may be necessary for phosphorylase action, it is not itself sufficient, thus not predictive of phosphorylase activity\(^{41}\). They showed that a GH3 \(\beta\)-N-acetylglicosaminidase from *Herbaspirillum seropedicae* SmR1 (Hsero1941) bearing the His-Asp dyad functions predominantly as a hydrolase and observed no activity stimulation or phosphorylated products in the presence of phosphate. They also observed that Hsero1941 showed greater activity toward pNPGlcNAc \((k_{cat} = 1.2 \text{ s}^{-1})\) than pNPGlc \((k_{cat} = 3.3 \times 10^{-3} \text{ s}^{-1})\), consistent with the fact that the residue sandwiched between Asp and His in their signature sequence is indeed Ser.

**DISCUSSION**

The screen, developed and validated with three known \(\beta\)-glycoside phosphorylases, proved to be robust, simple to implement, and useful in identifying active clones within a metagenomics library. From a relatively small number of initial hits a single \(\beta\)-glucoside phosphorylase belonging to CAZy family GH3 was discovered. This represents the first reported high-throughput functional metagenomic screen for glycoside phosphorylases, and the approach used may well be amenable to a range of other phosphorylases. Given successful implementation of the screening paradigm on known \(\beta\)-glucoside phosphorylases from GH94, and the relatively high number of initial hits, it was somewhat surprising that GH94 enzymes were not identified in the BCR library. It is unlikely that this is due to the activity stimulation by phosphate being masked through some other step being rate-limiting, as can be seen for the retaining enzyme studied here, since GH94 phosphorylases are inverting, with a single chemical step. A possible explanation why no GH94s were discovered is simply that none are present within the BCR library. This is consistent with the partial sequence information available from this library (from end-sequencing of fosmids), which contained no predicted GH94 ORFs. A second possible explanation is that those GH94s potentially present in the library have limited expression, which in many cases is due to the *E. coli* host RNA polymerase’s inability to recognize...
GH3 β-glycoside phosphorylase from a metagenomic library

foreign promoter sequences on the fosmid DNA. This limitation can be compensated for by equipping the host strain with additional sigma factors that help it recognize a wider range of promoter sequences, however the host strain used in this study contained only the native E. coli sigma factors. A third possibility is that those GH94s that are expressed are not capable of cleaving DNPGlc. To overcome this issue, a new functional screen would need to be devised that would utilize natural substrates as opposed to aryl glycosides. This could involve detection of sugar phosphate products when assayed in the degradative direction, or oligosaccharides when assayed in the synthetic mode.

The reason for the false positives (clones that showed phosphate-dependent DNPGlc cleavage, but failed to produce G1P) is not known at this stage. The most probable is that these are glycosidases whose activity is enhanced by high phosphate levels either through simple salt effects or through some form of allosteric interaction. Indeed, apart from 29K06 and 31P01, each clone analyzed by TLC had been previously identified through functional screens targeting glycoside hydrolases and subsequently sequenced (Supporting Table 3). Given that two of the non-GP GH3 ORFs were each independently found on three separate fosmids (see Supporting Figure 4), it is likely these are simple β-glucosidases whose activity is in some way stimulated by the presence of phosphate.

Although the screen of 17168 fosmid clones identified only one new GP, this enzyme did turn out to be interesting, as this represents only the second β-retaining glycoside phosphorylase of any kind to be identified and characterized, and the first for which a three-dimensional structure has been determined. BglP seems to be specialized towards β-glucoside substrates, generating a βG1P product, with much lower activities against N-acetylhexosaminides than against N-acetylhexosaminidases. Mayer et al. had speculated that this difference in substrate specificity (in the case of Nag3) was due to residues Asp193 and Gln200 since they differed from those in the conserved sequence motif (K-H-(FI)-P-G-D-G-X4-D-Q-H; altered residues underlined), although at the time the group had no structures of enzymes from this sub-group and was unaware that Nag3 was a phosphorylase. Since BglP contains the same altered sequence motif, our structural study provided insight into the functional significance of BglP’s homologous residues (Asp198 and Gln205). By comparing BglP’s structure to that of two other NagZ structures having the fully conserved motif (K-H-(FI)-P-G-(HL)-G-X4-D-(ST)-H), we concluded that Gln205 was likely causing steric hindrance with a C2 N-acetyl group, while at the same time the NH2 moiety of the Gln side chain was in a good position to hydrogen bond with a C2 hydroxyl group. Indeed, by mutating Gln205 to a serine (the amino acid present in the fully conserved GH3 N-acetylhexosaminidase motif) the substrate preference was cleanly switched from glucosides to N-acetylhexosaminides. Thus, through the development of a novel screen for glycoside phosphorylases, a new class of enzyme has been discovered and characterized both structurally and mechanistically. This in turn has provided new insights into simple ways in which enzyme specificity can be encoded and swapped while opening wide a screening paradigm for the discovery of wide ranging glycoside phosphorylases with biotechnological potential.

EXPERIMENTAL PROCEDURES

Materials, reagents and general methods – All buffers and reagents were purchased from Sigma-Aldrich, unless noted otherwise. Substrates used in this study were either purchased from suppliers or donated by members of this laboratory. p-Nitrophenyl N-acetyl-β-D-glucosaminide...
GH3 β-glycoside phosphorylase from a metagenomic library

(pNP-GlcNAc) was purchased from Calbiochem, p-nitrophenyl β-D-glucopyranoside (pNP-Glc) was purchased from Sigma-Aldrich. 2,4-Dinitrophenyl β-D-glucopyranoside (DNPGlc), 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (DNPF2Glc), 6,8-difluoromethylumbelliferoyl β-D-glucopyranoside (DFMU-Glc) and 6,8-difluoromethylumbelliferoyl N-acetyl-β-D-glucosaminide (DFMU-GlcNAc) were generously donated by H. Chen and F. Liu. Enzyme quantification was performed using the Bradford method or by measuring enzyme absorbance at 280 nm and using a molar absorption coefficient (ε280) of 76945 M⁻¹ cm⁻¹ to calculate concentration using the Beer-Lambert law. All TLC assays were done on silica gel 60 F₂₅₄ TLC plates (EMD Millipore Corporation, Billerica, MA, USA). Plates were eluted with a mobile phase of 1-butanol, methanol, ammonium hydroxide and water in a 5:4:4:1 ratio. Following elution, plates were dried then dipped in a molybdate TLC stain (2.5 % ammonium molybdate (w/v), 1 % ceric ammonium sulfate (w/v) and 10 % H₂SO₄ (v/v)) or p- anisaldehyde TLC stain (92.5 % ethanol, 4 % H₂SO₄ (v/v), 1.5 % acetic acid (v/v), 2 % p-anisaldehyde (v/v)). Visualization of TLC plates was done by heating until the product spots became visible. PCR primers and cycle conditions can be found in Supporting Table 1 and buffer recipes can be found in Supporting Table 2.

**Plasmid construction** — pET28-RtCBP.h6, pET28-RtCDP.h6 and pET28-BglP.h6 expression constructs were made according to the partially incomplete polymerase extension (PIPE) cloning method. pET28-BglP.h6.Q205S mutant was constructed using the QuikChange® PCR site directed mutagenesis method. RtCBP and RtCDP ORFs were amplified from *R. thermocellum* genomic DNA and the BglP ORF from *C. fimi* Nag3 was expressed and purified, as reported previously. 2 L (RtCBP, RtCDP and BglP.Q205S) or 4 L (BglP) of LB media containing 25 µg/mL kanamycin was inoculated with 1/100 of overnight culture. Expression cultures were grown at 37 °C until OD₆₀₀ = 0.5 (~3 h). Cells were induced with 0.5 mM IPTG and grown for an additional 3 h at 37 °C (RtCBP and RtCDP) or 18 h at 16 °C (BglP and BglP.Q205S). Cells were harvested by centrifuging at 6000 x g for 6 min in a Beckman Coulter Avanti J-E floor centrifuge (J-A10 rotor) followed by resuspension (20 mL per 1 L of original culture volume) in loading buffer A (RtCBP and RtCDP) or loading buffer B (BglP and BglP.Q205S) (Supporting Table 2). Cells were lysed with an Avestin C3 homogenizer with an average cell pressure of 16 000 psi. The soluble fraction was isolated by centrifuging the lysate at 15 000 rpm for 30 min (JA-20 rotor). The soluble fraction was either stored at -20 °C until needed or immediately purified. Purification was carried out by immobilized metal affinity chromatography (IMAC) on a GE ÄKTA FPLC equipped with a UV and conductance detector and an automatic fraction collector. Buffers used for protein purification are detailed in Supporting Table 2. For each enzyme a separate 5 mL HisTrap™ FF column (GE) was equilibrated with 10 column volumes of loading buffer A (RtCBP and RtCDP) or loading buffer B (BglP and BglP.Q205S). Soluble cell lysates were applied to the columns using a P-1 peristaltic pump (GE) followed by a wash step of 10 column volumes (CV) of the respective loading buffers. The columns were then transferred to the ÄKTA and washed again with 10 CV of wash buffer A (RtCBP and RtCDP) or wash buffer B (BglP and BglP.Q205S). Equilibration was performed by washing buffer B (BglP and BglP.Q205S) followed by equilibration with 10 CV of the respective loading buffers. Proteins were eluted using a 20 mL gradient (0 - 100 %) of loading buffer to elution buffer with the automatic fraction collector set to collect 1 mL fractions. Fractions were analyzed on SDS PAGE and those yielding the biggest band at 94 kDa for RtCBP, 113 kDa for RtCDP or 65 kDa for BglP and BglP.Q205S were combined and concentrated using an Amicon®
Ultra-4 MWCO 30 kDa centrifugal filter (Sigma-Aldrich), then dialysed against storage buffer A (RtCBP and RtCDP) or storage buffer B (BglIP and BglP.Q205S) and stored at -70 °C.

RtCBP, RtCDP and Nag3 DNPGlc phosphorolysis assay – Spectroscopic assay: DNPGlc (2 mM) was incubated with 25 µg of purified RtCBP, RtCDP or Nag3 and 0 or 50 mM phosphate in storage buffer A (200 µL reaction volume) and incubated at 37 °C for 1 h. A₄₀₀ values were performed in matched 1 cm path length quartz cuvettes using a Varian Cary 300 Bio UV-visible spectrophotometer with an automatic cell changer and circulating water bath. 

TLC assay: DNPGlc (20 mM) was incubated with 37.5 µg of purified RtCBP, RtCDP or Nag3 and 0 or 50 mM phosphate in storage buffer A (10 µL reaction volume) and incubated at 37 °C for 2 h. Samples (0.5 µL) of each reaction were spotted onto TLC plates and run and visualized as described above.

BCR library functional screen – The screening methodology was modeled on the functional metagenomic screen reported by Mewis et al. A generalized workflow of the screening method is shown in Figure 3. BCR library plates were replicated into fresh 384-well plates (Nunc™ 384-well clear polystyrene plates, non-treated) containing 50 µL LB media per well with 100 µg/mL arabinose and 12.5 µg/mL chloramphenicol using a QPix2 robot. The replicated plates were grown for 18 h at 37 °C. The assay was performed by adding 50 µL of 2x assay buffer S (Supporting Table 2) and incubating at 37 °C for 6 h, then absorbance measurements were taken at 400 nm. 56 fosmid clones that displayed an A₄₀₀ value greater than mean + 4 SD were re-arrayed to a 96-well GP master plate.

BCR GP master library screen and validation – GP master library screen: The BCR GP master plate was replicated to six identical 96-well plates (Costar® 96 well flat bottom polystyrene plate) containing 100 µL LB media per well with 100 µg/mL arabinose and 12.5 µg/mL chloramphenicol, then grown for 18 h at 37 °C. The plates were then screened in the presence and absence of 50 mM phosphate, each condition in triplicate. Either 100 µL 2x assay buffer P or 2x assay buffer H (Supporting Table 2) was added to each well and incubated at 37 °C for 6 h. After the incubation period absorbance measurements were taken at 400 nm in a BioTek Synergy H1 Hybrid microtiter 96 well plate reader. A Student’s t-test was performed between the averages of the A₄₀₀ values of each clone in the absence and presence of phosphate using Prism GraphPad v6.0 software. Clones possessing a significantly (p < 0.05) higher activity with phosphate present were analysed further by TLC. GP TLC validation: 2 x 100 µL of overnight cell culture from each of the 12 fosmid clones were spun down and the supernatant removed. Each cell pellet pair was resuspended with 10 µL 2x assay buffer H, 2 µL 100 mM DNPGlc and 8 µL dH₂O or 8 µL 125 mM potassium phosphate (pH 7.0) and incubated for 2 h at 37 °C. Samples (0.5 µL) of each reaction condition were spotted onto TLC plates and run as described above.

Full fosmid sequencing – Fosmid DNA was extracted from clones FOS62_29K06 and FOS62_31P21 using the GeneJET Plasmid Miniprep Kit (Thermo Scientific) according to the manufacturer’s instructions. Fosmid preparations were further treated with PlasmidSafe DNase (Epicentre) to degrade contaminating E. coli chromosomal DNA. DNA concentrations were measured with Quant-iT™ dsDNA HS Assay Kit (Invitrogen) using a Qubit® fluorometer (Invitrogen). For full fosmid sequencing, 2.4 ng of each fosmid was sent to UBC Sequencing Centre (Vancouver, Canada). Each fosmid was individually barcoded and sequenced using the MiSeq system.

Contig assembly, open reading frame prediction and gene identification – All Illumina MiSeq raw sequence data were trimmed and assembled using a python script available on GitHub at https://github.com/hallamlab/FabFos. Briefly, Trimomatic was used to remove adapters and low-quality sequences from the reads. These reads were screened for vector and host sequences using BWA, then filtered using samtools and a bam2fastq script to remove contaminants. These high-quality and purified reads were assembled by MEGAHIT with k-mer values ranging between 71 and 241, increasing by increments of 10. Since these libraries often had in excess of 20,000 times coverage, to prevent the accumulation of sequencing errors interfering with proper sequence assembly, the minimum k-mer multiplicity was calculated by 1% of the estimated coverage of a fosmid. Outside of the python script, assemblies
GH3 β-glycoside phosphorylase from a metagenomic library

which yielded more than one contig were then scaffolded using minimus\textsuperscript{53}. Parameterized commands can be found in both documentation on the GitHub page and in the python script itself. Fosmid ORFs were identified using the metagenomic version of Prodigal\textsuperscript{54} and compared to the CAZy database using BLASTP as part of the MetaPathways v2.5 software package\textsuperscript{55}. MetaPathways parameters: length > 60, BLAST score > 20, blast score ratio > 0.4, EValue < 1 x 10\textsuperscript{-6}.

**Kinetic analysis** – Kinetic parameters were determined by measuring enzymatic reaction rates with the chromogenic substrates and monitoring the change in absorbance at 400 nm. Measurements were performed in triplicate in matched 1 cm path length quartz cuvettes using a Varian Cary 300 Bio UV-visible spectrophotometer with an automatic cell changer and temperature controller, at 25 °C in Buffer B (Supporting Table 2) containing the indicated concentration of substrate and potassium phosphate (pH 7.0). 150 µL Buffer B was premixed with 20 µL 10× phosphate solution and 10 µL of 20× substrate solution. The reaction was initiated by the addition of 20 µL of 2 µM BglP or BglP,Q205S for DNPGlc, 20 µM BglP or BglP,Q205S for pNPGlc or 50 µM BglP or BglP,Q205S for pNPGlcNAc reactions and the change in absorbance at 400 nm was measured for 5 min. Hydrolysis/phosphorolysis rates were calculated by measuring absorbance changes as a function of time and converting these to concentration with the following extinction coefficients: 7280 M\textsuperscript{-1} cm\textsuperscript{-1} (pNP) and 12460 M\textsuperscript{-1} cm\textsuperscript{-1} (2,4DNP) at 25 °C, pH 7.0. Substrate final concentrations: DNPGlc: 2.5, 10, 100, 500, 1000 and 1500 µM. pNPGlc: 0.5, 1, 2.3, 3.6, 5, 10 and 25 mM. pNPGlcNAc: 0.1, 0.25, 0.5, 1, 2, 5 and 10 mM. Concentrations of phosphate were chosen to encompass apparent K\textsubscript{m} values where possible. Non-linear regression was performed using GraphPad Prism v6.0.

**Electrospray Mass Spectrometry** – 1 mg/mL BglP (in 20 mM Tris, pH 7.0, 300 mM NaCl, 10% glycerol (v/v), 5 mM MgSO\textsubscript{4} and 0.1 mM DTT) was mixed, on ice, with 10 mM DNPGlc or 10 mM DNPGlc and 50 mM sodium phosphate (pH 7.0). 10 µL reaction samples were immediately diluted 1 in 1000 (5 % acetonitrile, 0.1 % formic acid, pH 2.2) then injected onto the M.S. Samples were analyzed using methods described previously\textsuperscript{33,56}.

**Pre-Steady State Kinetics** – Pre-steady state phases of the reaction of BglP with 6,8-difluoromethylumbelliferyl β-D-glucopyranoside (DFMU Glc) and (separately) 6,8-difluoromethylumbelliferyl N-acetyl-β-D-glucosaminide (DFMU GlcNAc) were monitored in a temperature controlled fluorimeter (Varian Eclipse fluorescence spectrophotometer) cooled to 12 °C. Excitation of 353 nm (slit length 2.5 nm) was used and the fluorimeter was set to continuously monitor emission at 451 nm (slit length 5 nm) for the duration of the experiment. A solution containing 980 µL of pre-chilled (12 °C) assay buffer (50 mM HEPES pH 7.0, 300 mM NaCl, 10% glycerol (v/v), 5 mM MgSO\textsubscript{4}, 0.1 mM DTT) and either 10 µM DFMU Glc or DFMU GlcNAc in a 4.5 mL plastic cuvette (Fisherbrand® Disposable Plastic Cuvette, UV-VIS, CLR SIDE, Methacrylate) was used to establish a baseline. After approx. 1 min, 20 µL of BglP (2.1 µM for DFMU Glc or 21 µM for DFMU GlcNAc) was added to the assay buffer and the reaction was allowed to proceed for 15 min. The fluorescence signal intensity was converted to DFMU concentration using a standard curve of the fluorophore in the same assay buffer.

**Inactivation and reactivation of BglP** – **Inactivation:** In individual 50 µL reactions 5 µM BglP was combined with 0, 2, 5, 10, 20, 25, 50 and 75 mM DNP2FGlc in Buffer C and incubated in a 25 °C circulating water bath. At 0, 10, 30, 60, 90 and 120 min after beginning incubation 5 µL of each reaction was transferred to a matched 1 cm path length quartz cuvette containing 50 mM pNPGlc in 195 µL Buffer C and the change in absorbance at 400 nm was measured over 5 min. Turnover rates were calculated using the extinction coefficient for pNP (7280 M\textsuperscript{-1} cm\textsuperscript{-1}), plotted against time and fitted to a first order expression using Prism GraphPad v6.0 to give apparent rate constants for inactivation at each concentration of DNP2FGlc. A replot of these versus DNP2FGlc phases of the reaction of BglP with 6,8-difluoromethylumbelliferyl β-D-glucopyranoside (DFMU Glc) and (separately) 6,8-difluoromethylumbelliferyl N-acetyl-β-D-glucosaminide (DFMU GlcNAc) were monitored in a temperature controlled fluorimeter (Varian Eclipse fluorescence spectrophotometer) cooled to 12 °C. Excitation of 353 nm (slit length 2.5 nm) was used and the fluorimeter was set to continuously monitor emission at 451 nm (slit length 5 nm) for the duration of the experiment. A solution containing 980 µL of pre-chilled (12 °C) assay buffer (50 mM HEPES pH 7.0, 300 mM NaCl, 10% glycerol (v/v), 5 mM MgSO\textsubscript{4}, 0.1 mM DTT) and either 10 µM DFMU Glc or DFMU GlcNAc in a 4.5 mL plastic cuvette (Fisherbrand® Disposable Plastic Cuvette, UV-VIS, CLR SIDE, Methacrylate) was used to establish a baseline. After approx. 1 min, 20 µL of BglP (2.1 µM for DFMU Glc or 21 µM for DFMU GlcNAc) was added to the assay buffer and the reaction was allowed to proceed for 15 min. The fluorescence signal intensity was converted to DFMU concentration using a standard curve of the fluorophore in the same assay buffer.

**Inactivation and reactivation of BglP** – **Inactivation:** In individual 50 µL reactions 5 µM BglP was combined with 0, 2, 5, 10, 20, 25, 50 and 75 mM DNP2FGlc in Buffer C and incubated in a 25 °C circulating water bath. At 0, 10, 30, 60, 90 and 120 min after beginning incubation 5 µL of each reaction was transferred to a matched 1 cm path length quartz cuvette containing 50 mM pNPGlc in 195 µL Buffer C and the change in absorbance at 400 nm was measured over 5 min. Turnover rates were calculated using the extinction coefficient for pNP (7280 M\textsuperscript{-1} cm\textsuperscript{-1}), plotted against time and fitted to a first order expression using Prism GraphPad v6.0 to give apparent rate constants for inactivation at each concentration of DNP2FGlc. A replot of these versus DNP2FGlc concentration was fit to a Michaelis Menten-like expression to yield values of k\textsubscript{i} and K\textsubscript{i}. **Reaction:** In 50 µL, 100 µM BglP was incubated with 50 mM DNP2FGlc for 210 min at 25 °C. Inactivation was confirmed by monitoring pNPGlc hydrolysis throughout the duration of the incubation period. Following inactivation, excess DNP2FGlc was diluted to ~ 0.1 µM by sequential dilutions with reactivation (RA) buffer (50 mM
HEPES pH 7.0, 300 mM NaCl, 10 % glycerol (v/v), 5 mM MgSO₄, 0.1 mM DTT) and re-concentration steps with Amicon® Ultra – 0.5 mL centrifugal filters (30,000 MW) (Millipore, Darmstadt, Germany). After the final re-concentration step inactive BglP was adjusted to 100 μL final volume with RA buffer. Inactive BglP (5 μL) was incubated at 25 °C with 45 μL RA buffer supplemented with 0, 1, 5, 10, 25, 50 or 100 mM sodium phosphate pH 7.0. At 0, 12, 25, 60, 98, 130 and 170 min after beginning incubation 5 μL of each reaction was transferred to a quartz cuvette containing 50 mM pNPGlc in 195 μL of RA buffer. Turnover rates were calculated as described above.

**X-ray crystallographic structure determination** – To produce BglP in quantities sufficient for crystalization trials, *E. coli* BL21-Gold (DE3) harboring the pET28-BglP.h6 expression plasmid were grown in 500 mL of LB media supplemented with 35 μg/mL kanamycin at 37 °C to an OD₆₀₀ of 0.5-0.6, then induced with 1 mM IPTG and grown for an additional 18-20 hrs at 16 °C with shaking. The culture was pelleted by centrifugation, resuspended in lysis buffer (50 mM HEPES pH 7.0, 500 mM NaCl, 10% glycerol (v/v) and 5mM imidazole), and lysed using a French pressure cell press (Aminco). The lysate was centrifuged at 17,000 x g for 1 h at 4 °C. The supernatant was mixed with 1.5 mL of Ni-NTA resin (Qiagen) for 1 h at 4 °C, then poured into a gravity column and washed with 25 mL of lysis buffer, followed by 25 mL of lysis buffer supplemented with 25 mM imidazole. Recombinant BglP was eluted from the washed resin using lysis buffer supplemented with 250 mM imidazole and dialyzed overnight against 500 mM NaCl, 50 mM Tris-Cl (pH 7.5), 10 % glycerol (v/v), and 1 mM DTT. The protein was concentrated and loaded onto a size exclusion gel filtration column (Superdex 75) pre-equilibrated with dialysis buffer. Fractions containing BglP were pooled and concentrated using an Amicon Ultra Centricon spin cartridge (Merck Millipore).

When concentrated to 5 mg/mL, BglP crystallized after one day in a mother liquor composed of 27% PEG 1000 and 100 mM MES (pH 6.5), by hanging-drop vapor diffusion. To obtain a complex of BglP with DNP2FGlc, crystals of the protein were soaked for 1 h with the ligand at a final concentration of 20 mM. Both native and ligand-bound BglP were subsequently flash-cooled in liquid N₂ using the mother liquor above as a cryosolution.

X-ray diffraction data were collected in-house at 100 K using a Rigaku MicroMax HF X-ray generator and R-AXIS IV++ image plate detector. Data were indexed using MOSFLM⁵⁷ and scaled using Aimless⁵⁸. A molecular replacement (MR) search model was generated by SCULPTOR⁵⁹ using the crystal structure of a family GH3 N-acetylglucosaminidase (PDB ID: 3BMX) and a pairwise sequence alignment of the search model with BglP generated in GENEIOUS 8.1.⁷⁰. MR was carried out using PHENIX.PHASER⁶¹ to generate initial phase estimates for reflections collected from a crystal of BglP bound to 2FGlc, followed by model building using PHENIX.AUTOBUILD⁶¹ and iterative model improvement using COOT⁶² and PHENIX.REFINE⁶¹. A model of 2FGlc covalently linked to BglP was fitted into its ascribed density using COOT and restrained during refinement using geometric restraints generated by PHENIX eLBOW⁶¹. Initial phase estimates for native BglP were obtained using the refined 2FGlc-bound model from which 2FGlc and solvent had been removed as an MR search model in PHENIX.PHASER followed by iterative model building and refinement using COOT and PHENIX.REFINE.

A model of bound 2FGlc1P was built based on the GlcNAc-MurNAc substrate bound to a GH3 NagZ from Bacillus subtilis (PDB ID: 4GYJ) and 2-deoxy-2-fluoro-α-D-glucose-1-phosphate (gfp_msd.pdb) from the HIC-up database.⁶³ MurNAc was replaced in the GlcNAc-MurNAc substrate with the phosphate group of gfp while maintaining the distorted conformation of GlcNAc and pseudoaxial orientation of the glycosyl ester linkage connecting C1 of the sugar to phosphate. The 2-acetamide group of GlcNAc was then replaced with a fluorine to complete the model, which was fitted into the active site using the experimentally determined BglP-2FGlc complex as a guide.

**Accession numbers** – Fosmid sequences for 29K06 (MF625023) and 31P01 (MF625024) can be found in GenBank. Coordinate files and structure factors have been deposited to the Protein Data Bank.
GH3 β-glycoside phosphorylase from a metagenomic library

Bank under accession codes 5VQD for native BglP and 5VQE for BglP bound to 2FGlc, respectively.

Acknowledgements – We thank Alexandra Nastasa, Jason Rogalski and Sam Kheirandish for technical assistance and Drs Hongming Chen and Feng Liu for reagents and Zachary Armstrong for numerous helpful discussions on this project. S.S.M is supported by Ecosystem Services, Commercialization Platforms and Entrepreneurship (ECOSCOPE), funded through NSERC CREATE. B.L.M is supported by Natural Sciences and Engineering Research Council of Canada (NSERC) grant RGPIN-2015-05310 and A.P. holds a studentship award from Research Manitoba. S.G.W. and S.J.H. are supported by Discovery grants from NSERC as well as grants from the Canada Foundation for Innovation. S.G.W. thanks the Canada Council for the Arts and the Killam foundation for salary support.

Conflict of interest statement – The authors declare that they have no conflicts of interest with the contents of this article.
REFERENCES

1. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490-D495
2. Moure, A., Guillón, P., Domínguez, H., and Parajó, J. C. (2006) Advances in the manufacture, purification and applications of xylo-oligosaccharides as food additives and nutraceuticals. Process Biochem. 41, 1913–1923
3. Buchholz, K., and Seibl, J. (2008) Industrial carbohydrate biotransformations. Carbohydr. Res. 343, 1966-1979
4. Kumar, V., Marín-Navarro, J., and Shukla, P. (2016) Thermostable microbial xylanases for pulp and paper industries: trends, applications and further perspectives. World J. Microbiol. Biotechnol. 32, 34
5. Valdés, A., and Garrigós, M. C. (2016) Carbohydrate-based advanced biomaterials for food sustainability: A review. Mater. Sci. Forum 842, 182-195
6. Palcic, M. M. (2011) Glycosyltransferases as biocatalysts. Curr. Opin. Chem. Biol. 15, 226-233
7. O'Neill, E. C., and Field, R. A. (2015) Enzymatic synthesis using glycoside phosphorylases. Carbohydr. Res. 403, 23-37
8. Luley-Goedl, C., and Nidetzky, B. (2010) Carbohydrate synthesis by disaccharide phosphorylases: reactions, catalytic mechanisms and application in the glycosciences. Biotechnol. J. 5, 1324-1338
9. Nakai, H., Kitaoka, M., Svensson, B., and Ohtsubo, K. (2013) Recent development of phosphorylases possessing large potential for oligosaccharide synthesis. Curr. Opin. Chem. Biol. 17, 301-309
10. Kitaoka, M., and Hayashi, K. (2002) Carbohydrate-Processing Phosphorolytic Enzymes. Trends Glycosci. Glycotechnol. 14, 35-50
11. Zhang, Y.-H. P. H., and Lynd, L. R. (2004) Kinetics and relative importance of phosphorylolytic and hydrolytic cleavage of cellodextrins and cellobiose in cell extracts of Clostridium thermocellum. Appl. Environ. Microbiol. 70, 1563-1569
12. Kitaoka, M., Sasaki, T., and Taniguchi, H. (1992) Phosphorolytic Reaction of Cellvibrio gilvus Cellobiose Phosphorylase. Biosci. Biotechnol. Biochem. 56, 652-655
13. Ladevèze, S., Tarquis, L., Cecchini, D. A., Bercovici, J., André, I., Topham, C. M., Morel, S., Laville, E., Monsan, P., Lombard, V., Henrissat, B., and Potocki-Véronèse, G. (2013) Role of glycoside phosphorylases in mannose foraging by human gut bacteria. J. Biol. Chem. 288, 32370-32383
14. Cori, G. T., Colowick, S. P., and Cori, C. F. (1938) The formation of glucose-1-phosphoric acid in extracts of mammalian tissues and of yeast. J. Biol. Chem. 123, 375-380
15. Kitaoka, M. (2015) Diversity of phosphorylases in glycoside hydrolase families. Appl. Microbiol. Biotechnol. 99, 8377-8390
16. Zhang, Y.-H. P. H., and Lynd, L. R. (2005) Cellulose utilization by Clostridium thermocellum: bioenergetics and hydrolysis product assimilation. Proc. Natl. Acad. Sci. U.S.A. 102, 7321-7325
17. de Kok, S., Yilmaz, D., Suir, E., Pronk, J. T., Daran, J.-M. M., and van Maris, A. J. (2011) Increasing free-energy (ATP) conservation in maltose-grown Saccharomyces cerevisiae by expression of a heterologous maltose phosphorylase. Metab. Eng. 13, 518-526
18. Handelsman, J., Rondon, M. R., Brady, S. F., and Clardy, J. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5, R245-R249
19. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68, 669-685
20. Riesenfeld, C. S., Schloss, P. D., and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. Annu. Rev. Genet. 38, 525-552
21. Nyysönen, M., Tran, H. M., Karaoz, U., Weihe, C., Hadi, M. Z., Martiny, J. B., Martiny, A. C., and Brodie, E. L. (2013) Coupled high-throughput functional screening and next generation
GH3 β-glycoside phosphorylase from a metagenomic library

sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries. Front. Microbiol. 4, 1-14

22. Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., Cayouette, M., McHardy, A. C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S. G., Podar, M., Martin, H. G., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., Szeto, E., Salamov, A., Barry, K., Mikhailova, N., Kyrpides, N. C., Matson, E. G., Ottesen, E. A., Zhang, X., Hernández, M., Murillo, C., Acosta, L. G., Rigoutsos, I., Tamayo, G., Green, B. D., Chang, C., Rubin, E. M., Mathur, E. J., Robertson, D. E., Hugenholtz, P., and Leadbetter, J. R. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450, 560-565

23. Hess, M., Sezýrba, A., Egan, R., Kim, T.-W. W., Chokhawala, H., Schroth, G., Luo, S., Clark, D. S., Chen, F., Zhang, T., Mackie, R. I., Pennacchio, L. A., Tringe, S. G., Visel, A., Woyke, T., Wang, Z., and Rubin, E. M. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331, 463-467

24. Berini, F., Casciello, C., Marcone, G. L., and Marinelli, F. (2017) Metagenomics: novel enzymes from non-culturable microbes. FEMS Microbiol. Lett. 364

25. Mewis, K., Armstrong, Z., Song, Y. C., Baldwin, S. A., Withers, S. G., and Hallam, S. J. (2013) Biomining active cellulases from a mining bioremediation system. J. Biotechnol. 167, 462-471

26. Mazzella, J. L., Parkin, D. W., Tyler, P. C., Furneaux, R. H., and Schramm, V. L. (1996) Mechanistic diagnoses of N-ribohydrolases and purine nucleoside phosphorylase. J. Am. Chem. Soc. 118, 2111-2112

27. Lairson, L. L., Wakachuk, W. W., and Withers, S. G. (2007) Alternative donor substrates for inverting and retaining glycosyltransferases. Chem. Commun. (Camb.), 365-367

28. Pearce, P. J., and Simkns, R. J. J. (1968) Acid strengths of some substituted picric acids. Can. J. Chem.

29. Sih, C. J., Nelson, N. M., and McBee, R. (1957) Biological synthesis of celllobiose. Science 126, 1116-1117

30. Sheth, K., and Alexander, J. K. (1969) Purification and properties of beta-1,4-oligoglucan:orthophosphate glucosyltransferase from Clostridium thermocellum. J. Biol. Chem. 244, 457-464

31. Mayer, C., Vocadlo, D. J., Mah, M., Rupitz, K., Stoll, D., Warren, R. A., and Withers, S. G. (2006) Characterization of a beta-N-acetylhexosaminidase and a beta-N-acetylglucosaminidase/beta-glucosidase from Cellulomonas fimi. FEBS J. 273, 2929-2941

32. Macdonald, S. S., Blaukopf, M., and Withers, S. G. (2015) N-Acetylglucosaminidases from CAZy Family GH3 Are Really Glycoside Phosphorylases, Thereby Explaining Their Use of Histidine as an Acid/Base Catalyst in Place of Glutamic Acid. J. Biol. Chem. 290, 4887-4895

33. Vocadlo, D. J., Mayer, C., He, S., and Withers, S. G. (2000) Mechanism of action and identification of Asp242 as the catalytic nucleophile of Vibrio furnisii N-acetyl-beta-D-glucosaminidase using 2-acetamido-2-deoxy-5-fluoro-alpha-L-idopyranosyl fluoride. Biochemistry 39, 117-126

34. Ulusu, N. N. (2015) Evolution of enzyme kinetic mechanisms. J. Mol. Evol. 80, 251-257

35. Sun, W. C., Gee, K. R., and Haugland, R. P. (1998) Synthesis of novel fluorinated coumarins: Excellent UV-light excitable fluorescent dyes. Bioorg. Med. Chem. Lett.

36. Withers, S. G., Rupitz, K., and Street, I. P. (1988) 2-Deoxy-2-fluoro-D-glycosyl fluorides. A new class of specific mechanism-based glycosidase inhibitors. J. Biol. Chem. 263, 7929-7932

37. Litzinger, S., Fischer, S., Polzer, P., Diederichs, K., Welte, W., and Mayer, C. (2010) Structural and kinetic analysis of Bacillus subtilis N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. J. Biol. Chem. 285, 35675-35684

38. Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. Acc. Chem. Res. 33, 11-18

39. Davies, G. J., Planas, A., and Rovira, C. (2012) Conformational analyses of the reaction coordinate of glycosidases. Acc. Chem. Res. 45, 308-316
GH3 β-glycoside phosphorylase from a metagenomic library

40. Bacik, J.-P. P., Whitworth, G. E., Stubbs, K. A., Vocadlo, D. J., and Mark, B. L. (2012) Active site plasticity within the glycoside hydrolase NagZ underlies a dynamic mechanism of substrate distortion. *Chem. Biol.* **19**, 1471-1482

41. Ducatti, D. R. B., Carroll, M. A., and Jakeman, D. L. (2016) On the phosphorylase activity of GH3 enzymes: A β-N-acetylg glucosaminidase from Herbaspirillum seropedicae SmR1 and a glucosidase from Saccharopolyspora erythraea. *Carbohydr. Res.* **435**, 106-112

42. Gabor, E. M., and Alkema, W. B. L. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* **6**, 879-886

43. Warren, R. L. L., Freeman, J. D., Levesque, R. C., Smailus, D. E., Flibotte, S., and Holt, R. A. (2008) Transcription of foreign DNA in Escherichia coli. *Genome Res.* **18**, 1798-1805

44. Gaida, S. M., Sandoval, N. R., Nicolaou, S. A., Chen, Y., Venkataramanan, K. P., and Papoutsakis, E. T. (2015) Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. *Nat. Commun.* **6**, 7045

45. Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* **5**, 593-599

46. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254

47. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411-2423

48. Klock, H. E., Koesema, E. J., Knuth, M. W., and Lesley, S. A. (2008) Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* **71**, 982-994

49. Chung, C. T., Niemela, S. L., and Miller, R. H. (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2172-2175

50. Bolger, A. M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120

51. Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **00**, 1-3

52. Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676

53. Treangen, T. J., Sommer, D. D., Angly, F. E., Koren, S., and Pop, M. (2011) Next generation sequence assembly with AMOS. *Curr. Protoc. Bioinformatics* **11**, 11.18

54. Hyatt, D., Chen, G.-L. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119

55. Konwar, K. M., Hanson, N. W., Bhatia, M. P., Kim, D., Wu, S.-J. J., Hahn, A. S., Morgan-Lang, C., Cheung, H. K., and Hallam, S. J. (2015) MetaPathways v2.5: quantitative functional, taxonomic and usability improvements. *Bioinformatics* **31**, 3345-3347

56. Vocadlo, D. J., and Withers, S. G. (2000) Identification of active site residues in glycosidases by use of tandem mass spectrometry. *Methods Mol. Biol.* **146**, 203-222

57. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr.* **D67**, 271-281

58. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr.* **D62**, 72-82

59. Bunkóczi, G., and Read, R. J. (2011) Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr.* **D67**, 303-312

60. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012)
GH3 β-glycoside phosphorylase from a metagenomic library

Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649

61. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* **D66**, 213-221

62. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr.* **D66**, 486-501

63. Kleywegt, G. J. (2007) Crystallographic refinement of ligand complexes. *Acta Crystallogr.* **D63**, 94-100
Table 1: Kinetic parameters for the reaction of BglP with pNPGlcNAc, pNPGlc and DNPGlc. Reactions were carried out in storage buffer B and the indicated concentration of potassium phosphate, pH 7.0, at 25 °C. Molar extinction coefficients at 400 nm: pNP = 7280 M⁻¹ cm⁻¹ and DNP = 12,460 M⁻¹ cm⁻¹. Parameters were calculated using the following: \( V_0 = [E] \times k_{cat} \times [S] / (K_m + [S]) \). Standard error was calculated from 3 replicates. ND: not determined.

| Substrate  | [Pi] mM | \( K_m \) mM | \( k_{cat} \) s⁻¹ | \( k_{cat}/K_m \) mM⁻¹ s⁻¹ |
|------------|---------|---------------|-----------------|---------------------|
| pNPGlcNAc  | 0       | 0.3 ± 0.1     | 0.002 ± 0.0001  | 0.006               |
|            | 5       | 0.4 ± 0.1     | 0.002 ± 0.0003  | 0.005               |
|            | 10      | 0.6 ± 0.1     | 0.002 ± 0.0004  | 0.004               |
|            | 25      | 1.1 ± 0.4     | 0.004 ± 0.001   | 0.003               |
|            | 50      | 0.8 ± 0.2     | 0.003 ± 0.0005  | 0.004               |
|            | 100     | 0.9 ± 0.3     | 0.003 ± 0.0007  | 0.003               |
|            | 200     | 1.5 ± 0.5     | 0.004 ± 0.001   | 0.003               |
| pNPGlc     | 0       | 3 ± 1         | 0.22 ± 0.03     | 0.065               |
|            | 0.01    | 5 ± 1         | 0.32 ± 0.02     | 0.072               |
|            | 0.05    | 9 ± 1         | 0.55 ± 0.03     | 0.065               |
|            | 0.5     | 53 ± 5        | 2.5 ± 0.2       | 0.047               |
|            | 5       | 67 ± 11       | 3.2 ± 0.4       | 0.048               |
|            | 10      | 55 ± 5        | 2.4 ± 0.2       | 0.045               |
|            | 100     | 56 ± 7        | 2.3 ± 0.2       | 0.041               |
| DNPGlc     | 0       | 0.06 ± 0.01   | 0.7 ± 0.01      | 10.7                |
|            | 0.01    | 0.06 ± 0.01   | 0.7 ± 0.02      | 11.1                |
|            | 0.05    | 0.17 ± 0.01   | 1.7 ± 0.03      | 9.8                 |
|            | 0.1     | 0.29 ± 0.02   | 3 ± 0.1         | 9.1                 |
|            | 0.5     | 2.1 ± 0.2     | 15 ± 1          | 7.1                 |
|            | 5       | ND            | ND              | 7.6                 |
|            | 25      | ND            | ND              | 7.4                 |
|            | 50      | ND            | ND              | 7.4                 |
Table 2: Kinetic parameters for the reaction of BglP.Q205S with pNPGlcNAc, pNPGlc and DNPGlc. Reactions were carried out in storage buffer B supplemented with 0 or 50 mM phosphate, pH 7.0, at 25 °C. Molar extinction coefficients at 400 nm: pNP = 7280 M⁻¹ cm⁻¹ and DNP = 12,460 M⁻¹ cm⁻¹. Parameters were calculated using the following: $V_0 = [E] \cdot k_{cat} \cdot [S]/(K_m + [S])$. ND: not determined.

| Substrate | [Pi] | $K_m$ | $k_{cat}$ | $k_{cat}/K_m$ |
|-----------|------|-------|-----------|---------------|
| pNPGlcNAc | 0    | 0.40 ± 0.02 | 0.030 ± 0.0005 | 0.074 |
|           | 50   | 0.41 ± 0.02 | 0.030 ± 0.0004 | 0.072 |
| pNPGlc    | 0    | ND     | ND        | 0.005 |
|           | 50   | ND     | ND        | 0.006 |
| DNPGlc    | 0    | 0.05 ± 0.02 | 0.03 ± 0.002 | 0.6 |
|           | 50   | 2.5 ± 0.1 | 1.94 ± 0.03 | 0.8 |
Table 3: BglP crystallographic statistics table.

| Crystal | BglX (PDB ID: 5VQD) | BglX-2FGlc (PDB ID: 5VQE) |
|---------|----------------------|----------------------------|
| **X-ray source** | Rigaku R-AXIS IV++ | Rigaku R-AXIS IV++ |
| **Crystal geometry** | | |
| Space group | C222₁ | C222₁ |
| a=80.52 b=84.15 | a=80.39 b=83.95 |
| c=161.19; | c=161.27; |
| α=β=γ=90° | α=β=γ=90° |
| **Crystallographic data** | | |
| Wavelength (Å) | 1.5418 | 1.5418 |
| Resolution range (Å) | 54.72-2.10 (2.16-2.10) * | 33.12-1.89 (1.96-1.89) |
| Total observations | 310027 (27157) | 292494 (16456) |
| Unique reflections | 38928 (2647) | 43229 (2693) |
| Multiplicity | 10.1 (10.3) | 6.8 (6.1) |
| Completeness (%) | 94.6 (99.9) | 98.4 (96.1) |
| Rmerge | 0.112 (0.590) | 0.073 (0.235) |
| CC1/2 | 0.99 (0.95) | 0.99 (0.96) |
| I/σI | 15.7 (4.0) | 15.0 (6.9) |
| Wilson B-factor (Å²) | 13.89 | 16.79 |
| **Refinement statistics** | | |
| Reflections in test set | 1964 (208) | 1996 (196) |
| Non-hydrogen atoms: | | |
| Protein atoms | 4305 | 4334 |
| 2FGlc atoms | - | 11 |
| Water | 468 | 366 |
| Glycerol | 12 | - |
| Rwork/Rfree | 0.21/0.26 | 0.17/0.21 |
| RMSDs | | |
| Bond lengths/angles (Å/°) | 0.002/0.57 | 0.005/0.81 |
| Ramachandran plot | | |
| Favored/allowed/outliers (%) | 95.45/3.45/1.09 | 96.03/3.25/0.72 |
| Average B factor (Å²) | 16.30 | 20.31 |
| Macromolecules | 15.88 | 19.77 |
| Ligands | 23.41 | 13.04 |
| Solvent | 19.92 | 26.94 |

*Values in parentheses refer to the highest resolution shell
FIGURE LEGENDS

Figure 1: (A) Generalized phosphorolysis (upper pathway) and hydrolysis (lower pathway) of β1–4 linked glycans. R = H or glucose_n (n = number of glucose residues). GPs are distinguished from GHs by their use of a phosphate molecule to cleave the glycosidic linkage thereby producing a free sugar-1-phosphate. (B) Phosphorolysis of cellobiose or cello-oligosaccharides performed by the inverting phosphorylases RtCBP (R = H) and RtCDP (R = glucosene). (C) Proposed mechanistic scheme for a retaining β-glycosidase/phosphorylase. BglP and Nag3 both employ a double-displacement β-retaining mechanism involving a glycosyl-enzyme intermediate and act as preferential phosphorylases (k_{3p} > k_{3w}). Both enzymes possess the same active site residues which act as the catalytic nucleophile (Asp) and the acid/base catalytic dyad (Asp and His).

Figure 2: Phosphate-dependent cleavage of DNPG by RtCBP, RtCDP and Nag3. (A) RtCBP, RtCDP or Nag3 was incubated with 2 mM DNPGlc in 0 and 50 mM phosphate for 1 h at 37 °C then absorbance was measured at 400 nm. Error bars indicate standard deviation (n = 3). (B) RtCBP, RtCDP or Nag3 was incubated with 20 mM DNPGlc in 0 and 50 mM phosphate and incubated for 2 h at 37 °C then samples spotted on TLC. The glycerol visible on the TLC plate is from the enzyme storage buffer. The higher activity with phosphate present were analysed for the production of glucose-1-phosphate using TLC.

Figure 3: Overview of the GP functional metagenomic screening method. WHOLE LIBRARY: Metagenomic library in 384-well plate format was screened in the presence of DNPGlc and phosphate. Clones that gave an A_{400} value greater than mean + 4 SD were consolidated and re-arrayed into a master library. CONSOLIDATION: Master library was screened with DNPGlc in the presence (x3) and absence (x3) of phosphate. ANALYSIS: A Student’s t-test was performed between the averages of the A_{400} values of each clone in the absence and presence of phosphate. Clones possessing a significantly (p < 0.05) higher activity with phosphate present were analysed for the production of glucose-1-phosphate using TLC.

Figure 4: (A) Reaction underlying the functional metagenomic screen. (B) 400 nm absorbance measurements of 17168 metagenomic clones assayed in the presence of phosphate. White dashed line indicates mean (0.66); blue dashed line indicates mean + 4 SD (0.98). 54 clones displayed an A_{400} value greater than the mean + 4 SD cutoff, indicating cleavage of the DNPGlc substrate. (C) 400 nm absorbance measurements of 54 active clones in the presence and absence of 50 mM phosphate. Each clone was assayed in triplicate with 0 (red) or 50 mM (blue) potassium phosphate. Student’s t-test was performed and those clones meeting the 95 % confidence threshold and showing positive phosphate-dependent activity toward DNPGlc were selected for validation (black asterix). Error bars indicate standard deviation. (D) TLC analysis of the 12 clones displaying phosphate-dependent activity towards DNPG. Cell lysates were treated with 20 mM DNPGlc and 0 or 50 mM potassium phosphate then incubated for 2 h at 37 °C. S lane: 20 mM G1P, 20 mM Glc, 20 mM DNPG.

Figure 5: Amino acid alignment of Cellulomonas fimi Nag3 and BglP. The conserved GH3 hexosaminidase subgroup sequence motif is indicated (*), bold letters) and the catalytic nucleophile residues are denoted (●). Black shading indicated highly conserved residues and grey shading indicates conserved similar residues. Sequence alignment was generated using T-Coffee (http://tcoffee.crg.cat/apps/tcoffee/do:regular) and shading was done in BoxShade v3.21 (http://www.ch.embnet.org/software/BOX_form.html) by K. Hofmann and M. Baron.

Figure 6: Phosphate activation of BglP-mediated cleavage of (A) DNPGlc, (B) pNPGLc and (C) pNPGLcNAc. Michaelis-Menten plots of initial hydrolysis/phosphorolysis rates of (A) 1.5 mM DNPGlc, (B) 50 mM pNPGLc and (C) 5 mM pNPGLcNAc in the presence of increasing concentrations of phosphate. Data were fit to the Michaelis-Menten equation \( V_0 = [E]\cdot k_{cat}\cdot [S]/(K_m + [S]) \). Error bars indicate standard error of the mean; where bars are not present, the error is smaller than the symbols used. Errors were derived from the fit to the experimental data provided by GraphPad. Kinetic parameters for cleavage
of the aryl glycoside substrates at each concentration of phosphate can be found in Table 1. (D) Pre-
steady state burst phase analysis of the cleavage of DFMUGlc (red) and DFMUGlcNAc (grey).
DFMUGlc (10 μM) or DFMUGlcNAc (10 μM) were incubated in a fluorimeter pre-chilled to 12 °C.
DFMU release was measured as a function of time after the addition of 42 nM (for DFMUGlc) or 420 nM
(for DFMUGlcNAc) BglP at approx. 1 min. Fluorescent signal was converted to DFMU concentration
using a standard curve of free DFMU at pH 7.0.

Figure 7: Trapping the BglP covalent intermediate (A) Electrospray mass spectra of BglP, BglP + 10 mM
DNPGlc and BglP + 10 mM DNPGlc + 50 mM phosphate (pH 7.0). The single peak in the spectrum
containing untreated BglP at 64592.9 Da is consistent with the calculated mass of BglP (64589.2 Da).
The mass shift of 162.8 Da between the two peaks in the BglP + DNPGlc panel is consistent with the heavier peak representing the glycosyl-enzyme intermediate (BglP + 163 Da). When phosphate is present no peak shift can be detected, presumably because phosphate induces the intermediate to turn over too quickly.
(B) BglP inactivation with DNP2FGlc. BglP was incubated with increasing concentrations of the
inactivator, DNP2FGlc (●, 0 mM; ■, 2 mM; ▲, 5 mM; ▼, 10 mM; ○, 20 mM; □, 25 mM; △, 50 mM;
▽, 75 mM) for 120 min at 25 °C. Activity was measured at the indicated time points in 50 mM pNPGlc and measuring ABS @ 400 nm over 5 min. (C) Reactivation of the 2FGlc-BglP intermediate with
phosphate. Inactive BglP (50 mM DNP2FGlc for 210 min at 25 °C) was reactivated by incubating the
2FGlc-BglP intermediate with increasing concentrations of phosphate (●, 0 mM; ■, 1 mM; ▲, 5 mM;
▼, 10 mM; ○, 25 mM; □, 50 mM; △, 100 mM). Activity was measured at the indicated time points as
described above. Errors were derived from the fit to the experimental data provided by GraphPad.

Figure 8: Crystal structure of BglP covalently bound to 2FGlc. (A) Overall structure of BglP. The
catalytic N-terminal TIM-barrel domain (a.a. 1-392) is coloured grey, whereas the C-terminal domain (a.a. 393-562) is coloured green. (B) 2FGlc (green carbons) covalently bound to Asp288 within the BglP active site (grey carbons). The electron density Sigma-A-weighted 2Fobs-Fcalc density map contoured at 0.87 e/Å³. (C) Superimposition of the BglP active site (grey) with the active sites of NagZ enzymes from
Bacillus subtilis (BsNagZ) (orange carbons) (PDB: 3NVD)³⁷ and Burkholderia cenocepacia (BcNagZ) (pink carbons) (PDB: 4GNV). The overlay reveals that Gln205 in BglP is replaced by Ser residues in the NagZ enzymes, enabling them to accommodate the C2 N-acetyl group of GlcNAc.

Figure 9: Phosphate activation of BglP.Q205S mediated cleavage of (A) DNPGlc, (B) pNPGlc and (C)
pNPGlcnAc. Michaelis-Menten plots of initial hydrolysis/phosphorolysis rates of (A) 1.5 mM DNPGlc,
(B) 50 mM pNPGlc and (C) 5 mM pNPGlcnAc in the presence of increasing concentrations of
phosphate. The rate data were fit using the Michaelis-Menten equation (A) \( V_0 = [E] \cdot k_{cat} \cdot [S]/(K_{m} + [S]) \) or the Michealis-Menten equation with substrate inhibition (B and C) \( V_0 = [E] \cdot k_{cat} \cdot [S]/(K_m + [S]) \cdot (1 + [S]/K_i) \).
Error bars indicate standard error of the mean; where bars are not present, the error is smaller than the symbols used. Errors were derived from the fit to the experimental data provided by GraphPad. Kinetic parameters for cleavage of the sugar phosphates in 0 and 50 mM phosphate can be found in Table 2. (D) Histogram showing \( k_{cat}/K_m \) values for hydrolysis of pNPGlc and pNPGlcnAc by BglP and BglP-Q205S.

Figure 10: Modeling G1P (yellow carbons) into the BglP active site (grey carbons). Repulsive interactions between Met292 and the phosphate of G1P, in conjunction with the favourable binding interactions between the substrate glucosyl and phosphate moieties is predicted to force the phosphate of bound G1P product into a pseudo-axial position, where the oxygen atom of the phosphoester linkage approaches within ~3.2 Å of the imidazole of His206, which is predicted to be where the oxygen of the glycosidic bond of the substrate would reside when it is protonated during bond cleavage. The experimentally determined 2FGlc molecule covalently bound to Asp288 is shown (green carbons).
FIGURES

Figure 1
Figure 2

A

\[ A_{400} \]

\begin{tabular}{c|c|c|c|c|c}
\hline
\text{\( P_i \)} & \text{no enz.} & \text{CBP} & \text{CDP} & \text{Nag3} & \text{GH94} & \text{GH3} \\
\hline
\text{-} & 0 & \, & \, & \, & \, & \, \\
\text{+} & 2 & 3 & 2 & 1 & 4 & 3 \\
\hline
\end{tabular}

B

- DNPG
- Glycerol
- Glc
- G1P

\begin{tabular}{c|c|c|c|c|c}
\hline
\text{\( P_i \)} & \text{no enz.} & \text{CBP} & \text{CDP} & \text{Nag3} & \text{GH94} & \text{GH3} \\
\hline
\text{-} & 0 & \, & \, & \, & \, & \, \\
\text{+} & 2 & 3 & 2 & 1 & 4 & 3 \\
\hline
\end{tabular}
GH3 β-glycoside phosphorylase from a metagenomic library

Figure 3

50 mM phosphate

Rearray Clones: $A_{400} > \text{mean} + 4 \text{ SD}$

0 mM phosphate

50 mM phosphate

$A_{400}[0]$ vs. $A_{400}[50]$

TLC Validation

Sequence GP Positive Clones
GH3 β-glycoside phosphorylase from a metagenomic library

Figure 4

A

B

C

D
Figure 5

Nag3  1  MI--LAAE---SLDEGGIAMRVTAEABGDEKLGLOCLTTIDSDPEYLDGTEGLHLV
BglX  1  VYTSIIKEKSLNSKQTVNNTLVSDDFKLGCQGHDIKFPGDPLDWMHFLPG

Nag3  59  VNYLTMTRADAATVTTLSTATPLLANSLEDGQ---QTGTFTHQSNADAAAT
BglX  61  VNYLPFKRKRMEFTQRGKASKPLLACNLQEGGGNQQTRVWSPSAAT

Nag3  116  STDHVRPAATPAEAEAPVANWATEKEDKPNPITNVIKGAAPAATV
BglX  121  DERSAEELNCIISSAAANIWTPETKIDKPNPITNVIKSIFPERITMPKAC

Nag3  176  ETIQACGIAAAKHFGDGDVDEEQHLSVNTDVDWDDGTVVRDIVAAAVHTTVV
BglX  181  RCRKVECFLTKPVFGDGYRAQIDKDTRPLLEAVRILKAKAGHH

Nag3  236  CSHLSSYFARESIAKRIIIKELGCYQVESITMACLASVTS
BglX  241  HSCQSSNRNSSLKSQKELMGKLGFRPHNGLCQACQXXVTCSYF

Nag3  296  RSCAQFRV1AAACMFLKTENLEDFGARTGDRCCRPRLDEAVRILKEAKKDH
BglX  301  RHEAPSTSLQCNWLTPTLNPEDKQEOQSCSQTRBLDEAVRILKAKKLP

Nag3  356  ESTNFAQGGRAGVAPDPLSATEERASSIVLAK--EGLFPRREYRYRDLQN
BglX  361  EKRSTPPLHAMERQKSSKKKWDKADESIIVLAKGKLSPQCTRILVQAT

Nag3  414  GGSPIFICGGRAGAVECGVDAIFEAIHVTPEPQGWDGMAPTTDTERHLTVYLANL
BglX  421  EKF---EQLYSEARIKGLREKEFIVHFPEV-PRPSTYSIRDKRTDFLVMANF

Nag3  474  STSNSQTVREAWAPCVRANVKEAVHVSPTQFSSNPYHLFQREVPLINTQGSPVV
BglX  477  KVSSNQTTRVAGFLODSSPGCVPTIFSSNPYHLFQREVPLINTQGSPVV

Nag3  534  ETITIAACOAPNFSSPVDAACQNTELE
BglX  537  VMRMIEKLEKSSVESKSPVDPARAIAA
GH3 β-glycoside phosphorylase from a metagenomic library

Figure 6
Figure 7

A

B

C

GH3 β-glycoside phosphorylase from a metagenomic library
Figure 8
Figure 9

A

B

C

D

Figure 9: GH3 β-glycoside phosphorylase from a metagenomic library.
GH3 β-glycoside phosphorylase from a metagenomic library

Figure 10
Structural and mechanistic analysis of a β-glycoside phosphorylase identified by screening a metagenomic library
Spencer S Macdonald, Ankoor Patel, Veronica L.C. Larmour, Connor Morgan-Lang, Steven J Hallam, Brian L. Mark and Stephen G. Withers

J. Biol. Chem. published online January 9, 2018

Access the most updated version of this article at doi: 10.1074/jbc.RA117.000948

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts