JBC Papers in Press. Published on January 9, 2018 as Manuscript RA117.000936
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.RA117.000936

Identification of *Euglena gracilis* β-1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149

Sakonwan Kuhaudomlarp[a], Nicola J. Patron[b], Bernard Henrissat[c,d,e], Martin Rejzek[a], Gerhard Saalbach[a] & Robert A. Field[a,1]

[a]Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. [b]Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK. [c]Aix-Marseille University, Architecture et Fonction des Macromolécules Biologiques, 163 Avenue de Luminy, 13288 Marseille, France. [d]Centre National de la Recherche Scientifique, UMR 7257, 163 Avenue de Luminy, 13288 Marseille, France. [e]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

Running title: New β-1,3-glucan phosphorylase family

[1]To whom correspondence should be addressed: Robert A. Field: Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. rob.field@jic.ac.uk.

Keywords: *Euglena gracilis*, β-1,3-glucan, phosphorylase, Gram-negative Bacteria, orthologs, PULs

Abstract

Glycoside phosphorylases (EC 2.4.x.x) carry out the reversible phosphorolysis of glucan polymers, producing the corresponding sugar 1-phosphate and a shortened glycan chain. β-1,3-Glucan phosphorylase activities have been reported in the photosynthetic Euglenozoan *Euglena gracilis*, but the cognate protein sequences have not been identified to date. Continuing our efforts to understand the glycobiology of *E. gracilis*, we identified a candidate phosphorylase sequence, designated EgP1, by proteomic analysis of an enriched cellular protein lysate. We expressed recombinant EgP1 in *E. coli* and characterized it *in vitro* as a β-1,3-glucan phosphorylase. BLASTP identified several hundred EgP1 orthologs, most of which belong to Gram-negative bacteria, and have 37-91% sequence identity to EgP1. We heterologously expressed a bacterial metagenomic sequence, Pro_7066 in *E. coli* and confirmed it as a β-1,3-glucan phosphorylase, albeit with kinetics parameters distinct from those of EgP1. EgP1, Pro_7066, and their orthologs are classified as a new glycoside hydrolase (GH) family, designated GH149. Comparisons between GH94, EgP1 and Pro_7066 sequences revealed conservation of key amino acids required for the phosphorylase activity, suggesting a phosphorylase mechanism that is conserved between GH94 and GH149. We found bacterial *GH149* genes in gene clusters containing sugar transporter and several other GH family genes, suggesting that bacterial GH149 proteins have roles in the degradation of complex carbohydrates. The Bacteroidetes *GH149* genes located to previously identified polysaccharide utilization loci (PULs), implicated in the degradation of complex carbohydrates. In summary, we identified a eukaryotic and a bacterial β-1,3-glucan phosphorylases and uncovered a new family of phosphorylases we name GH149.

Glycoside phosphorylases (EC 2.4.x.x) carry out the phosphorolysis of sugar polymers to produce the corresponding sugar 1-phosphates and shortened glycan polymer chains. However, the reaction is freely reversible *in vitro*, enabling the production of oligosaccharides from a simple sugar 1-phosphate and suitable carbohydrate acceptors with strict regio-, stereo- and chain length specificity (1, 2). Despite the potential for use of the phosphorylases for oligosaccharide production, a relatively small number of these enzymes have been identified, limiting their range of applications. Nevertheless, the number of enzyme accessions are gradually increasing due to the substantial increase in available genome sequences and research efforts into characterization of novel phosphorylase activities.

Linear β-1,3-D-glucan polysaccharides are found across the prokaryotes and eukaryotes (3). The most well-known examples are bacterial curdlan (4), plant callose and microalgal (Euglena) paramylon, all of which are high molecular weight

and water insoluble. Paramylon, curdlan and related compounds have been used in a wide range of applications: as part of anti-tumor and anti-HIV treatments (5, 6), as immune stimulants (7), and as alternative materials to fossil fuel products (8). While the water insolubility of paramylon enables it to accumulate to up to 90% of Euglena cell mass (9), it creates complications for subsequent chemical modification, such as sulfation (10). Therefore, research into the generation of linear, soluble β-1,3-D-glucan polymers *in vitro* is of interest. Enzymatic transglycosylation using a mutant β-1,3-D-endoglucanase (11) or a partially purified β-1,3-D-glucan phosphorylase preparation from *Euglena gracilis* have been used for β-1,3-D-glucan production (12). The latter approach can be coupled with sucrose phosphorylase to generate β-1,3-D-glucan disaccharide, laminaribiose, from sucrose (13). Enzymatic synthesis of β-1,3-D-glucan using Euglena phosphorylases is attractive because of the relatively cheap substrates required (glucose and glucose 1-phosphate). However, the generation of the phosphorylase catalyst is time consuming and laborious because it requires large volume of Euglena cell culture that takes several days to grow. Moreover, the partial purification of the phosphorylase may result in loss of protein yield and attenuation of the activity.

Continuing our efforts to explore the glycobiology of *Euglena gracilis* (14–18) and our investigation into glycoside phosphorylases and their applications in enzymatic synthesis (19, 20), we focused on the established capability of the organism for β-1,3-D-glucan biochemistry, in particular its known β-1,3-D-glucan phosphorylase activities (EC 2.4.1.31) (Fig.1). The enzymes have been categorized into two subgroups, based on the chain length specificity for their acceptor substrates; laminaribiose (β-1,3-glucobiose) phosphorylases (LBPs), laminarin (laminaridextrin) or β-1,3-D-glucan phosphorylases (LDPs). LBP was previously considered to catalyze only a reversible phosphorolysis of laminaribiose (21), whereas LDPs have strong substrate preferences for β-1,3-D-glucans with DP 3 or greater (22). LBP activities were originally discovered in *Euglena gracilis* (23, 24) and *Astasia ocellata* (25), and subsequently in bacteria such as *Paenibacillus* sp. strain YM-1 (26) and *Acholeplasma laidlawii* (27). In contrast, LDP activities have only been described in the Euglenozoan *Euglena gracilis* (28) and in the heterokonts *Ochromonas malhemensis* (29) and *Ochromonas danica* (30). Despite the various reported β-1,3-D-glucan phosphorylase activities, LBP sequences are the only β-1,3-D-glucan phosphorylases that have been successfully cloned to date; they have been classified as members of glycoside hydrolase family GH94 in the CAZy database (www.cazy.org) (31). However, the eukaryote phosphorylases have not, to date, been classified due to the lack of sequence information, which compromises further biochemical and structural studies of this fascinating group of enzymes.

The physiological role of eukaryotic β-1,3-D-glucan phosphorylases remains inconclusive. In contrast, the biological function of the bacterial phosphorylases can be predicted from the identity of other genes with which they are clustered within the genome. LBP from *Paenibacillus* sp. strain YM-1 was hypothesized to be involved in degradation of laminaribiose, based on the position of LBP genetic locus next to ABC sugar transporter components (26). Organization of glycoside phosphorylase genes with other sugar transporters have also been observed. For instance, a sucrose phosphorylase gene in human gut *Bifidobacterium lactis*, a sucrose transporter and a transcriptional regulator form a sucrose-utilization gene cluster which are transcriptionally upregulated upon induction by sucrose and raffinose (32). A β-1,4-D-mannosyl-*N*-acetyl-D-glucosamine phosphorylase was reported as part of a gene cluster involved in *N*-glycan metabolism in *Bacteroidetes thetaiotaomicron* VPI-5482 and was shown to be involved in the phosphorolysis of β-1,4-D-mannosyl-*N*-acetyl-D-glucosamine into α-D-mannose 1-phosphate and *N*-acetyl-D-glucosamine (33). Two GH130 β-1,2-mannoside phosphorylases (Teth514_1788 and Teth514_1789) from *Thermoanaerobacter* sp. X-514 showed different chain length specificities on phosphorolysis of β-1,2-oligomannan. The genes encoding Teth514_1788 and Teth514_1789 are located in a predicted GDP-D-mannose biosynthetic gene cluster. This finding demonstrates a possible role of Teth514_1788 and Teth514_1789 in a novel biosynthetic pathway, in contrast to the predicted degradative roles in other phosphorylases (34).

In the present work, a new family of phosphorylases was uncovered, designated GH149. Two GH149 sequences, a eukaryotic *Euglena gracilis* phosphorylase 1 (EgP1) and a bacterial

metagenome sequence (Pro_7066), have been characterized and confirmed as β-1,3-D-glucan phosphorylases. GH149 contains several hundred sequences, most of which belong to Gram- negative marine bacteria from the phyla Proteobacteria and Bacteroidetes. The Bacteroidetes *GH149* genes map to previously predicted polysaccharide utilization loci (PULs) (35, 36), strongly suggesting a role for these GH149 enzymes in polysaccharide degradation by marine bacteria. Possible evolutionary relationships between the bacterial GH149 and the eukaryote GH149 are discussed based on phylogenetic analysis of the GH149 amino acid sequences and GC content analysis of the *GH149* genes.

Results

Identification of Euglena phosphorylase candidates through proteomic analysis.

It has previously been hypothesized that Euglena β-1,3-glucan phosphorylase sequences may belong to GH94 family based on an activity somewhat similar to the GH94 bacterial laminaribiose phosphorylases (26, 27). To test this hypothesis, we used 20 characterized GH94 amino acid sequences as queries (Supplemental Table S1) to interrogate the translated Euglena transcriptome (14) using tBLASTn (threshold = 0.0001). No sequences were recovered, suggesting that the Euglena phosphorylases may not be closely related to the members of GH94 family as previously hypothesized. This is supported by previous comprehensive Carbohydrate active enzyme (CAZyme) analyses on Euglena transcriptomes (14, 37), which did not identify any GH94 candidates. To identify the Euglena proteins with β-1,3-glucan phosphorylase activity, we performed a fractionation of Euglena protein extracts followed by proteomic analysis of the fractions with active phosphorylase activity**.** Cell-free extracts of 7-day old, dark-grown *E. gracilis* were fractionated through Anion Exchange Chromatography (AIEX) as the first step of partial purification of β-1,3-glucan phosphorylase, based on previous publications (21, 24). The AIEX fractions were screened for phosphorylase activity in a reverse phosphorolysis reaction (Fig. 1) using a phosphate release assay with glucose (Glc) and α-glucose 1-phosphate (G1P) as an acceptor and a donor respectively. The activity of the phosphorylase was detected in two sets of fractions, eluted at 18% w/v NaCl (Peak I) and 24-34 % w/v NaCl (Peak II) (Supplemental Fig. S1A). The oligosaccharide products of the reverse phosphorolysis reactions were analyzed by TLC, which showed production of a disaccharide only from the Peak II catalyzed reactions (Fig. 2A). The phosphorylase activity in Peak II could use laminaribiose (G2) and laminaritriose (G3) as acceptors (Supplemental Fig. S1B), forming higher DP oligosaccharides, which would suggest that this fraction contained an LDP activity.

Iminosugars such as 1-deoxynojirimycin (DNJ) show weak competitive inhibition ($K_i/K_{i'}$ = 570 μM, where K_i and $K_{i'}$ represent the inhibitions against the free enzyme and the enzyme-substrate complex respectively) against the phosphorylase activity of a GH94 cellobiose phosphorylase (CBP) from *Cellvibrio gilvus* (38). To investigate whether there is an inhibition of the Euglena phosphorylase(s) by DNJ, the partially purified enzyme activity in Peak II was incubated with Glc and G1P in the presence of DNJ. DNJ showed weak inhibition in 100-1000 μM range in the reverse phosphorolysis reaction (Fig. 2B, Supplemental Fig. S1B), in keeping with the inhibition reported for the *C. gilvus* CBP.

Immobilized glycomimetic molecules such as *N*-butyldeoxynojirimycin (NB-DNJ) have been used in affinity-enrichment proteomics to identify iminosugar-interacting proteins from mammalian tissue (39). Affinity-enrichment proteomics was performed using *N*-5-carboxypentyl-DNJ immobilized on agarose beads as a glycomimetic probe to identity DNJ-interacting phosphorylase candidates in the Euglena Peak II sample. The sample was applied to the *N*-5-carboxypentyl-DNJ matrix followed by washing with buffer to remove unbound proteins. The binding proteins were eluted with DNJ-containing buffer. The eluted protein fractions were subjected to proteomic LC-MS/MS analysis. From a total of 253 proteins with unique peptide counts > 3 and 100% protein identification probability, 240 were identified in the translated transcriptome database of *Euglena gracilis* (14) using Mascot (Matrix Science, Boston MA, USA). Putative functions of these proteins were assigned by comparison to protein sequences with annotated functions in the non-redundant database NCBI by BLASTP pairwise alignment (Supplemental Data File 1). One protein (m. 14570_dark) showed similarity to a deposited sequence, which was annotated by the depositor as a laminaridextrin phosphorylase-like protein from *Euglena gracilis*

(BAV19384.1, E-value = 0, % identity = 53). The predicted molecular weight of m. 14570_dark is 130 kDa, which is similar to what has been reported for the partially purified Euglena laminaribiose phosphorylase (24). The translated Euglena transcriptome was interrogated using the m. 14570_dark sequence (designated *Euglena gracilis* phosphorylase 1; EgP1) as a BLASTP query and three further sequences were identified with >70% sequence identity to m.14570_dark (m. 14571_dark, 14576_dark and 14578_dark). These three candidate phosphorylases were designated EgP2, EgP3 and EgP4, respectively.

Biochemical characterization of EgP1

To confirm the enzymatic activity of the EgP1-4, the recombinant proteins were produced for biochemical characterization. It was anticipated that heterologous expression of Euglena CAZymes might not be straightforward, given the problems encountered by the Suzuki group when trying to achieve productive expression of Euglena β-1,3-glucanase (40). An *E. coli* protein expression system was chosen for its simplicity and relatively short protein production time. Coding sequences were amplified from synthetic cDNA of EgP1-4 and individually cloned, into the Popin F plasmid vector (41), which were transformed into *E. coli* (Lemmo (DE3)) and expression induced using 0.2 mM IPTG. Only expression of EgP1 was detectable. Initially, EgP1 formed a completely insoluble aggregate, which was purified by washing with buffer containing 6 M urea (Supplemental Fig. S2A). The washed insoluble EgP1 suspension (Supplemental Fig. S2B) unexpectedly showed phosphorylase activity when assayed with Glc and G1P in the reverse phosphorolysis reaction (Supplemental Fig. S2C).

Analysis of the EgP1 amino acid sequence using TargetP version 1.1 (http://www.cbs.dtu.dk/services/TargetP/), predicted the presence of an N-terminal mitochondrial targeting peptide at residues 1-31 (Reliability class = 2), which we hypothesized could be the cause of protein aggregation. A truncated version of the EgP1 coding sequence was re-cloned without the predicted target peptide but with an N-terminal His_6-tag and the recombinant protein expressed as previously described. A soluble EgP1 recombinant protein was detected, which was purified by IMAC and gel filtration (Fig. 3A and Supplemental Fig. S3) with a final yield of ~30 mg per a liter *E. coli* culture.

The soluble EgP1 protein was able to catalyze the conversion of Glc and G1P into oligosaccharides with DP up to 12 glucose units that could be detected by MALDI-ToF after 1 hour (Supplemental Fig. S4 and Table S3). The EgP1 activity was dependent on both Glc and G1P as substrates: excluding either from the reaction mixture completely abolished oligosaccharide production (Fig. 3B). The recombinant EgP1 showed similar catalytic capability over a range of pH (5-8), and a range of temperature (30-40 °C) (Supplemental Fig. S5).

To investigate whether EgP1 activity is specific to β-1,3-linked glucan, phosphorolysis assays (Fig. 1) were performed against a variety of glucose disaccharides with various glycosidic linkages and anomeric configurations. Analysis of these reactions by TLC and High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) showed that EgP1 was active exclusively on the Glc-β-1,3-Glc oligosaccharides, thus confirming its function as a β-1,3-D-glucan phosphorylase (Fig. 3C and D). Kinetic data of the reverse phosphorolysis with Glc, laminaribiose (G2), laminaritriose (G3), laminaritetraose (G4), laminaripentaose (G5) or laminarihexaose (G6) as acceptors and G1P as a donor indicated that the enzyme preferred Glc and G2 as acceptors with similar catalytic efficiency (k_{cat}/K_m = 1.99 and 1.62 for Glc and G2 respectively). There was nearly a two-fold reduction in k_{cat}/K_m as the chain length increased from G2 to G3, which is reflected in a two-fold increase in the K_m values (0.67 for G2 and 1.26 for G3). The catalytic efficiency continued to decrease with increasing chain length from G4 to G6, accompanied by an increase in the K_m values for the respective acceptors (Table 1).

Identification and characterization of a bacterial ortholog of EgP1

A proprietary metagenome database was interrogated with tBLASTN using EgP1 as a query (communication with Prozomix Limited). The search identified a putative ortholog, designated Pro_7066, with 45% sequence identity to EgP1 (E-value = 0). To investigate whether Pro_7066 has similar activity to EgP1, a recombinant form of Pro_7066 was produced in *E. coli* in the same manner as EgP1 (Fig. 4A and Supplemental Fig.

S6). Characterization of Pro_7066 was performed as previously described for EgP1, which showed that the enzyme was also active as a β-1,3-D-glucan phosphorylase (Fig. 4B and C, Supplemental Fig. S7A and S7B), generating oligosaccharide detectable by MALDI-ToF up to DP 11 (Supplemental Fig. S7C and Table S3). Pro_7066 showed similar pH and temperature preferences to EgP1 (Supplemental Fig. S8). Interestingly, the kinetic parameters for the reverse phosphorolysis reaction with Glc-G6 as acceptors and G1P as a donor showed similar catalytic efficiency for all acceptors investigated, which is strikingly different from that observed in EgP1 (Table 2).

Multiple sequence alignment and phylogeny of EgP1 and Pro_7066

The non-redundant protein sequence database (https://www.ncbi.nlm.nih.gov/protein/), the Microbial Eukaryote Transcriptome Sequencing Project (MMET, http://marinemicroeukaryotes.org/) (7.8 x 10^6 sequences), and the *E. gracilis* strain Z transcriptome (http://euglenadb.org/) have been interrogated with PSI-BLASTP and tBLASTN (threshold 0.0001) using the EgP1 amino acid sequence as a query (Supplemental Data File 2). Several hundred sequences of bacterial origin with >30% sequence identity to EgP1 were identified in the non-redundant protein sequence NCBI database. Protein sequences of eukaryotic origin with 50-62% sequence identity to EgP1 (93-97% query coverage, E-value = 0) were identified in the MMET database and *E. gracilis* Z strain transcriptome.

A reconstructed molecular phylogeny showed that these EgP1 orthologous sequences formed a discrete clade (Fig. 5A, red branches), separate from GH94 sequences (Fig. 5A, black branches), thus forming a new glycoside hydrolase family, denominated GH149. In order to generate hypotheses about the active site residues of GH149, we aligned the amino acid sequences of EgP1-4 and Pro_7066 with those of members of family GH94. Interestingly the amino acids that constitute the active site of GH94, including the -1 subsites (G1P binding), Asp catalytic residues and residues for phosphate recognition (Fig. 6, Supplemental Table S2 and Fig. S9) appeared conserved in GH149, suggesting that GH149 and GH94 form a clan that we term GH-Q.

Within the GH149 clade, all sequences from Euglenophyceae share a common node (Fig. 5B, red label, branch support value = 1) comprising EgP1-4, two sequences from *E. gracilis* strain Z (EuglenaDB IDs 750 and 1135, with 50 and 53% sequence identity to EgP1 respectively; E-value = 0), the laminaridextrin phosphorylase-like protein from *E. gracilis* strain NIES-47 and three sequences from *Eutreptiella gymnastica* CCMP1594 identified from MMET (CCMP1594 IDs 200380720, 200416338 and 200418968 with 62, 50 and 51% sequence identity to EgP1, respectively; E-value = 0).

Most bacterial GH149 family members were identified from sequenced and translated genomes of phylum Proteobacteria, dominated by class γ-proteobacteria (Fig. 5B, 134 sequences, purple labels), and phylum Bacteroidetes (Fig. 5B, 125 sequences, green labels). The majority of these sequences are currently annotated as predicted hypothetical proteins with unknown function. Pro_7066 was found to be most closely related to sequences from Bacteroidetes species (Fig. 5B, a black arrow head with Pro_7066 label), in particular *Flavobacteria bacterium* GWF1-32-7 (OGS61795.1, %identity = 76, E-value = 0), suggesting that Pro_7066 might belong to a *Flavobacteria* species. The Bacteroidetes species from which GH149 were found were predominantly isolated from marine environments; 13 have been described to be associated with algae and phytoplankton.

Discussion

β-1,3-D-glucan phosphorylase activities have previously been detected in both bacterial LBPs and eukaryote microalgal LDPs. However, only bacterial LBPs were fully characterized; their sequences identified and categorized into the GH94 family. The lack of sequence information for eukaryotic β-1,3-D-glucan phosphorylases prevented heterologous expression and further characterization. It also prevented enzyme classification based on CAZy criteria and prohibited investigation of their relationship to other GH families. Identification of genetic sequences encoding proteins with β-1,3-D-glucan phosphorylase activity is therefore a crucial step in this study. DNJ-affinity enrichment of *Euglena* protein extract with active β-1,3-D-glucan phosphorylase activity coupled with proteomic analysis was employed to identify a sequence

candidate, designated EgP1. The EgP1 sequence was then used to interrogate a propriety metagenome database, identifying a second bacterial β-1,3-D-glucan phosphorylase candidate, designated Pro_7066. Both proteins were successfully expressed as recombinant proteins in *E. coli*, which allowed us to characterize their functions. Although more closely related to GH94 than to other GH families, EgP1 and Pro_7066 have no significant sequence identity to any characterized GH94 sequences, which suggests that they belong to a new CAZyme family, namely GH149.

In vitro characterization confirmed the function of EgP1 and Pro_7066 as β-1,3-D-glucan phosphorylases. EgP1 and Pro_7066 used Glc and longer β-1,3-linked gluco-oligosaccharides as acceptors in the reverse phosphorolysis reaction. This substrate specificity is slightly different from the previously described β-1,3-D-glucan phosphorylases where substrate chain length was used to distinguish between different groups of the phosphorylases (i.e. LBPs and LDPs). The preference of EgP1 for Glc resembles the activity reported for LBP from semi-purified *Euglena gracilis* extracts (23, 24). However, our findings are in agreement with more recent works by Ogawa *et al.* (12) and Muller *et al.* (13), which described a more relaxed chain length specificity of *Euglena* LBP. Therefore, the chain length specificity of β-1,3-glucan phosphorylases might be more flexible than previously observed. This has also been observed in a recently characterized thermophilic cellodextrin phosphorylase (β-1,4-glucan phosphorylase), which uses Glc and longer cello-oligosaccharide as acceptors (42) and is distinct from GH94 cellodextrin phosphorylases, which have no detectable activity on Glc as an acceptor (19, 43).

Comparisons of the kinetic parameters of EgP1 and Pro_7066 showed distinct preferences for the chain length of sugar acceptors; EgP1 preferred Glc and G2 and the k_{cat}/K_m values declined as the chain length rose to G6. Increasing acceptor chain length led to an increase in K_m values while k_{cat} values remained comparable. This result indicates that substrate binding (K_m) dictates the acceptor chain length preference in EgP1 rather than the turnover (k_{cat}) of the enzyme-substrate complex to product. In contrast, Pro_7066 showed comparable catalytic efficiency as well as k_{cat} and K_m towards all acceptors. The contrasting kinetic parameters of EgP1 and Pro_7066 were surprising considering that the two proteins are 45% identical. Structural studies of the two enzymes will help decipher the mechanistic details that may explain the difference in enzyme behaviors.

GH149 and GH94 members share conserved amino acids that are required for the phosphorylase activity but, overall, are not significantly similar. It is likely that phosphorylase activity in EgP1 and Pro_7066 sequences proceeds via a similar mechanism to that deciphered in GH94 enzymes (44–48). Therefore, GH149 and GH94 form a clan of related families, likely to have evolved from a common ancestor and diversified while retaining its key catalytic apparatus and substrate binding sites.

The GH149 family contains several hundred EgP1 orthologs from bacteria in phylum Bacteroidetes and Proteobacteria and from eukaryotes in phylum Euglenophyta, class Euglenophyceae. Phylogenetic analysis of GH149 sequences revealed a clade of Euglenophyceae sequences from *Euglena gracilis* and *Eutreptiella gymnastica* under a common node. Phylogenetic analyses suggest a common origin for GH149 sequences found in *E. gracilis* (fresh water living) and *E. gymnastica* (sea water living), indicated that these sequences were inherited from a common Euglenophyceae ancestor before topographic isolation of *Euglena* and *Eutreptiella*. The origin of Euglenophyceae GH149 ancestral is unclear, however, the position of the Euglenophyceae clade within bacteria, might suggest a lateral gene transfer from bacteria to the Euglenophyceae. A similar horizontal gene transfer from a bacterium to *E. gracilis* was predicted for mitochondrial trans-2-enoyl-CoA reductase (49). Nevertheless, it is important to note that bacteria are overrepresented in most available sequence databases compared to eukaryotes and therefore the prevalence of the bacterial orthologs in our study is unsurprising. Increasing the availability of sequences from the diversity of unicellular eukaryotes is essential for investigating evolutionary origins.

A predicted targeting peptide in EgP1 caused aggregation of the EgP1 recombinant protein when expressed in *E. coli*. Removing this peptide from the sequence enabled solubility without affecting the expression level or activity of the protein. The targeting peptide length is consistent with the average length of predicted mitochondrial targeting peptides in *Euglena*

previously reported by Krnáčová *et al*. (50), but it is still not known whether the targeting peptide is functional *in vivo*. Further investigation of the sub-cellular localization of the full length EgP1 in *E. gracilis* cells is required to elucidate its biological function. In contrast, no target peptide was predicted for Pro_7066 and the length of this sequence was similar to the putative mature EgP1 protein (Fig. 6). Similarly, no targeting peptide was predicted for the GH149 sequences found in *Eutreptiella*.

Examination of the genomic location of *GH149* open reading frames (ORF) in Proteobacteria revealed their presence in putative gene clusters; Type I clusters containing ORFs for an ABC transporter cassette (a substrate binding protein (COG0747), two permeases (COG0601 and 1173) and two ATP-binding proteins (COG0444 and 4608) and a transcriptional regulator LacI (COG1609) or AraC (COG2207)) (Fig. 7A). The same type of ABC transporter has been reported to be used by the thermophilic Gram-negative bacteria *Thermotoga maritima* for the utilization of extracellular laminarin after degradation by an extracellular laminarase. Oligosaccharide products from laminarase reactions are predicted to be taken up by the ABC transporter into the cytoplasm, where the oligosaccharides could be broken down into glucose by a laminaribiase (51). The lack of gene encoding extracellular laminarase and an outer membrane transporter in Type I clusters suggests that this cluster may be involved in scavenging low molecular weight β-1,3-gluco-oligosaccharides such as laminaritriose/biose using the putative phosphorylase as a catalyst. In our analysis, GH149 sequences were found in many *Vibrio* species including *Vibrio campbellii* HY01, which has recently been reported to express a novel GH3 family β-glucosidase (LamN) capable of digesting laminaribiose (52). The co-occurrence of a GH149 and LamN in *V. campbellii* supports the utilization of β-1,3-glucan by *Vibrio* spp., reinforcing its importance in the carbon cycle of marine environments.

A second genetic co-localization pattern was also observed (designated Type II) mainly in genomes of bacteria in class γ-proteobacteria, order Alteromonadales. Type II clusters contain the respective *GH149* genes co-localized with gene encoding a TonB-dependent receptor (cd01347), which is an outer membrane transporter involved in the transport of iron siderophore, vitamin B_{12}, nickel complexes, and carbohydrates into the periplasmic space (53). Type II clusters also contain genetic components encoding other CAZymes [extracellular GH16 (pfam11721), cytoplasmic GH17 (COG5309)], regulators [histidine kinase (pfam06580) and DNA-binding response regulator (COG3279)], and an inner membrane Major Facilitator Superfamily (MFS) transporter (pfam13347) (Fig. 7B). A more complicated organization of type II cluster suggests that they might be involved in degrading more complex β-1,3-glucan structures than that utilized by type I cluster. Most importantly, the majority of *GH149* genes from Bacteroidetes map to the previously predicted PULs (clusters containing *susCD*-like gene pairs) and CAZyme clusters (clusters containing CAZymes but lack *susCD*-like gene pairs) (Fig. 7C) (35, 54). The common features that are found in these PULs and CAZyme clusters are the presence of genes encoding β-glucosidases (GH3 and/or GH30_1) and glycanases able to hydrolyse β-1,3-glucans: GH17 and/or GH16, thus emphasize the potential contribution of the GH149 in the degradation of complex β-1,3-glucan structures.

The phylogenetic clustering of the bacterial GH149 amino acid sequences on Fig. 5B agreed with the taxonomic relationships of the species, suggesting that these bacterial *GH149* genes are likely inherited through a vertical gene transfer within each species. Comparison of the GC content of 52 bacterial *GH149* genes (Supplemental Data File 3) showed limited deviation from the median GC content of the corresponding complete genome sequences, supporting the vertical gene transfer hypothesis.

In summary, the genetic loci containing *GH149* revealed co-localization of *GH149* with genes encoding several families of glycoside hydrolase as well as sugar transporters, suggesting the biological roles of GH149 protein in polysaccharide degradation in particular β-1,3-glucan from external sources. Similar GC content of the *GH149* genes and its associated bacterial genomes suggests that the *GH149* genes were inherited through vertical gene transfer. Variation in the architectures of gene clusters containing the *GH149* genes suggests that the inherited *GH149* genes were independently associated into the gene clusters and PULs of individual species.

Our work supports the involvement of novel phosphorylases in β-1,3-glucan degradation

and emphasizes the importance of gene clusters containing CAZymes and their roles in carbohydrate metabolism in marine Gram-negative bacteria. Our discovery of the eukaryote GH149 sequences would enable further determination of their physiological function in Euglenophyceae using genetic manipulation, which would subsequently aid our understanding in the involvement of the GH149 phosphorylase in β-1,3-glucan metabolism.

Experimental procedures

Euglena culture

E. gracilis was grown in 2 l of EG:JM media (replacing peptone and tryptone with casein hydrolysate (Sigma-Aldrich)) supplemented with 1% glucose for 7 days at 30 °C with gentle agitation in the dark.

Anion exchange chromatography (AIEX) purification

Approximately 20 g wet weight of 7-day old Euglena cells was harvested by centrifugation (500 x g, 5 minutes). The cell pellet was washed twice with MilliQ water and re-suspended in a lysis buffer (50 ml, 20 mM HEPES pH 7.0, 1 mg ribonuclease A (Sigma), 1 tablet complete protease inhibitor cocktail (Roche)). The cells were lysed on ice by sonication, followed by centrifugation (32,914 x g, 30 minutes) to remove the cell debris. The supernatant was collected and filtered through 0.2 µm disc filter (Millipore). A 5 ml AIEX column (Hitrap Sepharose Q, GE Healthcare) was pre-equilibrated with AIEX buffer (25 mM Tris-HCl pH 8.5). The supernatant containing Euglena proteins (5 ml) was mixed with an equal volume of the AIEX buffer, and loaded onto the pre-equilibrated AIEX column. The proteins were eluted with NaCl gradient (0-700 mM) in the AIEX buffer over 20 column volumes and collected as 2 ml fractions.

Enzyme assays

For the Euglena lysate or AIEX fractions, phosphate release assay (55) was carried out in 20 µl of an assay buffer (100 mM HEPES pH 7.0, 20 mM α-glucose 1-phosphate (G1P), 10 mM acceptor (glucose (Glc), laminaribiose (G2), laminaritriose (G3)) and 200 mM sodium molybdate). The Euglena clear lysate or AIEX fractions (10 µl) were incubated with the assay buffer for 2 hours at 30 °C. The reaction was stopped by boiling (5 minutes) and left to cool to room temperature. A color solution (90 µl, 0.1 M HCl, 13.6 M sodium ascorbate) was added to the boiled reaction mixture and incubated for 30 minutes at room temperature to allow color development. A stop solution (90 µl, 68 mM sodium citrate tribasic dihydrate, 2% acetic acid) was added to the mixture to stop the color development. The absorbance of final solution was measured at 620 nm on a 96-well plate reader. The amount of phosphate release was calculated from the absorbance by comparing to a phosphate standard curve ranging between 0-10 mM. All assays were performed in triplicates. Kinetic parameters of EgP1 and Pro_7066 were determined using the phosphate release assay (20 µl) with the enzymes (25 µg/ml) in the presence of 0.2-10 mM of Glc laminaribiose (G2), laminaritriose (G3), laminaritetraose (G4), laminaripentaose (G5), or laminarihexaose (G6) and 10 mM G1P. The amount of phosphate release from the assays were measured and the values were fitted on non-linear regression with Michaelis-Menten model using GraphPad Prism to determine V_{max} and K_m.

Phosphorolysis assay was carried out using 20 µl of 20 mM oligosaccharides, 10 mM KH_2PO_4 in 20 mM HEPES pH 7.0. and 10 µl of AIEX fractions, which was desalted to remove any endogenous phosphate by passing through a PD-10 column (GE Healthcare) or with 1 µl of EgP1 or Pro_7066 (10 mg/ml stock solution). The reaction mixture was incubated for 1 hour at 30 °C. The reaction was stopped by boiling (5 minutes) and oligosaccharide products were analyzed by TLC analysis or High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD).

DNJ affinity column preparation

The protocol was modified from (39); *N*-5-carboxypentyl-DNJ (50 µmol, Toronto Research Chemicals) was conjugated to resin (1 ml, Bio-Rad Affidex 102 resin) overnight in anhydrous methanol using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDAC, 10 mg, Bio-Rad) as a catalyst. The reaction was performed at room temperature with gentle agitation. The conjugated resin was poured into a glass column and washed with 10 x 4 ml of water to remove methanol. The column was stored at 4 °C in aqueous 0.02 % sodium azide solution before use.

DNJ affinity chromatography

The DNJ column was pre-equilibrated with 4 x 4 ml water and 4 x 2 ml washing buffer (20 mM HEPES, pH 7.0). Protein solution (800 µl) was loaded onto the column. Unbound proteins were washed with 2 ml of the washing buffer and 2 ml of 100 mM NaCl in the washing buffer. Bound proteins were eluted with an increasing concentration of DNJ (50 nM, 1 µM, 50 µM and 1 mM) in the washing buffer (2 ml per each concentration) and collected as 2 ml fractions. The protein fractions were concentrated by freeze-drying. The concentrated fractions were pooled and analyzed by LC-MS/MS.

Proteomic analysis

Proteins were purified from the supernatant using OMIX C4 tips (Agilent, Cheadle, UK). The amount of protein was determined using the Direct Detect® Spectrometer (Merck Millipore, Watford, UK). Protein mixture (8 µg) was dissolved in 30 µl of 0.1 M TEAB buffer (Sigma Aldrich, Dorset, UK) and 0.2% Rapigest (Waters, Manchester, UK), reduced with DTT and alkylated with iodoacetamide, and digested with 0.4 µg Sequencing Grade Trypsin (Promega, Southampton, UK) at 37 °C for 16 hours. The reaction was stopped by adding TFA, and aliquots were analyzed by nanoLC-MS/MS on an Orbitrap Fusion™ Tribrid™ Mass Spectrometer coupled to an UltiMate® 3000 RSLCnano LC system (Thermo Scientific, Hemel Hempstead, UK). The mixtures were separated on a PepMap™ 100 C18 LC Column (C18, 2 µm, 500 x 0.75 mm, Thermo) using a gradient of 0.75% min^{-1} acetonitrile from 6% to 40% in water/0.1% formic acid at a flow rate of 0.3 µl min^{-1} and infused directly into the mass spectrometer. The mass spectrometer was run in positive ion mode, with no quadruple isolation, at 120K resolution over the mass range 350-1800 (m/z) for the precursor scans (orbitrap). One microscan of 50 ms with an AGC target of $2e^5$ was used. MS2 threshold was set to $1.5e^4$ and precursors fragmented by both CID and HCD with CE=30 and an isolation window of 1.6 Da (quadrupole) using the automatic maximum speed option with ion injection for all available parallelizable time. Dynamic exclusion was set to 1 count and 30 s. Recalibrated peak lists were generated using MaxQuant 1.5.2.8 (www.MaxQuant.org) and the database search was performed with the merged HCD and CID peak lists using Mascot 2.4 (Matrixscience, London, UK). The search was performed on a *Euglena gracilis* protein database (14) with a precursor tolerance of 6 ppm and a fragment tolerance of 0.6 Da. The enzyme was set to trypsin/P with a maximum of 2 allowed missed cleavages. Carbamidomethyl (C) was set as fixed modification, and oxidation (M) and acetylation (Protein N-terminus) were used as variable modifications. The Mascot search results were imported into Scaffold 4.4.1.1 (www.proteomsoftware.com) using identification probabilities of 99% and 95% for proteins and peptides.

Bioinformatics analyses

Orthologous sequences to EgP1 were obtained from the non-redundant protein sequence database (https://www.ncbi.nlm.nih.gov/protein/), the Marine Microbial Eukaryote Transcriptome (MMET) Sequencing Project (http://marinemicroeukaryotes.org/), and the *E. gracilis* Z strain transcriptome (http://euglenadb.org/) using BLASTP or tBLASTn with E-value score of 0.0001 or more . Multiple sequence alignments of amino acid sequences were performed using Clustal Omega (56) (www.clustal.org/omega, version 1.2.2) with the default settings and edited with trimAl v.1.2 using a heuristic automated method (57). The alignments were visualized by Jalview (version 14.6.4) (58). Phylogenetic trees were reconstructed from a matrix of 331 unambiguously aligned amino acids from 325 species using PhyML 3.0 (59) with the best fit model as inferred by a smart model selection (SMS). Bootstrap values were determined from a population of 100 replicates. Tree annotation and visualization were performed using iTOL v.3.4.3 (60). The GenBank IDs of the GH149 members can be found in Supplemental Data File 4.

Recombinant protein production

The *EgP1* cDNA sequence was synthesized and optimized for *E. coli* expression (custom DNA synthesis by Gen9, Inc.). The sequence was amplified by PCR and cloned into the PopinF plasmid vector (41) using In-Fusion™ (TakaraBio, Mountain View, CA, USA) following the manufacturer's protocol. The recombinant *PopinF-EgP1* was transformed into *E. coli* (Lemmo21) and a one l culture of the transformant was grown at 25 °C in LB media with agitation (180 rpm) overnight. Heterologous protein expression

was induced by adding IPTG to a final concentration of 0.2 mM and incubated for 2 days at 18˚C. A recombinant plasmid pET28a containing *Pro_7066* cDNA (Prozomix Limited), Pro_7066, was transformed into *E. coli* (BL21 (DE3)) and grown as previously described for EgP1. The cells were harvested (6721 x g, 10 min) and lysed by sonication in buffer A (20 mM HEPES pH 7.0, 250 mM NaCl) supplemented with 1 mg/ml DNase (Sigma) and 1 tablet complete protease inhibitor cocktail (Roche). Supernatant containing the recombinant proteins was separated from cell debris by centrifugation (32,914 x g, 30 min). Proteins were purified with ÄKTA pure FPLC system (GE Healthcare) at 4 °C. The supernatant containing either His_6-tagged EgP1 or Pro_7066 was loaded to a 1-ml HisTrap™ HP column (GE Healthcare) pre-equilibrated with buffer A (10 mM HEPES pH 7.5, 250 mM NaCl). The column was washed with buffer A and bound proteins were eluted in one step with 10 mM HEPES pH 7.5, 250 mM NaCl, 500 mM imidazole. The proteins were further purified by gel filtration using a Superdex S200 16/600 column (GE Healthcare) eluted with 20 mM HEPES pH 7.5, 150 mM NaCl, 1 ml/min. Fractions containing the proteins were pooled and concentrated to 10 mg/ml using Amicon Ultra-15 30 kDa MW cut off concentrator. The proteins were stored in aliquots at -80 °C until required. EgP1 nucleotide sequence was deposited to GenBank with accession number MG516599.

Oligosaccharide analyses

TLC was performed by spotting 0.5 µl of the recovered reaction mixture onto a silica plate (10 cm x 5 cm), then eluted using a mobile phase containing NH_4OH : H_2O : *iso*-propanol (3:1:4) in a sealed glass container for 2 hour to allow oligosaccharide separation. The plate was air-dried and stained with orcinol, which was prepared by adding concentrated sulfuric acid (20 ml) to ice cold solution of 3,5-dihydroxytoluene (360 mg) in ethanol (150 ml) and water (10 ml). The stained plate was then heated until oligosaccharide spots were visible.

HPAEC-PAD analyses were performed by diluting the reaction mixtures in MilliQ water to a final volume of 500 µl and desalted by mixed bed ion exchange resin (Sigma). The desalted mixtures were filtered through a disposable PTFE 0.45 µm filter disc (Merck Millipore), and subjected to HPAEC-PAD analysis using a Dionex ICS3000 chromatography system equipped with PAD and controlled by Chromeleon® software. A PA100 CarboPac column (analytical: 4 x 250 mm, guard: 4 x 50 mm) was used for all analyses. The solutions for elution of the oligosaccharides were as follows; solution A: 100 mM sodium hydroxide and solution B: 100 mM sodium hydroxide + 400 mM sodium acetate. The separation was achieved by gradient elution (0-100% solution B) from 1-30 min, followed by 30-50 min of 100% B then 50-60 min re-equilibration of the column with solution A. The solutions were delivered to the column at the rate of 0.25 ml/min.

Acknowledgements

This work was supported by the UK BBSRC Institute Strategic Program on Understanding and Exploiting Metabolism (MET) [BB/J004561/1]; the UK BBSRC, EPSRC and InnovateUK: IBCatalyst [BB/M02903411]; the John Innes Foundation and the Royal Thai Government Scholarship program We thank Ellis O'Neill for his advice on *Euglena gracilis* transcriptomics, Giulia Pergolizzi for advice on analytical biochemistry, Michael Rugen for advice on heterologous expression, Sergey Nepogodiev for insight into β-glucan chemistry and Prozomix Limited for their help with the identification of Pro_7066 sequence.

Conflict of interest

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

Author contributions

S.K and R.A.F designed the experiments, S.K, M.R and G.S performed the experiments and analyzed the data. S.K. and N.J.P designed and performed bioinformatics analyses, augmented by B.H. input on CAZyme analyses. S.K, M.R, G.S, N.J.P, B.H. and R.A.F wrote the manuscript.

References

1. O'Neill, E. C., and Field, R. A. (2014) Enzymatic synthesis using glycoside phosphorylases. *Carbohydr. Res.* **403**, 23–37

2. Field, R., Kuhaudomlarp, S., Pergolizzi, G., and Kalita, E. (2017) Glycan phosphorylases in multi-enzyme synthetic processes. *Protein Pept. Lett.* **24**, 696–709

3. Bacic, A., Fincher, G. B., and Stone, B. A. (2009) *Chemistry, biochemistry, and biology of (1→3)-β-glucans and related polysaccharides* (Bacic, A., Fincher, G. B., and Stone, B. A. eds), Academic Press/Elsevier, San Diego

4. Buller, C. S. (1990) Water insoluble polysaccharide polymer and method thereof. U.S. patent number 4,908,310 A.

5. Jagodzinski, P. P., Wiaderkiewicz, R., Kurzawski, G., Kloczewiak, M., Nakashima, H., Hyjek, E., Yamamoto, N., Uryu, T., Kaneko, Y., Posner, M. R., and Kozbor, D. (1994) Mechanism of the inhibitory effect of curdlan sulfate on HIV-1 infection *in vitro*. *Virology.* **202**, 735–745

6. Watanabe, T., Shimada, R., Matsuyama, A., Yuasa, M., Sawamura, H., Yoshida, E., and Suzuki, K. (2013) Antitumor activity of the β-glucan paramylon from *Euglena* against preneoplastic colonic aberrant crypt foci in mice. *Food Funct.* **4**, 1685–1690

7. Kankkunen, P., Teirila, L., Rintahaka, J., Alenius, H., Wolff, H., and Matikainen, S. (2010) (1,3)-β-glucans activate both dectin-1 and NLRP3 inflammasome in human macrophages. *J. Immunol.* **184**, 6335–6342

8. Shibakami, M., Tsubouchi, G., Sohma, M., and Hayashi, M. (2015) One-pot synthesis of thermoplastic mixed paramylon esters using trifluoroacetic anhydride. *Carbohydr. Polym.* **119**, 1–7

9. Barsanti, L., Vismara, R., Passarelli, V., and Gualtieri, P. (2001) Paramylon (β-1,3-glucan) content in wild type and WZSL mutant of *Euglena gracilis*. Effects of growth conditions. *J. Appl. Phycol.* **13**, 59–65

10. Trouvelot, S., Varnier, A.-L., Allègre, M., Mercier, L., Baillieul, F., Arnould, C., Gianinazzi-Pearson, V., Klarzynski, O., Joubert, J.-M., Pugin, A., and Daire, X. (2008) A β-1,3 glucan sulfate induces resistance in grapevine against *Plasmopara viticola* through priming of defense responses, including HR-like cell death. *Mol. Plant. Microbe. Interact.* **21**, 232–243

11. Hrmova, M., Imai, T., Rutten, S. J., Fairweather, J. K., Pelosi, L., Bulone, V., Driguez, H., and Fincher, G. B. (2002) Mutated barley (1,3)-β-D-glucan endohydrolases synthesize crystalline (1,3)-β-D-glucans. *J. Biol. Chem.* **277**, 30102–30111

12. Ogawa, Y., Noda, K., Kimura, S., Kitaoka, M., and Wada, M. (2014) Facile preparation of highly crystalline lamellae of (1 → 3)-β-D-glucan using an extract of *Euglena gracilis*. *Int. J. Biol. Macromol.* **64**, 415–9

13. Müller, C., Ortmann, T., Abi, A., Hartig, D., Scholl, S., and Jördening, H.-J. (2017) Immobilization and characterization of *E. gracilis* extract with enriched laminaribiose phosphorylase activity for bienzymatic production of laminaribiose. *Appl. Biochem. Biotechnol.* **182**, 197–215

14. O'Neill, E. C., Trick, M., Hill, L., Rejzek, M., Dusi, R. G., Hamilton, C. J., Zimba, P. V., Henrissat, B., and Field, R. A. (2015) The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol. BioSyst.* **11**, 2808–2820

15. O'Neill, E. C., Trick, M., Henrissat, B., and Field, R. A. (2015) *Euglena* in time: evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product

biochemistry. *Perspect. Sci.* **6**, 84–93

16. Ivanova, I. M., Nepogodiev, S. A., Saalbach, G., O'Neill, E. C., Urbaniak, M. D., Ferguson, M. A. J., Gurcha, S. S., Besra, G. S., and Field, R. A. (2017) Fluorescent mannosides serve as acceptor substrates for glycosyltransferase and sugar-1-phosphate transferase activities in *Euglena gracilis* membranes. *Carbohydr. Res.* **438**, 26–38

17. Rejzek, M., Hill, L., Hems, E. S., Kuhaudomlarp, S., Wagstaff, B. A., and Field, R. A. (2017) Profiling of sugar nucleotides. in *Methods in enzymology*

18. O'Neill, E., Kuhaudomlarp, S., Rejzek, M., Fangel, J., Alagesan, K., Kolarich, D., Willats, W., and Field, R. (2017) Exploring the Glycans of *Euglena gracilis*. *Biology (Basel).* **6**, 45

19. O'Neill, E. C., Pergolizzi, G., Stevenson, C. E. M., Lawson, D. M., Nepogodiev, S. A., and Field, R. A. (2017) Cellodextrin phosphorylase from *Ruminiclostridium thermocellum*: X-ray crystal structure and substrate specificity analysis. *Carbohydr. Res.* **451**, 118–132

20. O'Neill, E. C., Rashid, A. M., Stevenson, C. E. M., Hetru, A.-C., Gunning, A. P., Rejzek, M., Nepogodiev, S. A., Bornemann, S., Lawson, D. M., and Field, R. A. (2014) Sugar-coated sensor chip and nanoparticle surfaces for the *in vitro* enzymatic synthesis of starch-like materials. *Chem. Sci.* **5**, 341–350

21. Goldemberg, S. H., Maréchal, L. R., Souza, B. C. De, Mar, L. R., and Souza, B. C. D. E. (1966) β-1,3-Oligoglucan:orthophosphate glucosyltransferase from *Euglena gracilis*. *J. Biol. Chem.* **241**, 45–50

22. Kitaoka, M. (2015) Diversity of phosphorylases in glycoside hydrolase families. *Appl. Microbiol. Biotechnol.* **99**, 8377–8390

23. Maréchal, L. R., and Goldemberg, S. H. (1963) Laminaribiose phosphorylase from *Euglena gracilis*. *Biochem. Biophys. Res. Commun.* **13**, 2–5

24. Kitaoka, M., Sasaki, T., and Taniguchi, H. (1993) Purification and properties of laminaribiose phosphorylase (EC 2.4.1.31) from *Euglena gracilis* Z. *Arch. Biochem. Biophys.* **304**, 508–514

25. Manners, D. J., and Taylor, D. (1967) Studies on carbohydrate metabolizing enzymes XVI. Specificity of laminaribiose phosphorylase from *Astasia ocellata*. *Arch. Biochem. Biophys.* **121**, 443–451

26. Kitaoka, M., Matsuoka, Y., Mori, K., Nishimoto, M., and Hayashi, K. (2012) Characterization of a bacterial laminaribiose phosphorylase. *Biosci. Biotechnol. Biochem.* **76**, 343–8

27. Nihira, T., Saito, Y., Kitaoka, M., Nishimoto, M., Otsubo, K., and Nakai, H. (2012) Characterization of a laminaribiose phosphorylase from *Acholeplasma laidlawii* PG-8A and production of 1,3-β-D-glucosyl disaccharides. *Carbohydr. Res.* **361**, 49–54

28. Marechal, L. R. (1967) β-1,3-oligoglucan: orthophosphate glucosyltransferases from *Euglena gracilis* I. Isolation and some properties of a β-1,3-oligoglucan phosphorylase. *Biochem. Biophys. Acta*. **146**, 417–430

29. Albrecht, G. J., and Kauss, H. (1971) Purification, crystallization and properties of a β-(1→3)-glucan phosphorylase from *Ochromonas malhamensis*. **10**, 1293–1298

30. Yamamoto, Y., Kawashima, D., Hashizume, A., Hisamatsu, M., and Isono, N. (2013) Purification and characterization of 1,3-β-D-glucan phosphorylase from *Ochromonas danica*. *Biosci. Biotechnol. Biochem.* **77**, 1949–54

31. 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**, 490–495

32. Trindade, M. I., Abratt, V. R., and Reid, S. J. (2003) Induction of sucrose utilization genes from *Bifidobacterium lactis* by sucrose and raffinose. *Appl. Environ. Microbiol.* **69**, 24–32

33. Nihira, T., Suzuki, E., Kitaoka, M., Nishimoto, M., Ohtsubo, K., and Nakai, H. (2013) Discovery of β-1,4-D-mannosyl-*N*-acetyl-D-glucosamine phosphorylase involved in the metabolism of *N*-glycans. *J. Biol. Chem.* **288**, 27366–27374

34. Chiku, K., Nihira, T., Suzuki, E., Nishimoto, M., Kitaoka, M., Ohtsubo, K., and Nakai, H. (2014) Discovery of two β-1,2-mannoside phosphorylases showing different chain-length specificities from *Thermoanaerobacter* sp. X-514. *PLoS One*. **9**, 10.1371/journal.pone.0114882

35. Terrapon, N., Lombard, V., Gilbert, H. J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics*. **31**, 647–655

36. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., and Brumer, H. (2017) Polysaccharide Utilization Loci: Fuelling microbial communities. *J. Bacteriol.* 10.1128/JB.00860-16

37. Yoshida, Y., Tomiyama, T., Maruta, T., Tomita, M., Ishikawa, T., and Arakawa, K. (2016) *De novo* assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics*. **17**, 10.1186/s12864-016-2540-6

38. Fushinobu, S., Hidaka, M., Hayashi, A. M., Wakagi, T., Shoun, H., and Kitaoka, M. (2011) Interactions between glycoside hydrolase family 94 cellobiose phosphorylase and glucosidase inhibitors. *J. Appl. Glycosci.* **58**, 91–97

39. Cruz, I. N., Barry, C. S., Kramer, H. B., Chuang, C. C., Lloyd, S., van der Spoel, A. C., Platt, F. M., Yang, M., and Davis, B. G. (2013) Glycomimetic affinity-enrichment proteomics identifies partners for a clinically-utilized iminosugar. *Chem. Sci.* **4**, 3442–3446

40. Takeda, T., Nakano, Y., Takahashi, M., Konno, N., Sakamoto, Y., Arashida, R., Marukawa, Y., Yoshida, E., Ishikawa, T., and Suzuki, K. (2015) Identification and enzymatic characterization of an endo-1,3-β-glucanase from *Euglena gracilis*. *Phytochemistry*. **116**, 21–27

41. Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I., and Owens, R. J. (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, 10.1093/nar/gkm047

42. Wu, Y., Mao, G., Fan, H., Song, A., Zhang, Y.-H. P., and Chen, H. (2017) Biochemical properties of GH94 cellodextrin phosphorylase THA_1941 from a thermophilic eubacterium *Thermosipho africanus* TCF52B with cellobiose phosphorylase activity. *Sci. Rep.* **7**, 10.1038/s41598-017-05289-x

43. Sawano, T., Saburi, W., Hamura, K., Matsui, H., and Mori, H. (2013) Characterization of *Ruminococcus albus* cellodextrin phosphorylase and identification of a key phenylalanine residue for acceptor specificity and affinity to the phosphate group. *FEBS J.* **280**, 4463–4473

44. Hidaka, M., Kitaoka, M., Hayashi, K., Wakagi, T., Shoun, H., and Fushinobu, S. (2006) Structural dissection of the reaction mechanism of cellobiose phosphorylase. *Biochem. J.* **398**, 37–43

45. Hidaka, M., Honda, Y., Kitaoka, M., Nirasawa, S., Hayashi, K., Wakagi, T., Shoun, H., and Fushinobu, S. (2004) Chitobiose phosphorylase from *Vibrio proteolyticus*, a member of glycosyl transferase family 36, has a clan GH-L-like (α/α)(6) barrel fold. *Structure*. **12**, 937–947

46. Van Hoorebeke, A., Stout, J., Kyndt, J., De Groeve, M., Dix, I., Desmet, T., Soetaert, W., Van

Beeumen, J., and Savvides, S. N. (2010) Crystallization and X-ray diffraction studies of cellobiose phosphorylase from *Cellulomonas uda*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 346–351

47. Bianchetti, C. M., Elsen, N. L., Fox, B. G., and Phillips, G. N. (2011) Structure of cellobiose phosphorylase from *Clostridium thermocellum* in complex with phosphate. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67**, 1345–1349

48. Nam, Y.-W., Nihira, T., Arakawa, T., Saito, Y., Kitaoka, M., Nakai, H., and Fushinobu, S. (2015) Crystal structure and substrate recognition of cellobionic acid phosphorylase, which plays a key role in oxidative cellulose degradation by microbes. *J. Biol. Chem.* **290**, 18281–18292

49. Hoffmeister, M., Piotrowski, M., Nowitzki, U., and Martin, W. (2005) Mitochondrial trans-2-enoyl-CoA reductase of wax ester fermentation from *Euglena gracilis* defines a new family of enzymes involved in lipid synthesis. *J. Biol. Chem.* **280**, 4329–4338

50. Krnáčová, K., Vesteg, M., Hampl, V., Vlček, Č., and Horváth, A. (2012) *Euglena gracilis* and Trypanosomatids possess common patterns in predicted mitochondrial targeting presequences. *J. Mol. Evol.* **75**, 119–129

51. Conners, S. B., Montero, C. I., Comfort, D. A., Shockley, K. R., Johnson, M. R., Chhabra, S. R., and Kelly, R. M. (2005) An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Bacteriol.* **187**, 7267–7282

52. Wang, Z., Robertson, K. L., Liu, C., Liu, J. L., Johnson, B. J., Leary, D. H., Compton, J. R., Vuddhakul, V., Legler, P. M., and Vora, G. J. (2015) A novel Vibrio beta-glucosidase (LamN) that hydrolyzes the algal storage polysaccharide laminarin. *FEMS Microbiol. Ecol.* **91**, 1–10

53. Noinaj, N., Guillier, M., Barnard, T. J., and Buchanan, S. K. (2010) TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* **64**, 43–60

54. Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H. J., and Henrissat, B. (2017) PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* 10.1093/nar/gkx1022

55. De Groeve, M. R. M. R. M., Tran, G. H. H., Van Hoorebeke, a., Stout, J., Desmet, T., Savvides, S. N. N., and Soetaert, W. (2010) Development and application of a screening assay for glycoside phosphorylases. *Anal. Biochem.* **401**, 162–7

56. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2014) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539

57. Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**, 1972–1973

58. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **25**, 1189–1191

59. Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005) PHYML online - a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **33**, 557–559

60. Letunic, I., and Bork, P. (2016) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*. **23**, 127–128

Tables

Table 1. Kinetics parameters of EgP1 for the reverse phosphorolysis using G1P as a donor.

Acceptors	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1}mM^{-1}$)
Glc	1.10 ± 0.03	0.56 ± 0.06	1.99
G2	1.08 ± 0.03	0.67 ± 0.08	1.62
G3	1.12 ± 0.02	1.26 ± 0.09	0.89
G4	1.12 ± 0.03	1.41 ± 0.13	0.79
G5	1.13 ± 0.03	2.29 ± 0.19	0.50
G6	1.10 ± 0.03	2.88 ± 0.23	0.38

Table 2. Kinetic parameters of Pro_7066 for the reverse phosphorolysis reaction using G1P as a donor.

Acceptors	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1}mM^{-1}$)
Glc	1.66 ± 0.04	0.29 ± 0.03	5.79
G2	1.54 ± 0.01	0.25 ± 0.02	6.03
G3	1.53 ± 0.02	0.37 ± 0.03	4.16
G4	1.39 ± 0.01	0.36 ± 0.02	3.89
G5	1.27 ± 0.01	0.32 ± 0.02	4.04
G6	1.18 ± 0.04	0.26 ± 0.04	4.62

Figures and figure legends

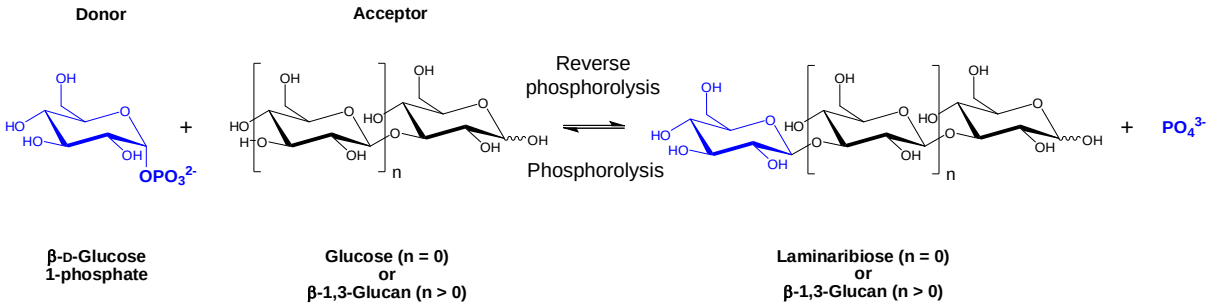

Fig. 1. Reaction carried out by β-1,3-D-glucan phosphorylase

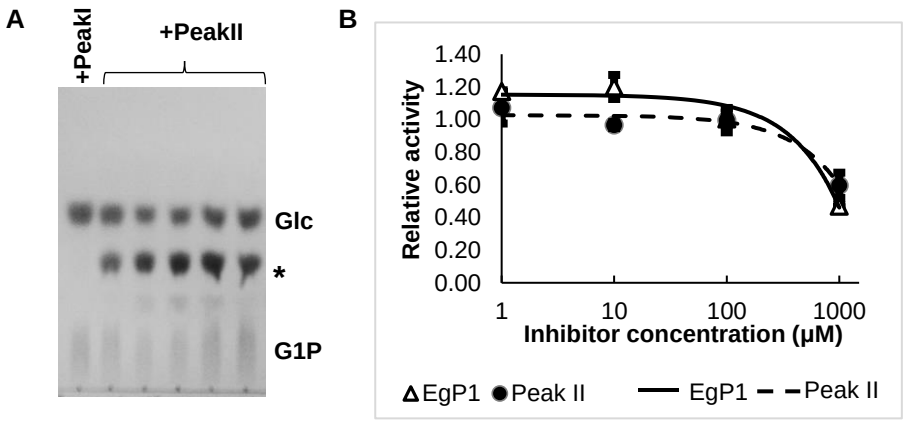

Fig. 2. Partial purification of Euglena phosphorylase from the cell-free extract. (A) TLC analysis of Euglena phosphorylase mediated reaction between Glc and G1P. Disaccharide product was detected (asterisk). (B) relative reverse phosphorolysis activity carried out by partially purified phosphorylase (PeakII) using Glc and G1P as substrates (dotted line) or recombinant EgP1 (solid line) in the presence of DNJ. Glc = glucose, G1P = α-glucose 1-phosphate, DNJ = 1-deoxynojirimycin.

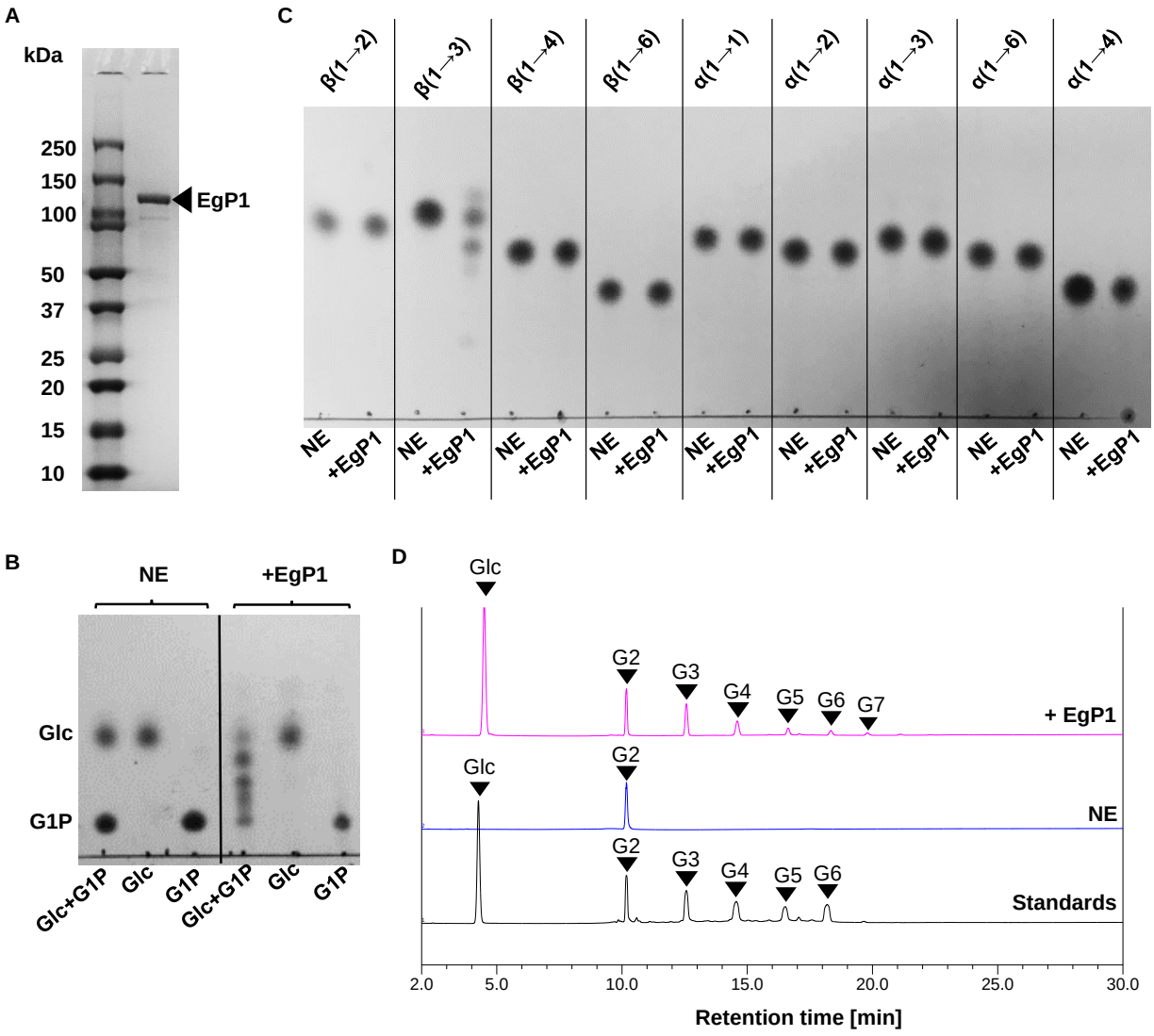

Fig. 3. Characterization of EgP1. (A) SDS-PAGE analysis of purified recombinant EgP1 after IMAC and gel filtration. (B) TLC analysis of the reverse phosphorolysis reaction catalyzed by EgP1. (C) TLC analysis of the phosphorolysis reactions with glucose disaccharides linked by various glycosidic bonds. The linkages of the substrates are indicated at the top. (D) HPAEC-PAD analysis of the phosphorolysis of G2 (Glc-β-1,3-Glc) to confirm the breakdown to glucose. However, due to the reversibility of the reaction, longer oligosaccharides (G3 to G7) were detected. NE = no enzyme control, G2 = laminaribiose, G3 = laminaritriose, G4 = laminaritetraose, G5 = laminaripentaose, G6 = laminarihexaose.

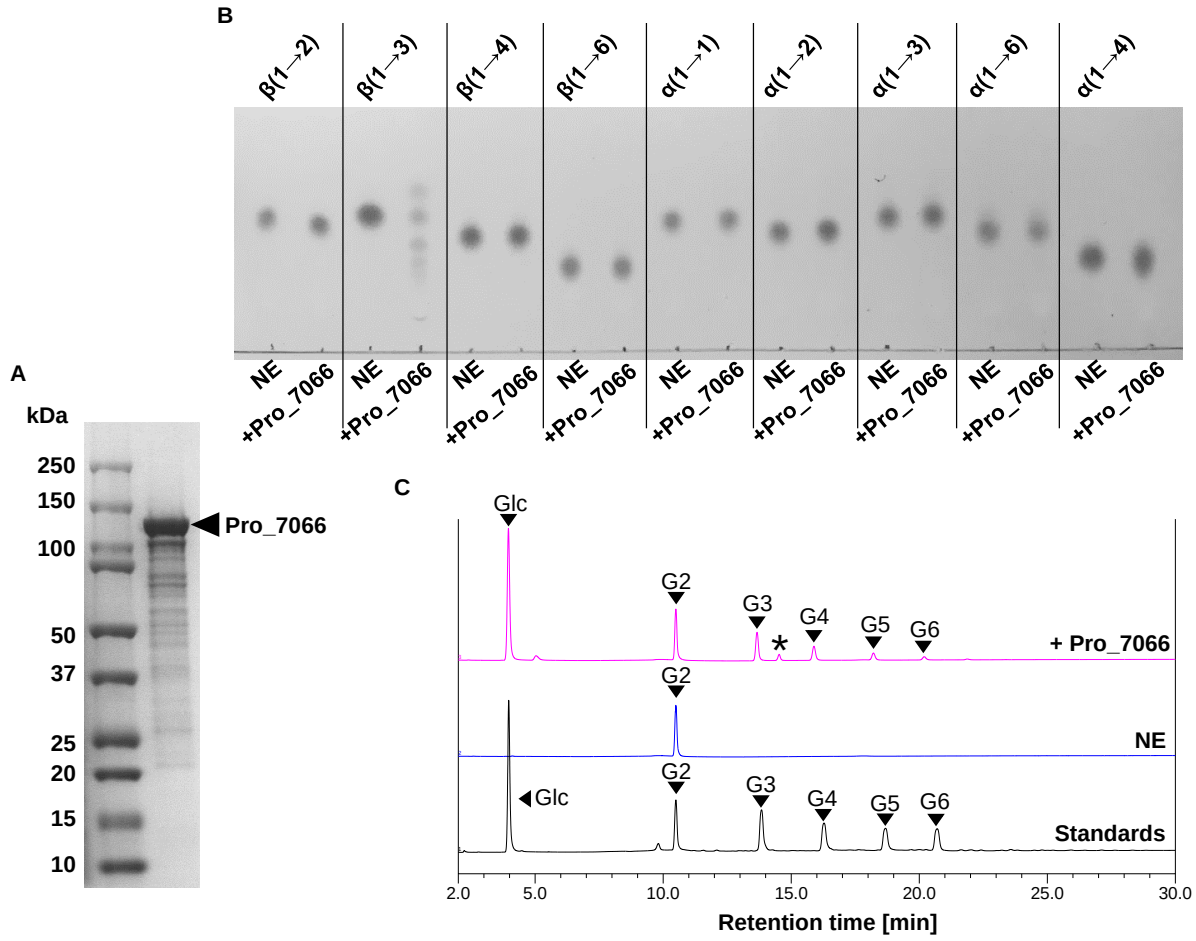

Fig. 4. Characterization of Pro_7066. (A) SDS-PAGE analysis of Pro_7066 after gel filtration. (B) TLC analysis of the phosphorolysis reactions with glucose disaccharides with different linkages. The linkages of the substrates are indicated at the top. (C) HPAEC-PAD analysis of G2 (Glc-β-1,3-Glc) phosphorolysis to confirm the breakdown to glucose. However, due to the reversibility of the reaction, the formation of G3 to G6 oligosaccharides was also detected. * = small trace of G1P due to incomplete desalting by mixed bed ion exchange resin. NE = no enzyme control, G2 = laminaribiose, G3 = laminaritriose, G4 = laminaritetraose, G5 = laminaripentaose, G6 = laminarihexaose.

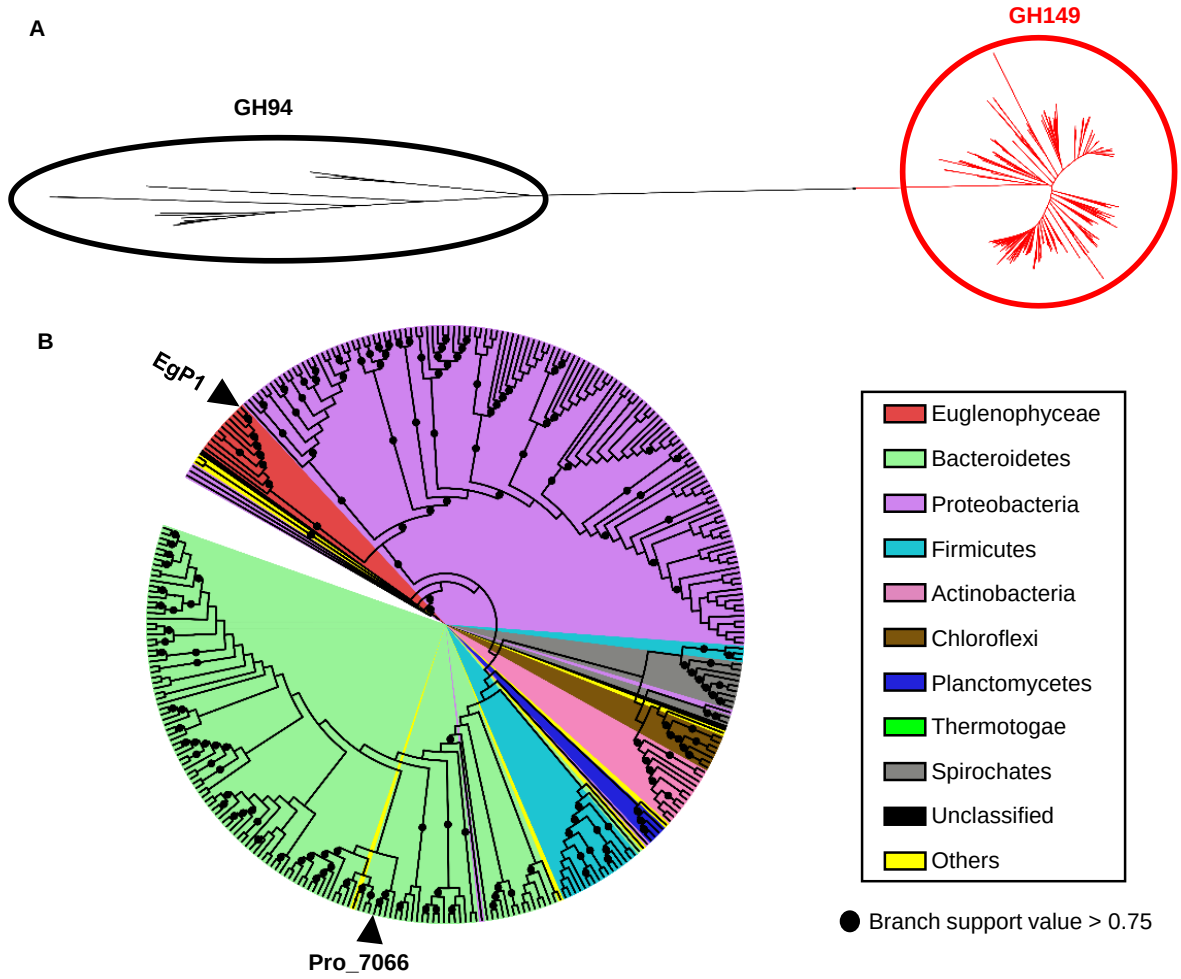

Fig. 5. Phylogenetic analysis of GH149. (A) an unrooted tree representing he phylogenetic relationship between GH94 queries (black branches) and GH149 (red branches). (B) Phylogenetic analysis of GH149. The label colors represent the taxonomic phyla of the species containing GH149. The positions of EgP1 and Pro_7066 are indicated by arrow heads.

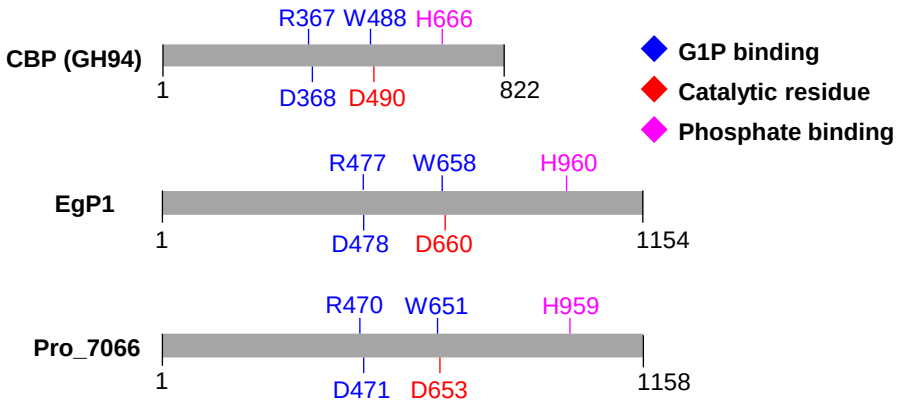

Fig. 6. Conserved amino acids required for phosphorylase activity. EgP1 sequence represents the truncated version where the predicted target peptide (residue 1-31) has been removed. CBP = cellobiose phosphorylase from *Cellvibrio gilvus* (BAA28631.1).

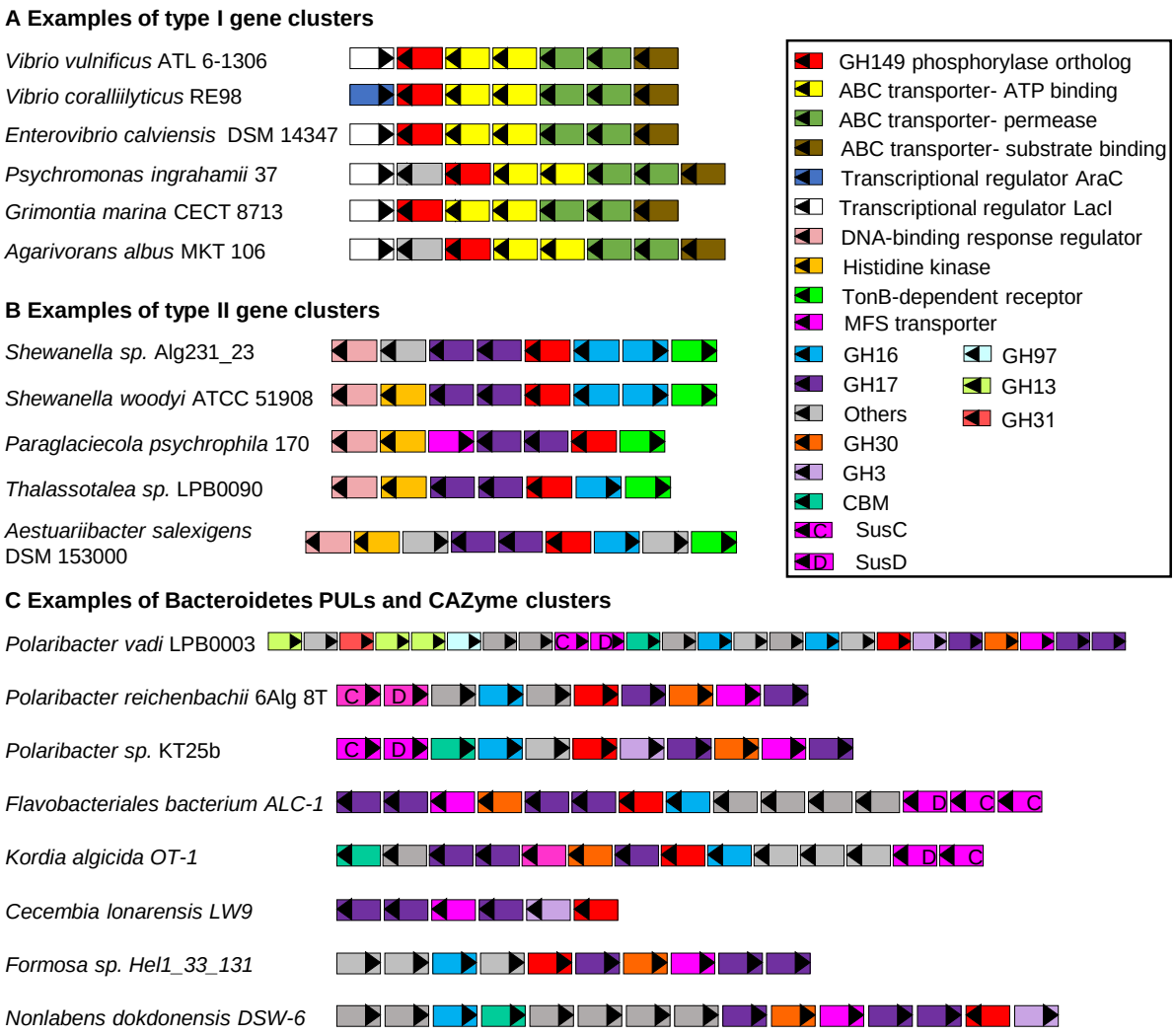

Fig. 7. Genetic organization of putative gene clusters containing the bacterial GH149 genes. (A) Type I gene clusters contain ABC-transporter cassette, the phosphorylase, and a transcriptional regulator. (B) Type II gene clusters contain a TonB-dependent receptor together with GHs and transcriptional regulators. (C) Predicted PULs and CAZyme clusters containing the putative phosphorylases from Bacteroidetes species.

Supplemental Data

Identification of *Euglena gracilis* β-1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149

Sakonwan Kuhaudomlarp[a], Nicola J. Patron[b], Bernard Henrissat[c,d,e], Martin Rejzek[a], Gerhard Saalbach[a] & Robert A. Field[a,1]

[a]Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. [b]Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK. [c]Aix-Marseille University, Architecture et Fonction des Macromolécules Biologiques, 163 Avenue de Luminy, 13288 Marseille, France. [d]Centre National de la Recherche Scientifique, UMR 7257, 163 Avenue de Luminy, 13288 Marseille, France. [e]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

Running title: New β-1,3-glucan phosphorylase family

[1]To whom correspondence should be addressed: Robert A. Field: Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. rob.field@jic.ac.uk.

Supplemental Table S1. Characterized GH94 sequences that were used for BLAST analysis with EgP1 and Pro_7066

Sequence names*	**Genbank accession**
CBP from *Cellvibrio gilvus* ATCC 13127	BAA28631.1
ChBP from *Vibrio proteolyticus*	BAC87867.1
CBP from *Cellulomonas uda*	AAQ20920.1
CBP from *Ruminiclostridium thermocellum*	AAL67138.1
CBP from *Saccharophagus degradans*	ABD80168.1
CDP from *Ruminiclostridium thermocellum*	BAB71818.1
LBP from *Halorhabdus tiamatea.*	WP_020936056.1
CBP from *Clostridium stercorarium*	AAC45510.0
CDP from *Clostridium stercorarium*	AAC45511.1
CBP from *Cellvibrio gilvus* ATCC 13127	BAA28631.1
CBP from *Thermotoga maritima* MSB8	AAD36910.1
CBP from *Thermotoga neapolitana*	AAB95491.2
N,N'-diacetylchitobiose phosphorylase from *Vibrio furnissii*	AAG23740.1
ChBP from *Vibrio proteolyticus*	BAC87867.1
LBP from *Acholeplasma laidlawii* PG-8A	ABX81345.1
LBP from *Paenibacillus* sp. YM1	BAJ10826.1
N,N'-diacetylchitobiose phosphorylase from *Ruminococcus albus*	WP_013499018.1
CBP from *Halorhabdus tiamatea* SARL4B	CCQ33375.1
NdvB protein from *Neurospora crassa* OR74A	EAA28929.1
CellAP from *Xanthomonas campestris*	WP_011039146.1

*CBP = cellobiose phosphorylase, ChBP = chitobiose phosphorylase, CDP = cellodextrin phosphorylase, LBP = laminaribiose phosphorylase, CellAP = cellobionic acid phosphorylase

Supplemental Table S2. Summary of the observed $[M+Na]^+$ of glucans in MALDI-ToF analysis.

DP of glucan	Calculated $[M+Na]^+$	The closest observed $[M+Na]^+$
G4	689.221	689.203
G5	851.274	851.223
G6	1013.326	1013.287
G7	1175.379	1175.330
G8	1337.432	1337.374
G9	1499.485	1499.483
G10	1661.538	1661.517
G11	1824.581	1823.591
G12	1986.723	1985.649

Supplemental Table S3. Summary of conserved amino acid residues that are involved in GH94 catalysis and G1P binding that have been found in the EgP1 and Pro_7066

GenBank accession	**CAZy families**	**Catalytic residues**	**G1P binding (-1 subsite)**	**Phosphate recognition**	**References**	**PDB structures**
BAA28631.1	GH94	D490	W488, R367, D368	H666	(44)	2CQS, 2CQT
BAC87867.1	GH94	D492	W490, R349, D350	H644	(45)	1V7V, 1V7W, 1V7X
AAQ20920.1	GH94	D490	W488, R367, D368	H666	(46)	3RRS, 3RSY, 3S4A, 3S4B
AAL67138.1	GH94	D483	W481, R360, D361	H653	(47)	3QDE
ABD80168.1	GH94	D472	W470, R349, (N350)	H626	(48)	4ZLE, 4ZLF, 4ZLG, 4ZLI
BAB71818.1	GH94	D624	W622, R486	H817	(19)	5NZ7, 5NZ8
WP_020936056.1	GH94	D535	W533, R384, D385	H749	Inferred from Figure S9	NA
EgP1 (in this study)	GH149	D660	W658, R477, D478	H960	Inferred from Figure S9	NA
Pro_7066 (in this study)	GH149	D653	W651, R470, D471	H959	Inferred from Figure S9	NA

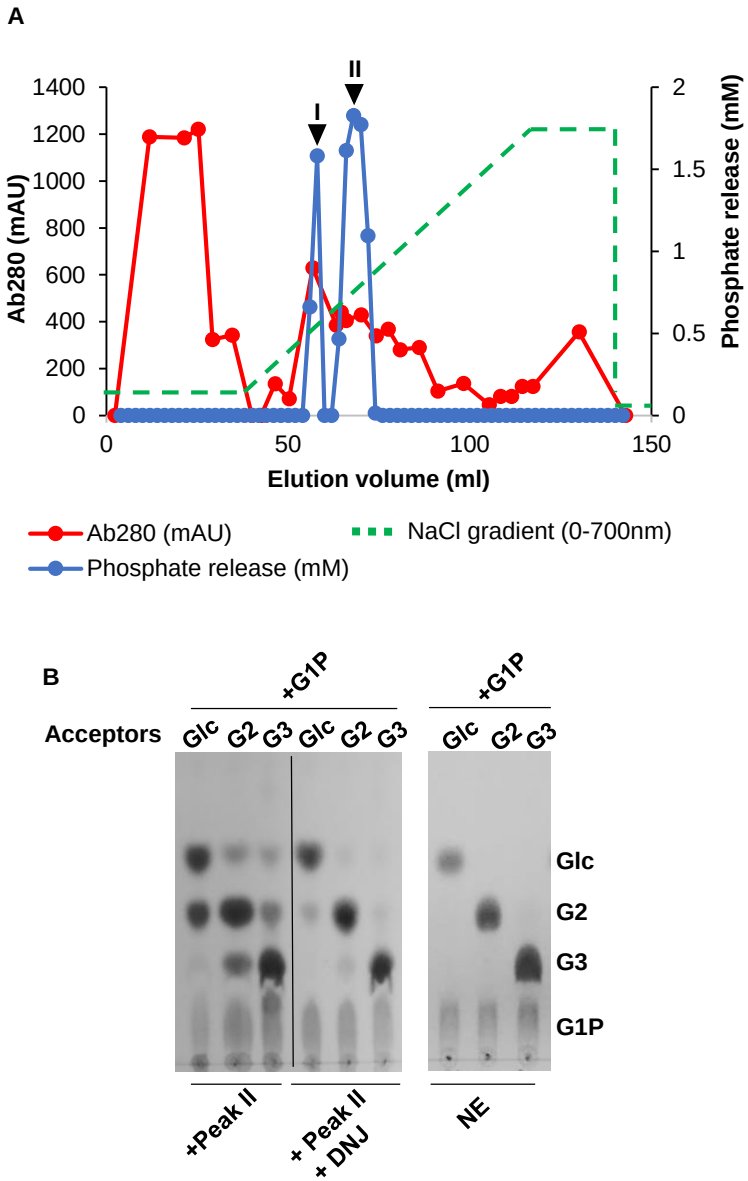

Fig. S1 Partial purification of Euglena phosphorylase from *E. gracilis* protein extract. (A) An overlay of AIEX elution profile (red) and the amount of phosphate release by the phosphorylase activity in each fraction in the phosphate release assay (blue). (B) TLC analysis of Euglena phosphorylase (PeakII) mediated reactions between G1P and laminaribiose (G2), or laminaritriose (G3) as substrates in the presence (+DNJ, 1000 μM) and absence of DNJ. NE = no enzyme control.

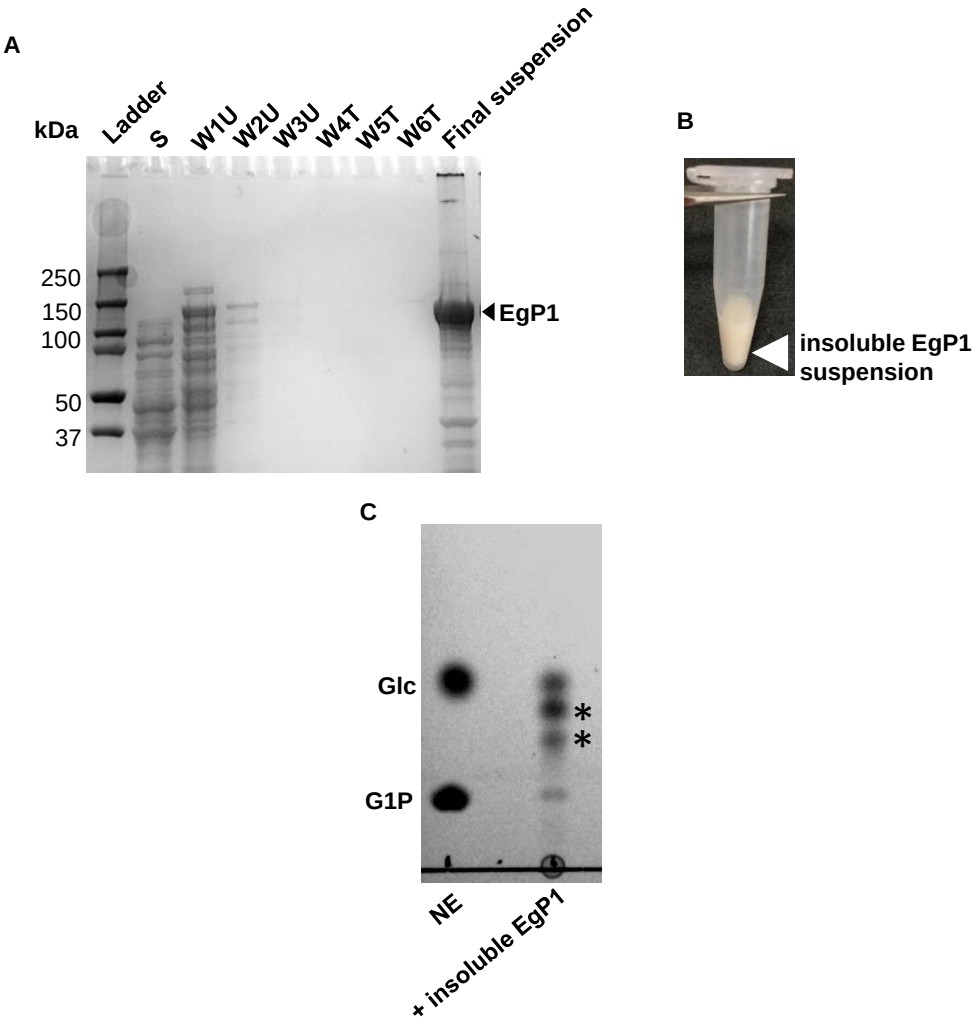

Fig. S2 Expression and characterization of insoluble recombinant EgP1. (A) SDS-PAGE analysis of insoluble EgP1after purification by washing with 50 mM Tris-HCl pH 7.5 containing 6 M urea (W1U-W3U) followed by washing with buffer alone (W4T-W6T, 50 mM Tris-HCl pH 7.5). (B) the final suspension of insoluble EgP1 in 50 mM Tris-HCl pH 7.5 buffer. (C) TLC analysis of the reverse phosphorolysis reaction carried out by the insoluble EgP1 with 10 mM Glc and 20 mM G1P. * = oligosaccharide products from the reverse phosphorolysis reaction.

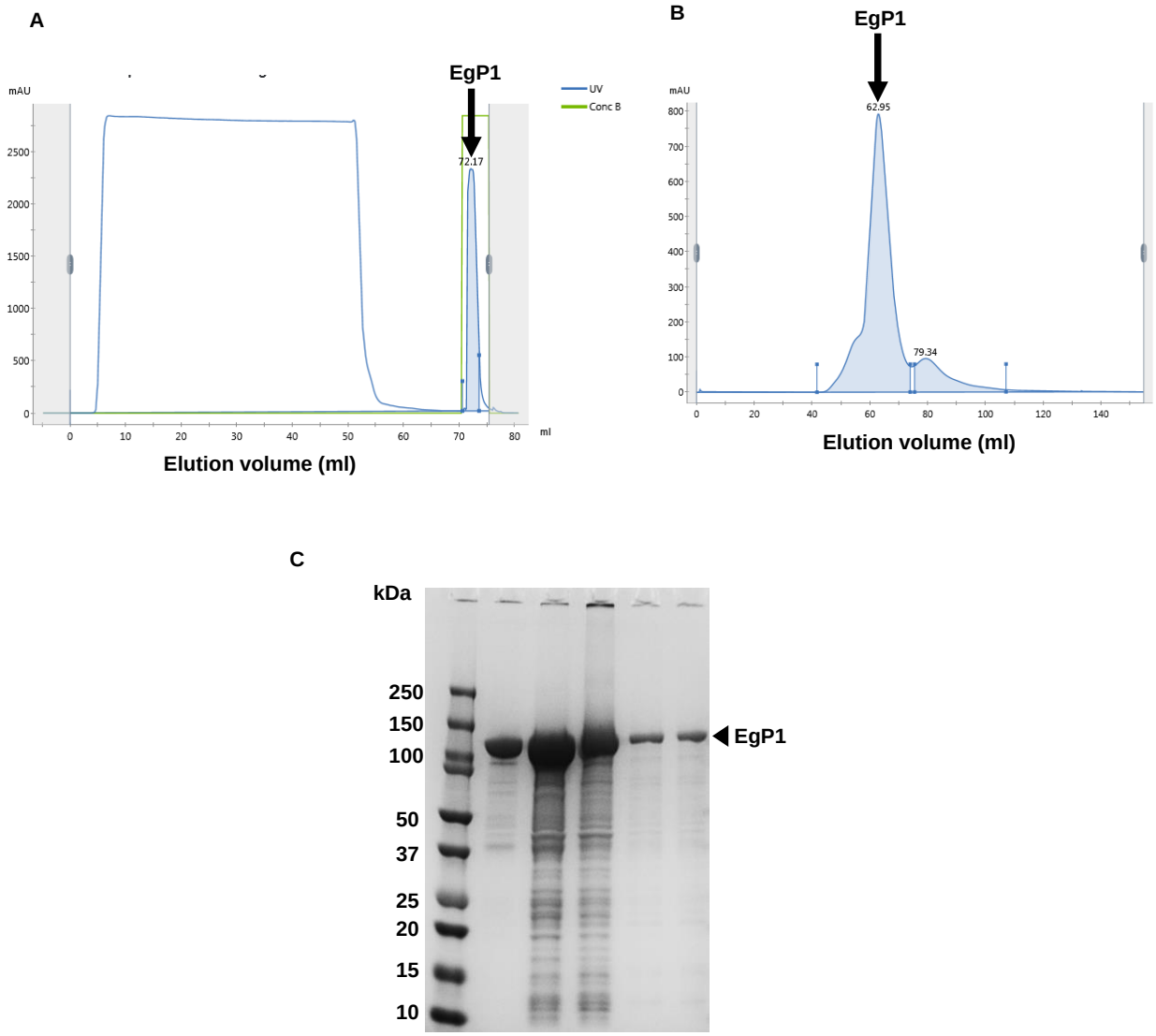

Fig. S3 IMAC and gel filtration purification of recombinant EgP1. (A) IMAC purification with UV detection (blue line). EgP1 was eluted in 100% of 10 mM HEPES pH 7.5, 250 mM NaCl, 500 mM imidazole (green line)). (B) Gel filtration of EgP1. The protein was eluted at 63 ml elution volume in 20 mM HEPES pH 7.5, 150 mM NaCl. (C) SDS-PAGE of EgP1 after IMAC purification.

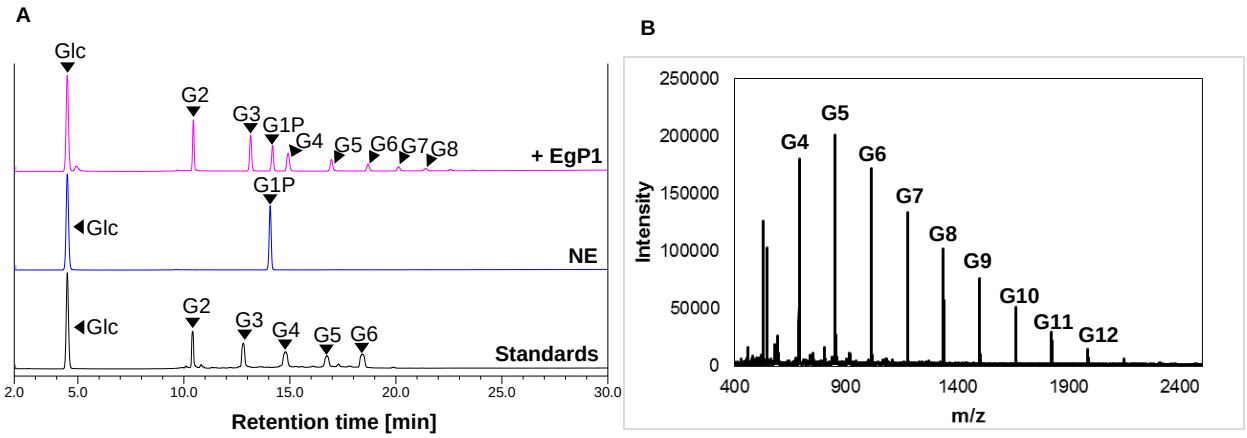

Fig. S4 Reverse phosphorolysis reaction catalyzed by EgP1. (A) HPAEC-PAD analysis of the reverse phosphorolysis carried out by EgP1 in the presence of Glc and G1P as substrates. (B) MALDI-ToF analysis after 1 hour.

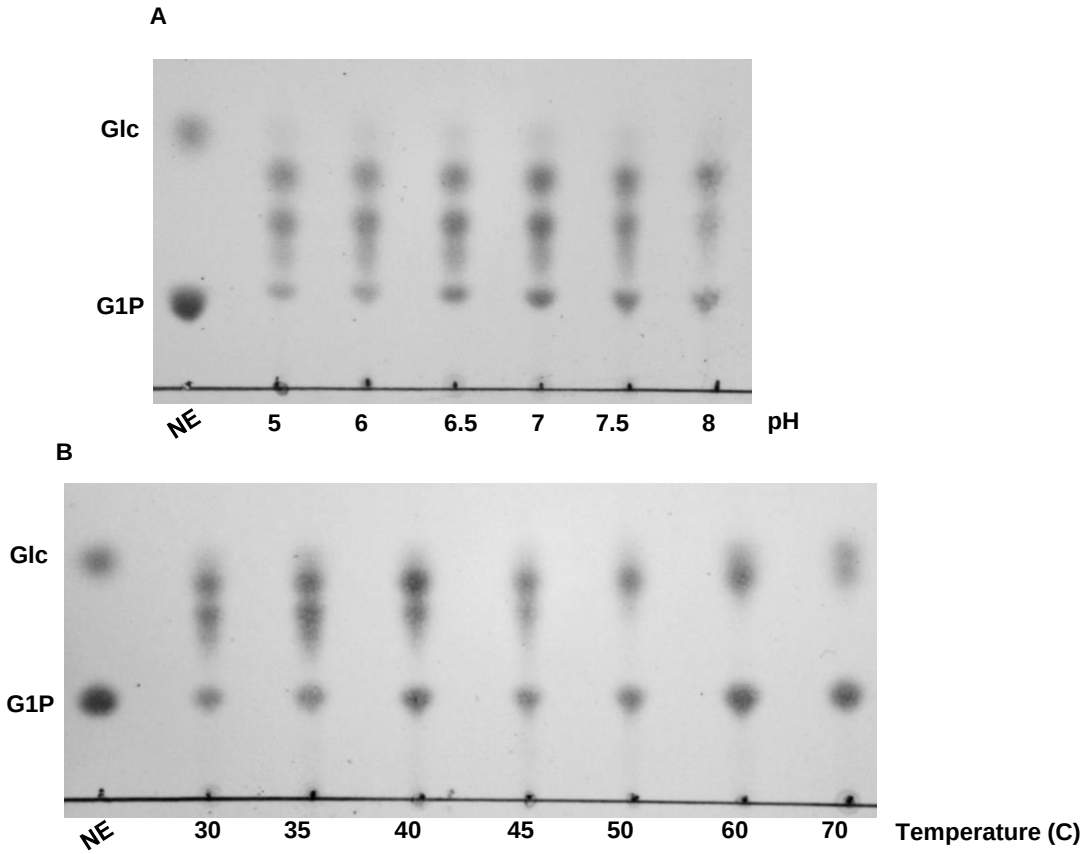

Fig. S5 TLC analysis of the reverse phosphorolysis carried out by EgP1 over a range of pH (A) at 30 °C or a range of temperatures (B) at pH 7.0.

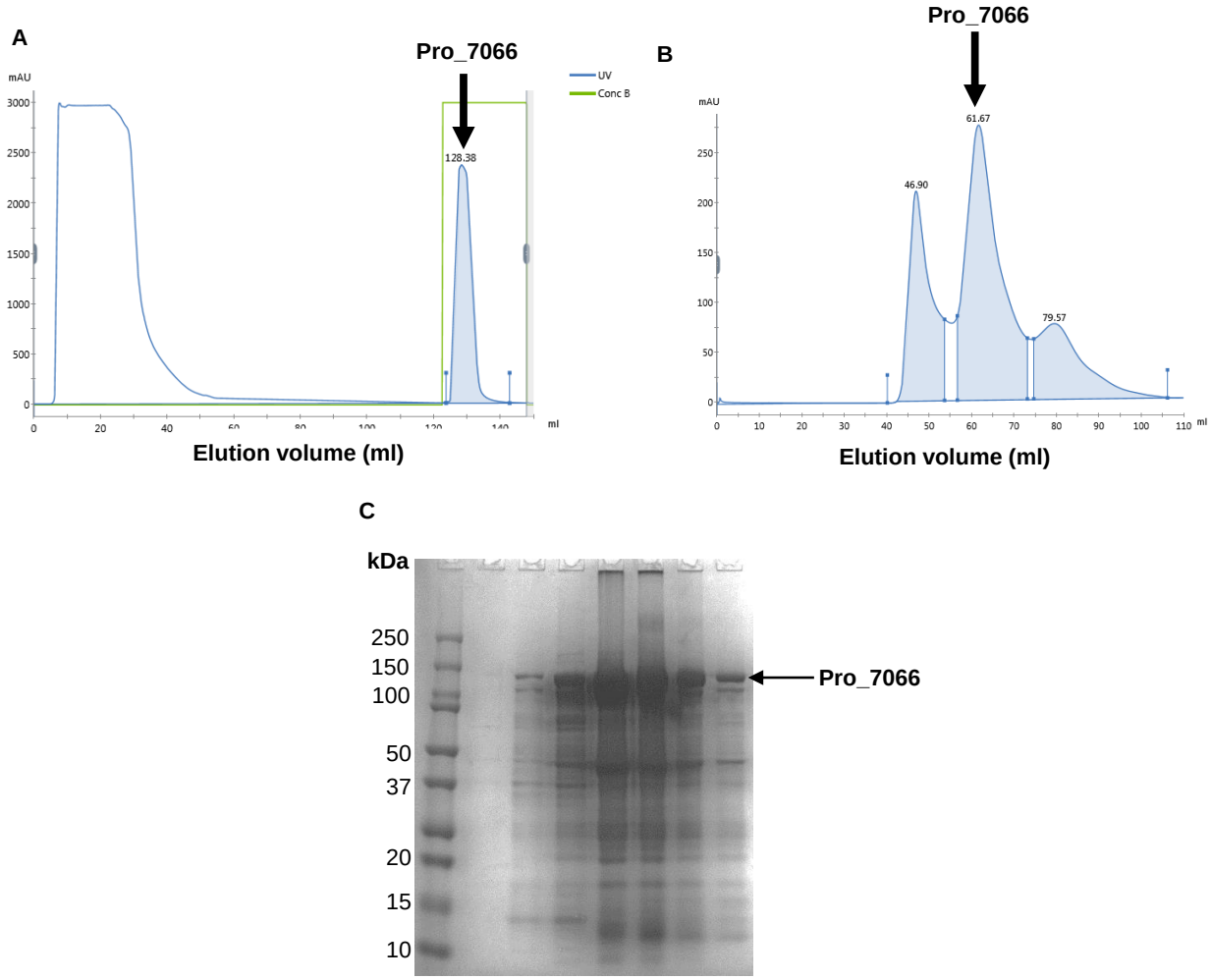

Fig. S6 IMAC and GF purification of recombinant Pro_7066. (A) IMAC purification with UV detection. Pro_7066 was eluted in 100% of 10 mM HEPES pH 7.5, 250 mM NaCl, 500 mM imidazole. (B) Gel filtration of Pro_7066. The protein was eluted at 63 ml elution volume in 20 mM HEPES pH 7.5, 150 mM NaCl. (C) SDS-PAGE of Pro_7066 after IMAC purification.

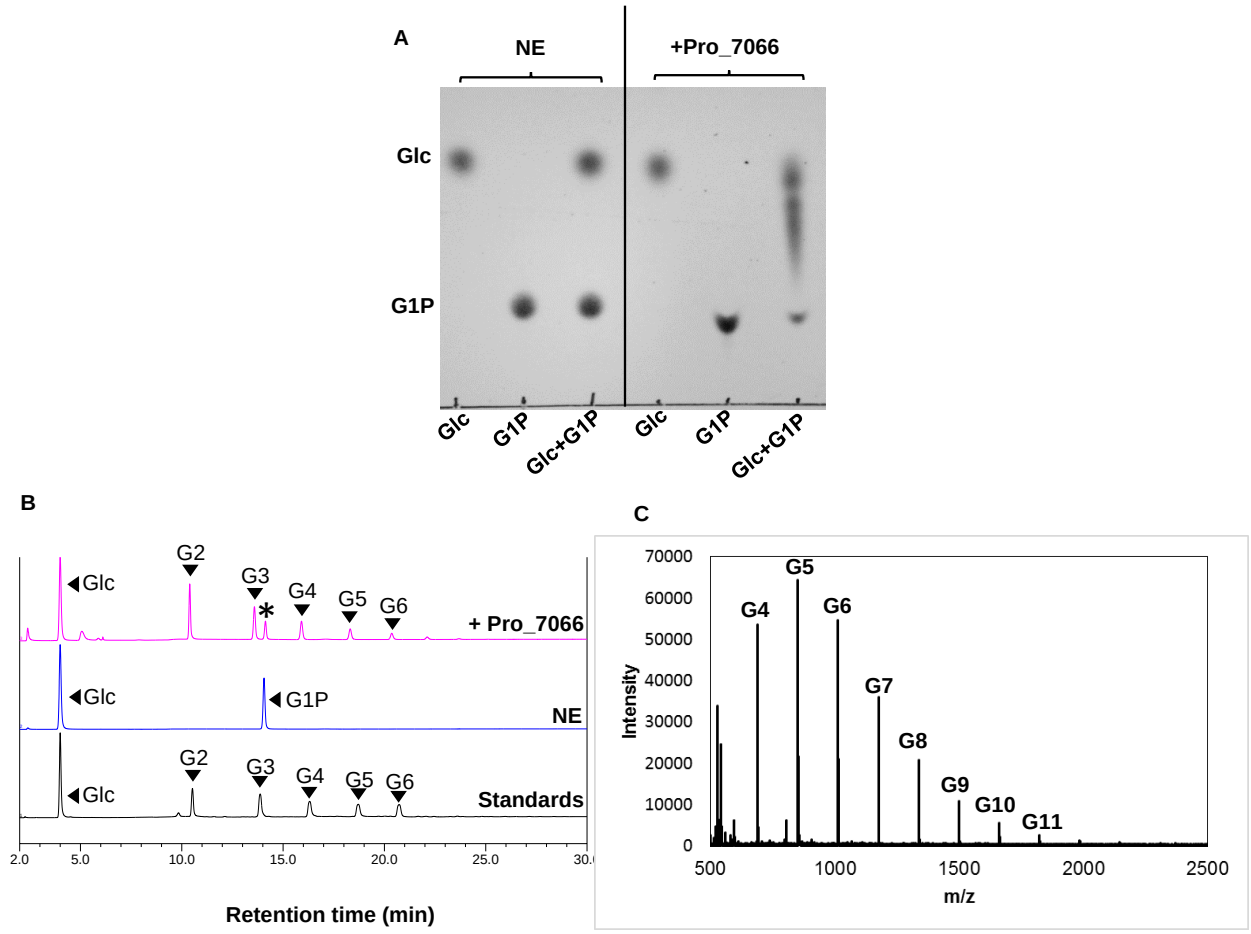

Fig. S7. Reverse phosphorolysis reaction catalyzed by Pro_7066 (A) TLC analysis of the reverse phosphorolysis carried out by Pro_7066. (B) HPAEC-PAD analysis of the reaction. (C) MALDI-ToF analysis of the reverse phosphorolysis after 1 hr. Asterisk = G1P.

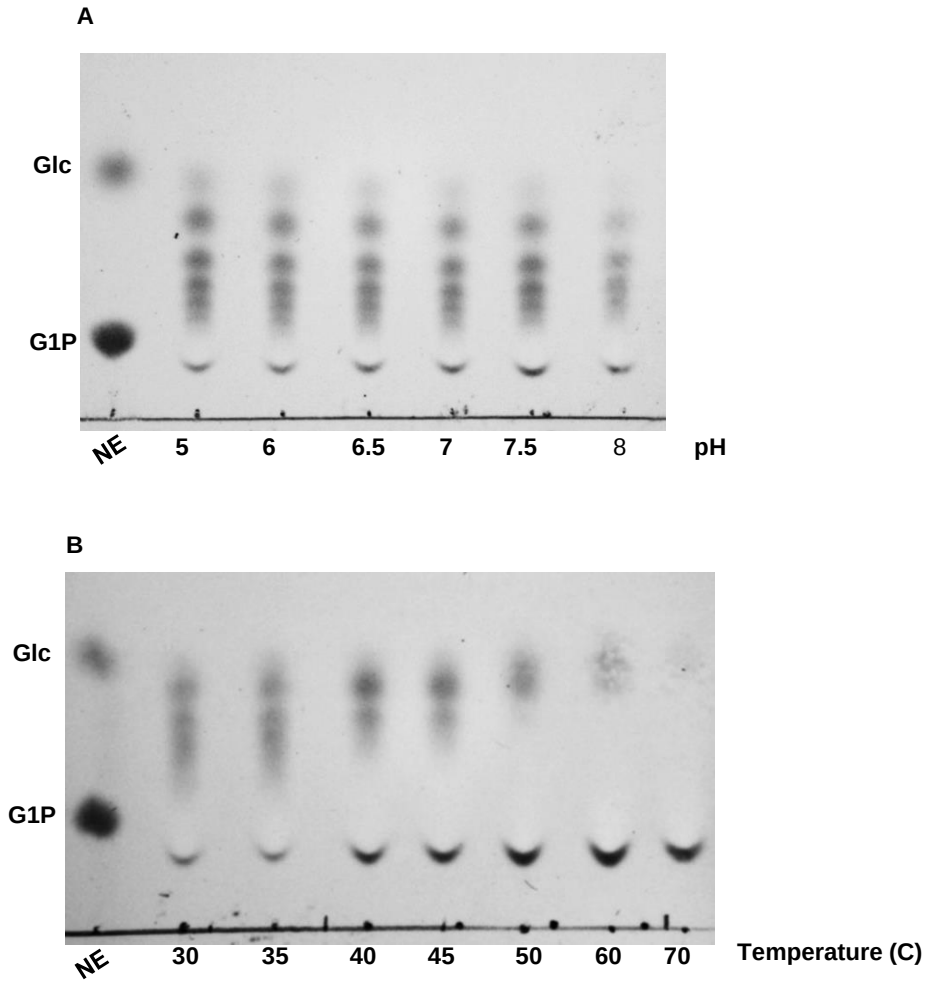

Fig. S8 TLC analysis of the reverse phosphorolysis carried out by Pro_7066 (A) over a range of pH at 30 °C or a range of temperatures (B) at pH 7.0.

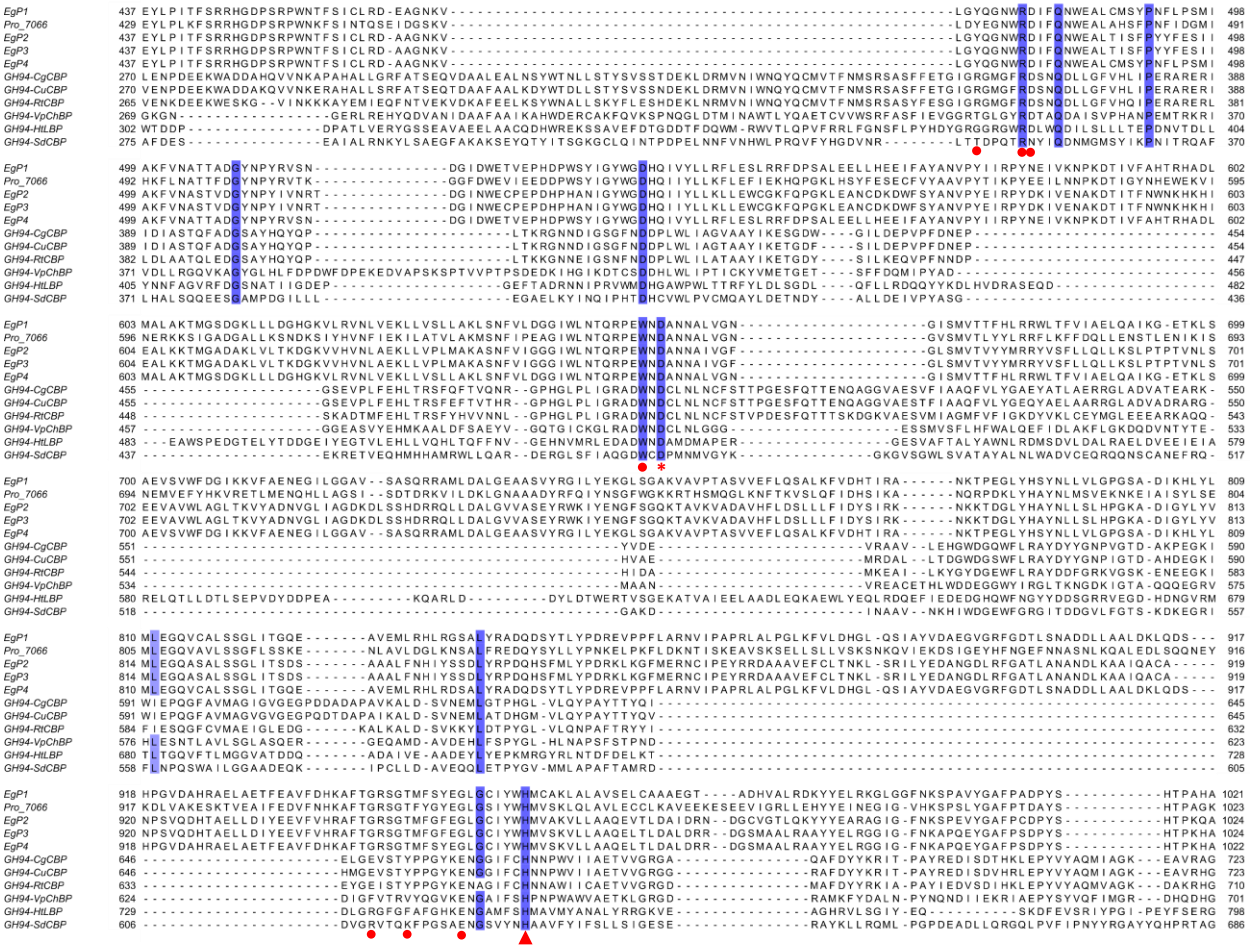

Fig. S9 Multiple alignment of EgP1-4, and Pro_7066 and GH94 enzymes revealed conservation of key amino acids required for the phosphorylase activity. Only a snap shot of the alignment is presented showing key amino acid residues. 100% conserved amino acid residues are highlighted in dark blue. Red circles indicated the residues involved in the subsite formation for sugar and sugar phosphate binding. Red triangle represents the residue involved in phosphate binding. Red asterisk represents the Asp catalytic residue.

List of supplemental data files

Supplemental Data File 1 – List of Euglena proteins identified from affinity proteomics and their predicted functions

Supplemental Data File 2 – BLAST analyses of GH149 amino acid sequences against EgP1

Supplemental Data File 3 – GC content analysis of bacterial GH149 DNA sequences

Supplemental Data File 4 – Sequence IDs for GH149 members

Identification of Euglena gracilis β-1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149

Sakonwan Kuhaudomlarp, Nicola J Patron, Bernard Henrissat, Martin Rejzek, Gerhard Saalbach and Robert A Field

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