

A EUROPEAN JOURNAL OF CHEMICAL BIOLOGY

CHEM **BIO** CHEM

SYNTHETIC BIOLOGY & BIO-NANOTECHNOLOGY

Accepted Article

Title: Unravelling the relaxed specificity of laminaribiose phosphorylase from *Paenibacillus* sp. strain YM-1 towards mannose 1-phosphate

Authors: Robert Andrew Field, Sakonwan Kuhaudomlarp, Samuel Walpole, Clare Stevenson, Sergey Nepogodiev, David Lawson, and Jesus Angulo

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *ChemBioChem* 10.1002/cbic.201800260

Link to VoR: <http://dx.doi.org/10.1002/cbic.201800260>

WILEY-VCH

www.chembiochem.org

A Journal of



Unravelling the specificity of laminaribiose phosphorylase from *Paenibacillus* sp. YM-1 towards donor substrates glucose/mannose 1-phosphate using X-ray crystallography and STD NMR spectroscopy

Sakonwan Kuhadomlarp,^[a] Samuel Walpole,^[b] Clare E.M. Stevenson,^[a] Sergey A. Nepogodiev,^[a] David M. Lawson,^[a] Jesus Angulo,^[b] Robert A. Field ^{*[a]}

Abstract: Glycoside phosphorylases (GPs) carry out a reversible phosphorolysis of carbohydrates into oligosaccharide acceptors and the corresponding sugar 1-phosphates. The reversibility of the reaction enables the use of GPs as biocatalysts for carbohydrate synthesis. Glycosyl hydrolase family 94 (GH94), which only comprises GPs, is one of the most studied GP families that have been used as biocatalysts for carbohydrate synthesis, in academic research and in industrial production. Understanding the mechanism of GH94 enzymes is a crucial step towards enzyme engineering to improve and expand the applications of these enzymes in synthesis. In this work with a GH94 laminaribiose phosphorylase from *Paenibacillus* sp. YM1 (PsLBP), we have demonstrated an enzymatic synthesis of disaccharide **1** (β -D-mannopyranosyl-(1 \rightarrow 3)-D-glucopyranose) using natural acceptor glucose and non-cognate donor substrate α -mannose 1-phosphate (Man1P). To investigate how the enzyme recognizes different sugar 1-phosphates, we solved the X-ray crystal structures of PsLBP in complex with Glc1P and Man1P, providing the first molecular detail of the recognition of a non-cognate donor substrate by GPs, which revealed the importance of hydrogen bonding between the active site residues and hydroxyl groups at C2, C4 and C6 of sugar 1-phosphates. Furthermore, we used STD NMR to support the crystallographic studies on the sugar 1-phosphates, as well as to provide further insights into the PsLBP recognition of the acceptors and the disaccharide products.

Introduction

Glycoside phosphorylases (GPs) are a group of carbohydrate-active enzymes that catalyse the reversible cleavage of glycosidic linkages in di- or oligo-saccharides by

transferring non-reducing end glycosyl residue to inorganic phosphate.^[1–4] The reverse reaction (synthetic reaction) of GPs is of practical importance because it can be used as an alternative method of enzymatic glycosylation utilizing sugar 1-phosphates as donor substrates. GPs have been classified based on their sequence identity into glycosyl hydrolase (GH) and glycosyl transferase (GT) families, or categorised into retaining and inverting phosphorylases, depending on the anomeric configuration in *O*-glycoside with respect to the sugar 1-phosphate substrates. Substrates for GP-catalyzed glycosylations are more readily available in comparison to that for GT-catalyzed reactions, making GPs attractive biocatalysts for carbohydrate syntheses. The use of GP biocatalysts have been demonstrated in academic research such as in the synthesis of homogeneous crystalline cellulose;^[5] self-assembled structures of alkylated cellulose;^[6] cellulose nanoribbon with primary amino groups;^[7] and formation of oligo(ethylene glycol)-bearing cellulose hydrogels;^[8] and more widely at industrial scale, such as for the synthesis of 2-*O*-(α -D-glucopyranosyl)-*sn*-glycerol, a cosmetic ingredient, by sucrose phosphorylase;^[9] kilogram scale synthesis of lacto-*N*-biose, a prebiotic made with lacto-*N*-biose phosphorylase;^[10] and the synthesis of disaccharide sweetener kojibiose, produced with a sucrose phosphorylase variant from *Bifidobacterium adolescentis*.^[11]

One of the most studied GP families is found in GH94 which includes GPs acting on β -1,2 (sophorose),^[12] β -1,3 (laminaribiose)^[13] and β -1,4-linked glycans (cellobiose,^[14,15] cellodextrin,^[16] chitobiose^[17] and cellobionic acid^[18]). Several characterised GH94 GPs show broad specificity towards non-physiological acceptor substrates, including cellodextrin phosphorylase (CDP) from *Clostridium sterococarium* has been used to produce cellobiose-containing antioxidants.^[19,20] Cellobiose phosphorylase (CBP) is capable of using xylose as an acceptor, to produce glucopyranosyl-xylose,^[21,22] and simple

[a] S. Kuhadomlarp, Dr C.E.M Stevenson, Dr S.A. Nepogodiev, Dr D.M. Lawson, Prof. R.A. Field
Department of Biological Chemistry
John Innes Centre
Norwich Research Park, Norwich, NR4 7UH, UK
E-mail: rob.field@jic.ac.uk

[b] S. Walpole, Dr J. Angulo
School of Pharmacy
University of East Anglia
Norwich, NR4 7TJ, UK

Supporting information for this article is given via a link at the end of the document.

FULL PAPER

alcohols, producing alkyl β -glucosides.^[23] While the relaxed acceptor specificity has been demonstrated, the specificity of GPs for sugar 1-phosphate is relatively narrow. The majority of GH94 GPs use Glc1P as a donor substrate, with the exception of chitobiose phosphorylase (ChBP), which uses α -D-GlcNAc 1-phosphate (GlcNAc1P) as its 'natural donor',^[17,24] although it can also use Glc1P with 20 times reduction in efficiency.^[17] Relaxed donor specificity has also been demonstrated for CDP from *Clostridium stercoararium*, which can use either Glc1P or α -D-galactose 1-phosphate (Gal1P) as its glycosyl donor for glycolipid synthesis, albeit with 10 times less efficiency on Gal1P.^[25] Both CBP and CDP from *Clostridium thermocellum* are capable of using α -D-glucosyl fluoride as a donor for the synthesis of cellobiose and cellodextrin.^[26] ChBP, which normally uses GlcNAc1P as its cognate donor.

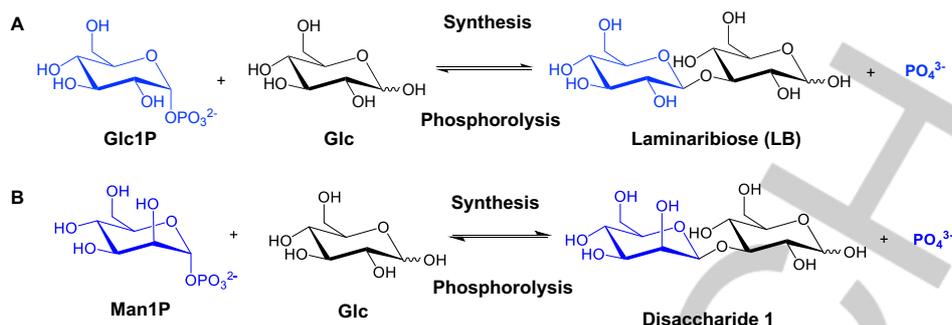
Numerous X-ray crystal structures are available for GH94 enzymes, either in the presence of phosphate or sulphate (PDB code 2CQS, 3QDE, 3RSY, 2CQT),^[27,28] the acceptors (PDB code 3S4B, 1V7X, 5H40, 4ZLG and 5NZ8),^[29–32] iminosugar inhibitors (PDB code 3QFY, 3QFZ, and 3QG0, 5H41),^[31,33] or disaccharide products (PDB code 3S4A, 4ZLF).^[32] These structures provide valuable resources that can be used to guide the engineering of GPs for non-cognate substrates. Structure-guided site-directed mutagenesis based has been performed extensively on cellobiose phosphorylase from *Cellvibrio gilvus* (CgCBP), including its 'conversion to a lactose phosphorylase'.^[34] In addition, a single mutation (E649C) in CgCBP created an enzyme variant that is capable of using methyl β -glucoside, ethyl β -glucoside and phenyl β -glucoside as acceptors.^[35,36] Another CgCBP variant was created by mutation of five amino acids within and around the entrance to the enzyme active, which broaden acceptor range to include both β - and α -glucosides.^[35] In contrast to the situation of acceptor substrate studies, the number of reported GP structures in complex with the sugar donors are relatively limited, with only a β -1,2-glucan phosphorylase from *Lachnoclostridium phytofermentans* (LpSOGP) in complex with Glc1P being reported (PDB code 5H42),^[31] which limits our understanding of the recognition of the sugar 1-phosphate donors by the GH94 family.

Whilst crystallographic studies provide valuable 'snapshots' of enzyme active sites, they do not capture the dynamics of the

enzyme-ligand interaction in solution. Therefore, other techniques to study protein-ligand interactions in solution are needed to complement the crystallographic data. Saturation transfer difference nuclear magnetic resonance spectroscopy (STD NMR) was developed to study the protein-ligand binding interaction in solution, based on the transfer of magnetization from the protein protons to the protons of the ligand whilst the ligand is bound.^[37] Those ligand protons in close contact with the protein exhibit the strongest STD NMR intensities, thus allowing the mapping of the ligand binding epitope.^[38] STD NMR can be used to facilitate the study of protein-glycan interactions, which is often difficult to study in solution due to the weak affinity as well as the flexibility and complexity of the carbohydrate ligands. STD NMR has been used to elucidate the specificity of protein-glycan interactions on different sialoglycan structures,^[39] and to reveal the importance of glycan polarity, which determines the interaction and subsequent biological activation of its receptor.^[40] This technique has also been used to study enzyme-carbohydrate interactions to elucidate recognition features that can be used for inhibitor design, such as the study of ligand recognition by enzymes that are involved in mycobacterial cell wall biosynthesis, including UDP-galactopyranose mutase^[41,42] and galactofuranosyltransferases.^[43] The same technique has been used to elucidate the binding of human blood group glycosyltransferases to their substrates, a process that is crucial to the biosynthesis of human blood antigen.^[44,45]

Following on from our efforts to understand GP structure-function relationships and their application in carbohydrate syntheses^[30,46,47], herein we investigated the GH94 laminaribiose phosphorylase from *Paenibacillus* sp. YM-1 (*PsLBP*), which has previously been reported for its specificity towards laminaribiose (β -D-glucopyranosyl-(1 \rightarrow 3)-D-glucopyranose) (Scheme 1A).^[13] We have evaluated the activity of *PsLBP* on a non-cognate donor, α -D-mannose 1-phosphate (Man1P), and cognate acceptor, glucose, in the production of β -D-mannopyranosyl-(1 \rightarrow 3)-D-glucopyranose (disaccharide **1**), (Scheme 1B). Furthermore, we used X-ray crystallography in conjunction with STD NMR to investigate the interaction between *PsLBP* and its substrates, in order to understand structural features that contribute to its donor substrate specificity.

FULL PAPER



Scheme 1. Reactions carried out by *PsLBP*. **(A)** Glc1P and Glc as a donor and acceptor respectively. **(B)** Man1P and Glc as a donor and acceptor respective

Results

Recombinant protein expression and *PsLBP* activity on the native donor and acceptor.

To obtain recombinant *PsLBP* protein for *in vitro* characterisation and X-ray crystallography, the gene encoding sequence of *PsLBP* was obtained from GenBank (accession number AB568298.2), codon-optimised for *E. coli* expression and synthesised by Gen9. The gene was amplified by PCR and cloned into a PopinF expression vector.^[48] The recombinant plasmid containing the *PsLBP* gene was introduced into BL21 (DE3) for protein expression. His₆-tagged recombinant protein was then produced and purified by immobilised affinity chromatography, followed by gel filtration. The gel filtration trace showed 3 different main peaks with different elution times (Figure 1A, peaks a, b and c). In order to investigate whether the 3 main peaks have similar activities, the phosphorylase activity of each peak was individually characterised in the synthetic direction (Scheme 1A) by phosphate release assays. No significant difference in activity was observed between these 3 peaks (data not shown) and that the enzyme in peak a and b are likely the higher oligomeric forms of *PsLBP*. Therefore, only peak c was used for further experiments based on its highest protein yield. SDS-PAGE analysis of peak c showed a major band of protein with an approximate size of 100 kDa, in agreement with the calculated mass of *PsLBP* monomer (101.6 kDa) (Figure 1B). However, gel filtration analysis of peak c against standard proteins showed that *PsLBP* formed a dimer in non-denaturing conditions with an estimated molecular mass of 240kDa.

To further confirm that the recombinant *PsLBP* was active, the enzyme was assayed in the synthetic direction (Scheme 1A) in the presence of its natural substrates (Glc and Glc1P) and the

reaction mixture was subjected to HPAEC-PAD analysis. The analysis showed that laminaribiose (LB) was produced (Figure 1C), complemented by the release of inorganic phosphate, which could be detected by a phosphate release assay (Figure S1).^[35]

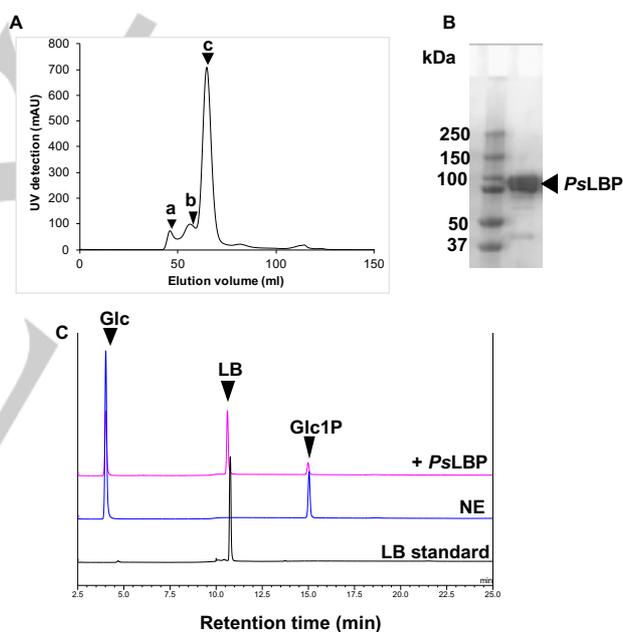


Figure 1. Expression and characterisation of recombinant *PsLBP*. **(A)** gel filtration analysis to determine the size of *PsLBP*. Elution volume of peak c (64.7 ml) was used to estimate the mass of the protein from a calibration curve constructed from standard proteins with known molecular mass. **(B)** SDS-PAGE analysis of the recombinant protein after IMAC and gel filtration. **(C)** HPAEC-PAD analysis of the synthetic reaction carried out by *PsLBP* when incubated the enzyme with 10 mM Glc and 10 mM Glc1P for 30 minutes at 45 °C. NE = no enzyme control.

Activity towards non-cognate donors.

PsLBP activity has previously been screened on several non-cognate acceptors, including mannose, methyl β -glucoside, 2-deoxyglucose and 6-deoxyglucose, with 50-100 fold reduction in

FULL PAPER

the activity compare to that of Glc.^[13] However, the specificity towards non-cognate sugar 1-phosphate donors has not been reported. To assess the donor specificity of *PsLBP*, the enzyme was assayed in the presence of α -D-galactosamine 1-phosphate (GalN1P), α -D-glucosamine 1-phosphate (GlcN1P), α -galacturonic-acid 1-phosphate (GalA1P), Gal1P or Man1P as donors and Glc as an acceptor. TLC and HPAEC-PAD analysis of the reactions showed that the enzyme can use Man1P as a donor, as indicated by the presence of an additional spot on TLC, corresponding to a generation of disaccharide **1** (Figure 2A and B). Kinetic parameters for the synthetic reaction using either Glc1P or Man1P as donors and Glc as a receptor showed comparable K_M values for Glc1P and Man1P, while the k_{cat} for Glc1P is more than 100-fold higher than that for Man1P (Table 1).

Table 1. Kinetic data of *PsLBP* for Glc1P and Man1P as donors in the presence of 10 mM Glc as an acceptor and for Glc in the presence of 10 mM Glc1P.

Donors	k_{cat} (s ⁻¹)	K_M (mM)	k_{cat}/K_M (s ⁻¹ mM ⁻¹)
Glc1P	13.0 ± 1.4	4.20 ± 1.5	3.07
Man1P	0.08 ± 0.01	3.80 ± 1.0	0.02
Glc	15.4 ± 1.3	6.04 ± 1.3	2.55

A large scale enzymatic reaction was carried out in 5-ml reaction containing 700 μ g of the enzyme, 10 mM Glc and 20 mM Man1P, incubating for 15 hr to produce milligram quantities of disaccharide **1** (Figure 2C and S2A). Unreacted Man1P and inorganic phosphate by-product were then removed from the reaction mixture by anion exchange chromatography and the disaccharide **1** was isolated by GPC (Figure S2A). The isolated disaccharide **1** was analysed on TLC, which showed that only one product was obtained with no Glc contamination (Figure S2B). Mass spectrometry analysis of disaccharide **1** on the TLC plate showed a major peak with m/z of 364.9, corresponding to mass of a disaccharide with sodium adduct (Figure S2C). These data are in line with those for reported by Awad *et al.* for the same disaccharide arising from GH130 β -1,3-mannan phosphorylase-mediated synthesis.^[49]

Assignment of NMR signals of disaccharide **1** (Figure S3 and Figure S4) was possible with the use of 2D experiments: COSY (Figure S5), HSQC (Figure S6) and 2D non-decoupled HSQC (Figure S7), as well as literature data for model methyl β -mannoside (Table S1).^[50] Assignment was also helped by

simulated spectra of disaccharide **1** generated by CASPER program (Table S2).^[51] Downfield positions of resonances of C-3 of β -Glc (84.6 ppm) and C-3 of α -Glc (82.1 ppm) residues with respect to corresponding signals in D-glucose (73.8 and 77.0 for α - and β -anomers respectively)^[52] indicated presence of 3-O-glycosylated glucopyranose unit. Coupled HSQC experiment revealed $^1J_{C-H}$ of 163 Hz for anomeric signals of mannopyranose residue, the value is characteristic of β -mannopyranosides (Figure S7).^[52] Most of carbon signals of the mannosyl residue in ^{13}C NMR of **1** are split into two very close peaks, due to the presence of α/β -anomers of Glc residue (Figure S3). Anomeric signals of non-reducing β -Glc residue are expected to appear at 103–104 ppm^[50,53] but there are no peaks in that region, therefore presence β -glucosides can be excluded. By comparison with the previously reported Glc- β -1,3-Glc ^{13}C NMR spectral,^[54] signals at 95.7 and 92.1 ppm can be assigned to C-1 of reducing β -Glc and α -Glc moieties respectively, whereas signals at 84.6 and 82.1 ppm can be assigned to C-3 β -Glc and C-3 α -Glc respectively.

Overall crystal structures of *PsLBP*

Three *PsLBP* structures were determined and designated according to the ligands found in their active sites; SO_4^{2-} complex, Glc1P complex and Man1P complex. All structures belong to the same space group (P4₁2₁2) and contain two subunits per asymmetric unit, which are related by a non-crystallographic 2-fold axis that superposes them with an R.M.S.D of 0.651 Å. The two copies of the molecule in the asymmetric unit formed a biological homodimer with an interfacial area of ~3360 Å² as calculated by jsPISA.^[55] The formation of homodimer observed in crystal structures is in agreement with the gel filtration analysis, where *PsLBP* was eluted as a dimer. Other GH94 enzymes also form homodimer with the exception to *LpSOGP* which is the only reported monomeric GH94.

Each *PsLBP* monomer consists of 4 domains (Figure 3A and B), which are an N-terminal β -sandwich (residues 1-297; yellow), a helical linker region (residues 298-327; lilac), an (α/α)₆ catalytic domain (residues 328-808; green) and a C-terminal domain (residues 809-911; red). The domain organisation in *PsLBP* is similar to that observed in other GH94 disaccharide phosphorylases, which include cellobiose phosphorylases from *Cellomonas uda* (*CuCBP*),^[56] *CgCBP*,^[27] chitobiose phosphorylase from *Vibrio proteolyticus* (*VpCBP*)^[29] and cellobionic acid phosphorylase from *Saccharophagus degradans*

FULL PAPER

(SdCBAP).^[57] PsLBP lacks the extended N-terminal α/β domain (Figure 3B, purple) that is only present in GH94 oligosaccharide phosphorylases including cellodextrin phosphorylase from *Ruminococcus albus* has Gln646 instead of the conserved His residue that is found in other GH94 phosphorylases.^[58]

The role of the extended N-terminal domain in *RtCDP* was proposed to be involved in the interaction of the lower portions of the homodimer, causing the upper portions to move apart, which leads to a widening of the active site to accommodate a larger acceptor (i.e. cellodextrin) in *RtCDP*.^[30]

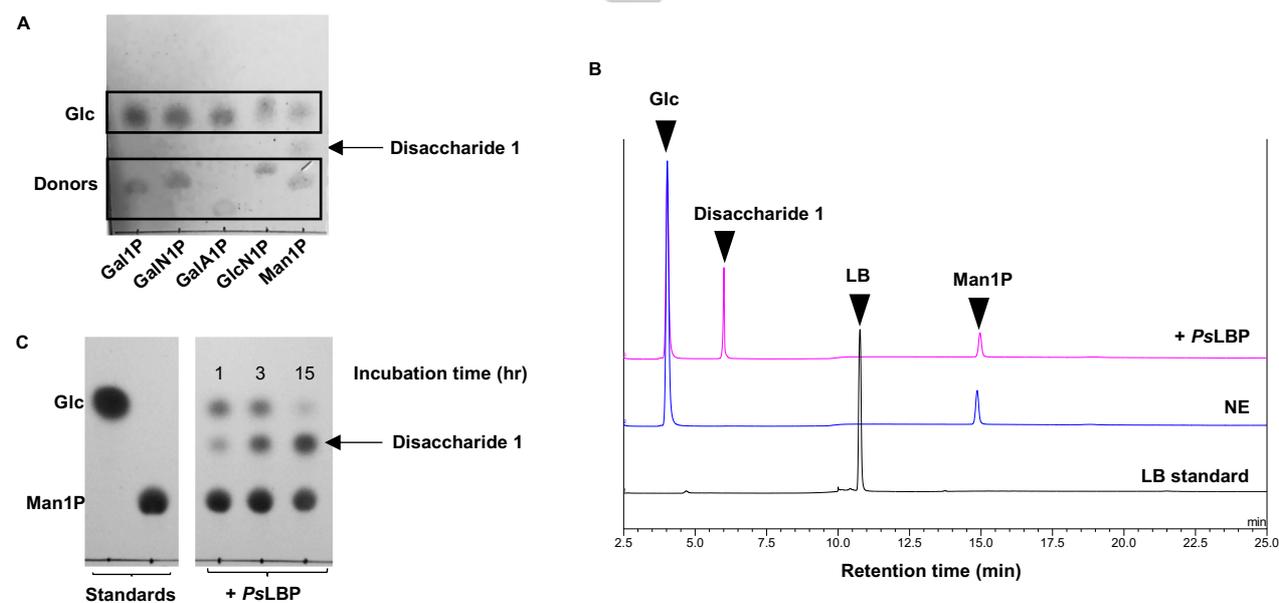
Phosphate recognition by PsLBP

In the SO_4^{2-} complex (Figure S9A), SO_4^{2-} which was derived from the precipitant used for crystallization, occupied a similar position to phosphate and sulphate in other GH94 disaccharide phosphorylases such as in *CgCBP* (Figure S9B). The SO_4^{2-} molecule forms hydrogen bonds with R353, T731 and E782 side chains. From the amino acid sequence alignment of *PsLBP*, *CgCBP* and *RtCDP*, H739 in *PsLBP* aligned with the conserved histidine residues in *CgCBP* and *RtCDP* that forms a hydrogen bond with phosphate in the active site (Figure S9B). However, the distance between H739 side chain and SO_4^{2-} in *PsLBP* structure is greater than hydrogen bonding distance, suggesting that H739 may not be essential for phosphate recognition. This is further supported by three evidences. Firstly, when the corresponding histidine (H666) in *CgCBP* was mutated to Asn, a phosphate molecule still bound to the active site of the enzyme (PDB code 3ACT). Secondly, the phosphate moiety in Glc1P in complex with *LpSOGP* does not form hydrogen bond with the corresponding

histidine (H924) (Figure 4B).^[31] Lastly, a wild-type cellodextrin phosphorylase from *Ruminococcus albus* has Gln646 instead of the conserved His residue that is found in other GH94 phosphorylases.^[58]

Glc1P and Man1P recognition by PsLBP

The overall structure of SO_4^{2-} complex and Glc1P complex are very similar (R.M.S.D of 0.23 Å for a dimer on dimer superposition). Glc1P was bound with the pyranose ring in ${}^4\text{C}_1$ conformation and α -anomeric configuration of phosphate at C1 position, supporting the enzyme specificity for sugar 1-phosphate in α -anomeric configuration. Glc1P is completely buried within a donor subsite (-1 subsite), which is formed entirely within a single subunit of *PsLBP*. Comparison between Glc1P complex of *PsLBP* with that of *LpSOGP* (Figure 4B, PDB code 5H42) showed that Glc1P recognition by the two proteins is different. In *PsLBP*, the hydroxyl group on C3 form a hydrogen bond with R374 side chain (Figure 4A), whereas in *LpSOGP*, the same hydroxyl group forms hydrogen bonds with R630 and D631. In *LpSOGP* structure, D631 also involves in the recognition of the hydroxyl group on C2 via hydrogen bonding, whereas in *PsLBP*, the hydroxyl group forms hydrogen bond with R353. The equivalence of D631 in *PsLBP* (D375) form hydrogen bond with neither the hydroxyl groups on C2 or C3 because the distance between the hydroxyl groups on C2 and C3 and D375 is greater than the hydrogen bonding distance. The phosphate moiety in Glc1P also interact differently with the enzyme active sites. In *PsLBP*, the phosphate moiety



FULL PAPER

Figure 2. *PsLBP* activity on Man1P and production of disaccharide 1. (A) Screening of the synthetic activity of *PsLBP* (8 μ g) on 4 different donors (10 mM) in the presence of 10 mM Glc as an acceptor. The reactions were incubated at 45 °C for 30 minutes. (B) HPAEC-PAD analysis of the reaction containing Glc and Man1P as substrates. (C) TLC analysis of the Glc + Man1P large scale reaction at time intervals.

forms hydrogen bonds with the side chains of R353 and E782, similar to those found in SO_4^{2-} complex, whereas in *Lp*SOGP, the phosphate moiety in Glc1P forms hydrogen bonds with S1005 and Y922.

The Man1P complex (Figure 4C) represents the first GP structure in complex with a non-cognate sugar 1-phosphate donor. The overall structure of Glc1P complex and Man1P complex are very similar (R.M.S.D of 0.14 Å for a dimer to dimer superposition). Man1P position almost overlaps completely with that of Glc1P, suggesting similar binding mode to Glc1P. However, the

hydrogen bond between axial hydroxyl group on C2 and R353 side chain cannot form since the distance between the hydroxyl group on C2 and R353 side chain is 5.1 Å (in contrast to 3.2 Å between the equatorial hydroxyl group on C2 and R353 in Glc1P complex). The loss of hydrogen bond between C2-hydroxyl group and R353 side chain in Man1P complex may have a negative impact on the conversion of Man1P and Glc to disaccharide 1, which is reflected in the reduction in k_{cat}/K_m from 3.07 (when Glc1P was used as a donor) to 0.02 (when Man1P was used) that were previously observed in the kinetic studies.

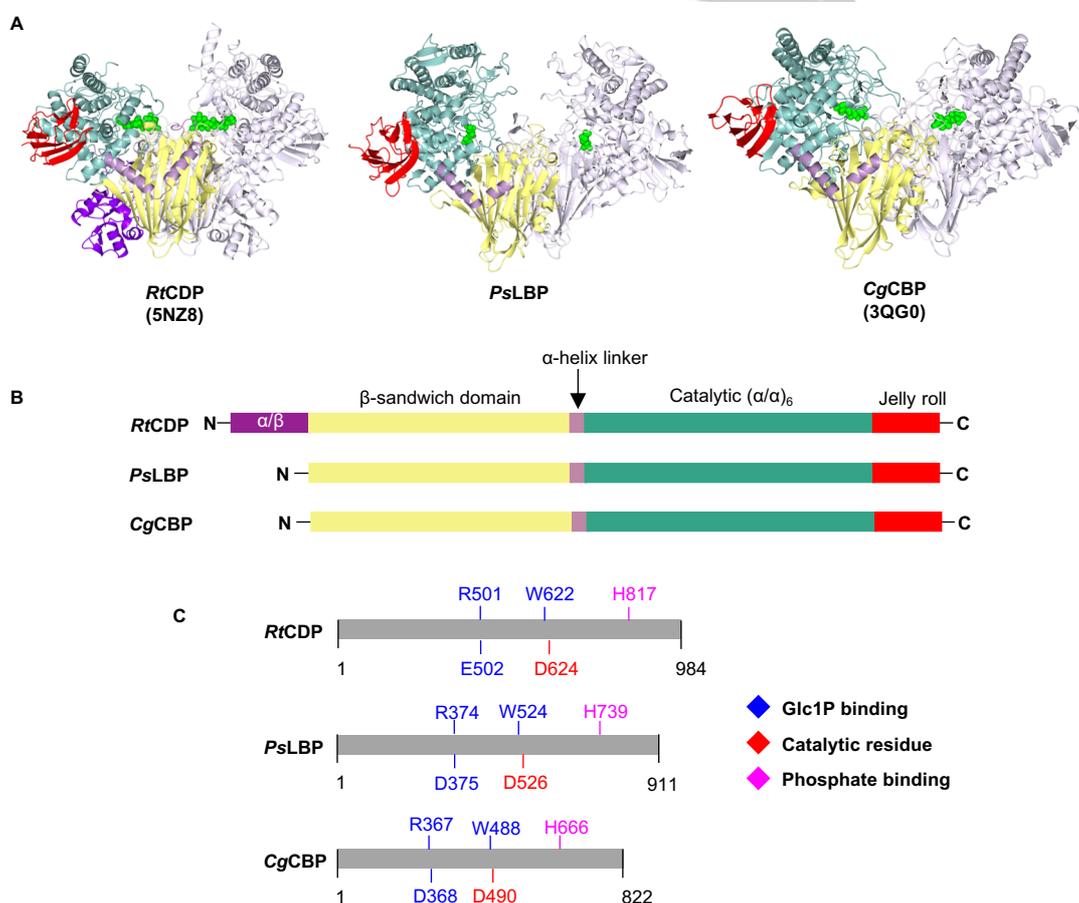


Figure 3. Structural comparison between *CgCBP*, *RtCDP* and *PsLBP*. (A) Comparison between the GH94 structures. All proteins are in dimeric form. The domains are coloured in one subunit only, while the other subunit is coloured in grey. Green spheres represent ligands bound in the active sites. (B) Sequences of *RtCDP*, *PsLBP* and *CgCBP* coloured according to the domains. (C) Summary of conserved residues found in the amino acid sequence alignment of *CgCBP*, *PsLBP* and *RtCDP*. The conserved residues are involved in Glc1P binding (blue), phosphate binding (magenta) or the predicted catalytic aspartate (red). Detail of the alignment can be found in Figure S8.

FULL PAPER

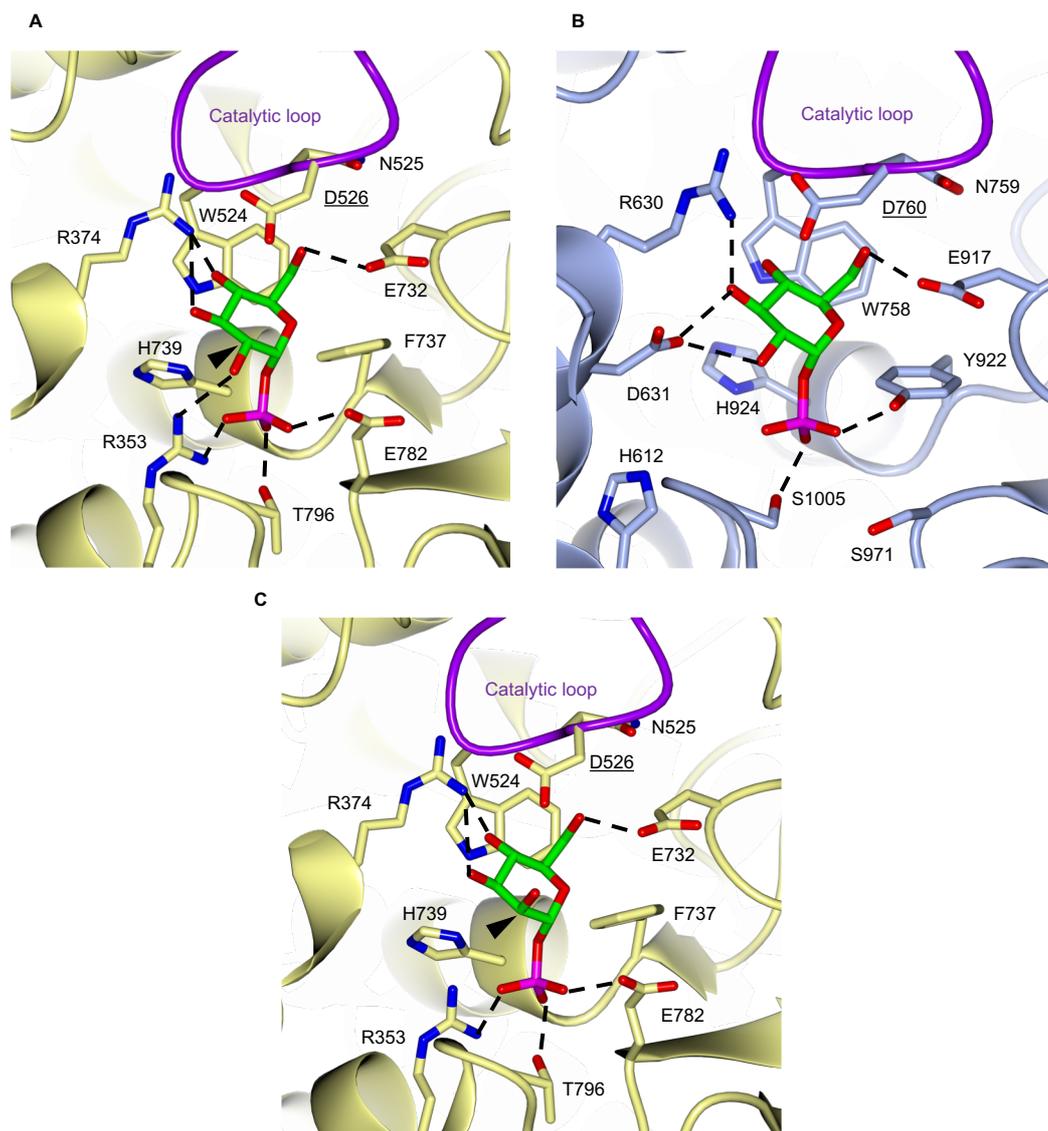


Figure 4. Comparison between *PsLBP* (A) and *LpSOGP* (B) in complex with *Glc1P*. (C) The active site of *PsLBP* in complex with *Man1P*. The C2 position on the pyranose ring is indicated with black arrow head. The protein backbone is shown in cartoon representation and the neighbouring side chains in stick representation. The conserved catalytic loop is presented in purple. The catalytic residue (D526) is underlined.

Catalytic loop in GH94 enzymes

The catalytic loop (purple, Figure 4) is a highly conserved feature among GH94 structures. In *PsLBP*, this loop consists of WND motif (W524, N525 and D526), with D526 as a predicted catalytic residue. The tryptophan residue (W524) is structurally conserved among GH94 members and provides a hydrophobic platform for the binding of the donor.^[30] Comparison between the *CgCBP* and *PsLBP* structures showed that the *PsLBP* active site is slightly more closed in comparison to *CgCBP* due to several different structural features. Firstly, the length of the *PsLBP* catalytic loop is 10-amino acid residues shorter than that of *CBP*, and only partially occludes the active site. In contrast, the extended

catalytic loop in *CgCBP* forms a lid-like structure that extends over the active site (Figure S10A and B, purple). Secondly, the position of the 'adjacent loop' (Figure S10A and B, brown) runs in parallel to a significant proportion of catalytic loop in *CBP*, forming a zipper-like interface that further encloses the *CgCBP* active site. The absence of the extended catalytic loop in *PsLBP* may be compensated for the presence of a loop in the β -sandwich domain of the opposing subunit, which is six residues longer than the equivalent loop in *CgCBP*. This forms a β -hairpin 'gate' loop, which overlaps with the extended catalytic loop in *CgCBP* (Figure S10BC and D, cyan).

FULL PAPER

Moreover, the opposing loop in CgCBP projects into the active site more than that of PsLBP (Figure S10, red). Whereas in PsLBP, the active site is relatively open, since the opposing and adjacent loops are located further away from the active site with respect to those in the CgCBP structure. The relatively “open” state of the PsLBP active site may be representative of the initial binding of the sugar 1-phosphate donor before the synthetic reaction occurs or the final stage of phosphorolysis where sugar 1-phosphate is about to be released. On the other hand, the relatively “closed” state observed in CgCBP structure likely represents the intermediate state of the catalysis where the glycosidic bond is either being broken during phosphorolysis or being formed during the synthetic reaction.

STD NMR and CORCEMA-ST suggests different binding affinity of Glc1P and Man1P to PsLBP

To study the interaction between PsLBP and Glc1P or Man1P, binding epitopes for Glc1P and Man1P were constructed (Figure 5A and B) using the initial rates approach (SI 1) to avoid overestimation of slow relaxing protons at large saturation times and to eliminate any effect of ligand rebinding.^[59] The epitope maps provide a qualitative measure of proximity of the protons of the ligands to the protein surface, with larger values indicating more intimate contact.^[38]

For Glc1P (Figure 5A) all the ligand protons received strong saturation from the protein indicating that the sugar is intimately recognised making contacts all along the ring. Nevertheless, the exocyclic H6 protons exhibit the strongest STD intensities, followed by H4, whilst H1 and H2 exhibit the weakest STD intensities (Table S3). This suggests that the H4/H6 area is buried within the binding cavity, whilst the H1/H2 is further from it, and most likely being more exposed to the solvent. This is in very good agreement with the X-ray structure (Figure 4A) and with the role of Glc1P as a donor substrate, since an open cavity exists adjacent to the anomeric proton in order to accommodate the acceptor substrate.

The binding epitope of Man1P (Figure 5B) follows a similar ranking as seen for Glc1P, suggesting that Man1P binds to the same subsite as Glc1P, with a rather similar binding mode. This is in agreement with the X-ray structure (Figure 4C) and the observation that Man1P can act as donor substrate. Importantly, the larger dispersion of STD values observed for Man1P,

compared to Glc1P, is an indicative of a shorter residence time, and therefore weaker affinity of Man1P. Since the only difference between the two substrates is the configuration at C2, it follows that a favourable interaction, observed for the native Glc1P, is broken by the inversion in Man1P.

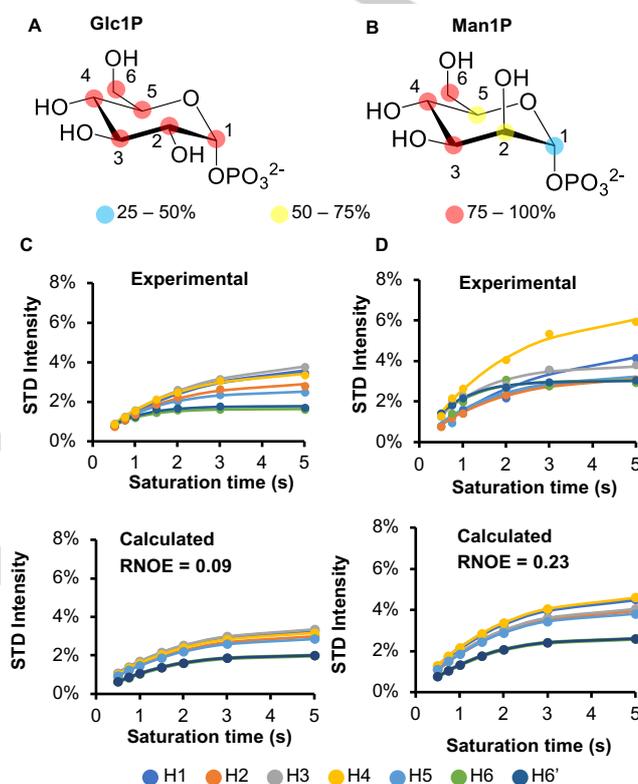


Figure 5. STD NMR experiments on Glc1P and Man1P. (A) Binding epitope of Glc1P in the presence of PsLBP. (B) Binding epitope of Man1P in the presence of PsLBP. Colours represent normalised values of STD₀ at each position depicted in the structure. For each ligand, the values of STD₀ are normalised against the largest value (see Table S3 for the normalised STD₀). (C) Experimentally determined STD build-up curves for Glc1P binding to PsLBP (top) and CORCEMA-ST-calculated STD intensities (bottom). The NOE R-factor (RNOE) between the experimental and calculated data is 0.09. (D) Experimentally determined STD build-up curves for Man1P binding to PsLBP (top) and CORCEMA-ST-calculated STD intensities (bottom). The NOE R-factor between the experimental and calculated data is 0.23. For experimental data, circles show observed STD intensities, whilst curves are determined from least squares fitting to the equation in SI 1.

CORCEMA-ST^[60] can be used to predict STD intensities based on a three-dimensional model of the protein-ligand complex, e.g. derived from X-ray crystallography, NMR spectroscopy, or molecular modelling. These intensities can then be compared against experimental STDs to determine whether the model of the complex is able to quantitatively explain the experimental STD NMR data. This comparison is carried out using the so-called NOE R-factor (RNOE) (SI 2).^[61] Generally, a RNOE of less than 0.3 is considered a good fit to the data.

FULL PAPER

For the crystal structure of the Glc1P complex (*vide supra*), the RNOE was 0.09 (Figure 5C), indicating an excellent fit between the crystal-derived model complex and the solution-state complex observed by NMR. For the crystal structure of the Man1P complex, the RNOE was 0.23 (Figure 5D). The poorer, yet still very good, fit indicates a less well-defined binding mode in solution, given the lower affinity.

STD NMR identified the binding subsite for Glc and the directional binding of LB and disaccharide 1 to *PsLBP*

In the study of Glc binding, the exchange between its α - and β -anomers precluded such a detailed study, due to the combination of different concentrations of the species and the significant peak overlap observed from both anomers. However, it is clear from the spectra that although similar concentrations of each anomer are present, the STD intensities from the α -anomer are very much weaker than those from the β -anomer (Figure 6A and C and S11). This indicates that the β -anomer is preferentially recognised by the enzyme. This suggests several things. Firstly, the subsite occupied by Glc1P and Man1P must require a sugar 1-phosphate, otherwise it would be expected that α -Glc would bind well. This then means that Glc must bind to a separate subsite, in agreement with its role as the acceptor substrate. Finally, the structure of this subsite must be such that α -Glc, with its axial C1-hydroxyl group, is unable to bind, perhaps due to steric interactions.

As in Glc, the reducing end of LB exists as an equilibrium between both of its α - and β -anomers. Again, it is observed that the β -anomer is preferentially recognised by the enzyme (Figure 6B and S12). Given the proposed reaction mechanism and the polarity of the disaccharide, it appears that the non-reducing ring of LB (Glc1) binds to the same subsite as Glc1P/Man1P, whilst the corresponding reducing sugar (Glc2 β) binds to the same subsite as Glc. The saturation transferred to the reducing end of LB (Figure 6D, bottom) is much stronger than that received by the non-reducing ring (Figure 6D, top), suggesting that the key interaction between the ligand and the enzyme is formed with the reducing sugar hexopyranose moiety. This agrees with the previous observation that the -1 subsite can only recognise sugar 1-phosphates effectively.

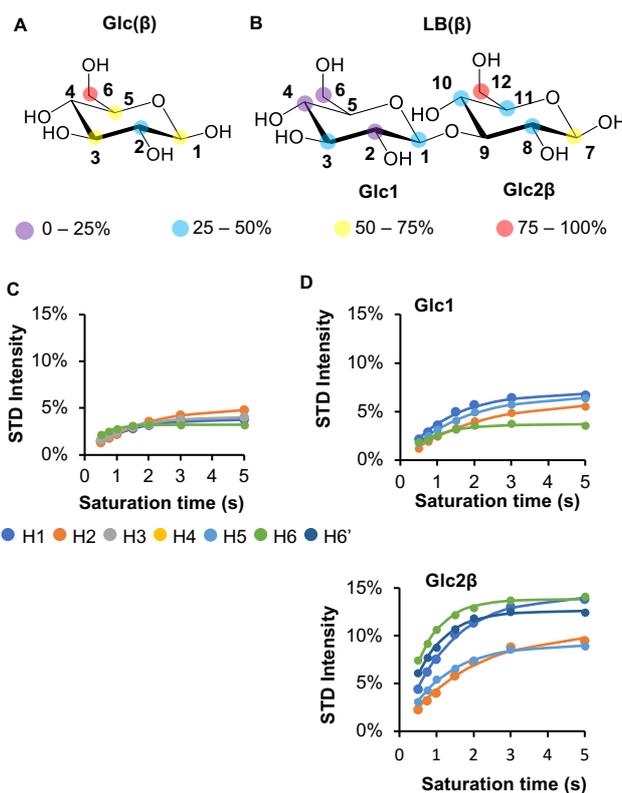


Figure 6. STD NMR experiments on Glc and LB. (A) Binding epitope of Glc(β) in the presence of *PsLBP*. **(B)** Binding epitope of LB(β) in the presence of *PsLBP*. Colours represent normalised values of STD₀ at each position depicted in the structure. For each ligand, the values of STD₀ are normalised against the largest value (see Table S3 for the normalised STD₀). **(C)** Experimentally determined STD build-up curves for Glc(β) binding to *PsLBP*. **(D)** Experimentally determined STD build-up curves for LB(β) binding to *PsLBP*. Circles show observed STD intensities, whilst curves are determined from least squares fitting to the equation in SI 1.

STD NMR was also performed on disaccharide **1** and *PsLBP* to determine the interaction between the non-cognate product and the protein. Overall interaction between *PsLBP* and disaccharide **1** is similar to that described in LB, with the main contacts appearing to be with Glc2, in particular with the H6's as previously mentioned for LB (Figure 7A and 7B). The magnitude of the STD intensities is comparable to those of LB, suggesting the affinity is similar. This would make sense given that the Glc2, common to both, seems to be the most important for recognition. Moreover, only the β -anomer binds strongly to *PsLBP*, with the STD intensities for the α -anomer being very much weaker (Figure 7C). However, the STDs for the H6 of Man1 are a lot stronger than in Glc1 in LB. This may be because Man binds in a slightly different orientation, which is probably to be expected given the different stereochemistry of the C2.

FULL PAPER

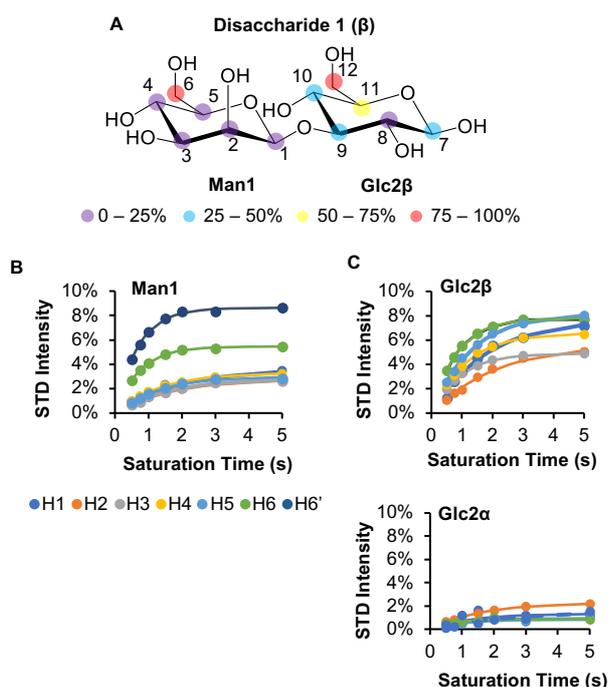


Figure 7. STD NMR experiments on disaccharide 1. (A) Binding epitope of disaccharide 1(β) in the presence of *PsLBP*. Colours represent normalised values of STD_0 at each position depicted in the structure. For each ligand, the values of STD_0 are normalised against the largest value (see Table S3 for the normalised STD_0). (B) Experimentally determined STD build-up curves for Man1 binding to *PsLBP*. (C) Experimentally determined STD build-up curves for Glc2 β and Glc2 α in disaccharide 1 binding to *PsLBP*. Circles show observed STD intensities, whilst curves are determined from least squares fitting to the equation in SI 1.

Discussion

GPs are attractive biocatalyst for oligo- and polysaccharides synthesis due to their broad specificity towards acceptor substrates and relatively low cost of donors compared to other enzymatic glycan syntheses. Understanding the mechanism of GP action on both natural and non-cognate substrates would therefore provide background knowledge that would underpin applications of GPs in carbohydrate synthesis, both in academic and industrial settings. Unlike other conventional substrate screening experiments which have been conducted by various groups on GPs, [13–15,26,30,35,49,62,63] we aimed to pinpoint the mechanism by which *PsLBP* recognised and utilised Man1P as its non-cognate donor, using X-ray crystallography and STD NMR spectroscopy.

The use of GPs for β -1,3-mannosylation has been previously conducted using a GH130 β -1,3-mannan mannoside phosphorylase (Zg0232) from *Zobellia galactanivorans* DSM 12802, which transfers mannose from Man1P to a variety of sugar acceptors, including a non-cognate acceptor, Glc.^[64] In contrast,

our study demonstrates the relaxed specificity of *PsLBP* towards the sugar 1-phosphate donor, Man1P, from which mannose was transferred onto a Glc acceptor, resulting in the production of Man- β -1,3-Glc disaccharide 1.

Significant interactions between Glc1P and *PsLBP*, as indicated by the crystal structure and STD NMR analysis, suggests that the specificity of this enzyme towards the donor substrate is more restricted compared to the acceptor site. Therefore, any manipulation to broaden the donor specificity may be challenging. Nevertheless, our work demonstrated a relaxed specificity of *PsLBP* towards Man1P, which indicates that alternative configurations of the hydroxyl group at C2 on the pyranose ring are tolerated. A strengthening of the interaction of Man1P with the active site could be the strategy to make the production of disaccharide 1 more efficient. However, the enzyme crystal structure showed that the axial configuration of the C2-OH of Man1P is pointing into empty space (at the dimer interface), suggesting that a simple mutation approach is unlikely to restore a hydrogen bond with C2-OH on Man1P. On the other hand, GlcN1P was not a substrate for *PsLBP* despite having the same configuration as Glc1P at C2. In this case, the C2-OH group is substituted by an NH_2 , which likely causes steric and/or electrostatic clashes with R353, thus disfavouring the binding of GlcN1P to the active site. The size restriction of the substituted group at C2 position has been reported in *CgCBP*, which cannot accommodate GlcNAc1P (OH is replaced by CH_3CONH at C2), whereas *VpChBP* can accommodate both GlcNAc1P and Glc1P, despite having the same amino acid for interaction with the hydroxyl group at C2.^[17] The difference is only in the placement of the Arg side chain that interacts with the C2 group, which is more distant in *VpChBP* to accommodate a larger substitution.^[27]

Crystallographic structures and STD NMR data generated in this study enabled us to explain the *PsLBP* preference for other sugar 1-phosphates (see Figure S13 for structures of all sugar 1-phosphate mentioned in discussion below). For instance, a hydrogen bond formed between C4-OH and the side chains of R374 and the strong saturation transfer signal from the protein to the hydrogen on C4 in the STD NMR experiment indicate that C4-OH and its configuration might be crucial for the recognition of Glc1P by *PsLBP* and therefore any modification at this position may compromise the enzyme activity on the donor. This hypothesis is supported by the fact that *PsLBP* could not use

FULL PAPER

Gal1P as a sugar donor (Figure 2A TLC enzyme screening), which has the C4-OH in an axial rather than equatorial position. The same explanation can be used to explain the lack of *PsLBP* activity towards GalN1P and GalU1P, both of which are derivatives of Gal1P. We can also predict whether *PsLBP* can work on donors that we have not screened in this work. For instance, glucuronic-acid 1-phosphate (GlcU1P) which has carboxyl group substitution at C6 would likely cause steric and electrostatic clashes with E732 and therefore may not permit binding of GlcU1P to *PsLBP*. The importance of C6 for binding to *PsLBP* was evident from the strong STD intensities at the germinal protons at this position in both Glc1P and Man1P.

Both *PsLBP* and CBPs work on disaccharides, but with different linkage specificity (i.e. β -1,3 vs β -1,4). However, each enzyme may employ a different mechanism to restrict the length of the substrate/product. In *CgCBP*, the extended catalytic loop is a unique characteristic to *CgCBP* that was not found in our *PsLBP* structure, nor in any other characterised disaccharide phosphorylases in GH94 family. It is likely to be involved in the substrate specificity of *CgCBP* with regards to the degree of polymerisation of the product.^[27] In contrast, *PsLBP* contains a unique β -hairpin 'gate', which when superposed with the CBP structure, overlaps with the position of the extended catalytic loop of the latter. Therefore, the β -hairpin gate in *PsLBP* may perform similar role to that of the *CgCBP* extended catalytic loop in restricting the degree of polymerization.

In summary, we have demonstrated the simple enzymatic synthesis of a non-natural disaccharide by utilising the promiscuity toward non-cognate sugar 1-phosphates of *PsLBP* and provide structural insight into the mechanisms whereby the enzyme distinguishes sugar 1-phosphate donors. Our work provides a stepping stone towards the design and engineering of GPs for tolerance towards other non-cognate sugar donors, which will help to expand the range of GP applications in carbohydrate synthesis.

Experimental section

Expression and purification of *PsLBP*: The *PsLBP* 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^[48] using In-Fusion™ (TakaraBio, Mountain View, CA, USA) following the manufacturer's protocol. The recombinant *PopinF-PsLBP* was transformed into *E. coli* (BL21 (DE3)) and a one l culture of the

transformant was grown at 37 °C in LB media with agitation (180 rpm) until $OD_{600} \sim 0.7$. Heterologous protein expression was induced by adding IPTG to a final concentration of 0.2 mM and incubated overnight at 18 °C. 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 DNase (1 mg/ml, Sigma). Supernatant containing the recombinant proteins was separated from cell debris by centrifugation (32,914 x g, 30 min), then purified with an ÄKTA pure FPLC system (GE Healthcare) at 4 °C. The supernatant containing His₆-tagged *PsLBP* 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 was eluted in one step with 10 mM HEPES pH 7.5, 250 mM NaCl, 500 mM imidazole. The sample was 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 protein was stored in 30 μ l aliquots at -80 °C until required. To estimate the assembly of *PsLBP* in solution, a calibration curve was generated using standard proteins from Gel Filtration Calibration Kit HMW (GE Healthcare). Approximately 2 mg/ml of the standard proteins were analysed individually using the same method described for *PsLBP* to determine the elution volumes. These values were then used to construct a calibration curve, following the manufacturer's protocol.

Enzymatic assays: Phosphate release assay^[35] was carried out in an assay buffer (20 μ l, 100 mM HEPES pH 7.0, 20 mM sugar 1-phosphate donors, 10 mM acceptors, 200 mM sodium molybdate). The reactions were incubated at 45 °C for 30 minutes. A color solution (90 μ l, 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 μ l, 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 *PsLBP* were determined using the phosphate release assay (20 μ l) with the enzymes (25 μ g/ml) in the presence of 0.2-10 mM of Glc and 10 mM Glc1P or Man1P. 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 .

Oligosaccharide analysis: TLC was performed by spotting the recovered reaction mixture (0.5 μ l) onto a silica plate (10 cm x 5 cm), then eluted using a mobile phase containing NH₄OH : H₂O : 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.

FULL PAPER

HPAEC-PAD analyses were performed by diluting the reaction mixtures in MilliQ water to a final volume of 500 μ l and desalted by mixed bed ion exchange resin (Sigma). The desalted mixtures were filtered through a disposable PTFE 0.45 μ m 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.

NMR spectroscopy. ¹H NMR were recorded at 298 K on a Bruker at 800 MHz and ¹³C NMR were recorded at 298 K on a Bruker Avance III 400 spectrometer at 100 MHz. Chemical shifts (δ) are reported in parts per million (ppm) with respect to residual HOD signal in D₂O (δ_{H} 4.79). Coupling constants (J) are reported in Hz. NMR signal assignments were made with the aid of COSY and HSQC experiments.

Mass spectrometry: disaccharide **1** was analysed by spotting 1 μ l of 2 mg/ml in water onto a silica gel plate. The spot was analysed by TLC/MS (Plate Express[™], Advion BioSciences, Ithaca, NY, USA), which subjects the compound to electrospray ionisation using spray voltage and sample delivery pressure of 3.5 kV and 3,000 psi for positive ion mode, with the flow rate of 0.3 ml/min. The sample was analysed at the capillary temperature of 250 $^{\circ}$ C, collision energy and the scan time of 1799 ms.

Optical rotation analysis: disaccharide **1** was dissolved in water to the final concentration of 2.2 mg/ml. The specific rotation of disaccharide **1** was recorded on a polarimeter model 341 polarimeter (PerkinElmer) at 20 $^{\circ}$ C, 589 nm.

Physical data for β -D-mannopyranosyl-(1 \rightarrow 3)-D-glucopyranose (disaccharide **1):** $[\alpha]_{\text{D}}^{20} +7^{\circ}$ (c 0.2, H₂O); ¹H NMR (800 MHz, D₂O) δ 5.172 (d, J =3.7, 1H, H-1 α -Glc), 4.839 (d, J =1.1, 1H, H-1 β -Man), 4.823 (d, J =1.0, 1H, H-1 β -Man), 4.611 (d, J =8.1, 1H, H-1 β -Glc), 4.081 – 4.049 (m, 2H, H-2 β -Man), 3.897 – 3.851 (m, 2H, H-6 β -Man), 3.865 – 3.808 (m, 2H, H-3 α -Glc, H-6 β -Glc), 3.801 (dddd, J =10.0, 5.0, 2.3, 0.6, 1H, H-5 α -Glc), 3.772 (dd, J =12.3, 2.3, 1H, H-6 α -Glc), 3.718 (dd, J =12.3, 5.0, 1H, H-6' α -Glc), 3.698 – 3.642 (m, 4H, H-6 β -Man, H-3 β -Glc, H-6' β -Glc), 3.633 – 3.579 (m, 3H, H-3 Man, H-2 α -Glc), 3.527 (t, J =9.7, 1H, H-4 β -Man), 3.522 (t, J =9.7, 1H, H-4 β -Man), 3.491 – 3.454 (m, 2H, H-4 α -Glc, H-4 β -Glc), 3.429 (ddd, J =10.0, 5.5, 2.2, 1H, H-5 β -Glc), 3.374 – 3.334 (m, 2H, H-5 β -Man), 3.314 (dd, J =9.3, 8.1, 1H, H-2 β -Glc); ¹³C NMR (101 MHz, D₂O) δ 100.56 and 100.51 C-1 Man), 95.7 (C-1 β -Glc), 92.1 (C-1 α -Glc), 84.6 (C-3 β -Glc), 82.1 (C-3 α -Glc), 76.35 and 76.31 (C-5 Man), 75.4 (C-5 β -Glc), 73.7 (C-2 β -Glc), 72.9 and 72.8 (C-3 Man), 71.1 (C-5 α -Glc), 70.9 (C-2 α -Glc), 70.5 (C-2 Man), 68.2 and 68.1 (C-4 α -Glc and C-4 β -Glc), 66.7 (C-4 Man), 61.0 (C-6 Man), 60.7, 60.53 (C-6 Glc); HRMS (ESI): calculated for C₁₂H₂₂NaO₁₁⁺ m/z 365.1054, found m/z 365.1062.

Crystallographic methods: Crystallization trials were set up for purified PsLBP (~10 mg/ml in 20 mM HEPES pH 7.0, 150 mM NaCl) using a range of commercial crystallisation screens (Molecular Dimensions) in MRC2 96-well sitting-drop vapour diffusion crystallization plates (Swissci) with a mixture of 0.3 μ l well solution and 0.3 μ l protein solution using an OryxNano robot (Douglas Instruments). After optimisation, the best crystals were obtained from drops containing 0.4 μ l of protein and 0.2 μ l of a crystallisation solution comprised of 18% (w/v) polyethylene glycol 3350, 0.1 M Tris-citrate buffer pH 6.0, 0.3 M ammonium sulphate. Crystals were cryo-protected with well solution containing 20% (v/v) ethylene glycol and flash-cooled in liquid nitrogen. For phasing, crystals were soaked for 30 minutes in a saturated solution of mercury(II) chloride made up in the cryoprotectant solution; the ligand bound complexes were obtained by soaking crystals for 5 minutes in cryoprotectant containing 20 mM of the compound (Glc1P or Man1P).

The pre-cooled crystals were transferred robotically to the goniostat on either beamline I03 or I04 at Diamond Light Source (Oxfordshire, UK) and maintained at -173 $^{\circ}$ C with a Cryojet cryocooler (Oxford Instruments). X-ray diffraction data were recorded using a Pilatus 6M hybrid photon counting detector (Dectris), then integrated and scaled using XDS^[65] via the XIA2 expert system^[66] and merged using AIMLESS^[67] All crystals belonged to space group P4₁2₁2 with approximate cell parameters of a = b = 147 Å , c = 222 Å (see Table S4 for a summary of data collection statistics).

Analysis of the likely composition of the asymmetric unit (ASU) suggested that it would contain two copies of the 102 kDa protein chain, giving an estimated solvent content of 58%. The structure was solved at 2.9 Å resolution by SAD phasing using the CRANK2 pipeline^[68] by combining data collected from two mercury soaked crystals at the L_{III} X-ray absorption edge of mercury (wavelength = 1.0052 Å). SHELXD^[69] located eleven sites in the ASU with occupancies >0.25 and BUCCANEER^[70] went on to build a model in which 59% of the sequence was fitted with R_{work} and R_{free} values of 0.342 and 0.400, respectively. This was then edited in COOT^[71] before refining in REFMAC5^[72] against native data processed to 1.95 Å resolution. Phases calculated from this model were used as input to second BUCCANEER job, which produced a model with 97% of the sequence fitted and R_{work} and R_{free} values of 0.282 and 0.323, respectively. The model was finalised by further iterations of manual rebuilding in COOT and restrained refinement in REFMAC5 using isotropic thermal parameters and TLS group definitions obtained from the TLSMD server (<http://skuld.bmsc.washington.edu/~tlsmd/>).^[73] In each of the expected active sites, residual density consistent with an oxanion was present. This was interpreted as sulphate derived from the precipitant solution. This sulphate-bound structure was used as the starting model for the Glc1P and Man1P complexes, which were built and refined as above.

The geometries of the final models were validated with MOLPROBITY^[74] before submission to the Protein Data Bank (see Table S4 for a summary of model statistics). Omit $mF_{\text{obs}}-dF_{\text{calc}}$ difference electron density maps were generated for the bound ligands using phases from the final model without the ligands after the application of small random shifts to the atomic

FULL PAPER

coordinates, re-setting temperature factors, and re-refining to convergence (Figure S14). All structural figures were prepared using CCP4MG.^[75]

STD-NMR: All samples were prepared in D₂O with 25mM Tris-d₁₁ pH 7.4 and contained final protein and ligand concentrations of 50 μM and 6 mM respectively. All experiments were performed at 278 K on a Bruker Avance III 800 MHz spectrometer equipped with a 5-mm TXI 800 MHz H-C/N-D-05 Z BTO probe. STD NMR experiments were performed using a train of 50 ms Gaussian pulses applied on the f2 channel at either 0.8 ppm (on-resonance) or 40 ppm (off resonance). A spoil sequence was used to destroy unwanted magnetisation and a spinlock was used to suppress protein signals (stdtiff.3). The recycle delay (d1) was set to 5 s. The total saturation time and number of scans were selected according to the following scheme:

Total Saturation Time (s)	0.5	0.75	1	1.5	2	3	5
No. Scans	512	512	256	256	128	128	128

Preparation of models: Crystal structures were imported into Schrödinger Maestro^[76] and prepared with the Protein Preparation Wizard. All non-protein or non-ligand atoms were removed. Protons were then added to the model, using PROPKA to predict the protonation state of polar sidechains at pH 7.^[77] The hydrogen-bonding network was automatically optimised by allowing asparagine, glutamine and histidine sidechains to be flipped. The model was then minimised using the OPLS3^[78] force field and a heavy atom convergence threshold of 0.3 Å. Since STD NMR experiments were performed in D₂O, polar protons were removed from the ligand prior to CORCEMA-ST.^[60]

CORCEMA-ST Calculations: Protein chemical shifts were calculated using the SHIFTX2^[79] webserver according to experimental conditions. All protein protons within 15 Å of the ligand were considered in the calculation. The instrument field strength, solvent type, ligand concentration, and protein concentration were set according to experimental values. The free and bound ligand correlation times were estimated to be 0.3 ns and 300 ns respectively, based on reasonable values for a monosaccharide binding to a 200 kDa protein. The non-specific leakage was also optimised to 0.8 s⁻¹. The internal correlation time was set to 10 ps and the methyl-X order parameter was set to 0.85, according to previously published values.^[60] All protein protons with resonances between 0.6 – 1 ppm were considered to be instantaneously saturated to account for line broadening. For glucose 1-phosphate, the equilibrium constant and k_{on} were optimised to 25000 M⁻¹ and 10⁵ M⁻¹ s⁻¹ respectively. For Man1P, the equilibrium constant was reduced to 16000 M⁻¹. Both values are in agreement with the affinities typically observed for carbohydrate-binding proteins.

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/M0290341/1]; the John Innes Foundation and the

Royal Thai Government Scholarship program. J. A. and S. W. acknowledge funding from the BSRC through a research grant (BB/P010660/1) and a DTP PhD studentship, respectively. We acknowledge the Diamond Light Source for access to beamlines I03 and I04 under proposal MX13467 with support from the European Community's Seventh Framework Program (FP7/2007–2013) under Grant Agreement 283570 (BioStruct-X).

Keywords: laminaribiose phosphorylase • enzymatic synthesis • structural biology • glycosyl hydrolase family 94

- [1] V. Puchart, *Biotechnol. Adv.* **2015**, *33*, 261–276.
- [2] E. C. O'Neill, R. A. Field, *Carbohydr. Res.* **2015**, *403*, 23–37.
- [3] G. Pergolizzi, S. Kuhadomlarp, R. A. Field, E. Kalita, *Protein Pept. Lett.* **2017**, *24*, 696–709.
- [4] H. Nakai, M. Kitaoka, B. Svensson, K. Ohtsubo, *Curr. Opin. Chem. Biol.* **2013**, *17*, 301–309.
- [5] M. Hiraishi, K. Igarashi, S. Kimura, M. Wada, M. Kitaoka, M. Samejima, *Carbohydr. Res.* **2009**, *344*, 2468–2473.
- [6] Y. Yataka, T. Sawada, T. Serizawa, *Langmuir* **2016**, *32*, 10120–10125.
- [7] T. Nohara, T. Sawada, H. Tanaka, T. Serizawa, *J. Biomater. Sci. Polym. Ed.* **2017**, *28*, 925–938.
- [8] T. Nohara, T. Sawada, H. Tanaka, T. Serizawa, *Langmuir* **2016**, *32*, 12520–12526.
- [9] C. Goedel, T. Sawangwan, M. Mueller, A. Schwarz, B. Nidetzky, *Angew. Chemie - Int. Ed.* **2008**, *47*, 10086–10089.
- [10] M. Nishimoto, M. Kitaoka, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 2101–2104.
- [11] K. Beerens, K. De Winter, D. Van De Walle, C. Grootaert, S. Kamiloglu, L. Miclotte, T. Van De Wiele, J. Van Camp, K. Dewettinck, T. Desmet, *J. Agric. Food Chem.* **2017**, *65*, 6030–6041.
- [12] M. Nakajima, H. Toyozumi, K. Abe, H. Nakai, H. Taguchi, M. Kitaoka, *PLoS One* **2014**, *9*, e92353.
- [13] M. Kitaoka, Y. Matsuoka, K. Mori, M. Nishimoto, K. Hayashi, *Biosci. Biotechnol. Biochem.* **2012**, *76*, 343–348.
- [14] M. Kitaoka, T. Sasaki, H. Taniguchi, *J. Biochem.* **1992**, *112*, 40–44.
- [15] Y.-K. Kim, M. Kitaoka, M. Krishnareddy, Y. Mori, K. Hayashi, *J. Biochem.* **2002**, *132*, 197–203.
- [16] M. Reichenbecher, F. Lottspeich, K. Bronnenmeier, *Eur. J. Biochem.* **1997**, *247*, 262–267.
- [17] Y. Honda, M. Kitaoka, K. Hayashi, *Biochem. J.* **2004**, *377*, 225–232.
- [18] T. Nihira, Y. Saito, M. Nishimoto, M. Kitaoka, K. Igarashi, K. Ohtsubo, H. Nakai, *FEBS Lett.* **2013**, *587*, 3556–3561.
- [19] K. De Winter, G. Dewitte, M. E. Dirks-Hofmeister, S. De Laet, H. Pelantová, V. Křen, T. Desmet, *J. Agric. Food Chem.* **2015**, *63*, 10131–10139.
- [20] K. De Winter, L. Van Renterghem, K. Wuyts, H. Pelantova, V. Kren, W. Soetaert, T. Desmet, *Adv. Synth. Catal.* **2015**, *357*, 1961–1969.
- [21] K. Hamura, W. Saburi, S. Abe, N. Morimoto, H. Taguchi, H. Mori, H. Matsui, *Biosci. Biotechnol. Biochem.* **2012**, *76*, 812–818.
- [22] K. Chomvong, V. Kordić, X. Li, S. Bauer, A. E. Gillespie, S.-J. Ha, E.

FULL PAPER

- Oh, J. M. Galazka, Y.-S. Jin, J. H. D. Cate, *Biotechnol. Biofuels* **2014**, *7*, DOI 10.1186/1754-6834-7-85.
- [23] K. Kino, R. Satake, T. Morimatsu, S. Kuratsu, Y. Shimizu, M. Sato, K. Kirimura, *Biosci. Biotechnol. Biochem.* **2008**, *72*, 2415–2417.
- [24] J. K. Park, N. O. Keyhani, S. Roseman, *J. Biol. Chem.* **2000**, *275*, 33077–33083.
- [25] H. G. Tran, T. Desmet, K. Saerens, H. Waegeman, S. Vandekerckhove, M. D'hooghe, I. Van Bogaert, W. Soetaert, *Bioresour. Technol.* **2012**, *115*, 84–87.
- [26] H. Nakai, M. A. Hachem, B. O. Petersen, Y. Westphal, K. Mannerstedt, M. J. Baumann, A. Dilokpimol, H. a. Schols, J. Ø. Duus, B. Svensson, *Biochimie* **2010**, *92*, 1818–1826.
- [27] M. Hidaka, M. Kitaoka, K. Hayashi, T. Wakagi, H. Shoun, S. Fushinobu, *Biochem. J.* **2006**, *398*, 37–43.
- [28] C. M. Bianchetti, N. L. Elsen, B. G. Fox, G. N. Phillips, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2011**, *67*, 1345–1349.
- [29] M. Hidaka, Y. Honda, M. Kitaoka, S. Nirasawa, K. Hayashi, T. Wakagi, H. Shoun, S. Fushinobu, *Structure* **2004**, *12*, 937–947.
- [30] E. C. O'Neill, G. Pergolizzi, C. E. M. Stevenson, D. M. Lawson, S. A. Nepogodiev, R. A. Field, *Carbohydr. Res.* **2017**, *451*, 118–132.
- [31] M. Nakajima, N. Tanaka, N. Furukawa, T. Nihira, Y. Kodutsumi, Y. Takahashi, N. Sugimoto, A. Miyanaga, S. Fushinobu, H. Taguchi, H. Nakai, *Sci. Rep.* **2017**, *7*, DOI 10.1038/srep42671.
- [32] Y. W. Nam, T. Nihira, T. Arakawa, Y. Saito, M. Kitaoka, H. Nakai, S. Fushinobu, *J. Biol. Chem.* **2015**, *290*, 18281–18292.
- [33] S. Fushinobu, M. Hidaka, A. M. Hayashi, T. Wakagi, H. Shoun, M. Kitaoka, *J. Appl. Glycosci.* **2011**, *58*, 91–97.
- [34] M. R. M. De Groeve, M. De Baere, L. Hoflack, T. Desmet, E. J. Vandamme, W. Soetaert, *Protein Eng. Des. Sel.* **2009**, *22*, 393–399.
- [35] M. R. M. De Groeve, G. H. Tran, A. Van Hoorebeke, J. Stout, T. Desmet, S. N. Savvides, W. Soetaert, *Anal. Biochem.* **2010**, *401*, 162–167.
- [36] M. R. M. De Groeve, T. Desmet, W. Soetaert, *J. Biotechnol.* **2011**, *156*, 253–260.
- [37] M. Mayer, B. Meyer, *Angew. Chemie - Int. Ed.* **1999**, *38*, 1784–1788.
- [38] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117.
- [39] C. D. Owen, L. E. Tailford, S. Monaco, T. Šuligoj, L. Vaux, R. Lallement, Z. Khedri, H. Yu, K. Lecointe, J. Walshaw, S. Tribolo, M. Horrex, A. Bell, X. Chen, G. L. Taylor, A. Varki, J. Angulo, N. Juge, *Nat. Commun.* **2017**, *8*, DOI 10.1038/s41467-017-02109-8.
- [40] J. C. Muñoz-García, M. J. García-Jiménez, P. Carrero, A. Canales, J. Jiménez-Barbero, M. Martín-Lomas, A. Imberty, J. L. de Paz, J. Angulo, H. Lortat-Jacob, P. M. Nieto, *Glycobiology* **2014**, *24*, 1004–1009.
- [41] Y. Yuan, X. Wen, D. A. R. Sanders, B. M. Pinto, *Biochemistry* **2005**, *44*, 14080–14089.
- [42] Y. Yuan, D. W. Bleile, X. Wen, D. A. R. Sanders, K. Itoh, H. W. Liu, B. M. Pinto, *J. Am. Chem. Soc.* **2008**, *130*, 3157–3168.
- [43] M. G. Szczepina, R. B. Zheng, G. C. Completo, T. L. Lowary, B. M. Pinto, *ChemBioChem* **2009**, *10*, 2052–2059.
- [44] J. Angulo, B. Langpap, A. Blume, T. Biet, B. Meyer, N. Rama Krishna, H. Peters, M. M. Palcic, T. Peters, *J. Am. Chem. Soc.* **2006**, *128*, 13529–13538.
- [45] S. M. L. Gagnon, M. S. G. Legg, N. Sindhuwinata, J. A. Letts, A. R. Johal, B. Schuman, S. N. Borisova, M. M. Palcic, T. Peters, S. V. Evans, *Glycobiology* **2017**, *27*, 966–977.
- [46] E. C. O'Neill, A. M. Rashid, C. E. M. Stevenson, A.-C. Hetru, A. P. Gunning, M. Rejzek, S. A. Nepogodiev, S. Bornemann, D. M. Lawson, R. A. Field, *Chem. Sci.* **2014**, *5*, 341–350.
- [47] S. Kuhaudomlarp, N. J. Patron, B. Henrissat, M. Rejzek, G. Saalbach, R. A. Field, *J. Biol. Chem.* **2018**, *293*, 2865–2876.
- [48] N. S. Berrow, D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D. I. Stuart, R. J. Owens, *Nucleic Acids Res.* **2007**, *35*, DOI 10.1093/nar/gkm047.
- [49] F. N. Awad, P. Laborda, M. Wang, A. M. Lu, Q. Li, Z. P. Cai, L. Liu, J. Voglmeir, *Biochim. Biophys. Acta - Gen. Subj.* **2017**, *1861*, 3231–3237.
- [50] P.-E. Jansson, J. Lindberg, G. Widmalm, *Acta Chem. Scand.* **1993**, *47*, 711–715.
- [51] R. Stenutz in *Bioinformatics for Glycobiology and Glycomics: An Introduction*, (Eds.: C.-W. von der Lieth, T. Lütteke, M. Frank), Wiley-Blackwell, New Jersey, **2009**, pp. 311–320.
- [52] P. Colson, R. R. King, *Carbohydr. Res.* **1976**, *47*, 1–13.
- [53] M. U. Roslund, E. Sävén, J. Landström, J. Rönnols, K. H. M. Jonsson, M. Lundborg, M. V. Svensson, G. Widmalm, *Carbohydr. Res.* **2011**, *346*, 1311–1319.
- [54] M. U. Roslund, P. Tähtinen, M. Niemitz, R. Sjöholm, *Carbohydr. Res.* **2008**, *343*, 101–112.
- [55] E. Krissinel, *Nucleic Acids Res.* **2015**, *43*, DOI 10.1093/nar/gkv314.
- [56] A. Van Hoorebeke, J. Stout, J. Kyndt, M. De Groeve, I. Dix, T. Desmet, W. Soetaert, J. Van Beeumen, S. N. Savvides, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2010**, *66*, 346–351.
- [57] Y.-W. Nam, T. Nihira, T. Arakawa, Y. Saito, M. Kitaoka, H. Nakai, S. Fushinobu, *J. Biol. Chem.* **2015**, *290*, 18281–18292.
- [58] T. Sawano, W. Saburi, K. Hamura, H. Matsui, H. Mori, *FEBS J.* **2013**, *280*, 4463–4473.
- [59] J. Angulo, P. M. Enríquez-Navas, P. M. Nieto, *Chem. Eur. J.* **2010**, *16*, 7803–7812.
- [60] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2002**, *155*, 106–118.
- [61] N. R. Krishna, D. G. Agresti, J. D. Glickson, R. Walter, *Biophys. J.* **1978**, *24*, 791–814.
- [62] D. Aerts, T. F. Verhaeghe, B. I. Roman, C. V. Stevens, T. Desmet, W. Soetaert, *Carbohydr. Res.* **2011**, *346*, 1860–1867.
- [63] T. Nihira, Y. Saito, M. Kitaoka, M. Nishimoto, K. Otsubo, H. Nakai, *Carbohydr. Res.* **2012**, *361*, 49–54.
- [64] F. N. Awad, P. Laborda, M. Wang, A. M. Lu, Q. Li, Z. P. Cai, L. Liu, J. Voglmeir, *Biochim. Biophys. Acta - Gen. Subj.* **2017**, *1861*, 3231–3237.
- [65] W. Kabsch, *Acta Crystallogr. Sect. D* **2010**, *66*, 125–132.
- [66] G. Winter, *J. Appl. Crystallogr.* **2010**, *43*, 186–190.
- [67] P. R. Evans, G. N. Murshudov, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 1204–1214.
- [68] P. Skubák, N. S. Pannu, *Nat. Commun.* **2013**, *4*, 1–6.
- [69] G. M. Sheldrick, *Acta Crystallogr. Sect. A Found. Crystallogr.* **2007**,

FULL PAPER

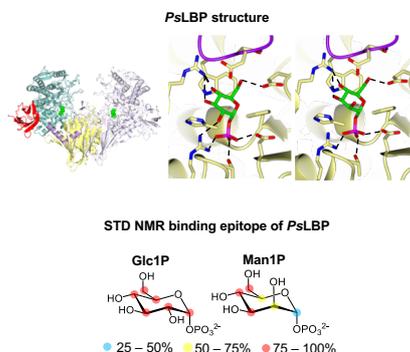
- 64, 112–122.
- [70] K. Cowtan, *Acta Crystallogr. Sect. D* **2006**, 62, 1002–1011.
- [71] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, 60, 2126–2132.
- [72] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. Sect. D* **1997**, 53, 240–255.
- [73] J. Painter, E. A. Merritt, *J. Appl. Crystallogr.* **2006**, 39, 109–111.
- [74] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, D. C. Richardson, *Nucleic Acids Res.* **2007**, 35, DOI 10.1093/nar/gkm216.
- [75] S. McNicholas, E. Potterton, K. S. Wilson, M. E. M. Noble, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, 67, 386–394.
- [76] Schrödinger Release 2016-4: Maestro, Schrödinger, LLC, New York, NY, **2016**.
- [77] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, 7, 525–537.
- [78] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilow, W. L. Jorgensen, R. Abel, R. A. Friesner, *J. Chem. Theory Comput.* **2016**, 12, 281–296.
- [79] B. Han, Y. Liu, S. W. Ginzinger, D. S. Wishart, *J. Biomol. NMR* **2011**, 50, 43–57.

FULL PAPER

Table of Contents

FULL PAPER

Glycoside phosphorylase (GP) as a tool for glycosylation: GPs are potential biocatalysts for oligo- and polysaccharide synthesis using an acceptor and a sugar 1-phosphate donor. Herein, we reported the synthesis of a disaccharide from glucose acceptor and mannose 1-phosphate as a non-cognate donor using a GP (*PsLBP*) as a catalyst and elucidated the binding of the non-cognate donor to the enzyme by crystallography and STD NMR.



*Sakonwan Kuhaudomlarp, Samuel Walpole, Clare E.M. Stevenson, Sergey A. Nepogodiev, David M. Lawson, Jesus Angulo, Robert A. Field**

Page No. – Page No.

Unravelling the specificity of laminaribiose phosphorylase from *Paenibacillus* sp. YM-1 towards donor substrates glucose/mannose 1-phosphate using X-ray crystallography and STD NMR spectroscopy