Characterization of a glycolipid glycosyltransferase with broad substrate specificity from the marine bacterium *Candidatus* Pelagibacter sp. HTCC7211

Tao Wei[a]*, Caimeng Zhao[a], Mussa Quareshy[b], Nan Wu[a], Shen Huang[a], Yuezhe Zhao[a], Pengfei Yang[a], Duobin Mao[a]，Yin Chen[b]*

a School of Food and Biological Engineering, Zhengzhou University of Light Industry, Zhengzhou, China

b School of Life Sciences, University of Warwick, Coventry, CV4 7AL United Kingdom

Running title: New glycolipid glycosyltransferase from SAR11

#Address correspondence to: Dr Wei Tao (School of Food and Biological Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China. Email, weit8008@zzuli.edu.cn or **Dr Yin Chen** (School of Life Sciences, University of Warwick, Coventry, CV4 7AL, United Kingdom. Email: Y.chen.25@warwick.ac.uk

Keywords Glycoglycerolipid, glycosyltransferase, GT_{cp}, MGlc-DAG, lipid remodelling

Submission to *Environmental Microbiology*

Abstract

In the marine environment, phosphorus availability significantly affects the lipid composition in many cosmopolitan marine heterotrophic bacteria, including members of the SAR11 clade and the Roseobacter clade. Under phosphorus stress conditions, non-phosphorus sugar-containing glycoglycerolipids are substitutes for phospholipids in these bacteria. Although these glycoglycerolipids play an important role as surrogates for phospholipids under phosphate deprivation, glycoglycerolipid synthases in marine microbes are poorly studied. In the present study, we biochemically characterized a glycolipid glycosyltransferase (GT_{cp}) from the marine bacterium *Candidatus* Pelagibacter sp. HTCC7211, a member of the SAR11 clade. Our results showed that GT_{cp} is able to act as a multifunctional enzyme by synthesizing different glycoglycerolipids with UDP-glucose, UDP-galactose, or UDP-glucuronic acid as sugar donors and diacylglycerol as the acceptor. Analyses of enzyme kinetic parameters demonstrated that Mg^{2+} notably changes the enzyme's affinity for UDP-glucose, which improves its catalytic efficiency. Homology modelling and mutational analyses revealed binding sites for the sugar donor and the diacylglycerol lipid acceptor, which provided insights into the retaining mechanism of GT_{cp} with its GT-B fold. A phylogenetic analysis showed that GT_{cp} and its homologs form a group in the GT4 glycosyltransferase family. These results not only provide new insights into the glycoglycerolipid synthesis mechanism in lipid remodelling, but also describe an efficient enzymatic tool for future synthesis of bioactive molecules.

Importance

The bilayer formed by membrane lipids serves as the containment unit for living microbial cells. In the marine environment, it has been firmly established that phytoplankton and heterotrophic bacteria can substitute phospholipids with non-phosphorus sugar-containing glycoglycerolipids in response to phosphorus limitation. However, little is known about how these glycoglycerolipids are synthesized. Here, we determined the biochemical characteristics of a glycolipid glycosyltransferase (GT_{cp}) from the marine bacterium *Candidatus* Pelagibacter sp. HTCC7211. GT_{cp} and its homologs form a group in the GT4 glycosyltransferase family, and can synthesize neutral glycolipids (MGlc-DAG and MGal-DAG) and an acidic glycolipid (MGlcA-DAG). We also uncover the key residues for DAG-binding through molecular docking, site-direct mutagenesis and subsequent enzyme activity assays. Our data provide new insights into the glycoglycerolipid synthesis mechanism in lipid remodelling.

Introduction

Phospholipids form the structural basis of all cells, but sugar-containing glycoglycerolipids are mainly restricted to marine microbes, cyanobacteria, and higher plants (1, 2). Glycoglycerolipids are found on the lipid bilayer of cell membranes and play critical roles in cell growth, cellular recognition, adhesion, neuronal repair, and signal transduction. These natural glycoglycerolipids often have unusual and sometimes unexpected biological activities, such as antitumor, antiviral, anti-inflammatory, antimalarial, immunostimulatory, and neuritogenic activities, which make them valuable molecular targets for research (3-5). The basic structure of glycoglycerolipids is characterized by a 1, 2-diacyl-*sn*-glycerol (DAG) moiety with different numbers and types of sugars (glucose, galactose, mannose, rhamnose, or charged sugars like glucuronic acid or sulfoquinovose) attached at the *sn*-3 position of the glycerol backbone in DAG. These sugar attachments have an α- or β-anomeric configuration, and are bound via (1→2), (1→3), (1→4), or (1→6) linkages (6, 7). The common glycoglycerolipid structures in marine heterotrophic microbes and cyanobacteria are 1,2-diacyl-3-O-(β-D-galactopyranosyl)-*sn*-glycerol (monogalactosyl diacylglycerol, MGal-DAG), 1,2-diacyl-3-O-(α-D-glucopyranosyl)-*sn*-glycerol (monoglucosyl DAG, MGlc-DAG),1,2-diacyl-3-O-(α-D-galactopyranosyl)-(1→6)-O-β-D-galactopyranosyl)-*sn*-glycerol (digalactosyl DAG, DGal-DAG), and 1,2-diacyl-3-O-(6-deoxy-6-sulfo-α-D-galactopyranosyl)-*sn*-glycerol (sulfoquinovosyl DAG, SQDG) (2,8).

Glycoglycerolipids are usually synthesized by glycosyltransferases (GTs), which are highly divergent and polyphyletic. The GTs can be categorized into 110 numbered families according to their sequence similarity and signature motifs, and the stereochemistry of the glycoside linkage formed (9). Of the 110 families, the families GT4, GT21, and GT28, which are known as glycoglycerolipid synthases, utilize sugar nucleotides as donors and contain a consensus sugar donor binding domain near the C-terminus (10, 11). Despite the wide variety of bacterial glycoglycerolipids, only a few bacterial lipid GTs have been identified and characterized so far. The GTs synthesizing MGlc-DAG and DGal-DAG have been isolated from the cell wall-less bacterium *Acholeplasma laidlawii*, and were found to belong to the GT4

family in the carbohydrate-active enzymes (CAZy) database (12). Other known members of bacterial GT4 include the MGlc-DAG synthases from *Deinococcus radiodurans* and *Thermotoga maritima* and the MGal-DAG synthase from *Borrelia burgdorferi*, which was the first cloned galactosyltransferase forming MGal-DAG with the α-anomeric configuration of the sugar (13, 14). A bifunctional GT (designated as Agt) from *Agrobacterium tumefaciens* was found to synthesize MGlc-DAG or MGlcA-DAG with UDP-glucose (UDP-Glc) or UDP-glucuronic acid (UDP-GlcA) as the sugar donor, respectively (7). This enzyme also belongs to the GT4 family and was the first glucuronosyl DAG synthase to be isolated. The processive GTs (Pgts), however, are members of the GT21 family, and show high sequence similarity to GT4 family GTs. The Pgts include enzymes from *Mesorhizobium loti* and *A. tumefaciens* that synthesize DGal-DAG, glucosylgalactosyl-DAG (GlcGal-DAG), and triglycosyl DAGs (15, 16). To the best of our knowledge, the structure function relationship of DAG-dependent GT4 glycosyltransferases has not been studied previously due to the lack of crystal structure of these glycoglycerolipids-producing enzymes and, as such, the binding pockets for UDP-sugars and DAG remain elusive.

Glycoglycerolipids play important roles in marine phytoplankton and heterotrophic bacteria under phosphate deprivation. Lipid remodelling reduces the cellular requirement for phosphorus, and the glycoglycerolipids MGlc-DAG/MGlcA-DAG and SQDG replace phospholipids in marine heterotrophic bacteria and marine phytoplankton and cyanobacteria, respectively (17-20). However, little is known about how these glycoglycerolipids are synthesized. Previous studies have shown that a manganese-dependent metallophosphoesterase, PlcP, is essential for lipid remodelling in marine heterotrophs, and that the *plcP* gene is organized in an operon-like structure and a putative glycosyltransferase was found down stream of *plcP* in numerous marine heterotrophic bacteria, such as members of the SAR11 clade (18, 21). Our previous work has shown that the GT from the marine bacterium SAR11 (*Candidatus* Pelagibacter sp. HTCC7211 and HTCC1062) is homologous to the Agt GT in *A. tumefaciens* (18). However, the activity of a SAR11 GT in the synthesis of glycoglycerolipids has not been characterized so far.

In this study, we report the detailed biochemical characterization of a glycoglycerolipid GT (GT_{cp}) from the marine bacterium *Candidatus* Pelagibacter sp. HTCC7211. Our results showed that GT_{cp} has a broad substrate specificity and can synthesize neutral glycolipids (MGlc-DAG and MGal-DAG) and an acidic glycolipid (MGlcA-DAG). GT_{cp} represents the first member of the GT4 family of lipid GTs from marine bacteria. In addition, homology modelling and site-directed mutagenesis analyses revealed details of its substrate recognition mechanism and identified key residues involved in the co-ordination of DAG in a GT4 family glycosyltransferase for lipid biosynthesis.

Results and discussion

GT_{cp} and its homologs form a new group in the GT4 family with a GT-B fold structure

The gene encoding a putative GT_{cp} (WP_008545403.1) consists of 1005 bp encoding a peptide of 334 amino acids, containing a GT4-like domain (Fig. 1A). This gene was previously hypothesized to be involved in lipid remodelling in *Candidatus* Pelagibacter sp. HTCC7211 for the synthesis of MGlc-DAG and MGlcA-DAG (17, 18). Sequence alignment (Fig. 1B) analyses showed that the GT_{cp} amino acid sequence has between 46-58%sequence identity with the putative GT from *Labrenzia aggregata*(GT_{la}, WP_040439323.1), *Thalassospira lucentensis* (GT_{tl}, WP_062950653.1), *Methylophaga nitratireducenticrescens* (GT_{mn}, WP_014706011.1), *Desulfobulbus mediterraneus* (GT_{dm}, WP_028584068.1), *Citromicrobium bathyomarinum* JL354 (GT_{cb}, WP_010239457.1), *Kordiimonas gwangyangensis* (GT_{kg}, WP_051078133.1), and the characterized Agt (locus tag, atu2297) from *A. tumefaciens* (7). In the phylogenetic analysis, the GT_{cp}, together with its close homologs (GT_{la}, GT_{tl}, GT_{mn}, GT_{dm}, GT_{cb}, GT_{kg} and Agt), formed a clade in the GT4 family (Fig. 1A). Sequences from this clade showed low sequence identity (< 25%) to other members of GT4 family, which includes more than 150,000 proteins with at least 22 different enzymatic activities at the time of writing. Purified Agt from *A. tumefaciens* has been found to synthesize MGlc-DAG or MGlcA-DAG with UDP-glucose or UDP-glucuronic acid as the sugar donor, respectively, and the expression of *agt* is known to

be induced under phosphate deficiency (18). Neither GT_{cp} nor any of its homologs from marine bacteria have been purified nor characterized to date.

Due to the lack of a three-dimensional structure of DAG-dependent GT4 glycosyltransferases to date, a model of GT_{cp} was generated by homology modelling using the X-ray structure of the GT MshA co-crystalized with UDP (Protein Data Bank [PDB] entry 3C4Q; 17% sequence identity) from *Corynebacterium glutamicum* as the template (Fig. 1C) (22). MshA, also a member of the GT4 family, catalyses the first step of the biosynthesis of mycothiol in actinobacteria using UDP-*N*-acetylglucosamine (UDP-GlcNAc) as the sugar donor (22). Using the PDBeFold server, we compared the crystal structure of MshA with the modelled structure of GT_{cp}. The predicted structure of GT_{cp} includes a GT-B fold consisting of two Rossmann-like *β-α-β* domains, with the N-terminal domain (residues 1–160 and residues 320–332) and C-terminal domain (residues 169–315), separated by a large cleft that includes the catalytic centre. The same GT-B fold is also found in several members of the GT4 family enzymes, including MshA and PimA (Fig. 1A) (22, 23).The substrate binding site of the sugar-donor is located mainly in the C-terminal domain, where the sugar-donor forms a number of hydrogen bonds with the protein. Given that the two enzymes (*i.e.* MshA and GT_{cp}) share only 17% overall sequence identity, the alignment was manually corrected by incorporating information such as predicted secondary structures and conserved functional residues. Multiple sequence alignment of GT_{cp} and its homologs revealed that GT_{cp} contains a catalytic dyad composed of His104–Asp256, two conserved UDP-sugar binding motifs (GRVAXEKN and FPSXTDTFG), and a conserved Gly-rich motif, all of which are commonly found in the GT4 family (Fig. 1B) (24).

Cloning of putative gene encoding GT_{cp} and functional verification

The recombinant plasmid pET22b–GT_{cp} was constructed to purify the enzyme for determination of its catalytic properties. Soluble expression of His-tagged GT_{cp} was achieved in *E. coli* BL21 (DE3) by adding 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The recombinant GT_{cp} was purified to homogeneity by Ni–NTA affinity followed by gel filtration

chromatography using the Superdex 200 column. In SDS-PAGE analyses, the purified recombinant His-tagged GT_{cp} was visible as a major band with a calculated mass of 38 kDa (Fig. 2A). The process used to purify GT_{cp} is summarized in Table 1. The enzyme was purified 3.5-fold with a yield of 17.1% and a specific activity of 47.1 U/mg. The predicted molecular mass of the monomeric form of GT_{cp} is 38 kDa, but GT_{cp} eluted as a peak corresponding to a molecular mass of approximately 70 kDa in the gel filtration chromatography experiments (Fig. 2B). These results suggested that GT_{cp} forms a dimer in solution, consistent with the proposed mechanism that oligomerization is a major factor contributing to the biochemical function and enzymatic activity of GTs (24).

We used two methods to determine the activity of purified GT_{cp}. Previous studies have shown that thin-layer chromatography (TLC) can resolve these selected glycoglycerolipid standards (15, 16). Indeed, as shown in Fig. 3A, the products of the enzymatic reaction with MGlc-DAG, MGal-DAG, and MGlcA-DAG were observed by staining with sulfuric acid/methanol/water, and corresponded to the standard markers. These results showed that GT_{cp} is able to transfer galactose, glucose, and hexuronic acid to the DAG acceptor using the respective UDP-sugars.

For structural identification, the glycoglycerolipids were analysed by liquid chromatography-mass spectrometry (LC-MS). LC-MS analyses (in the positive ion mode) detected an ammonium adduct (Fig. 3B-D) and fragmentation spectra for monohexuronosyl DAGs (MGlc-DAG and MGal-DAG) and MGlcA-DAG were obtained from the products of the GT_{cp}-catalysed reactions. The calculated m/z of the parental ion of MGlc-DAG, MGal-DAG, and MGlcA-DAG was 756.3, 756.3, and 770.4, respectively. The two species differed in the neutral loss corresponding to the polar head group (179.0 and 193.1 m/z for the loss of the hexosyl group and the hexuronic acid group, respectively). In each case, this loss yielded a DAG-16:0/18:1 (m/z 577.3 or 577.4). A further two peaks corresponded to monoacylglycerol with glyceryl-16:0 (m/z 313.3 or 313.5) and -18:1 (m/z 339.3 or 339.4) fatty acids, respectively. These results demonstrated that GT_{cp} shows high enzymatic activity towards the synthesis of MGlc-DAG, MGal-DAG, and MGlcA-DAG from DAG and UDP-sugars. Several bacterial lipid

GTs from *M. loti*, *A. tumefaciens*, *Mycoplasma pneumonia,* and *Mycoplasma genitalum* have been found to synthesize different glycoglycerolipids (DGal-DAG, GlcGal-DAG, and triglycosyl DAGs) using UDP-Glu and UDP-galactose (UDP-Gal) as sugar donors (15, 16, 25, 26). The *A. tumefaciens* GT Agt, which synthesizes neutral glycoglycerolipid (MGlc-DAG) and acidic glycoglycerolipid (MGlcA-DAG) with UDP-Glc or UDP-GlcA as the sugar donor, respectively, has been isolated and characterized (18). To the best of our knowledge, GT_{cp} is the first bacterial lipid GT acting as a multifunctional enzyme to synthesize MGlc-DAG, MGal-DAG, and MGlcA-DAG with different UDP-sugars as donors. It remains to be seen whether bi-functional/multifunctional GT4 enzymes involved in glycoglycerolipid synthesis are a common trait in this group. At least one member of this family, the GT4 homologue in *Pseudomonas* sp. appears to be specific for UDP-glucose and does not accept UDP-galactose nor UDP-glucuronic acid as the substrate (Supplementary Figure S1).

Catalytic properties of GT_{cp}

The catalytic activity of GT_{cp} was tested with UDP-Glc as the sugar donor substrate and DAG as the potential acceptor substrate. The effect of temperature on GT_{cp} activity was determined in the range of 10–50°C (Fig. 4A). GT_{cp} showed maximum activity at around 35°C and more than half of maximum activity at 20–45°C. Calculation of the activation energy E_a. using the Arrhenius plot in a semi-logarithmic form (ln $v \rightarrow 1/T$) was shown in Fig. 4B, and the slope of the plot was used to calculate the activation energy which was E_a = 25.1 kJ mol^{-1}. The thermostability of GT_{cp} was evaluated at three different temperatures (30°C, 40°C, and 50°C) with increasing incubation times up to 120 min. Most of the enzyme activity was maintained after incubation at 40°C for at least 120 min, whereas incubation at 50°C for 30 min reduced activity by approximately 50% (Fig. 4C). To investigate the effect of pH on the enzymatic activity of GT_{cp}, the enzymatic reaction was evaluated in different buffers (pH 7.0–11.0). The maximum activity of GT_{cp} was at pH 8.5, and it retained more than 50% of maximum activity between pH 7.5 and 9.0 (Fig. 4D). Considering the significance of NaCl for marine enzymes,

enzyme activity was determined in the presence of NaCl at different concentrations. The recombinant GT_{cp} maintained 54% of its maximum activity in the presence of 1.5 M NaCl and 30% of its maximum activity when the NaCl concentration was increased to 4 M (Fig. 4E).

To investigate the substrate specificity of GT_{cp} for DAGs, different species of varying chain length of DAGs were tested (Table 2). The k_m and k_{cat} values of GT_{cp} were calculated from Hanes-Wolff plots and the Michaelis-Menten equation. For the DAGs with saturated fatty acid chains (di8:0, di10:0, di12:0, di14:0, di16:0, and di18:0), the k_m, k_{cat}, and k_{cat}/k_m values increased with increasing acyl chain length. GT_{cp} showed the higher activities for the unsaturated DAGs than for the saturated DAGs. The most preferred DAG substrate for GT_{cp} was C16:0/C18:1 DAG, consistent with the fact that C16:0 and C18:1 fatty acids are common in marine bacteria (17, 18).

The Michaelis-Menten kinetic parameters for GT_{cp} were determined using UDP-Glc, UDP-Gal, and UDP-GlcA as the sugar donors (Table 3). The K_m value for UDP-Glc (82 μM) was higher than those for UDP-Gal and UDP-GlcA, consistent with the fact that UDP-Glc is the preferred substrate at physiological condition (27). The K_{cat}/K_m value for GT_{cp} toward different sugar donors followed the order UDP-Glc (71.4±2.7) >UDP-GlcA (59.6±3.6) >UDP-Gal (32.2±2.9). UDP-xylose, UDP-rhamnose, UDP-mannose and UDP-fructose were also tested with C16:0/C18:1 DAG, which showed no activity. Thus, GT_{cp} exhibited the highest enzymatic activity for UDP-Glc among the sugar donors tested. In contrast with GT_{cp}, the Pgts from *M. loti* and *A. tumefaciens* favour uridine UDP-Gal over UDP-Glc (15, 16). A comparison of GT_{cp} kinetics with other biochemically characterized GT4 glycosyltransferases is listed in Table 4 (22, 28-33).

Metal ions improve enzyme activity of GT_{cp}

The effect of various metal ions on the enzyme activity of GT_{cp} is shown in Fig. 4F. Among the tested metal ions (5 mM), Mg^{2+}, Ca^{2+}, and Mn^{2+} significantly stimulated GT_{cp} activity by up to 223%, 125%, and 138%, respectively, although the as-isolated enzyme is already active Furthermore, EDTA did not significantly affect the enzymatic activity of GT_{cp} after incubation

for 60 min at room temperature. The activity of GT_{cp} was decreased by Ba^{2+}, Cu^{2+}, and Fe^{2+} to 52%, 23%, and 35%, respectively. Moreover, Hg^{2+}, Cd^{2+}, Ni^{2+}, and Co^{2+} completely abolished enzyme activity. This may have resulted from the binding of these metal ions to the -SH, -CO, and -NH moieties of the amino acids of GT_{cp}, leading to structural changes and inactivation (34).

Given that Mg^{2+} markedly improved the activity of GT_{cp}, we determined metal content of the purified enzyme by inductively coupled plasma-mass spectrometry (ICP-MS) and found that Mg^{2+}: protein molar ratio was < 0.03 (supplementary table 1). Similarly, no substation amount of Ca, Mn or Zn was found in GT_{cp}, suggesting that this enzyme is unlikely a metalloprotein. The kinetic parameters were subsequently determined using purified GT_{cp} in the presence and absence of Mg^{2+} (Table 3). In the presence of excess DAG, the K_m value of GT_{cp} decreased from 82 μM to 58 μM with the addition of Mg^{2+}, indicating that GT_{cp} has a higher affinity for UDP-Glu in the presence of Mg^{2+}. The catalytic efficiency was nearly 2.0-fold higher than that in the absence of Mg^{2+}. In the presence of excess UDP-Glc, the K_m was almost unchanged with/without Mg^{2+}. Metal ions are important regulators of physiological functions and contribute to the preservation of the structural integrity of some proteins (35). In contrast to GT-A fold GTs, GT-B fold GTs, including GT_{cp}, are metal ion-independent (24). However, some studies have found that metal ions also change GT-B fold activity, such as GGT58A1 from *Absidia coerulea* (36), UGT59A1 from *Rhizopus japonicas,* Bs-PUGT from *Bacillus subtilis* P118 (37), and human POFUT2 (38). In some cases, metal ion simultaneously interacts with both the enzyme and the sugar donor in the active site, which causes the glucosyl donor to realign in the active site, and thus affects the activity of the enzyme (39). Interestingly, although human POFUT2 is not a metalloprotein, Mg^{2+}, Mn^{2+} and Ca^{2+} are also known to enhance the enzyme activity by facilitating the release of product from the enzyme (38). Therefore, it is tempting to speculate that the addition of Mg^{2+} may also help enhance GT_{cp} catalysis in a similar manner.

Insights into structure and glycosylation mechanism of GT_{cp}

279 GT_{cp}, along with the previously characterized Agt, form a group in the GT4 family and are
280 GT-B fold GTs (Fig. 1, Fig. 5A). The GT-B fold GTs are thought to employ the so-called
281 retaining glycosylation for adding the sugar unit to the substrate although some researchers
282 have proposed an alternative internal return (S_Ni-like) mechanism (24, 40). In the latter model,
283 the nucleophilic attack and the departure of the leaving group occur on the same face of the
284 sugar, and involve the formation of a short-lived oxocarbenium-like transition state with
285 asynchronous acceptor glycoside bond formation and phosphate bond breakdown (41-43).
286 According to this latter model, two conserved amino acid residues are important for catalysis
287 (*e.g.* Glu316 and His109 in MshA). In MshA, Glu316 and His109 act as catalytic nucleophiles
288 in the S_Ni-like mechanism, and form hydrogen bonds with the OH-3 and OH-6 of the glucosyl
289 moiety, respectively (22). Indeed, we also found the corresponding residues (Asp256, His104)
290 in GT_{cp} (Fig. 1B). Asp256 and His104 fulfil a critical role in GT_{cp} catalysis, because their
291 substitution to Ala completely abolished catalytic activity (Fig. 5B). Interestingly, mutation of
292 the conserved Asp residue at position 256 to Glu (D256E) reduced activity by more than 90%
293 even though both Asp and Glu residues have a carboxylic acid moiety. Asp256 is essential for
294 the enzymatic activity of GT_{cp} and cannot be replaced by Glu, indicating that the length of the
295 side chain of this residue is important for the activity of GT_{cp}. These results suggested that
296 Asp256 and His104 play important functions in glycosyl transfer while maintaining the α-
297 configuration of the anomeric carbon. To confirm the glycosylated position and the anomeric
298 stereoselectivity, MGlc-DAG was purified and analysed by ^{1}H-nuclear magnetic resonance
299 (NMR) (Fig. 5C). The glucosyl moiety of the glycolipid was suggested by the characteristic
300 signals for Glc-H1 (δH 5.17, 1H) and Glc-H2–6 (δH 3.20 to 3.70, 6H) by ^{1}H NMR. The location
301 of the glucosyl moiety was indicated by the correlation between Glc-H1 and C-3 (δH 136.0) in
302 the heteronuclear multiple bond correlation (HMBC) spectrum. The large coupling constant
303 (J=3.7Hz) of Glc-H1 revealed the α-D-configuration of the glucosidic linkage.

304 In terms of the donor binding pocket, GT4 family GTs have two highly conserved
305 sequence motifs in their C-terminal domain that are involved in the binding of the sugar
306 nucleotide with the UDP moiety (23, 29, 44). The sequence alignment analyses indicated that

GT_{cp} and its homologs have these two conserved motifs in the donor binding pocket (Fig. 1B), consistent with other GT4 proteins. The highly conserved sites were further investigated to illuminate the mechanism of sugar donor binding (Fig. 5A). Each of the potential active sites in the donor binding pocket (Gly16, Gly82, Gly85, Arg190, Lys195, Thr257, Phe258, Gly259, and Glu264) was mutated to Ala. Enzymatic activity of GT_{cp} in all the mutants was almost completely abolished, supporting our assumption from the homology model that these residues are integral to the sugar donor binding site (Fig. 5B). Thr257 in the structure of GT_{cp} corresponds to MshA residue Ser317, which interacts with the 4-OH of the sugar moiety and forms hydrogen bonds (22). When the residue at position 257 (Thr) in GT_{cp} was substituted with Ser (T257S), the T257S mutant exhibited approximately 1.6-times higher activity than that of wild-type GT_{cp}. Compared with the parental residue Thr, Ser has one fewer methyl group; therefore, there is more free space between the Ser residue and the substrate (45). The generated space might allow more room for the appropriate interaction between the Ser residue and the sugar donor. These findings indicate that the mechanism of the sugar donor recognition of GT_{cp} is similar to that of known GT-B fold GTs, such as MshA. That is, the conserved key residues would form hydrogen bonds with the UDP part and interact with the sugar moiety of the sugar donor.

In order to identify potential residues involved in binding of DAG, we docked C16:1/C18:0 DAG into the homology model of GT_{cp}. DAG is predicted to lay in the groove of the open structure in the model (Fig. 6A, B). Docked DAG interacts with several key resides primarily through hydrophobic interactions but also some polar interactions *e.g.* His104, Thr162 and Arg163 and Trp320 (Fig. 6C), of which His104, Arg163 and Trp320 are strictly conserved in GT_{cp} in a range of bacteria (Fig. 1B). Indeed, the His104, Arg163 and Trp320 mutants are inactive, supporting a key role in GT_{cp} catalysis (Fig. 5B). Subsequent independent docking of the three UDP-sugars predicted the identical binding pocket for all three (Fig.6D) in the presence of already docked DAG and with UDP-galactose as an example we observe the DAG and UDP-sugars positioned parallel to one another in the ‘open state’ (Fig. 6 E). To the

best of our knowledge, our study provides the first insight of the binding pocket of DAG in a GT4 glycosyltransferase involved in glycoglycerolipid biosynthesis.

In MshA, the hydrophobic residue Leu76 helps to stabilize the dimer (22). This residue corresponds to Leu56 in GT_{cp} (Fig 1B). To investigate the role of Leu56 in GT_{cp}, a site-directed mutant to Ala was made. The mutant L56A protein was loaded onto a Superdex 200 column to analyse its oligomeric state. The eluted peaks of L56A mutant corresponded to the monomer of GT_{cp} (38 kDa) according to their elution volumes (Fig. 2B). No difference was observed between the wild-type profile and the other mutants during protein purification (data not shown). The mutant L56A showed approximately 85% of wild-type GT_{cp} activity. Structural studies revealed that most of the residues involved in oligomerization are conserved with related GTs, and they appear to be primarily hydrophobic and aromatic residues that form an extensive hydrophobic interface between the monomers (46). These results demonstrated that residue Leu56 is essential for the stable dimerization of the protein, but does not play a direct role in the catalytic reaction of GT_{cp}.

To conclude，our data show that the activity of purified GT_{cp} from the marine bacterium *Candidatus* Pelagibacter sp. HTCC7211 is sufficient for the synthesis of several glycoglycerolipids, including MGlc-DAG, MGal-DAG and MGlcA-DAG. The ability to synthesize MGlc-DAG and MGlcA-DAG suggest that GT_{cp} may play an important role in lipid remodelling in natural marine systems. GT_{cp} and PlcP, a manganese-dependent metallophosphoesterase, are organized in an operon-like structure in numerous marine heterotrophic bacteria (17, 18). Upon phosphorus (P) deficiency, PlcP selectively degrades phospholipids such as phosphatidylglycerol (PG) and phosphatidylethanolamine (PE) to DAG, which then serves as the substrate for the biosynthesis of these glycolipids by GT_{cp} using UDP-Glc, UDP-Gal, or UDP-GlcA as sugar donors (Fig. 7).Both the phospholipid PG and the glycoglycerolipid MGlc-DAG are anionic under physiological conditions and it is likely that they could be interchangeable while maintaining the desirable biophysical properties of the membrane. Indeed, this has also been documented in the SAR11 strain HTCC7211 (17, 18).

Similarly, substitution of PG by the anionic sulfur-containing glycolipid SQDG has also been shown for marine cyanobacteria and phytoplankton (47). Together, our work thus points to the important role of glycosyltransferases as key enzymes in the synthesis of glycoglycerolipid in marine bacteria.

Materials and Methods

General materials and microorganisms

We purchased UDP-glucose (UDP-Glc), UDP-galactose (UDP-Gal), UDP-glucuronic acid (UDP-GlcA), UDP-xylose, UDP-rhamnose, UDP-mannose, UDP-fructose and diacylglycerols (DAGs) from Sigma-Aldrich Co. (St. Louis, MO, USA). Nickel column and Superdex 200 gel filtration columns were from GE Healthcare (Buckinghamshire, UK). All other chemicals were of the highest reagent grade and were obtained from Sangon (Shanghai, China). The *E. coli* strains JM109 for DNA manipulation and BL21-CodonPlus (DE3)-RIL for protein expression were obtained, respectively, from TaKaRa Bio, Inc. (Dalian, China) and Stratagene (La Jolla, CA, USA).

Cloning, expression, and purification of GT_{cp}

The gene GT_{cp} from *Candidatus* Pelagibacter sp. HTCC7211 was codon-optimized and chemically synthesized by Sangon (Shanghai, China). Several site-directed GT_{cp} mutants (encoding mutations G16A, L56A, G82A, G85A, H104A, T162A, R163A, R190A, K195A, D256A, D256E, T257A, T257S, F258A, G259A, E264A and W320A) were constructed using the overlapping PCR method with the common primers as for the wild-type GT_{cp} and two site-specific primers for each mutant (Table 5).The genes were then cloned into the pET22b expression vector using *Nco*I and *Sal*I restriction sites, and transformed into the host *E. coli* BL21-CodonPlus (DE3)-RIL for gene expression. The transformed cells were grown in LB medium containing 100 mg/l ampicillin and 34 mg/l chloramphenicol at 37°C with shaking at 180 rpm. When the cultures reached an OD 600 of 0.6, IPTG was added to a final

concentration of 0.5 mM. After a further 4 h of growth at 37°C, the cells were harvested by centrifugation and lyophilized by vacuum-freezing. The harvested cells were re-suspended in buffer A (50 mM Tris–HCl, pH 7.9, 50 mM NaCl), with 1% (w/v) of triton X-100, and then disrupted by sonication. The cell mixture was then centrifuged at 12,000×g for 30 min, and the soluble fraction was loaded onto a nickel column (GE Healthcare) pre-equilibrated with buffer A. The recombinant enzymes were eluted with elution buffer (20 mM Tris–HCl, pH 7.9, 500 mM NaCl, 300 mM imidazole) and dialyzed overnight in buffer A to remove imidazole. For further purification, the enzymes were loaded on a Superdex 200 (16/60) gel filtration column (GE Healthcare), which was pre-equilibrated with buffer B (50 mM Tris–HCl, pH 7.9, 200 mM NaCl). The fraction size was 0.5 ml and the flow rate was 0.5 ml/min. The peak fractions were collected, concentrated, and analysed by SDS-PAGE (12% polyacrylamide). The protein concentration was determined using the Bradford method. The purified protein was stored in buffer A containing 25% glycerol at −80°C. SDS-PAGE gels and circular dichroism (CD) spectroscopy analyses of the purified protein and the mutants are shown in supplementary Figure S2.

Bioinformatics and homology modelling

A putative GT gene encoding GT_{cp} was identified in the genome of *Candidatus* Pelagibacter sp. HTCC7211 (GenBank accession no. WP_008545403.1). ClustalW2 software (http://www.ebi.ac.uk/Tools/clustalw2/index.html) was used for multiple sequence alignment analysis of GT_{cp} (48). For the phylogenetic analysis, we used the neighbour-joining method and Molecular Evolutionary Genetic Analysis 7.1 software (MEGA, version 7.1) (49). The three-dimensional model structure of GT_{cp} was generated using tools at the Phyre 2 protein modelling server (50) and the crystal structure of MshA (Protein Data Bank [PDB] entry 3C4Q) from *C. glutamicum* as the template. Docking of UDP-Gla/UDP-Glc/UDP-GlcA with GT_{cp} was predicted using the Flexible Docking module in Accelrys Discovery studio (51, 52), and the protein model was imported into Flare (v3.0, Cresset) for docking the DAG substrate and firstly energy minimized with 2000 iterations with a cut off of 0.200 kcal/mol/A. The DAG lipid was

417 imported as a ligand and energy minimized in Flare before being docked into the active site
418 and the best scoring pose selected. Thereafter we utilised the DAG docked ligand with the
419 model as the basis for the follow up *in silico* docking of UDP-sugars in the presence of the
420 DAG.
421
422 **Enzyme activity assay**
423 The enzymatic activity of GT_{cp} was measured using 0.1 mM UDP-Glc (or UDP-Gal, or
424 UDP-GlcA) and 0.1 mM DAG as the substrate and 2.0 μM purified enzyme in 10 mM
425 Tricine/KOH, pH 8.5, 2 mM DTT. The resulting mixture (500 μl) was incubated at 35°C for 60
426 min with constant shaking at 200 rpm. The products (glycoglycerolipids and DAGs) were
427 extracted using the Floch method with methanol-chloroform-water at a ratio of 1:2:0.6
428 (vol/vol/vol). The lipid extract was dried under nitrogen as at room temperature. The dried
429 lipids were resuspended in acetonitrile and ammonium acetate (10 mM, pH 9.2) at a ratio of
430 95:5 (vol/vol) and analysed by LC-MS. One unit of enzymatic activity was defined as the
431 amount of enzyme required to catalyse the conversion of 1 μmol DAG per min under the
432 standard conditions. The measurements were corrected for background hydrolysis in the
433 absence of the enzyme. The K_m and V_{max} values were calculated using Hanes-Wolff plots with
434 various concentrations of substrate (0.02 to 1.0 mM) and three replicates.
435
436 **Lipid analysis by TLC and LC-MS**
437 The GT_{cp}-synthesized glycoglycerolipids were analysed by TLC using a Camag Automatic
438 TLC Sampler III (Camag, Muttenz, Switzerland) for spotting. Glycoglycerolipids were
439 separated on silica gel 60 (Merck, Darmstadt, Germany) with chloroform/methanol/water
440 (65:35:4 vol/vol), and stained with sulfuric acid/methanol/water (45:45:10 vol/vol) for
441 visualization. The resulting solutions were further analysed by high-performance liquid
442 chromatography using a 1290 Infinity II UPLC instrument (Agilent Corp., Santa Clara, CA,
443 USA) coupled with an AB SCIEX Triple Quad^{TM}5500 (AB SCIEX, Framingham, MA, USA)
444 equipped with an electrospray-ion (ESI) detector. A BEH Amide XP column (2.5-μm inner

diameter, 3 mm by 150 mm, Waters, Milford, MA, USA) was used for chromatographic separation. The mobile phase consisted of acetonitrile (solvent A) and 10 mM ammonium acetate, pH 9.2 (solvent B). The column was equilibrated for 10 min with 95% A: 5% B prior to sample injection. The separation was conducted using a stepwise gradient starting from 95% A: 5% B to 70% A: 30% B after 15 min with a constant flow rate of 150 µl min^{-1}. Mass spectrometric analysis was performed in the ESI positive ion mode with the ion spray voltage at 3500 V and temperature at 350°C. The nebulizer gas and heater gas were set at 40 psi. The analytical data were processed by Analyst software (version 1.6.3).

Inductively coupled plasma-mass spectrometry (ICP-MS)

The metal content of GT_{cp} was measured by using an ICP-MS (Agilent Technologies 7900 ICP-MS). The standards for calibration were freshly prepared by diluting Ca, Mg, Mn, Zn and S stock solution (at 1000 $mg \cdot L^{-1}$; Sigma-Aldrich, Saint Louis, MO, USA) with 1% (v/v) nitric acid with concentrations from 0.1 to 2.0 $mg \cdot L^{-1}$ for Ca, Mg, Mn, Zn, and from 1 to 25 $mg \cdot L^{-1}$ for S. About 3.0 mg protein was digested in 1% (v/v) nitric acid matrix for metal analyses. The content of S was quantified in order to determine the protein concentration. The contents of Ca, Mg, Mn, Zn and S were measured using the emission lines of 396.847 nm (Ca), 280.270 nm (Mg), 259.373 nm (Mn), 213.856 nm (Zn) and 180.669 nm (S) respectively.

Nuclear magnetic resonance spectroscopy (NMR) spectroscopy

NMR spectroscopy experiments were carried out in $CDCl_3$ with tetramethylsilane as an internal standard. ^{1}H, ^{13}C, DEPT, COSY, HSQC, HMBC and ROESY experiments were recorded at 298 K with 600 MHz spectrometer (Bruker Avance 600, Bruker). Bruker standard software Topspin 3.2 was applied to acquire and process all the spectra data. COSY and ROESY experiments were recorded using data sets (t1 by t2) of 2048 by 256 points, COSY with 4 and ROESY with 16 scans.

Characterization of recombinant GT_{cp}

The optimum temperature of GT_{cp} was measured in Tricine/KOH buffer (pH 8.5) in the range of 10–50 °C. The buffer was adjusted to pH 8.5 for each of the assayed temperatures. The activation energy of the cleavage reaction was calculated using the logarithmic form of the Arrhenius equation: $\ln K_{cat} = \ln K_0 - E_a/R \cdot T$. The effect of pH on enzymatic activity was tested at 35°C for pH values in the range of 7.0–11.0. The following buffers were used: sodium phosphate (pH 6.0–7.5), Tricine/KOH (pH 7.5–9.5), and N-cyclohexyl-3-aminopropanesulfonic acid (CAPS, pH 9.5–11.0). The thermostability of the purified GT_{cp} was examined by incubating the enzyme in 50 mM Tricine/KOH buffer (pH 8.5) at three different temperatures (30, 40, and 50 °C). Samples (80 µL) of the enzyme were collected after incubation periods of 30, 60, 90, and 120 min at each temperature and the residual activity of each sample was assayed under standard conditions. The enzymatic activity of GT_{cp} was also measured in the presence of various metal salts ($MnCl_2$, $ZnCl_2$, $MgCl_2$, $CaCl_2$, $BaCl_2$, $CdCl_2$, $HgCl_2$, $CuCl_2$, $FeSO_4$, $NiSO_4$, and $CoCl_2$) at 5 mM or 5 mM EDTA. To determine the salt stability of the enzyme, 0–4 M NaCl (final concentration) was added to the reaction mixture and enzyme activity was determined using optimum conditions.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 726116), Royal Society International Exchanges 2017 Cost Share (China) (IEC\NSFC\170213; grant agreement no. 170213) , Program for Science &Technology Innovation Talents in the Universities of Henan Province (18HASTIT040) and Projects of science and technology activities for overseas students from department of human resources and social security of Henan Province (2019-3)

References

1. Cheng-Sánchez I, Sarabia F. 2018. Chemistry and biology of bioactive glycolipids of marine origin. Mar drugs 16:294.

2. Hölzl G, Dörmann, P. 2007. Structure and function of glycoglycerolipids in plants and bacteria. Prog. Lipid Res 46:225–243.

3. Kolter, TA. 2011. View on sphingolipids and disease. Chem Phys Lipids 164:590-606.

4. Chirasuwan N, Chaiklahan R, Kittakoop P, Chanasattru W, Ruengjitchatchawalya M, Tanticharoen M, Bunnag B. 2009. Anti HSV-1 activity of sulphoquinovosyl diacylglycerol isolated from *Spirulina platensis*. Sci Asia 35:137–141.

5. Bergé JP, Debiton E, Dumay J, Durand P, Barthomeuf C. 2002. In vitro anti-inflammatory and anti-proliferative activity of sulfolipids from the red alga *Porphyridium cruentum*. J Agric Food Chem 50: 6227–6232.

6. Hölzl G, Dörmann P. 2007. Structure and function of glycoglycerolipids in plants and bacteria. Prog Lipid Res 46:225–243.

7. Semeniuk A, Sohlenkamp C, Duda K, Hölzl G. 2014. A bifunctional glycosyltransferase from *Agrobacterium tumefaciens* synthesizes monoglucosyl and glucuronosyl diacylglycerol under phosphate deprivation. J Biol Chem 289:10104–10114.

8. Zhang J, Li CX, Yu GL, Guan HS. 2014. Total synthesis and structure-activity relationship of glycoglycerolipids from marine organisms. Mar drugs 12:3634-3659.

9. Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem 77:521-55.

10. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490−D495.

11. Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol 328:307−317.

12. Berg S, Edman M, Li L, Wikström M, Wieslander, A. 2001.Sequence properties of the 1,2-diacylglycerol 3-glucosyltransferase from *Acholeplasma laidlawii* membranes. Recognition of a large group of lipid glycosyltransferases in eubacteria and archaea. J Biol Chem 276:22056–22063.

13. Hölzl G, Zähringer U, Warnecke D, and Heinz E. 2005. Glycoengineering of cyanobacterial thylakoid membranes for future studies on the role of glycolipids in photosynthesis. Plant Cell Physiol 46:1766–1778.

14. Ostberg Y, Berg S, Comstedt P, Wieslander A, Bergström S. 2007. Functional analysis of a lipid galactosyltransferase synthesizing the major envelope lipid in the Lyme disease spirochete *Borrelia burgdorferi*. FEMS Microbiol Lett 272:22–29.

15. Hölzl G, Leipelt M, Ott C, Zähringer U, Lindner B, Warnecke D, Heinz E. 2005. Processive lipid galactosyl/glucosyltransferases from *Agrobacterium tumefaciens* and *Mesorhizobium loti* display multiple specificities. Glycobiology 15:874–886.

16. Devers EA, Wewer V, Dombrink I, Dörmann P, Hölzl G. 2011. A processive glycosyltransferase involved in glycolipid synthesis during phosphate deprivation in *Mesorhizobium loti*. J Bacteriol 193:1377–1384.

17. Carini P, Van Mooy BA, Thrash JC, White A, Zhao Y, Campbell EO, Fredricks HF, Giovannoni SJ. 2015. SAR11 lipid renovation in response to phosphate starvation. Proc Natl Acad Sci U S A 112:7767–7772.

18. Sebastián M, Smith AF, González JM, Fredricks HF, Van Mooy B, Koblížek M, Brandsma J, Koster G, Mestre M, Mostajir B, Pitta P, Postle AD, Sánchez P, Gasol JM, Scanlan DJ, Chen Y. 2016. Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. ISME J 10:968–978.

19. Van Mooy BA, Fredricks HF, Pedler BE, Dyhrman ST, Karl DM, Koblízek M, Lomas MW, Mincer TJ, Moore LR, Moutin T, Rappé MS, Webb EA. 2009. Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. Nature 458:69–72.

20. Smith AF, Rihtman B, Stirrup R, Silvano E, Mausz MA, Scanlan DJ, Chen Y. 2019. Elucidation of glutamine lipid biosynthesis in marine bacteria reveals its importance under phosphorus deplete growth in *Rhodobacteraceae*. ISME J 13:39–49.

21. Wei T, Quareshy M, Zhang YZ, Scanlan DJ, Chen, Y. 2018. Manganese is essential for PlcP metallophosphoesterase activity involved in lipid remodelling in abundant marine heterotrophic bacteria. Appl Environ Microbiol 84:01109-18.

22. Vetting MW, Frantom PA, and Blanchard JS. 2008. Structural and enzymatic analysis of MshA from *Corynebacterium glutamicum*: substrate-assisted catalysis. J Biol Chem 283:15834–15844.

23. Guerin ME, Kordulakova J, Schaeffer F, Svetlikova Z, Buschiazzo A, Giganti D, Gicquel B, Mikusova K, Jackson M, Alzri PM. 2007. Molecular recognition and interfacial catalysis by the essential phosphatidylinositol mannosyltransferase PimA from *mycobacteria*. J Biol Chem 282:20705-20714.

24. Liang DM, Liu JH, Wu H, Wang BB, Zhu HJ, Qiao JJ. 2015. Glycosyltransferases: mechanisms and applications in natural product development. Chem Soc Rev 44: 8350–8374.

25. Andrés E, Martínez N, Planas A. 2011. Expression and characterization of a *Mycoplasma genitalium* glycosyltransferase in membrane glycolipid biosynthesis: Potential target against Mycoplasma infections. J Biol Chem 286:35367–35379.

26. Klement ML, Ojemyr L, Tagscherer KE, Widmalm G, Wieslander A. 2007. A processive lipid glycosyltransferase in the small human pathogen *Mycoplasma pneumoniae*: involvement in host immune response. Mol Microbiol 65:1444–1457.

27. Forget SM, Shepard SB, Soleimani E, Jakeman DL. 2019. On the catalytic activity of a GT1 family glycosyltransferase from *Streptomyces venezuelae* ISP5230. J Org Chem 84:11482-11492.

28. Gu XL, Chen M, Wang QZ, Zhang M, Wang BL, Wang HH. 2005. Expression and purification of a functionally active recombinant GDP-mannosyltransferase (PimA) from Mycobacterium tuberculosis H37Rv. Protein Expr Purif 42:47-53.

29. Batt SM, Jabeen T, Mishra AK, Veerapen N, Krumbach K, Eggeling L, Besra GS, Fütterer k. 2010. Acceptor substrate discrimination in phosphatidyl-myo-inositol mannoside synthesis:

609 structural and mutational analysis of mannosyltransferase Corynebacterium glutamicum
610 PimB'. J Biol Chem 285:37741-37752.
611
612 30. Sobhanifar S, Worrall LJ, Gruninger RJ, Wasney GA, Blaukopf M, Baumann L, Lameignere
613 E, Solomonson M, Brown ED, Withers SG, Strynadka NCJ. 2015. Structure and mechanism
614 of Staphylococcus aureus TarM, the wall teichoic acid α-glycosyltransferase. Proc Natl Acad
615 Sci USA 112:576-585.
616
617 31. Royer CJ, Cook PD. A structural and functional analysis of the glycosyltransferase BshA
618 from Staphylococcus aureus: insights into the reaction mechanism and regulation of bacillithiol
619 production. Proitein Sci 28:1083-1094.
620
621 32. Shi WW, Jiang YL, Zhu F, Yang YH, Shao QY, Yang HB, Ren YM, Wu H, Chen YX, Zhou
622 CZ. 2014. Structure of a novel O-Linked N-aetyl-D-glucosamine (O-GlcNAc) tansferase, GtfA,
623 reveals Insights into the gycosylation of pneumococcal serine-rich repeat adhesions. J Biol
624 Chem 289:20898-20907.

625 33. Troutman JM, Imperiali B. 2009. Campylobacter jejuni PglH is a single active site
626 processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer
627 reactions. Biochemistry 48:2807-2816.

628
629 34. Li Y, Li W, Zhang G, Lü X, Hwang H, Aker WG, Wang P. 2016. Purification and
630 characterization of polysaccharides degradases produced by *Alteromonas* sp. A321. Int J Biol
631 Macromol 86:96–104.
632
633 35. Zhang S, Zhou Z, Yao Z, He BF. 2013. Efficient production of skimmin and 6'-
634 succinylskimmin from umbelliferone by organic solvent-tolerant *Bacillus licheniformis* ZSP01
635 using nitrogen sources regulation strategy. Biochem Eng J 71:105-110.
636
637 36. Xie K, Dou X, Chen R, Chen D, Fang C, Xiao Z, Dai J. 2017. Two novel fungal phenolic
638 UDP glycosyltransferases from *Absidia coerulea* and *Rhizopus japonicus*. Appl Environ
639 Microbiol 83: e03103-16.
640
641 37. Kitayska T, Petrova P, Ivanova V, Tonkova AI. 2011. Purification and properties of a new
642 thermostable cyclodextrin glucanotransferase from *Bacillus pseudalcaliphilus* 8SB. Appl
643 Biochem Biotechnol 165:1285-1295.
644
645 38. Chen CI, Keusch JJ, Klein D, Hess D, Hofsteenge J, Gut H. 2012. Structure of human
646 POFUT2: insights into thrombospondin type 1 repeat fold and O-fucosylation. EMBO J
647 31:3183–3197.
648
649 39. Pruitt RN, Chumbler NM, Rutherford SA, Farrow MA, Friedman DB, Spiller B, and Lacy
650 DB. 2012. Structural determinants of *Clostridium difficile* toxin A glucosyltransferase activity. J
651 Biol Chem 287:8013–8020.
652
653 40. Albesa-Jove D, Guerin ME. 2016. The conformational plasticity of glycosyltransferases.
654 Curr Opin Struct Biol 40:23–32.
655
656 41. Errey JC, Lee SS, Gibson RP, Fleites CM, Barry CS, Jung PMI, O'Sullivan AC, Davis BG,
657 Davies GJ. 2010. Mechanistic insight into enzymatic glycosyl transfer with retention of
658 configuration through analysis of glycomimetic inhibitors. Angew Chem Int Ed Engl 49:1234-
659 1237.
660

661 42. Gómez H, Polyak I, Thiel W, LIuch JM, Masgrau L. 2012. Retaining glycosyltransferase
662 mechanism studied by QM/MM methods: lipopolysaccharyl-a-1, 4-galactosyltransferase C
663 transfers agalactose via an oxocarbenium ion-like transition state. J Am Chem Soc134:4743-
664 4752.
665
666 43. Lee SS, Hong SY, Errey JC, Izumi A, Davies GJ, Davis BG. 2011. Mechanistic evidence
667 for a front-side, S_Ni-type reaction in a retaining glycosyltransferase. Nat Chem Biol 7:631-638.
668
669 44. Guerin ME, Kaur D, Somashekar BS, Gibbs S, Gest P, Chatterjee D, Brennan PJ, Jackson
670 M. 2009. New insights into the early steps of phosphatidylinositol mannoside biosynthesis in
671 *mycobacteria*: PimB′ is an essential enzyme of *Mycobacterium smegmatis*. J Biol Chem
672 284:25687–25696.
673
674 45. Honda Y, Nakano S, Ito S, Dadashipour M, Zhang ZL, Kawarabayasi Y. 2018. Improvement
675 of ST0452 N-Acetylglucosamine-1-phosphate uridyltransferase activity by the cooperative
676 effect of two single mutations identified through structure-based protein engineering. Appl
677 Environ Microbiol84:e02213-18.
678
679 46. Chiu CP, Lairson LL, Gilbert M, Wakarchuk WW, Wither SG, Strynadka NC. 2007.
680 Structural analysis of the α-2, 3-Sialyltransferase Cst-I from *Campylobacter jejuni*in apo and
681 substrate-analogue bound forms'. Biochemistry 46:7196-7204.
682
683 47. Van Mooy, BAS, Rocap, G, Fredricks, HF, Evans, CT, Devol, AH. 2006. Sulfolipids
684 dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine
685 environments. Proc Natl Acad Sci USA 103: 8607–8612.
686
687 48. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin
688 F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and
689 Clustal X version 2.0. Bioinformatics 23: 2947-2948.
690
691 49. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis
692 version 7.0 for bigger datasets. Mol BiolEvol 33:1870–1874.
693
694 50. Kelly L, Mezulis S, Yates C, Wass M, Sternburg M. 2015. The Phyre2 web portal for protein
695 modelling, prediction and analyses. Nat Protoc 10:845–858.
696
697 51. Hodis E, Sussman JL. 2009. An encyclopedic effort to make 3D structures easier to
698 understand. Trends Biochem Sci 34:100–101.
699
700 52. Diller DJ, Merz KM, Jr. 2001. High throughput docking for library design and library
701 prioritization. Proteins 43:113–124.

Tables

Table 1 Purification of the recombinant GT_{cp} from *Candidatus* Pelagibacter sp. HTCC7211

Step	*Total protein (mg)*	*Total activity (U)*	*Specific activity (U/mg)*	*Purification fold*	*Yield (%)*
Crude cell extract	34.8	469	13.4	1	100
Ni-NTA affinity	1.9	85	44.7	3.3	18.1
Superdex-200 gel filtration	1.7	80	47.1	3.5	17.1

Table 2 Kinetic parameters of GT_{cp} using UDP-Glc, UDP-Gal and UDP-GlcA as sugar donors and different DAG as acceptors [a].

Substrate	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} mM^{-1})
UDP-Glc/ DAG (C16:0 /C18:1)	5.9±0.4	82.0±0.3	71.9±2.7
UDP-Gal/ DAG (C16:0 /C18:1)	1.0±0.3	32.0±0.4	31.3±2.9
UDP-GlcA/ DAG (C16:0 /C18:1)	4.1±0.1	68.0±0.5	60.3±3.6
UDP-Glc/DAG (C18:0 /C20:4)	4.8±0.2	77.1±0.4	62.2±0.9
UDP-Glc/DAG (C18:0 /C18:2)	6.2±1.5	102.5±0.8	60.5±1.8
UDP-Glc/DAG (di18:1)	5.0±0.8	90.5±1.6	55.2±2.2
UDP-Glc/DAG (di18:0)	3.6±0.7	75.3±1.2	47.8±1.3
UDP-Glc/DAG (di16:0)	1.7±1.4	55.2±0.9	30.8±2.5
UDP-Glc/DAG (di14:0)	1.0±2.2	45.9±1.1	21.7±2.8
UDP-Glc/DAG (di12:0)	0.5±0.4	21.6±0.8	13.9±1.7
UDP-Glc/DAG (di10:0)	ND	ND	ND
UDP-Glc/DAG (di8:0)	ND	ND	ND
UDP-Glc/DAG (C16:0 /C18:1) [b]	8.3±0.3	58.0±0.6	143.1±3.9

[a]The values are means of three independent experiments ± standard deviations. ND, not detectable.

[b] The kinetic parameters were determined using purified GT_{cp} in the presence of Mg^{2+} (5 mM).

Table 3 Kinetic parameters of GT_{cp} using UDP-Glc, UDP-Gal and UDP-GlcA as sugar donor and C16:0 /C18:1 DAG as acceptor [a].

Substrate	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} mM^{-1})
UDP-Glc	5.85±0.4	82±0.3	71.4±2.7
UDP-Gal	1.03±0.3	32±0.4	32.2±2.9
UDP-GlcA	4.05±0.1	68±0.5	59.6±3.6
UDP-Glc (with 5 mM Mg^{2+})	8.25±0.3	58±0.6	142.2±3.9

[a]The values are means of three independent experiments.

Table 4 Kinetic parameters of GT_{cp} in comparison with GT4 glycosyltransferases using different sugar donor [a].

Parameter	GT_{cp}	PimA	PimB	MshA	TarM	BshA	GtfA	PglH
Donor substrate	UDP-Glc	GDP-Man		UDP-GlcNAC				
K_m (μM)	82 ± 0.3	18 ± 2	19.0 ± 4.6	0.208 ± 0.017	65 ± 10	180 ± 50	11.8 ± 1.5	2.6 ± 0.3
k_{cat} (min^{-1})	5.8 ± 0.4				126 ± 10	78.6	7.35 ± 0.42	4.4 ± 0.2
k_{cat}/K_m ($min^{-1}mM^{-1}$)	71.4 ± 2.7					6.8	0.63	1.7 ± 0.2

a The values are means of three independent experiments. UDP-GlcNAC: UDP-N-acetylglucosamine; GDP-Man: GDP-mannose

Table 5 Primers used for PCR amplification in this study

Vectors	Upper primers [a]	Lower primers [b]
GT_{cp}	5′-GCCG CATATG AAAATTTTAATCGTAAC-3′ (*NcoI*)	5′-GGCGTCGACATTAGGTGATATTAAG-3′ *(SalI)*
G16A	5′- CCACTTGTGAATGCTGTAGTTCGAAC-3′	5′-GTTCGAACTACAGCATTCACAAGTGG-3′
L56A	5′-GAAATTAGAGCATCATTAAATGTTTG-3′	5′-CAAACATTTAATGATGCTCTAATTTC-3′
G82A	5′- CATATTGCAACAGAGGCACCTCTTGG-3′	5′- CCAAGAGGTGCCTCTGTTGCAATATG-3′
G85A	5′-GGGACCTCTTGCTTTTATGGCAAG-3′	5′-CTTGCCATAAAAGCAAGAGGTCCC-3′
H104A	5′- CAACAAGTTTTGCTACAAGATTTG-3′	5′- CAAATCTTGTAGCAAAACTTGTTG-3′
T162A	5′-GGTTACATGGGCTAGGGGAGGTAATC-3′	5′-CCATGATTACCTCCCCTAGCCCATGTAAC-3′
R163A	5′-GGTTACATGGACTAGCGGAGGTAATCATGG-3′	5′-CACCATGATTACCTCCGCTAGTCCATGTAAC-3′
R190A	5′- GGATATACGTCGGTGCAGTTGCAGTTG-3′	5′- CAACTGCAACTGCACCGACGTATATCC-3′
K195A	5′- GCAGTTGAAGCAAATATTAAAGC-3′	5′- GCTTTAATATTTGCTTCAACTGC-3′
D256A	5′- CCCTAGCAAAACGAATACTTTTGG-3′	5′- CCAAAAGTATGCGTTTTGCTAGGG-3′
D256E	5′- CCCTAGCAAAACGAATACTTTTGG-3′	5′- CCAAAAGTATTCGTTTTGCTAGGG-3′
T257A	5′-CCCTAGCAAAACCGATGCTTTTGGTATTG-3′	5′-CAATACCAAAAGCATCGGTTTTGCTAGGG-3′
T257S	5′-CCCTAGCAAAACCGATAGTTTTGGTATTG-3′	5′-CAATACCAAAACTATCGGTTTTGCTAGGG-3′
F258A	5′-CCGATACTGCTGGTATTGTGGTTTTGG-3′	5′- CCAAAACCACAATACCAGCAGTATCGG-3′
G259A	5′-CCGATACTTTTGCTATTGTGGTTTTGGAG-3′	5′- CTCCAAAACCACAATAGCAAAAGTATCGG-3′
E264A	5′-GTGGTTTTGGAGTCTTTAAGTTGTGG-3′	5′- CCACAACTTAAAGACTCCAAAACCAC-3′
W320A	5′-GCTAAAAAATATAGTTCGGAAGAAAC-3′	5′-CCTTGCTGTTTCTTCCGAACTATTTTTTAG-3′

[a, b] The sites of mutations are underlined.

723 **Figure legends**

724 **Figure 1** Multiple sequence alignment and functional domain analyses of GT_{cp} protein. (A) A
725 phylogenetic tree of GT_{cp} and its homologs with known 3-dimentionalX-ray structure of GTs.
726 Sequences and structures of GTs are obtained from the NCBI database and the PDB database,
727 including GT1 (calG3, SnogD, CalG1, SpnG, UrdGT2, OleD, CalG2, CalG4,GtfD, GtfA, Vinc,
728 UGT78G1, UGT78K6, UGT85H2, UGT71G1 and UGT72B1), GT3 (CeGs), GT5(Gbss1 and
729 SSI), GT9 (WaaC, WaaF and Vpar 0760), GT28 (MurG), GT30 (WaaA and KdtA), GT35
730 (AtPHS2, MalP and GlgP), GT68 (POFUT2), GT80 (Pdst and Bst), GT4 (PimA, PimB, WsaF,
731 GtfA, TarM, PglH, MshA, BshA and WaaG), GT_{cp} and its homologs (GT_{la}, GT_{tl}, GT_{mn}, GT_{dm},
732 GT_{cb}, GT_{kg}, GT_{al}, GT_{bb}, GT_{dr}, GT_{pa}, and Agt). GTpa is the GT4 glycosyltransferase of
733 *Pseudomonas* sp. PA14 (Supplementary Figure S1). (B) Multiple sequence alignment for GT_{cp}
734 and its homologs using Clustal W program with manual adjusting: GT_{cp} from *Candidatus*
735 Pelagibacter sp. HTCC7211 (WP_008545403.1); GT_{la} from *Labrenzia*
736 *aggregata*(WP_040439323.1); GT_{tl} from *Thalassospira lucentensis* (WP_062950653.1); GT_{mn}
737 from *Methylophaga nitratireducenticrescens* (WP_014706011.1); GT_{dm} from *Desulfobulbus*
738 *mediterraneus* (WP_028584068.1); GT_{cb} from *Citromicrobium bathyomarinum* JL354
739 (WP_010239457.1); GT_{kg} from *Kordiimonas gwangyangensis* (WP_051078133.1); the
740 characterized Agt from *A. tumefaciens* (atu2297) and MshA from *C. glutamicum*
741 (WP_143854623.1). Red bars represent α-helical regions and blue arrows represent β-sheet
742 from MshA (Protein Data Bank [PDB] entry 3C4Q). Residues interacting with UDP-sugar are
743 indicated by the closed black circles. Catalytic dyad composed of His104-Asp256 is shaded
744 in grey. Black boxed regions and red boxed regions in the alignment indicated two conserved
745 UDP-sugar binding motifs of GRVAXEKN and FPSXTDTFG and a conserved Gly-rich motif,
746 respectively. (C) Homology modelling showing the predicted structure of GT_{cp} and the catalytic
747 dyad composed of His104-Asp256. The signature glutamate residue in MshA (Glu316) is
748 substituted to an aspartate residue (Asp256) in GT_{cp}.
749
750 **Figure 2** (A) Over-expression and purification of GT_{cp} proteins from *Candidatus* Pelagibacter

751 sp. HTCC7211. M, protein molecular weight marker. Lane 1 cell-free supernatant induced with
752 Isopropyl β-D-1-thiogalactopyranoside (IPTG), lane 2 cell-free supernatant without IPTG
753 induction and lanes 3-4, purified GT_{cp} (molecular weight estimated to be ~ 38 kDa). (B) Gel
754 filtration analysis of the wild-type GT_{cp} and the mutant L56A. The red arrows indicate the eluted
755 position of GT_{cp} and the L56A mutant, the black arrows indicate protein markers (from left to
756 right): alcohol dehydrogenase (150 kDa, 12.35 ml), albumin (66 kDa, 15.83 ml) and carbonic
757 anhydrase (29 kDa, 17.95 ml).
758
759 **Figure 3** Functional characterization of recombinant GT_{cp}. (A) TLC of the enzymatic reaction
760 products with different UDP-sugar donors and diacylglycerol (DAG) as the acceptor by staining
761 with sulfuric acid/methanol/water (45:45:10). (B-D) LC-MS of fragmentation spectra for
762 monohexuronosyl DAGs (MGlc-DAG and MGal-DAG) and MGlcA-DAG were obtained from
763 the products of the GT_{cp}-catalysed reaction.
764
765 **Figure 4** Biochemical characterization of GT_{cp}. (A) Effect of temperature on the activities of
766 GT_{cp}. (B) Arrhenius plot of GT_{cp}. The activation energy of the reaction E_a = 25.1 kJ mol^{-1} could
767 be determined from the slope of the regression curve. (C) Thermostability of GT_{cp}. The residual
768 enzyme activity was measured after incubation of the purified enzyme at 30°C (diamonds),
769 40°C (triangles), and 50°C (boxes), respectively. (D) Effect of pH on the activities of GT_{cp}. (E)
770 Effect of NaCl on the activities of GT_{cp}. The enzyme was incubated in buffers containing
771 different concentrations of NaCl (0 to 4 M) at 4°C for 1 h. Residual activity was measured
772 under optimal conditions. F) Effect of metal ions (5 mM) on the activities of GT_{cp}. The values
773 are means of three independent experiments.
774
775 **Figure 5** Homology modelling prediction of the UDP-sugar donor binding pocket in the GT_{cp}.
776 (A) UDP-sugar binding site of MshA (22) and predicted UDP-sugar binding site of GT_{cp}. (B)
777 Mutational analysis of the key amino acids involved in the catalytic glycosylation reactions of
778 GT_{cp}. C) Overlay of ^{1}H, heteronuclear single-quantum correlation (HSQC) and heteronuclear

779 multiple bond correlation (HMBC) spectra of MGlc-DAG.

780

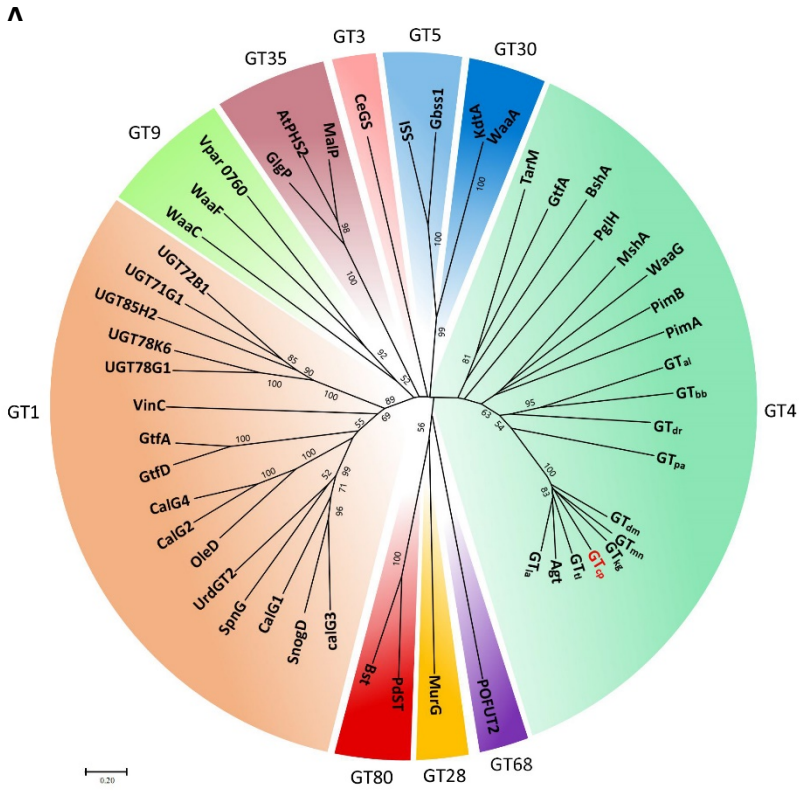
781 **Figure 6** Identification of key residues for diacylglycerol (DAG) binding in the GT_{cp}. **A)**
782 Homology modelling prediction of the open model based on template 2R60. The arrow
783 indicates the wide cleft. **B)** C16:1/C18:0-DAG (green) docked in GT_{cp} in the groove (shown as
784 a transparent surface). **C)** A detailed depiction of key coordinating hydrophobic (black) and
785 polar (light blue) residues for DAG. His104, Thr162, Arg163, Trp320 are crucial for enzyme
786 activity (Fig. 5B). **D)** An overlay of UDP-glucose (Blue), UDP-galactose (Cyan) and UDP-
787 glucuronic acid (Green) showing all three ligands can occupy the same binding site in similar
788 poses in the DAG-docked GT_{cp}. **E)** A top-down view of DAG and UDP-galactose shown docked
789 parallel to each other in their respective binding pockets/grooves.

790

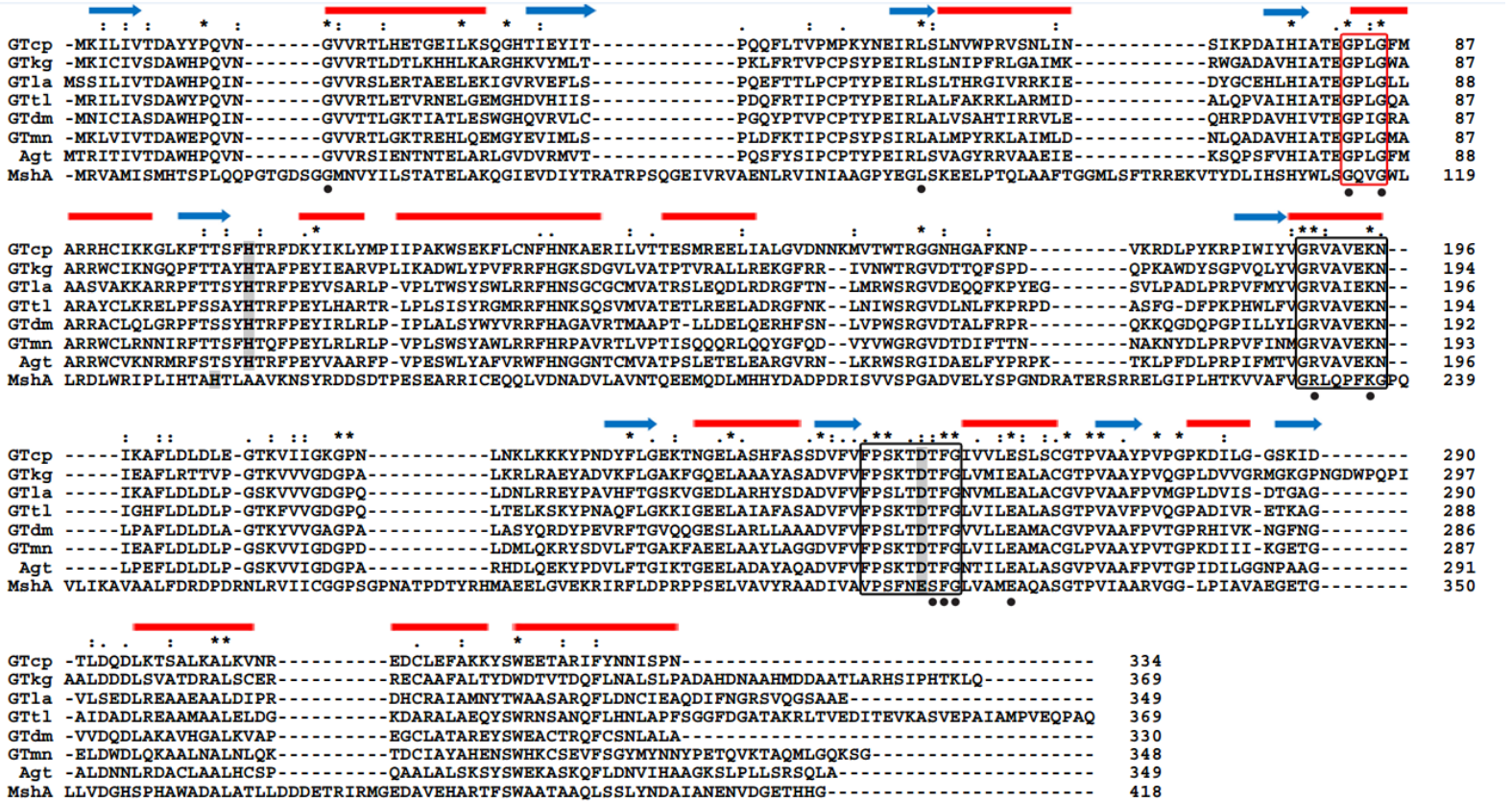
791

792 **Figure 7** Proposed pathway of the synthesis of non-phosphorus glycoglycerolipids through
793 PlcP and GT_{cp} in *Candidatus* Pelagibacter sp. HTCC7211. PlcP converts phosphatidylglycerol
794 (PG) or phosphatidylethanolamine (PE) to generate diacylglycerol (DAG), and GT_{cp} can
795 synthesize different glycoglycerolipids MGlc-DAG, MGal-DAG and MGlcA-DAG with UDP-Glc,
796 UDP-Gal, or UDP-GlcA as sugar donors and DAG as the acceptor.

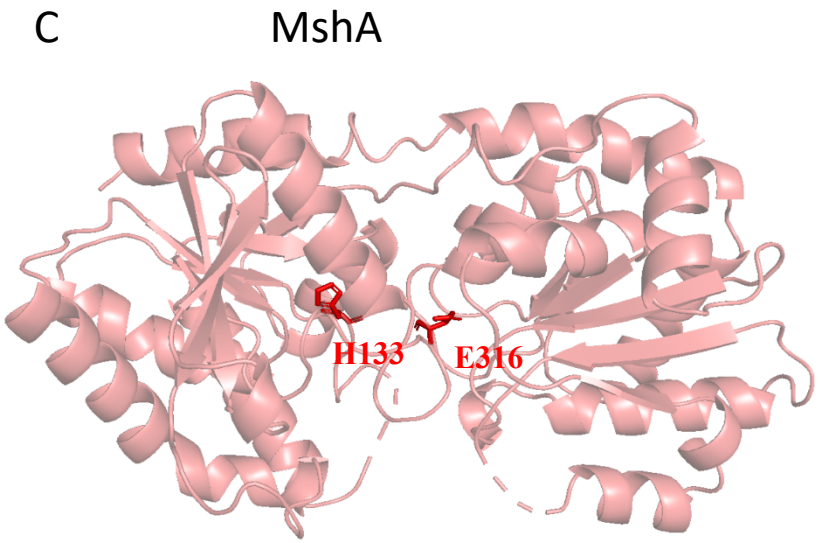
A

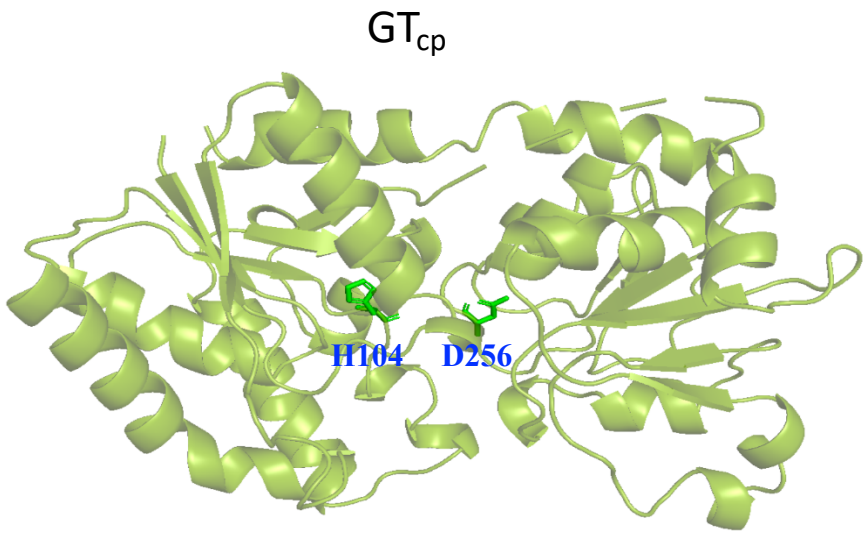

GT3
GT5
GT30
GT35
GT9
GT1
GT4
GT80
GT28
GT68
AtPHS2
MalP
CeGS
ISS
Gbss1
KdtA
WaaA
TarM
GtfA
BshA
PglH
MshA
WaaG
PimB
PimA
GTal
GTbb
GTdr
GTpa
GTdm
GTmn
GTkg
GTla
Agt
GTtl
GTcp
Vpar 0760
WaaF
WaaC
GigP
UGT72B1
UGT71G1
UGT85H2
UGT78K6
UGT78G1
VinC
GtfA
GtfD
CalG4
CalG2
OleD
UrdGT2
SpnG
CalG1
SnogD
calG3
BsF
PdST
MurG
POFUT2

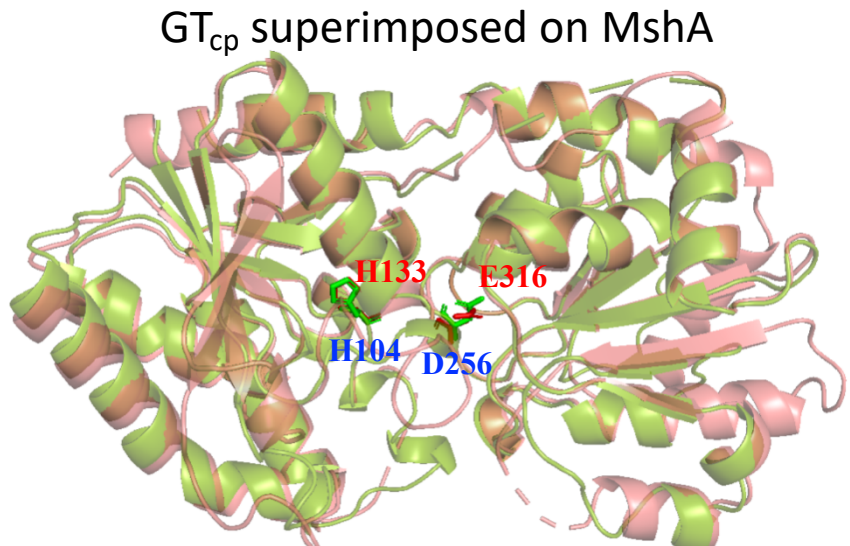
B


GTcp -MKILIVTDAYYPQVN-------GVVRTLHETGEILKSQGHTIEYIT-------------PQQFLTVPMPKYNEIRLSLNVWPRVSNLIN------------SIKPDAIHIATEGPLGFM 87
GTkg -MKICIVSDAWHPQVN-------GVVRTLDTLKHHLKARGHKVYMLT-------------PKLFRTVPCPSYPEIRLSLNIPFRLGAIMK------------RWGADAVHIATEGPLGWA 87
GTla MSSILIVTDAWHPQIN-------GVVRSLERTAEELEKIGVRVEFLS-------------PQEFTTLPCPTYPEIRLSLTHRGIVRRKIE------------DYGCEHLHIATEGPLGLL 88
GTtl -MRILIVSDAWYPQVN-------GVVRTLETVRNELGEMGHDVHIIS-------------PDQFRTIPCPTYPEIRLALFAKRKLARMID------------ALQPVAIHIATEGPLGQA 87
GTdm -MNICIASDAWHPQIN-------GVVTTLGKTIATLESWGHQVRVLC-------------PGQYPTVPCPTYPEIRLALVSAHTIRRVLE------------QHRPDAVHIVTEGPIGRA 87
GTmn -MKLVIVTDAWEPQVN-------GVVRTLGKTREHLQEMGYEVIMLS-------------PLDFKTIPCPSYPSIRLALMPYRKLAIMLD------------NLQADAVHIATEGPLGMA 87
Agt MTRITIVTDAWHPQVN-------GVVRSIENTNTELARLGVDVRMVT-------------PQSFYSIPCPTYPEIRLSVAGYRRVAAEIE------------KSQPSFVHIATEGPLGFM 88
MshA -MRVAMISMHTSPLQQPGTGDSGGMNVYILSTATELAKQGIEVDIYTRATRPSQGEIVRVAENLRVINIAAGPYEGLSKEELPTQLAAFTGGMLSFTRREKVTYDLIHSHYWLSGQVGWL 119
GTcp ARRHCIKKGLKFTTSFHTRFDKYIKLYMPIIPAKWSEKFLCNFHNKAERILVTTESMREELIALGVDNNKMVTWTRGGNHGAFKNP---------VKRDLPYKRPIWIYVGRVAVEKN-- 196
GTkg ARRWCIKNGQPFTTAYHTAFPEYIEARVPLIKADWLYPVFRRFHGKSDGVLVATPTVRALLREKGFRR--IVNWTRGVDTTQFSPD---------QPKAWDYSGPVQLYVGRVAVEKN-- 194
GTla AASVAKKARRPFTTSYHTRFPEYVSARLP-VPLTWSYSWLRRFHNSGCGCMVATRSLEQDLRDRGFTN--LMRWSRGVDEQQFKPYEG-------SVLPADLPRPVFMYVGRVAIEKN-- 196
GTtl ARAYCLKRELPFSSAYHTRFPEYLHARTR-LPLSISYRGMRRFHNKSQSVMVATETLREELADRGFNK--LNIWSRGVDLNLFKPRPD-------ASFG-DFPKPHWLFVGRVAVEKN-- 194
GTdm ARRACLQLGRPFTSSYHTRFPEYIRLRLP-IPLALSYWYVRRFHAGAVRTMAAPT-LLDELQERHFSN--LVPWSRGVDTALFRPR---------QKKQGDQPGPILLYLGRVAVEKN-- 192
GTmn ARRWCLRNNIRFTTSFHTQFPEYLRLRLP-VPLSWSYAWLRRFHRPAVRTLVPTISQQQRLQQYGFQD--VYVWGRGVDTDIFTTN---------NAKNYDLPRPVFINMGRVAVEKN-- 193
Agt ARRWCVKNRMRFSTSYHTRFPEYVAARFP-VPESWLYAFVRWFHNGGNTCMVATPSLETELEARGVRN--LKRWSRGIDAELFYPRPK-------TKLPFDLPRPIFMTVGRVAVEKN-- 196
MshA LRDLWRIPLIHTAHTLAAVKNSYRDDSDTPESEARRICEQQLVDNADVLAVNTQEEMQDLMHHYDADPDRISVVSPGADVELYSPGNDRATERSRRELGIPLHTKVVAFVGRLQPFKGPQ 239
GTcp -----IKAFLDLDLE-GTKVIIGKGPN----------LNKLKKKYPNDYFLGEKTNGELASHFASSDVFVFPSKTDTFGIVVLESLSCGTPVAAYPVPGPKDILG-GSKID-------- 290
GTkg -----IEAFLRTTVP-GTKVVVGDGPA----------LKRLRAEYADVKFLGAKFGQELAAAYASADVFVFPSKTDTFGLVMIEALACGTPVAAYPVQGPLDVVGRMGKGPNGDWPQPI 297
GTla -----IKAFLDLDLP-GSKVVVGDGPQ----------LDNLRREYPAVHFTGSKVGEDLARHYSDADVFVFPSLTDTFGNVMLEALACGVPVAAFPVMGPLDVIS-DTGAG-------- 290
GTtl -----IGHFLDLDLP-GTKFVVGDGPQ----------LTELKSKYPNAQFLGKKIGEELAIAFASADVFVFPSKTDTFGLVILEALASGTPVAVFPVQGPADIVR-ETKAG-------- 288
GTdm -----LPAFLDLDLA-GTKYVVGAGPA----------LASYQRDYPEVRFTGVQQGESLARLLAAADVFVFPSLTDTFGVVLLEAMACGVPVAAFPVTGPRHIVK-NGFNG-------- 286
GTmn -----IEAFLDLDLP-GSKVVIGDGPD----------LDMLQKRYSDVLFTGAKFAEELAAYLAGGDVFVFPSKTDTFGLVILEAMACGLPVAAYPVTGPKDIII-KGETG-------- 287
Agt -----LPEFLDLDLP-GSKVVIGDGPA----------RHDLQEKYPDVLFTGIKTGEELADAYAQADVFVFPSKTDTFGNTILEALASGVPVAAFPVTGPIDILGGNPAAG-------- 291
MshA VLIKAVAALFDRDPDRNLRVIICGGPSGPNATPDTYRHMAEELGVEKRIRFLDPRPPSELVAVYRAADIVAVPSFNESFGLVAMEAQASGTPVIAARVGG-LPIAVAEGETG-------- 350
GTcp -TLDQDLKTSALKALKVNR----------EDCLEFAKKYSWEETARIFYNNISPN---------- 334
GTkg AALDDDLSVATDRALSCER----------RECAAFALTYDWDTVTDQFLNALSLPADAHDNAAHMDDAATLARHSIPHTKLQ---------- 369
GTla -VLSEDLREAAEAALDIPR----------DHCRAIAMNYTWAASARQFLDNCIEAQDIFNGRSVQGSAAE---------- 349
GTtl -AIDADLREAAMAALELDG----------KDARALAEQYSWRNSANQFLHNLAPFSGGFDGATAKRLTVEDITEVKASVEPAIAMPVEQPAQ 369
GTdm -VVDQDLAKAVHGALKVAP----------EGCLATAREYSWEACTRQFCSNLALA---------- 330
GTmn -ELDWDLQKAALNALNLQK----------TDCIAYAHENSWHKCSEVFSGYMYNNYPETQVKTAQMLGQKSG---------- 348
Agt -ALDNNLRDACLAALHCSP----------QAALALSKSYSWEKASKQFLDNVIHAAGKSLPLLSRSQLA---------- 349
MshA LLVDGHSPHAWADALATLLDDDETRIRMGEDAVEHARTFSWAATAAQLSSLYNDAIANENVDGETHHG---------- 418

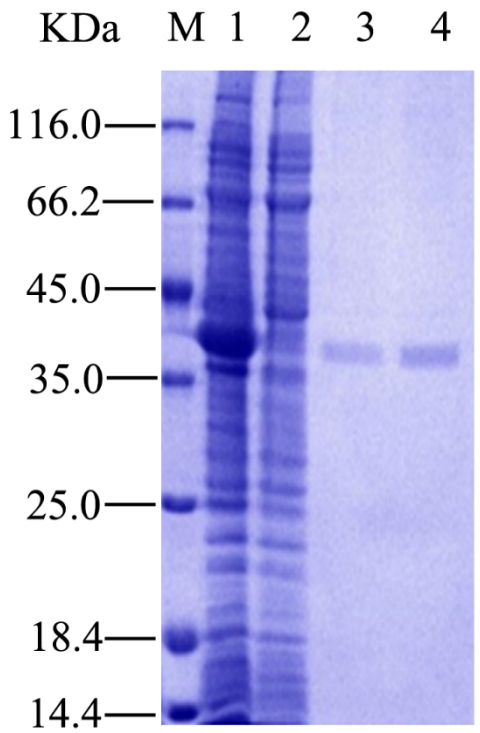
C


MshA
H133
E316

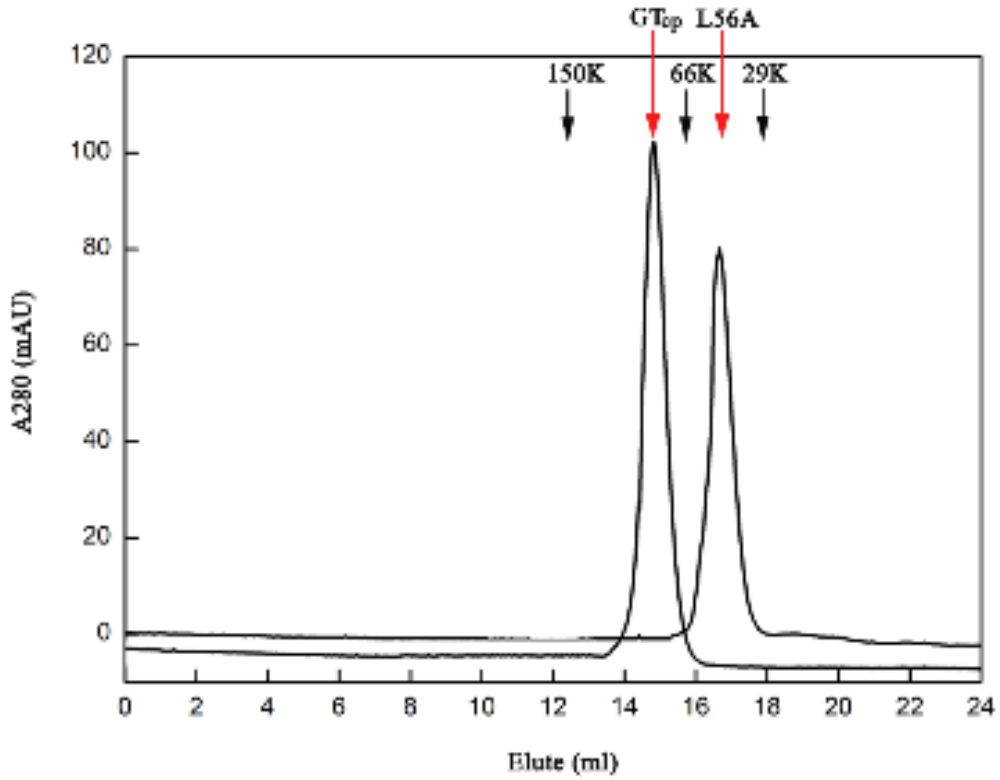

GTcp
H104
D256


GTcp superimposed on MshA
H133
E316
H104
D256

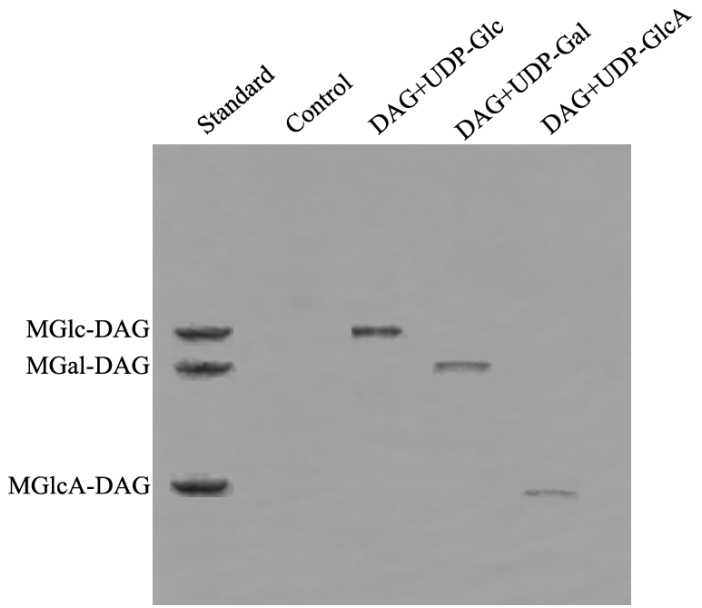
A


KDa
M
1
2
3
4
116.0
66.2
45.0
35.0
25.0
18.4
14.4

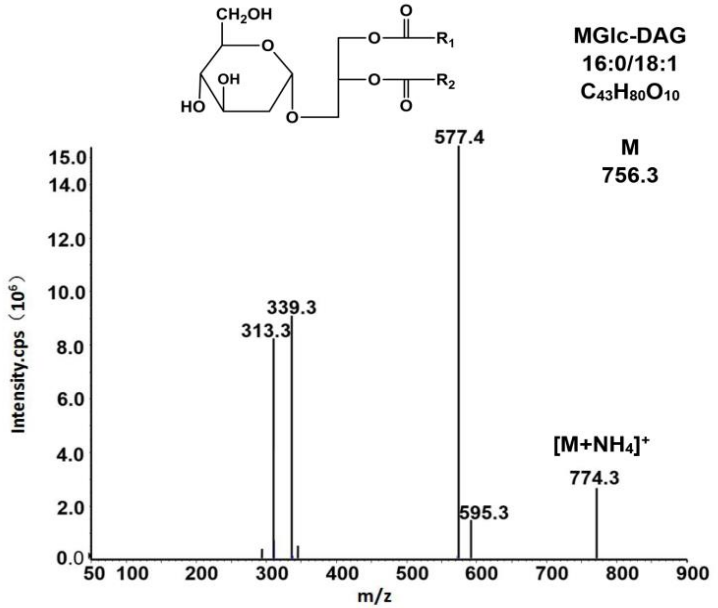
B


GTcp
L56A
150K
66K
29K
A280 (mAU)
Elute (ml)
0
20
40
60
80
100
120
0
2
4
6
8
10
12
14
16
18
20
22
24

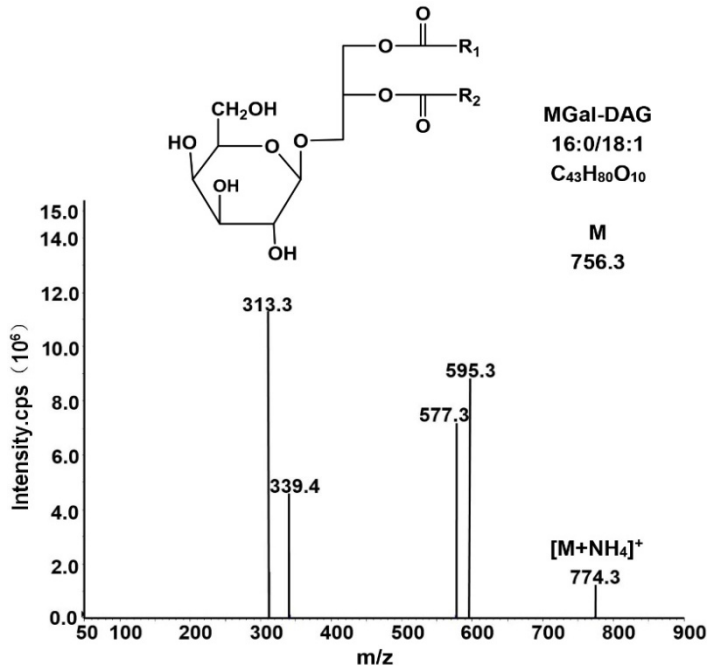
A


Standard
Control
DAG+UDP-Glc
DAG+UDP-Gal
DAG+UDP-GlcA
MGlc-DAG
MGal-DAG
MGlcA-DAG

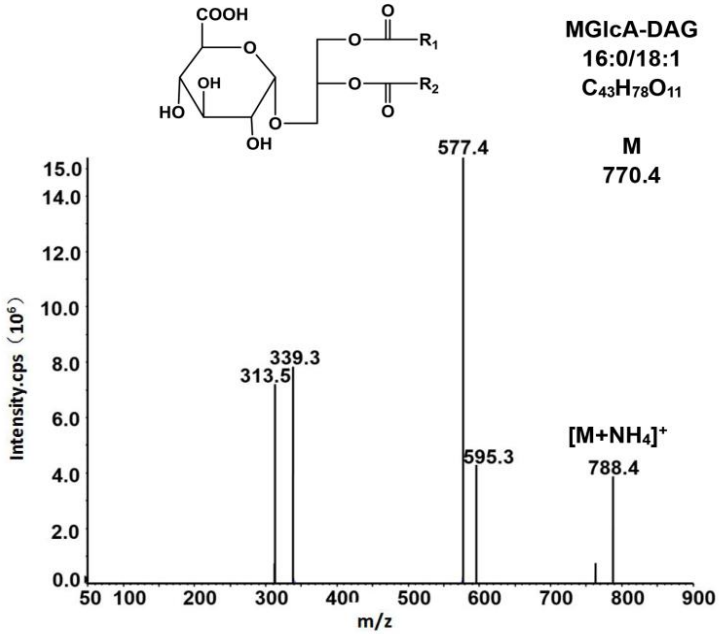
B

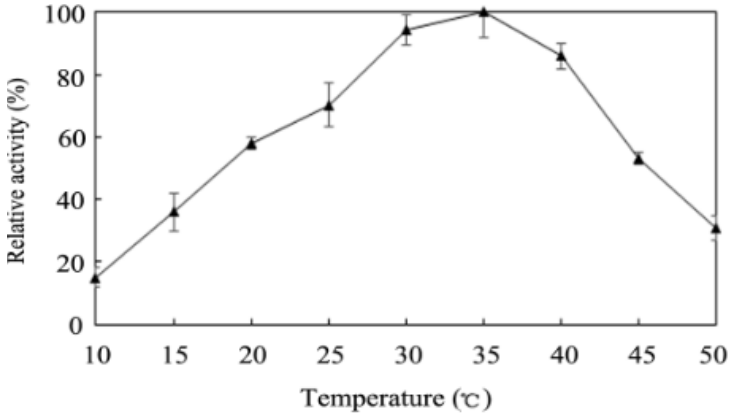

MGlc-DAG
16:0/18:1
C43H80O10
M
756.3
577.4
339.3
313.3
595.3
[M+NH4]+
774.3
Intensity.cps (10^6)
m/z

C

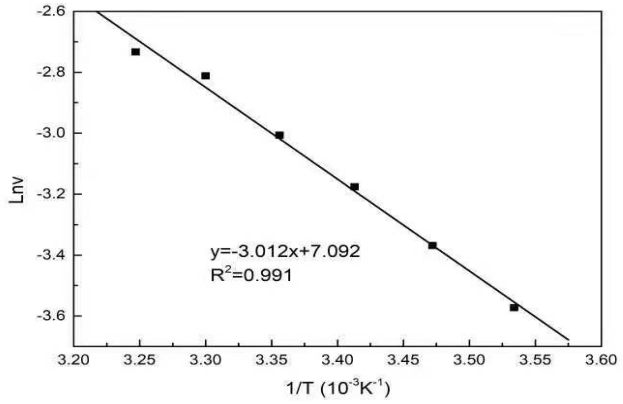

MGal-DAG
16:0/18:1
C43H80O10
M
756.3
313.3
595.3
577.3
339.4
[M+NH4]+
774.3
Intensity.cps (10^6)
m/z

D

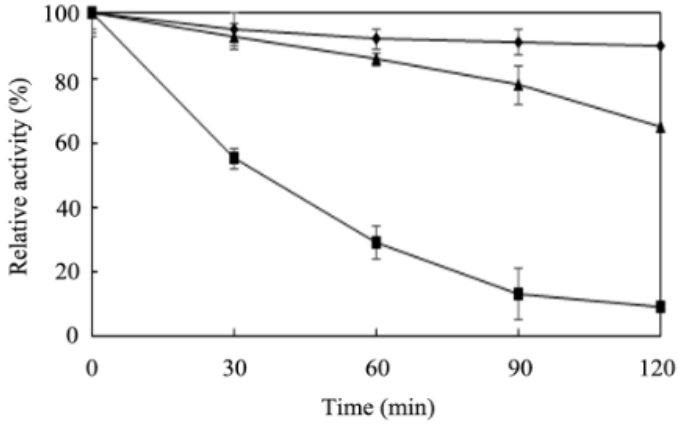

MGlcA-DAG
16:0/18:1
C43H78O11
M
770.4
577.4
339.3
313.5
595.3
[M+NH4]+
788.4
Intensity.cps (10^6)
m/z


Relative activity (%)
Temperature (℃)

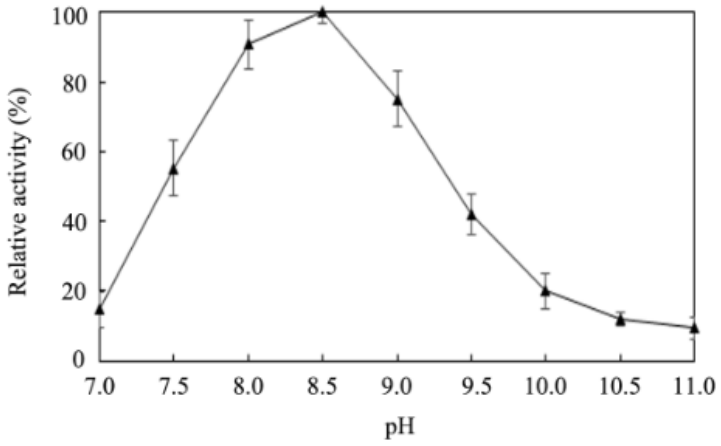
A


Lnv
y=-3.012x+7.092
R²=0.991
1/T (10⁻³K⁻¹)

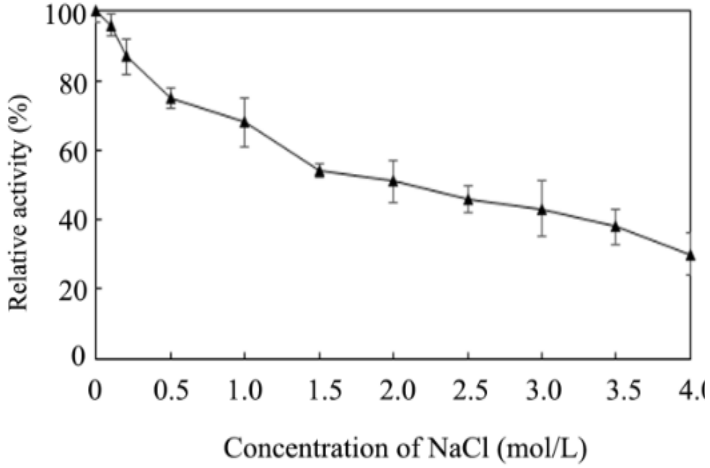
B


Relative activity (%)
Time (min)

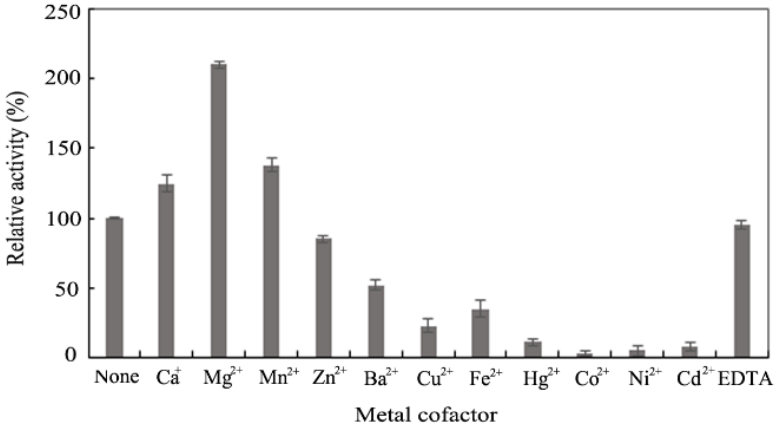
C


Relative activity (%)
pH

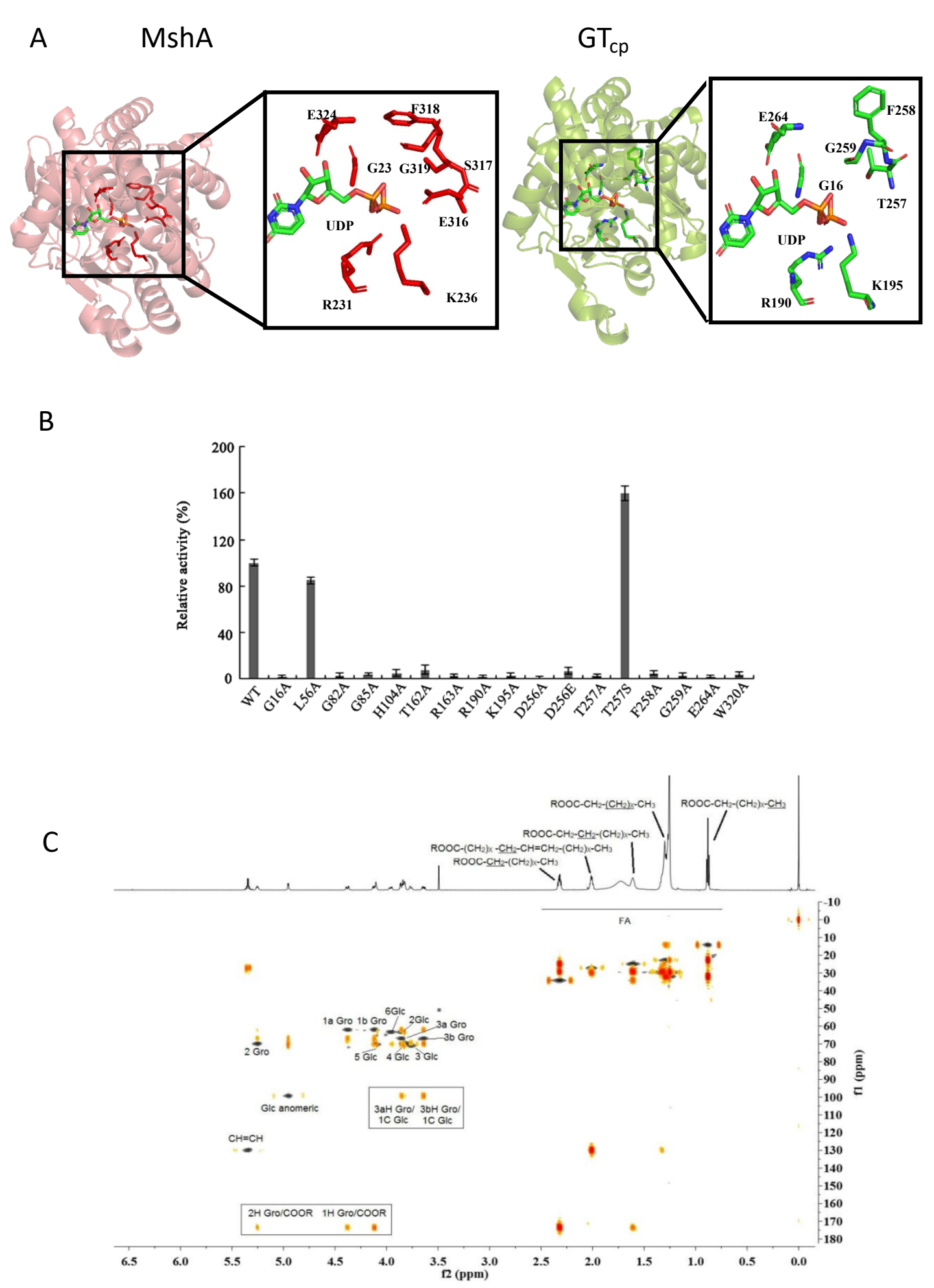
D


Relative activity (%)
Concentration of NaCl (mol/L)

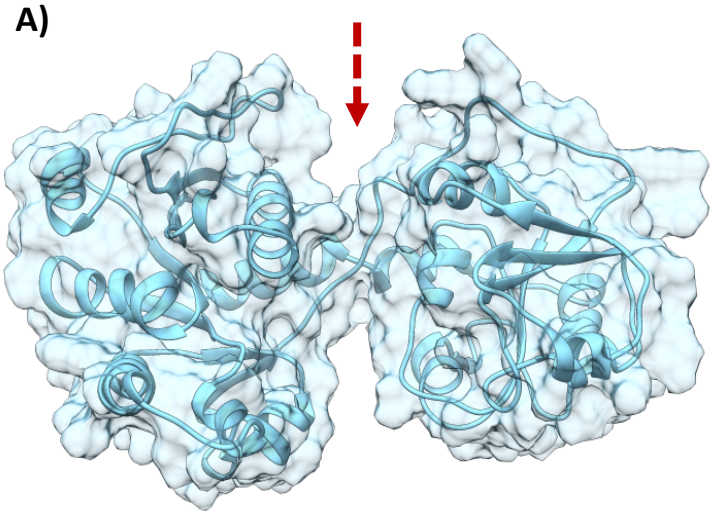
E


Relative activity (%)
None
Ca2+
Mg2+
Mn2+
Zn2+
Ba2+
Cu2+
Fe2+
Hg2+
Co2+
Ni2+
Cd2+
EDTA
Metal cofactor

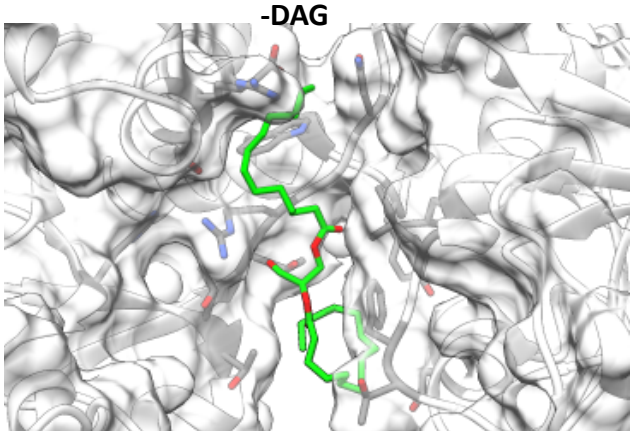
F


A
MshA
GTcp
E324
F318
G23
G319
S317
UDP
E316
R231
K236
E264
F258
G259
G16
T257
UDP
K195
R190
B
Relative activity (%)
200
160
120
80
40
0
WT
G16A
L56A
G82A
G85A
H104A
T162A
R163A
R190A
K195A
D256A
D256E
T257A
T257S
F258A
G259A
E264A
W320A
C
ROOC-CH2-(CH2)x-CH3
ROOC-CH2-CH2-(CH2)x-CH3
ROOC-(CH2)x -CH2-CH=CH2-(CH2)x-CH3
ROOC-CH2-(CH2)x-CH3
FA
6Glc
1a Gro
1b Gro
2Glc
3a Gro
3b Gro
2 Gro
5 Glc
4 Glc
3 Glc
Glc anomeric
3aH Gro/ 1C Glc
3bH Gro/ 1C Glc
CH=CH
2H Gro/COOR
1H Gro/COOR
f2 (ppm)
f1 (ppm)

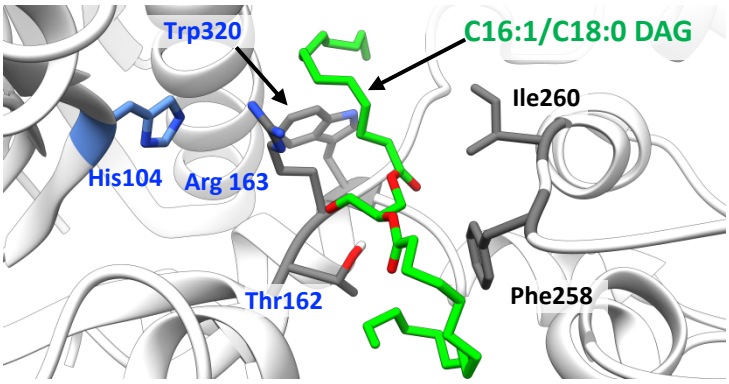
A)



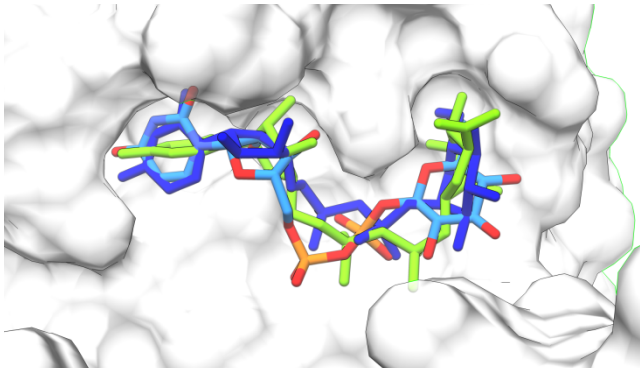
B)


C16:1/C18:0
-DAG

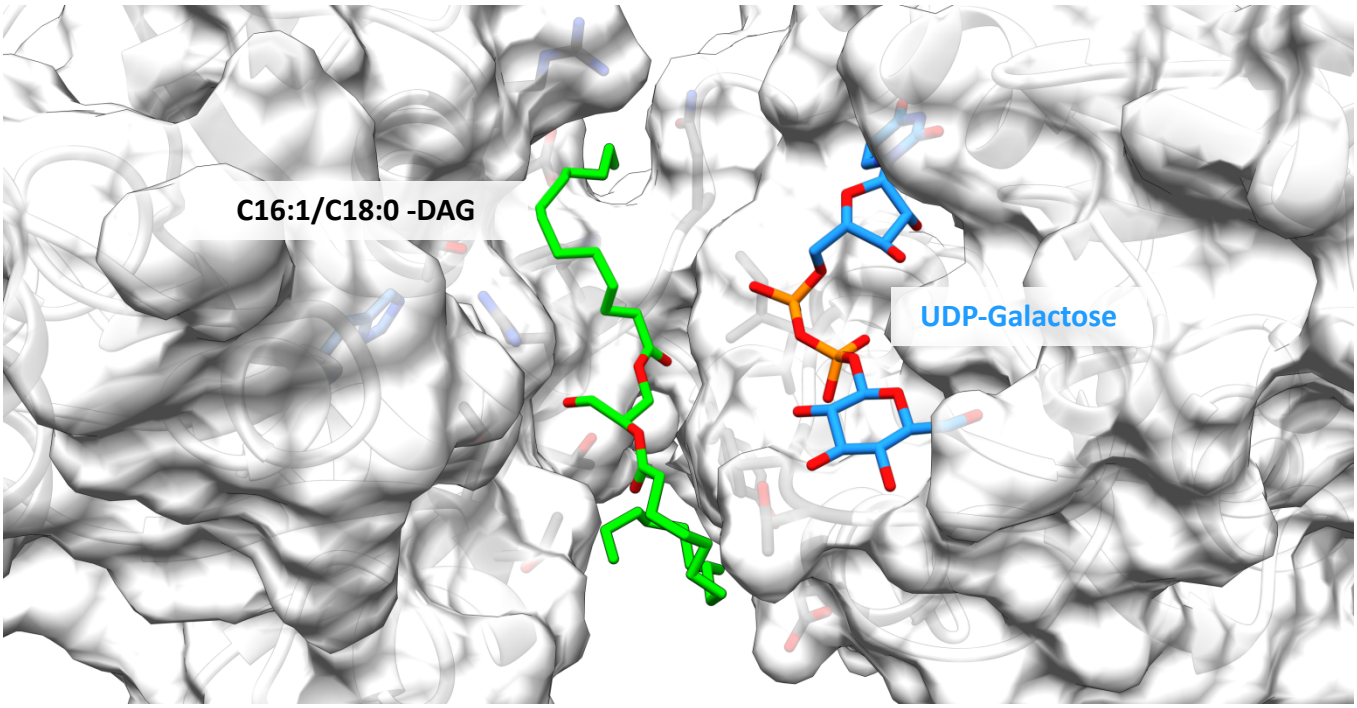
C)

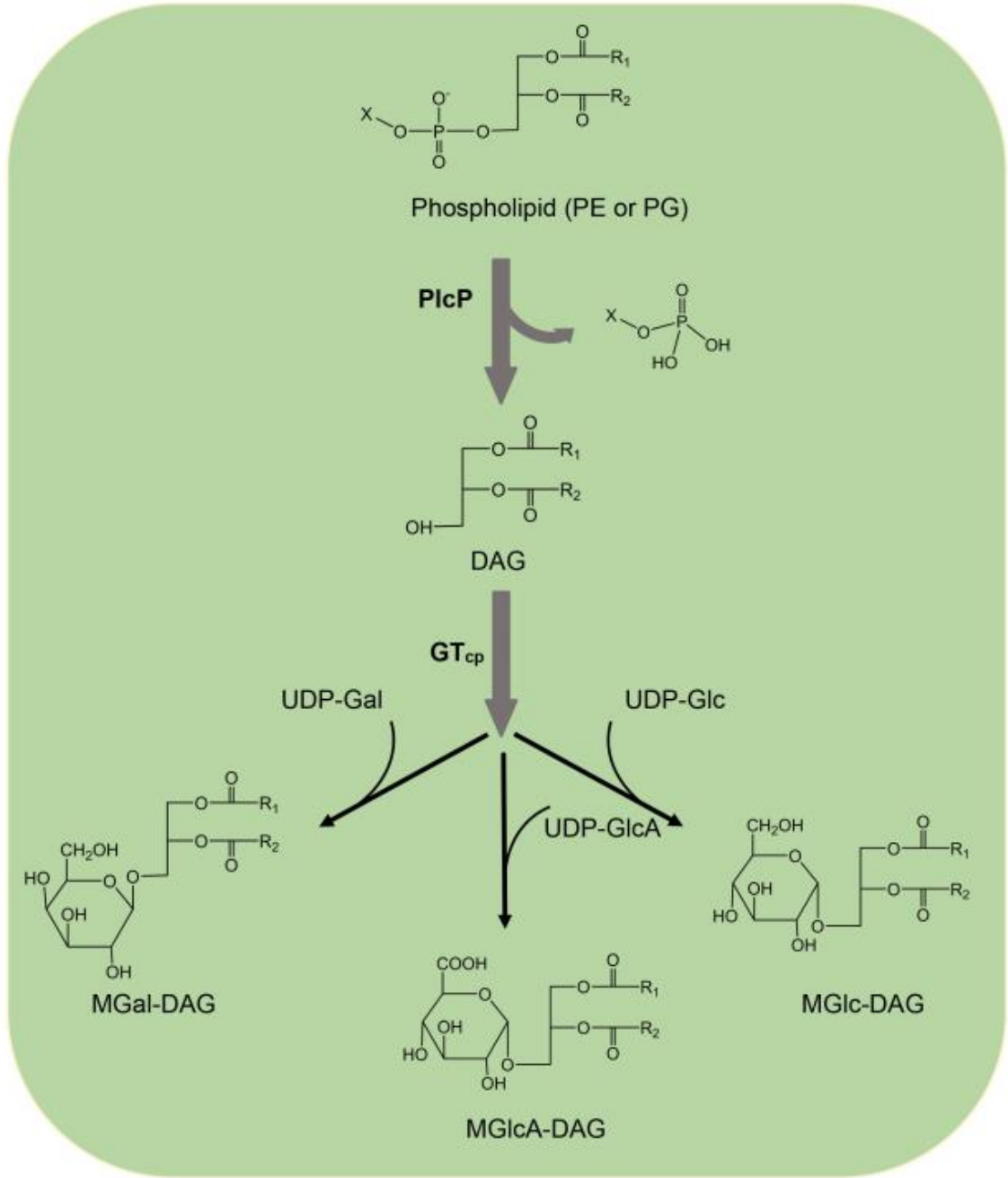

Trp320
C16:1/C18:0 DAG
Ile260
His104
Arg 163
Thr162
Phe258

D)



E)


C16:1/C18:0 -DAG
UDP-Galactose


Phospholipid (PE or PG)
PlcP
DAG
GTcp
UDP-Gal
UDP-Glc
UDP-GlcA
MGal-DAG
MGlc-DAG
MGlcA-DAG