Engineering and applications of glycosyltransferases for synthetic glycobiology : Two galactosyltransferases, MGD1 and LgtC, as case studies

Dylan Caccam Jabeguero

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Chimie Biologie

Arrêté ministériel : 25 mai 2016

Présentée par

Dylan Caccam JABEGUERO

Thèse dirigée par **Christelle BRETON**,Professeur, Université Grenoble Alpes

préparée au sein du **Laboratoire CEntre de Recherche sur les MAcromolécules Végétales**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Ingénierie et applications des glycosyltransférases pour la glycobiologie synthétique : Deux galactosyltransférases, MGD1 et LgtC, comme études de cas

Engineering and applications of glycosyltransferases for synthetic glycobiology: Two galactosyltransferases, MGD1 and LgtC, as case studies

Thèse soutenue publiquement le **30 juin 2022**,
devant le jury composé de :

Monsieur Sébastien MONGRAND
DIRECTEUR DE RECHERCHE, CNRS délégation Aquitaine, Rapporteur
Monsieur Tony LEFEBVRE
PROFESSEUR DES UNIVERSITES, Université de Lille, Rapporteur
Monsieur Romain VIVES
DIRECTEUR DE RECHERCHE, CNRS délégation Alpes, Examinateur
Madame Corinne MERCIER
PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes, Présidente
Monsieur Winfried RÖMER
PROFESSEUR, Albert-Ludwigs-Universität Freiburg , Examinateur

Acknowledgements

There is a long list of people whom I would like to give my heartfelt gratitude.

First and foremost is Christelle Breton, who has been an exceptional supervisor and overall amazing person. Displaying patience and understanding, she has been a wealth of knowledge and beacon of guidance throughout the past three years. I would not be the scientist I am today without you. I cannot thank you enough!

Next to her is the GBMS group including Anne Imberty, Serge Perez, Olivier, Annabelle Varrot, Emilie Gillon and Valérie Chazalet who have welcomed me warmly at CERMAV. In one way or another they have shared their knowledge, experience and expertise. It's no secret that I took a year off from science before starting the PhD. It's thanks to mentoring from Emilie and specially Valou that I was able to hit the ground running at the lab.

Without a doubt this project wouldn't be possible without the synBIOcarb consortium and the amazing people in it. The in-person meetings at Leeds, Graz and Prague are some of the best highlights. Through this network I was also fortunate to form close collaboration with two groups and eventually secondments at their respective labs. First at the University of Freiburg under the supervision of Winfried Romer and mentoring from Lina Siukstaite. Then later on at the University of Leeds with supervision from Ralf Richter and mentoring from Chunyue Wang. Truly, a wonderful experience and the skills and knowledge gained are invaluable.

Of course I wouldn't be writing this thesis without the members of my CSI committee: Eric Marechal, Romain Vives and Sylvie Armand, whom every year had showed care in my wellbeing as a PhD student, taken time to evaluate my work and allowed me to progress to the next year.

A very special thanks goes to the members of my thesis committee: Tony Lefebvre, Sébastien Mongrand, Winfried Römer, Romain Vives and Corinne Mercier. They have showed goodwill by taking interest in my project, taking time to read my thesis and joining me in the defense.

Naturally these three years wouldn't be as bright and enjoyable without all the friends and people I have met. First coming to CERMAV I was welcomed by Martin, Sue and Simona, the later whom would be a constant friend and confidante throughout the PhD. Later on, I was happy to meet new faces in Olga, Raphael, Margherita, Dania, Kanhaya and Jalaa, the last two whom I shared an office with. In one way or another, they have all motivated me to press on forward with this project. Also in this journey I also had the opportunity to mentor a master student, Sara Pullara. While brief she has helped in the project and we have both benefitted from this mentorship period.

While first settling in at Grenoble I was fortunate enough to meet staunch friends in Julio, Lara, Cil and Matt. This meeting eventually to led to a blossoming organized Filipino community in Grenoble, through which I was able to meet other Filipinos living and studying in Grenoble. I am still amazed at how a coincidental meeting have led to lifelong friends.

Finally, I would like to thank my parents and friends back in the UK, who have been very supportive and a constant presence, specially through the early days of the pandemic.

I will always remember this three-year journey fondly and a highlight of my life.

This PhD is part of the synBIOcarb European Training Network and has received funding from Horizon 2020.

Table of Contents

Abbreviations

Anti-His-HRP : His-tag specific antibody + Horseradish peroxidase

CaMGD1 : *Caldilinea aerophila* MonoGalactosylDiacylglycerol Synthase 1

CBPs : Carbohydrate-binding proteins

CHAPS : 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate

DAG : Diacylglycerol

DGDG : DiGalactosylDiacylglycerol

DLS : Dynamic light scattering

DOPC : 1,2-Dioleoyl-sn-glycero-3-phosphocholine

DOPE : 1,2-Dioleoyl-*sn*-glycero-3-phosphoethanolamine

DTT : Dithiothreitol

Gal : Galactose

GalT : Galactosyltransferase

Gb3 : Globotriaosylceramide

Glc : Glucose

GlcT : Glucosylsyltransferase

GT : Glycosyltransferase

GUV : Giant unilamellar vesicles

HBS : HEPES buffered saline

His-tag : 6- Histidine tag

IMAC : Immobilized Metal Affinity Chromatography

IPTG : Isopropyl β- d-1-thiogalactopyranoside

LacCer : Lactosylceramide

LB : Lysogeny broth

LgtC : Lipooligosaccharide α-1,4-galactosyltransferase C

LOS : Lipo-oligosaccharide

LUV : Large unilamellar vesicles

MD : Molecular Dynamics

MGD1 : *Arabidopsis thaliana* MonoGalactosylDiacylglycerol Synthase 1

MgdA *Chlorobaculum tepidum* MonoGalactosylDiacylglycerol Synthase

MGDG : MonoGalactosylDiacylglycerol

MGlcDG	:	MonoGlucosylDiacylglycerol
MurG	:	*Pseudomonas aeruginosa* UDP-GlcNAc transferase
MWCO	:	Molecular weight cut-off
NST	:	*Neisseria meningitidis* α-2,3-sialyltransferase
PA	:	Phosphatidic acid
PG	:	Phosphatidylglycerol
PimA	:	*Mycobacterium smegmatis* Phosphatidyl-myo-inositol mannosyltransferase
PmST1	:	*Pasteurella multocida* α-2,3-sialyltransferase
QCM-D	:	Quartz crystal microbalance with dissipation
RB	:	Reference buffer
RcMGlcD	:	*Roseiflexus castenholzii* MonoGlucosylDiacylglycerol synthase
RLU	:	Relative light units
SDS-PAGE	:	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	:	Size exclusion chromatography
SLB	:	Supported lipid bilayer
SQDG	:	Sulfoquinovosyl diacylglycerol
StxB	:	Shiga toxin β subunit
SUV	:	Small unilamellar vesicles
TCEP	:	Tris(2-carboxyethyl)phosphine
TEV	:	Tobacco etch virus
TLC	:	Thin layer chromatography
TMB	:	3,3',5,5'-Tetramethylbenzidine
UDP-Gal	:	Uridine diphosphate galactose
UDP-Glc	:	Uridine diphosphate glucose
UDR	:	UDP detection reagent
WT	:	Wild-type

Chapter 1

Introduction

I. Introduction to the synBIOcarb project

The synBIOcarb European Training Network brings together a diverse team of academic and industrial experts from across Europe who are pioneering the development of synthetic glycobiology. Collectively, the teams aim to provide 15 early-stage researchers (ESRs) with advanced training in enabling technologies for the development and exploitation of glycoscience for diagnostics and targeted drug delivery.

In the context of the synBIOcarb project, glycosyltransferases (GTs) are of great interest for the building of sophisticated glycolandscapes on the surface of artificial cells or tissues. They offer the possibility to add one or several sugars at specific positions but also to introduce unnatural sugars (i.e. azido sugars) that could serve for biorthogonal chemistry (Figure 1). Research undertaken in this project from me (ESR7) uses GTs to expand the toolbox of synthetic glycobiology. A small set of GTs selected by the consortium will be produced in recombinant forms and functionally characterised. Following this, the proteins will be tested on protocells such as giant unilamellar vesicles (GUVs) and supported lipid bilayers (SLBs).

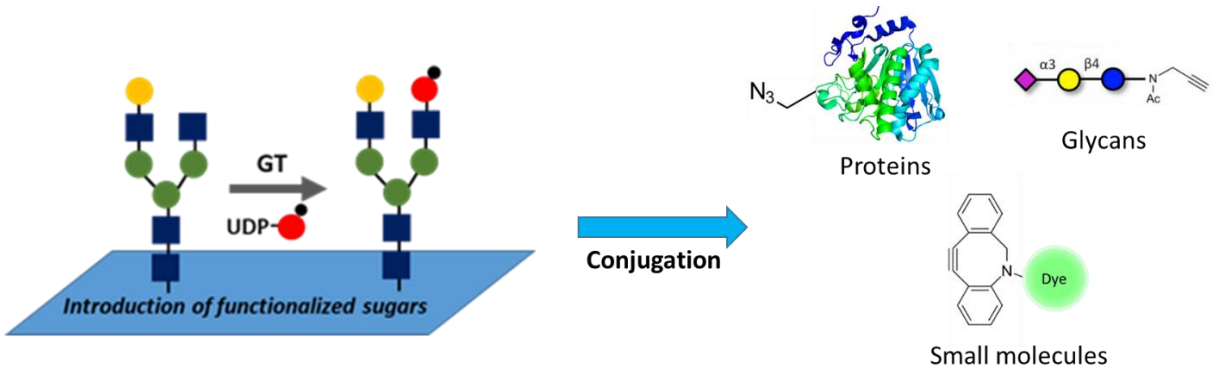


Figure 1: *Demonstrating the potential use of glycosyltransferases for the modification of solid surface-supported glycocalyx models with further application in conjugation of proteins, glycans and other small molecules.*

For this project two galactosyltransferases have been selected, MGD1 and LgtC. MGD1 is a monotopic membrane protein that adds a galactose residue from UDP-Gal to a lipid acceptor (diacylglycerol). It was chosen as a model protein for understanding substrate specificity,

enzyme mechanism and lipid membrane interaction. LgtC is a soluble bacterial GT that can add a galactose residue to lactose-containing acceptors. Its interest in the project was the possibility to generate on model membrane systems the Gb3 glycolipid antigen that is the receptor of the bacterial Shiga-toxin StxB. Overexpression of Gb3 is observed in a number of cancers and also in Fabry's disease, thus making the development of diagnostic and therapeutic strategies based on StxB attractive (Stimmer et al., 2014; Geyer et al., 2016; Kovbasnjuk et al., 2005; Aerts et al., 2008).

Within the synBIOcarb network, collaborations were established with two groups. From the University of Freiburg - BIOSS and Institute of Biology II, collaborations with Professor Winfried Roemer and Lina Siukstaite (ESR4) were established for their expertise in protocell synthesis and advanced microscopy techniques. From the University of Leeds, we established collaborations with Dr Ralf Richter and Chunyue Wang (ESR3) for biophysical studies on the interface of SLBs (supported lipid bilayers) using quartz crystal microbalance with dissipation (QCM-D). Secondments were also undertaken with the respective synBIOcarb partners from the University of Freiburg and University of Leeds in order to train in the respective techniques related to the collaborations established.

Finally, collaboration with Dr Olga Makshakova (Kazan Institute of Biochemistry and Biophysics) was essential for molecular modelling and simulations in this project.

II. State of the Art

1. The Glycocalyx

Centre to the synBIOcarb project is the glycocalyx, a forest of complex carbohydrates in the form of glycoproteins, glycolipids and proteoglycans that surrounds every cell (Figure 2). This sweet barrier between the cell and its environment extends several hundreds of nanometres in the extracellular matrix. This dense network of diverse glycans is synthesised by a group of enzymes called glycosyltransferases that reside primarily in the endoplasmic reticulum and Golgi apparatus. The glycocalyx regulates major biological events such as cellular protection, adhesion, signalling and differentiation (Varki et al, 2017; Zeng, 2017; McMorran et al., 2016; Mulivor and Lipowsky, 2002; Möckl, 2020). Molecular interactions between glycans and lectins (carbohydrate-binding proteins, CBPs) play central roles in these processes.

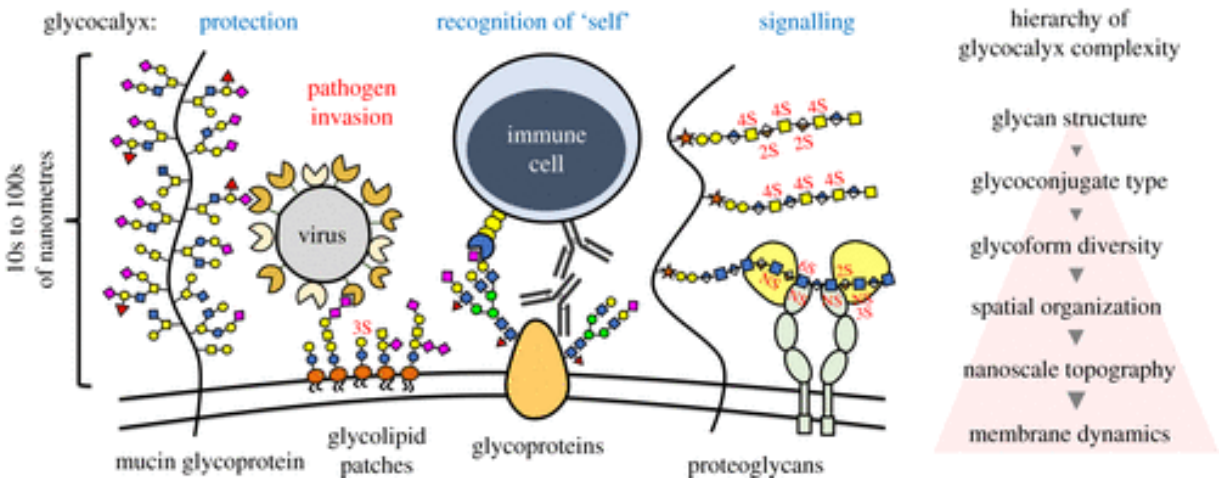


Figure 2: *Components and hierarchical complexity of the glycocalyx (from Purcell and Godula, 2019)*

The glycocalyx has multiple levels of complexity that determine its biological functions. From individual glycans providing affinity for carbohydrate binding proteins, to spatial and nanoscale organisation of glycans within glycoconjugates and across the glycocalyx which can determine the extent of activity and higher order interactions (Purcell and Godula, 2019). The

study of biological functions of cell surface glycans is also complicated by the microheterogeneity of glycoforms.

The glycolandscape present at the cell surface is ever evolving to protect from opportunistic pathogens (viruses, bacteria, parasites) that exploit cell surface glycans (Imberty and Varrot 2008; Castillo-Acosta et al., 2017; Formosa-Dague et al., 2018; Van Breedam et al., 2014). Changes in the glycocalyx must occur while maintaining its biological functions and identity ('self' signature), in order to prevent the cell from being rejected by the host immune response (Maverakis et al., 2015). Aberrant cell surface glycosylation is also a hallmark of severe pathologies such as cancer and genetic diseases arising from defects in the biosynthetic machinery (Pinho and Reis, 2015; Ng and Freeze, 2018). Glycan structures that are overexpressed, or uniquely present, on cancer cells are biomarkers that can be exploited for the development of new diagnostic methods and therapeutic strategies (Haga and Ueda, 2022; Diniz et al., 2022; Kailemia et al., 2017). Understanding the way cells use the glycocalyx to determine their biological functions is therefore of great interest, particularly in glycan based treatments, or also termed glycotherapy (Hudak and Bertozzi, 2014).

2. Advances in glycocalyx engineering and characterisation

The study of the glycocalyx requires characterisation of its components, their spatial organisation and their association with its biological functions. Altering glycan structures at the cell surface is a powerful approach to decipher glycan-mediated events. Some of the earliest approaches to the study of the glycocalyx were genetic techniques that targeted glycan biosynthesis machinery, chemoenzymatic methods that edited glycan structures, and the introduction of non-natural sugars by metabolic strategies (Mahal et al., 1997; Griffin and Hsieh-Wilson, 2016; Critcher et al., 2021).

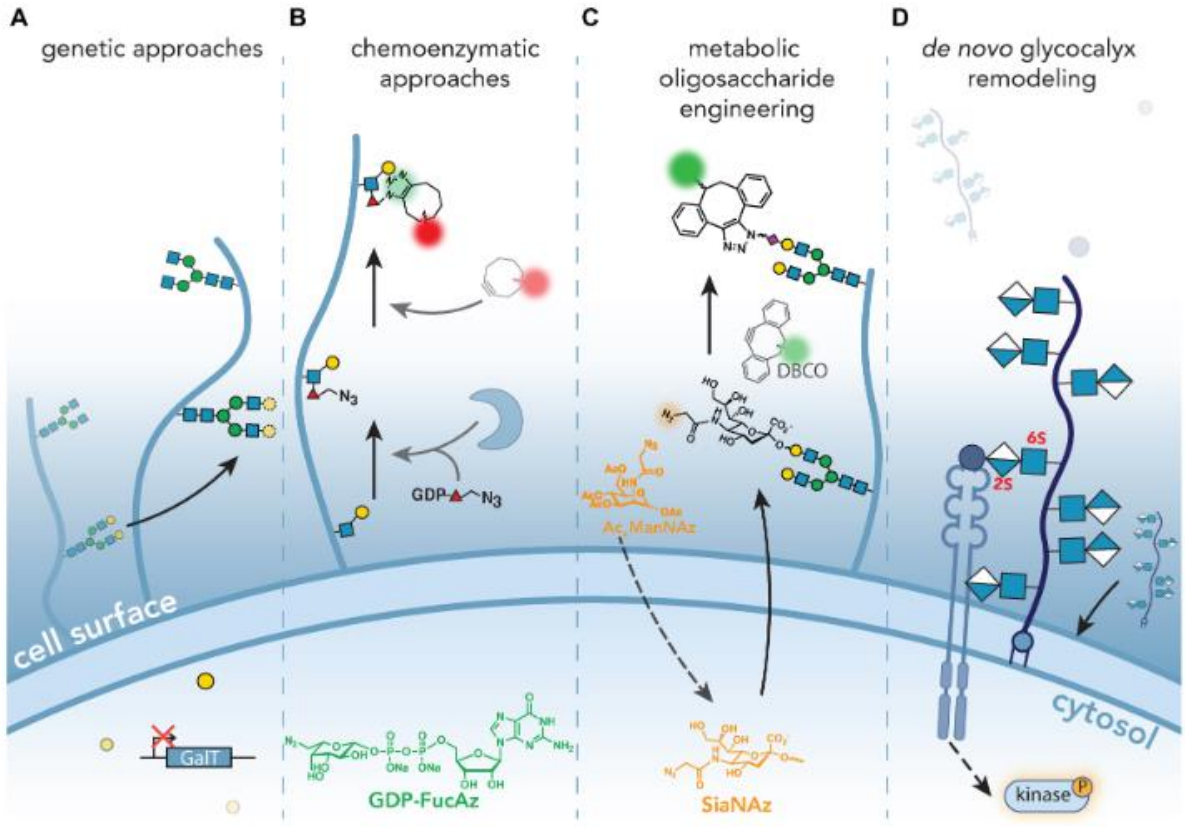


Figure 3: *Various approaches to modify the glycocalyx (from Critcher et al., 2021). The different approaches are genetic (A), chemoenzymatic (B), metabolic oligosaccharide engineering (C) and de novo glycocalyx remodelling (D).*

The targets of genetic approaches (Figure 3A) are mostly glycosyltransferases and glycan-modifying enzymes. The resulting mutant cell lines (i.e. CHO-Lec mutants with altered N-glycosylation from Pamela Stanley's group) have significantly advanced our understanding to assign glycan epitopes to specific CBP (Patnaik and Stanley, 2006). However, as the target enzymes typically operate on multiple substrates, it is difficult to study the impact of a modified glycan on a specific glycoprotein. In addition, perturbing glycan biosynthetic genes may have indirect, unexpected or undesirable effects such as protein unfolding or cellular lethality (Critcher et al. 2021). Modern genetic techniques (CRISPR/Cas9) have been recently used to build a panel of cell surface glycan arrays (Narimatsu et al. 2019; Qiu et al. 2018). This approach is likely to improve our understanding of the many biological roles of surface glycans.

Chemoenzymatic methods (Figure 3B) mostly rely on the use of exogenously added carbohydrate-active enzymes (i.e. GTs, glycosidases) to modify cell surface glycans. This

approach also offer the possibility to introduce unnatural functional groups (i.e. azido-sugars) that can be used to specifically label glycans or to overcome glycan microheterogeneity (Chaubard et al., 2012; Hong et al., 2019; Tang et al., 2020).

Metabolic oligosaccharide engineering (MOE) involves the use of modified analogues of monosaccharides that are taken up and metabolized by cells and finally incorporated into cell surface glycoconjugates (Figure 3C). These monosaccharides derivatives can then be functionalised using biorthogonal chemistry with a molecular probe or biomolecules (Mahal et al., 1997, Saxon and Bertozzi 2000). This approach has found many applications, notably in imaging and cell targeting, and greatly facilitated the *in vivo* study of glycans (Li N et al., 2019; Laughlin et al., 2008; Wu et al., 2018).

The glycocalyx can also be engineered for the display of large glycoconjugates using lipid-functionalised anchors that can insert into cell membranes (Figure 3D). This approach has been extensively used to introduce glycosaminoglycans (GAG) onto cell surfaces to investigate protein-glycan interactions. This is a powerful strategy to influence cellular signalling and determine their downstream effects (Pulsipher et al., 2015; Huang et al., 2018). Similarly, the introduction of mucin mimetics into cell membranes was helpful in understanding their roles in body's defence against pathogens (Wheeler et al., 2019).

However, although powerful and transformative these different strategies provide no control over cell surface presentation or nanoscale organisation, often resulting in perturbation of the glycocalyx.

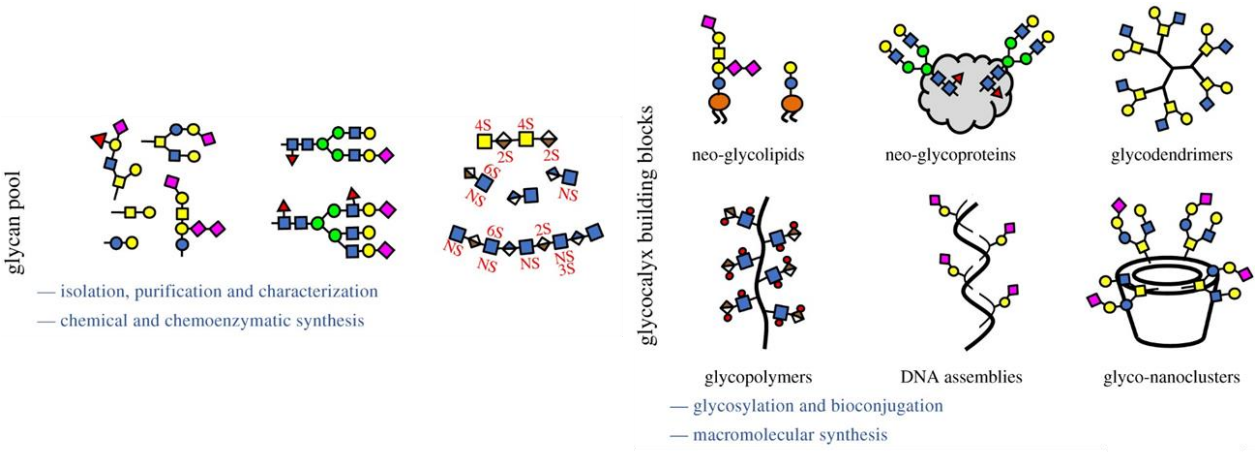


Figure 4: *Glycans and synthetic glycomaterials that mimic components of the glycocalyx (from Purcell and Godula, 2019).*

Other synthetic glycobiology approaches have been developed. A vast array of glycans, glycoconjugates and synthetic glycomaterials have been synthesized to mimic constituents of the glycocalyx (Figure 4). Initially these were used as soluble ligands to probe interactions with CBP. Nevertheless, this does not reflect the behaviour of glycans on cell surfaces. For example, glycoconjugates on surfaces can reorganise to enhance multivalency and hence avidity, alter molecular dynamics, as well as promote higher-order binding interactions (Purcell and Godula, 2019; Migliorini et al., 2014). Glycocalyx models such as solid-surface and lipid-bilayer supported models and even modified live cells, have been developed to elucidate the more intricate interactions observed in the glycocalyx (Figure 5). This approach, sometimes referred to as "precision glycocalyx editing", is becoming increasingly used to answer fundamental questions in glycobiology (Purcell and Godula 2019), and is the centre of the synBIOcarb European project.

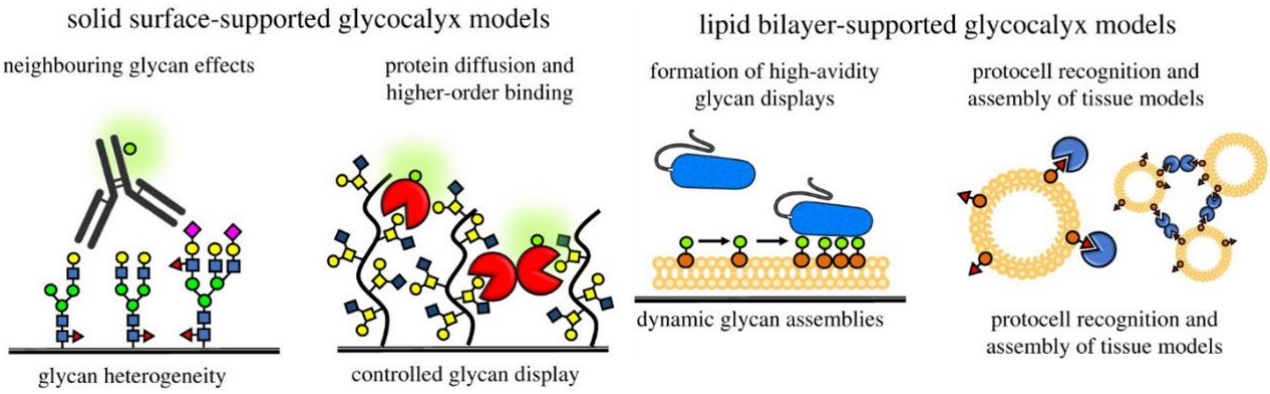


Figure 5: *Glycocalyx models displaying the effects of spatial organisation of glycoconjugates (from Purcell and Godula, 2019).*

The most widely used application of solid-surface models are glycan arrays, where surface immobilised glycans and synthetic glycoconjugates are used in high-throughput profiling of protein interactions (Oyelaran et al., 2009; Blixt et al., 2004; Parera Pera et al., 2010). Despite their utility, glycans immobilized onto solid supports fail to capture the dynamic nature of the glycocalyx.

On the other hand, membrane models more closely resemble the glycocalyx, providing conditions such as curvature and dynamic features, for example the movement of glycoconjugates across the membrane. The use of artificial systems in synthetic glycobiology has grown in recent years. The supported lipid bilayer (SLB) (Figure 5) is an appropriate

experimental model to study cellular processes associated with the glycocalyx, as they provide insight into the effects of glycan density and clustering for optimum presentation to CBP (Richter et al., 2006: Migliorini et al., 2014). The building of complex and well-defined glycocalyx into artificial membranes is still challenging. Lipid vesicles such as liposomes or giant unilamellar vesicles (GUVs) can be used in a bottom-up approach to build protocells of increased complexity (Figure 6). These minimal membrane models have been used to investigate glycolipid-mediated events such as endocytosis, cell adhesion and membrane fusion (Omidvar and Römer, 2019). Assembly of protocells to form prototissues has been obtained by combining glycan-decorated vesicles and multivalent lectins that act as cross-linking agents (Villringer et al., 2018; Ribeiro et al., 2018). This represents an important step towards synthetic minimal tissues.

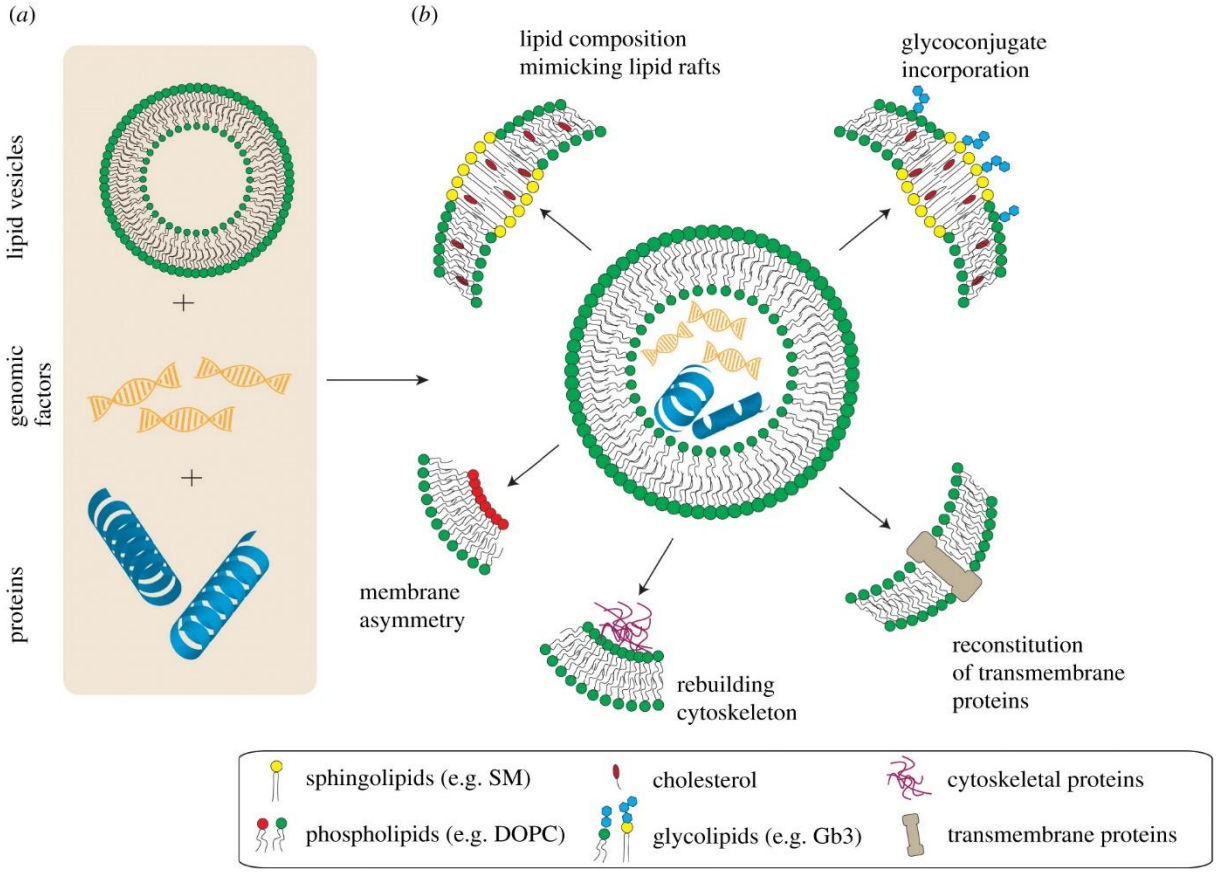


Figure 6: *Bottom-up approach to build a protocell from GUVs. Increasing protocell complexity achieved by incorporation of different lipids, glycoconjugates, proteins and cytoskeleton (from Omidvar and Römer 2019).*

Of interest for synBIOcarb and the study of the glycocalyx is the synthesis of specific glycan structures on the surface of synthetic membrane models. This approach has multiple advantages such as the synthesis of specific glycan epitopes and glycan structures that either are not commercially available or does not exist in nature.

Key to the engineering of glycocalyx components are the family of enzymes called glycosyltransferases which catalyse the synthesis of complex glycan structures. In this context, GTs have enormous potential as tools for building glycan structures on the surface of membrane models.

III. The large glycosyltransferase family

Glycosyltransferases (GTs) are a large group of enzymes responsible for the biosynthesis of glycoconjugates, disaccharides and polysaccharides that are found in nature. They are an ancient family of enzymes found in all domains of life. GTs catalyse the transfer of a sugar moiety from an activated donor to a specific acceptor. There is a large variety of biomolecules that can act as acceptors such as proteins, lipids, DNA, mono/oligosaccharides and a myriad of small molecules. For sugar donors, a vast majority (~90%) of GTs use nucleotide sugar donors, but some GTs can also use lipid-phosphate-sugar donors (Figure 7).

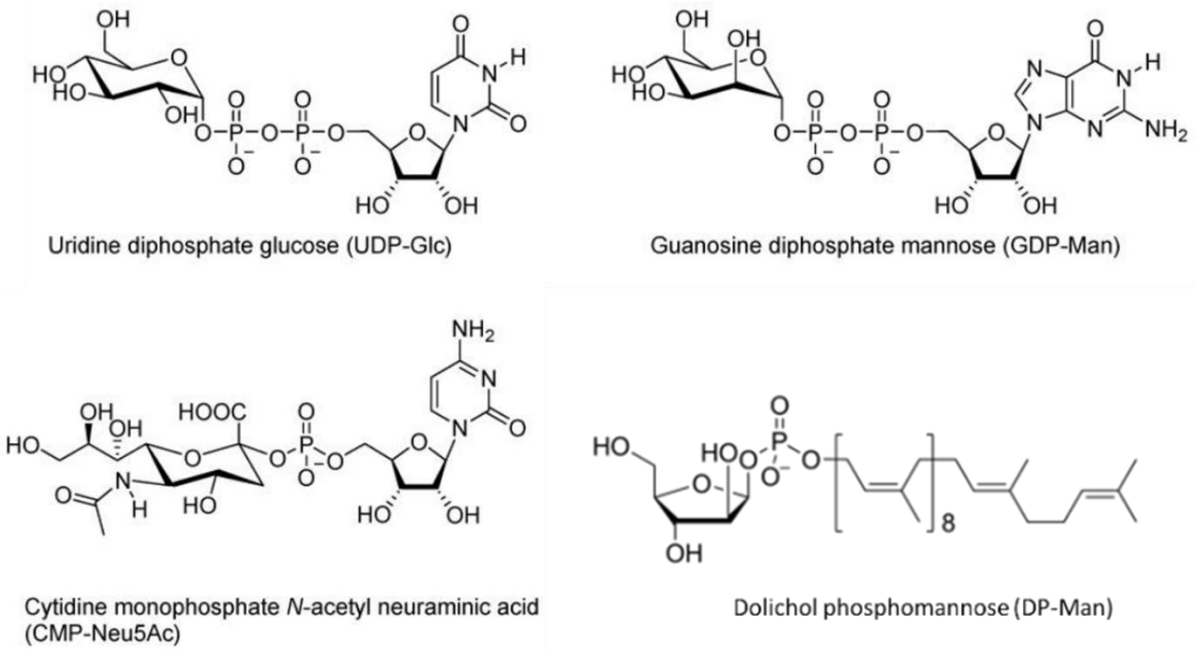


Figure 7: *Different types of activated sugar donors.*

Currently there are approximately more than 900,000 GT sequences identified and classified into 114 different families in the Carbohydrate-Active Enzymes (CAZy) database (http://www.cazy.org). CAZY classification into GT families is based on sequence similarity. New families are created when GTs with proven activities are discovered but that cannot be assigned to an existing GT family, as in the case of the β-Kdo glycosyltransferase from which the GT99 family originated (Ovchinnikova et al., 2016). The human genome contains over 200 genes coding for GTs distributed into 46 distinct GT families.

Another criterion for classification is the processive or non-processive characteristic of the GT, this is the ability of the enzyme to add only one sugar residue or to polymerise consecutive repeats of glycans.

Although a large number of GTs have been identified, less than 1% are biochemically characterised and even less with solved structures. Using sequence comparison from the same CAZY family shows highly conserved peptide motifs which are likely to be involved in structure and catalytic activity (Coutinho et al., 2003). Generally, GTs of the same family have the same overall protein fold and mechanism of action. Substrate specificity, however, is much more difficult to predict and often GTs of the same family with closely related sequences can have different donor and/or acceptor specificities. Examples of this can be seen in Blood Group A and B transferases, from the GT6 family, where a difference of only 4 amino acids results in different sugar donor specificities (GalNAc versus Gal) (Yamamoto et al., 1990). Determining the acceptor and sugar donor can be challenging, often requiring experimental characterisation of individual GTs, and this is the reason for the small number of characterised GTs compared to glycoside hydrolases.

1. Structures of Glycosyltransferases

To date, GTs discovered and structurally solved that use nucleotide sugars as donors are of the GT-A and GT-B fold (or variants of these two folds) (Figure 8). Currently, around 40 GT families can be classified as GT-A fold and 36 as GT-B fold on the basis of solved structures or structure prediction and these two folds are also the focus of the PhD project (Breton et al., 2012; Taujale et al., 2021; Moremen and Haltiwanger, 2019).

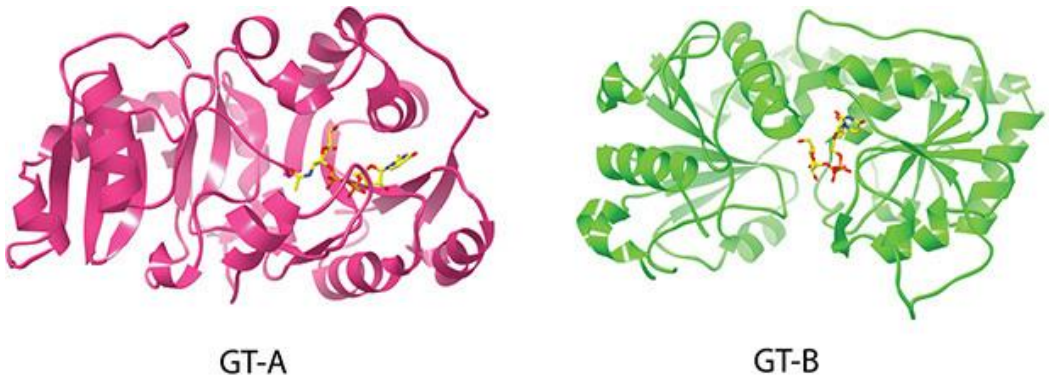


Figure 8: *Representative 3D structures of GT-A and GT-B glycosyltransferases corresponding to β1-2 N-acetylglucosaminyltransferase I (PDB ID 1FOA) and T4 phage β-glucosyltransferase (PDB ID 1j39) (from Varki et al., 2017).*

GT-A and GT-B enzymes comprise a Rossmann-like fold, an α/β/α sandwich, which is typically found in proteins that bind nucleotides. With a few exceptions, most GT-A enzymes are dependent on a divalent cation for activity and the catalytic domain contains a 'DxD' motif that interacts with the donor phosphate group through coordination with the metal-ion (Breton et al. 2006). In the few metal-ion independent GT-A enzymes identified, Lys and Arg side chains are utilised for phosphate interactions (Pak et al, 2011).

GT-B enzymes possess two distinct Rossmann domains connected by a linker region with the catalytic site in the cleft between the two domains. Unlike GT-A enzymes they are metal-ion independent and do not contain a DxD motif, instead employing Lys and Arg residues similarly to metal-ion independent GT-A (Breton et al., 2006; Li et al., 2017).

Completely different folds have been described for GTs that utilize lipid-phosphate donor substrates (Breton et al., 2012; Henrissat et al., 2008). These are integral membrane proteins, sometimes improperly referred as GT-C fold enzymes (Bohl et al., 2021).

2. Topology of glycosyltransferases

Most GTs that synthesize the diversity of glycoconjugates found in the glycocalyx reside in the endoplasmic reticulum and Golgi apparatus. Golgi-resident GTs are usually type II transmembrane proteins, consisting of a short N-terminal cytoplasmic domain followed by a transmembrane domain, a stem region of variable length and a large globular catalytic domain facing the luminal side. Proteolysis in the stem region can generate a soluble catalytic domain (Breton et al, 2001). Other membrane protein architectures have been described for GTs, including type I and integral transmembrane proteins (Magdalou et al., 2010; Oikawa et al., 2013). The topology of GTs may be more complex with the presence of add-on domains (Breton et al., 2012). These additional domains can confer new functionalities; i.e. mediate substrate recognition or protein-protein interactions (Qasba and Ramakrishnan, 2007; Lazarus et al., 2011). Some glycosylation reactions that occur in the cytosol are mediated by soluble GTs. This is for example the case for the numerous plant GTs involved in the glycosylation of secondary metabolites (Lim and Bowles, 2004).

In the context of the PhD project, we have a particular interest for membrane-associated GTs that are either 'peripheral' or 'integral' depending on the level of binding to the membrane (Albesa-Jové et al., 2014) (Figure 9). Peripheral GTs weakly bind to the lipid membrane or other membrane proteins through noncovalent interactions such as hydrogen bonding and electrostatic forces. An example of a peripheral GT is phosphatidyl-myo-inositol mannosyltransferase (PimA) from *Mycobacterium smegmatis* (Guerin et al., 2007). Integral GTs bind strongly and permanently to the membrane. They can be further classified as monotopic, bitopic or polytopic. Monotopic proteins bind only to one leaflet in the bilayer. For example, the monogalactosydiacylglycerol transferase (MGD1) from *Arabidopsis thaliana* is a monotopic protein that was shown to localize to the outer leaflet of the chloroplast inner envelope membrane (Miège et al., 1999; Rocha et al., 2018). Bitopic and polytopic proteins cross the bilayer by one or more transmembrane domains, such as the bitopic sialyltransferase (PmST1) from *Asteurella multocida* (Ni et al., 2006)

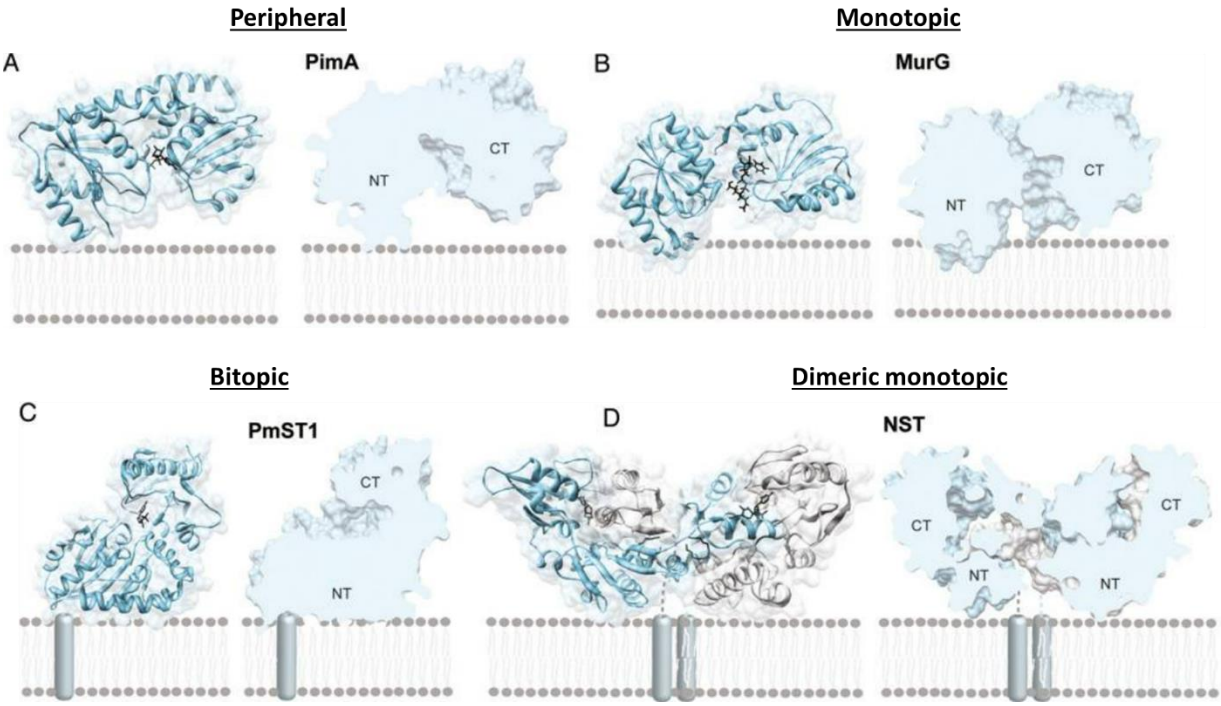


Figure 9: *Examples of membrane associated GT-B proteins. A) Peripheral GT, phosphatidyl-myo-inositol mannosyltransferase (PimA), from Mycobacterium smegmatis (Guerin et al., 2007). B) Monotopic GT, MurG, from Pseudomonas aeruginosa (Hu et al., 2003). C) Bitopic GT, sialyltransferase (PmST1), from Pasteurella multocida (Ni et al., 2006). D) Dimeric monotopic GT, α-2,3-sialyltransferase (NST) from Neisseria meningitidis (Lin et al., 2011). Figure from Albesa-Jové et al. (2014).*

3. Catalytic mechanisms of glycosyltransferases

Glycosyltransferases can be classified into 'inverting' or 'retaining' enzymes depending on the stereochemical configuration of the donor sugar glycosidic bond, when transferred (Figure 10). In retaining reactions the stereochemistry of the anomeric linkage is conserved after transfer while it is reversed in inverting reactions. This capacity for stereochemical inversion or retention leads to different catalytic mechanisms.

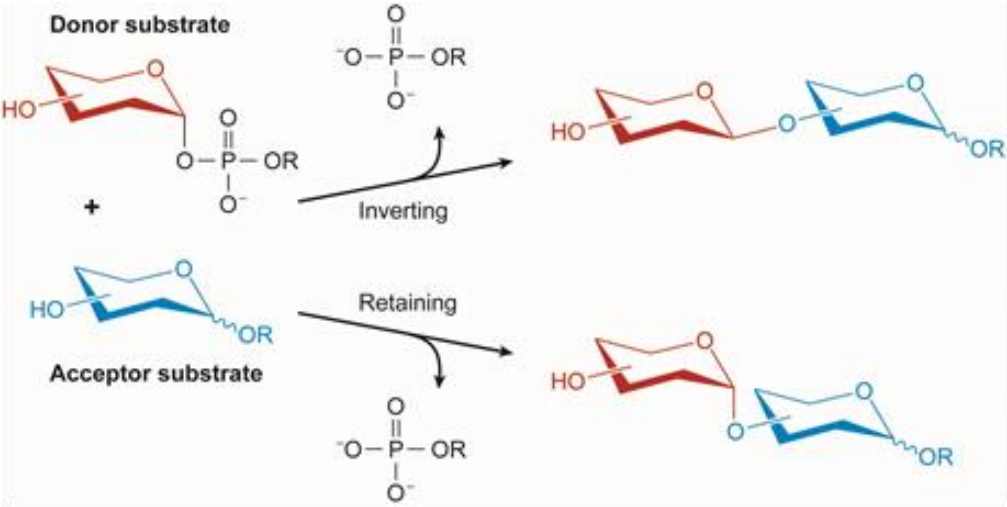


Figure 10: *Representation of outcomes from retaining and inverting reactions (Lairson et al., 2008).*

a) Inverting mechanisms

Inverting GTs primarily use a substitution nucleophilic bimolecular (S_N2) reaction mechanism with the formation of an oxocarbenium-ion transition state (Figure 11). This reaction uses an enzyme amino acid as a catalytic base (typically Glu, Asp or His) to deprotonate the acceptor nucleophilic hydroxyl group which then attacks the anomeric carbon of the donor sugar from the opposite side. The leaving group (nucleotide donor) is displaced, and its negative charge is stabilised by either a metal ion, positive amino acids or helix dipole, thus favouring its departure. The reaction results in the inversion of the anomeric carbon configuration (Breton et al, 2012; Lairson et al., 2008; Moremen and Haltiwanger, 2019).

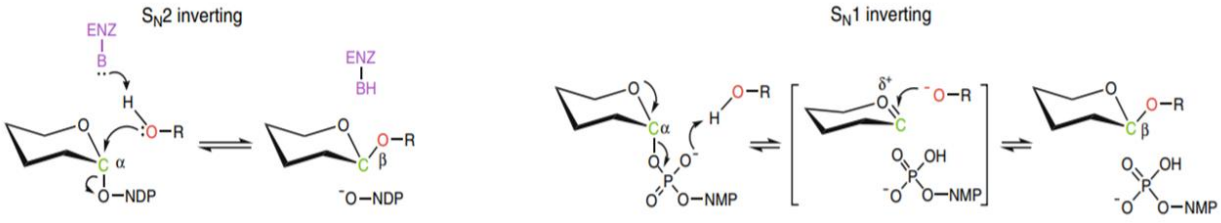


Figure 11: *Catalytic mechanisms for inverting glycosyltransferases* (from Moremen and Haltiwanger, 2019).

For several inverting GTs such as POFUT1 (a protein-O-fucosyltransferase), where no catalytic base could be identified, a S_N1-like mechanism has been proposed where the oxygen atom

from the donor phosphate deprotonates the acceptor hydroxyl, priming it for nucleophilic attack of the donor sugar anomeric carbon. In the case of POFUT1 a water molecule can promote catalysis by acting as a proton relay (Lira-Navarrete et al., 2011).

b) Retaining mechanisms

The reaction mechanism for retaining GTs is not as well understood and has been debated for years. Two mechanisms have been proposed, a double-displacement and S_Ni-like (internal return) mechanism (Figure 12). The double-displacement includes two S_N2 reactions with the formation of a transient covalent intermediate. The first reaction is the nucleophilic attack of the sugar donor anomeric carbon resulting in the formation of a covalent glycosyl-enzyme intermediate. In the second reaction, the glycosyl-enzyme intermediate is attacked by the acceptor hydroxyl group from the same side as the leaving nucleotide. This reaction is catalysed by the deprotonation of the acceptor hydroxyl by either an enzymatic base or the leaving nucleotide. This results in the overall retention of the anomeric configuration (Lairson et al, 2008). However, there is no convincing evidence for the existence of a covalent intermediate to date.

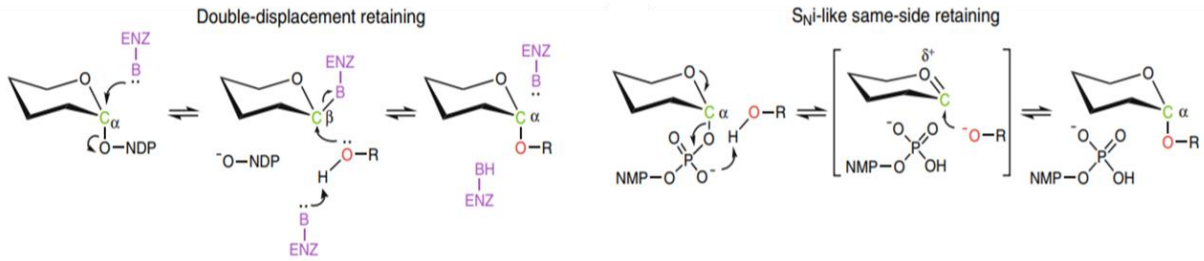


Figure 12: *Catalytic mechanisms for retaining glycosyltransferases (Moremen and Haltiwanger, 2019).*

In the S_Ni-like mechanism the acceptor hydroxyl group is deprotonated by an oxygen atom in the donor phosphate group. The nucleophilic hydroxyl then attacks the anomeric carbon of the donor sugar from the same side as the leaving group phosphate, thereby retaining the stereochemistry of the anomeric carbon (Moremen and Haltiwanger, 2019).

4. Applications of glycosyltransferases

a) Protein glycosylation

One of the most widely used applications of GTs is the glycosylation of proteins and small molecules. It is known that adding or changing the type of sugar bound to molecules can significantly influence their biological properties. In the case of pharmaceutical proteins, it has been shown that glycosylation can enhance their overall stability by protecting them from proteolytic degradation, aggregation, chemical and thermal denaturation, oxidation, chemical crosslinking, and it also increases their solubility (Sola and Griebenow, 2009).

More than 70% of therapeutic proteins are glycosylated and how they are glycosylated can profoundly affect their physical and biological properties and hence their therapeutic efficacy. This is particularly true for the production of therapeutic monoclonal antibodies (Wang et al., 2019) (Figure 13).

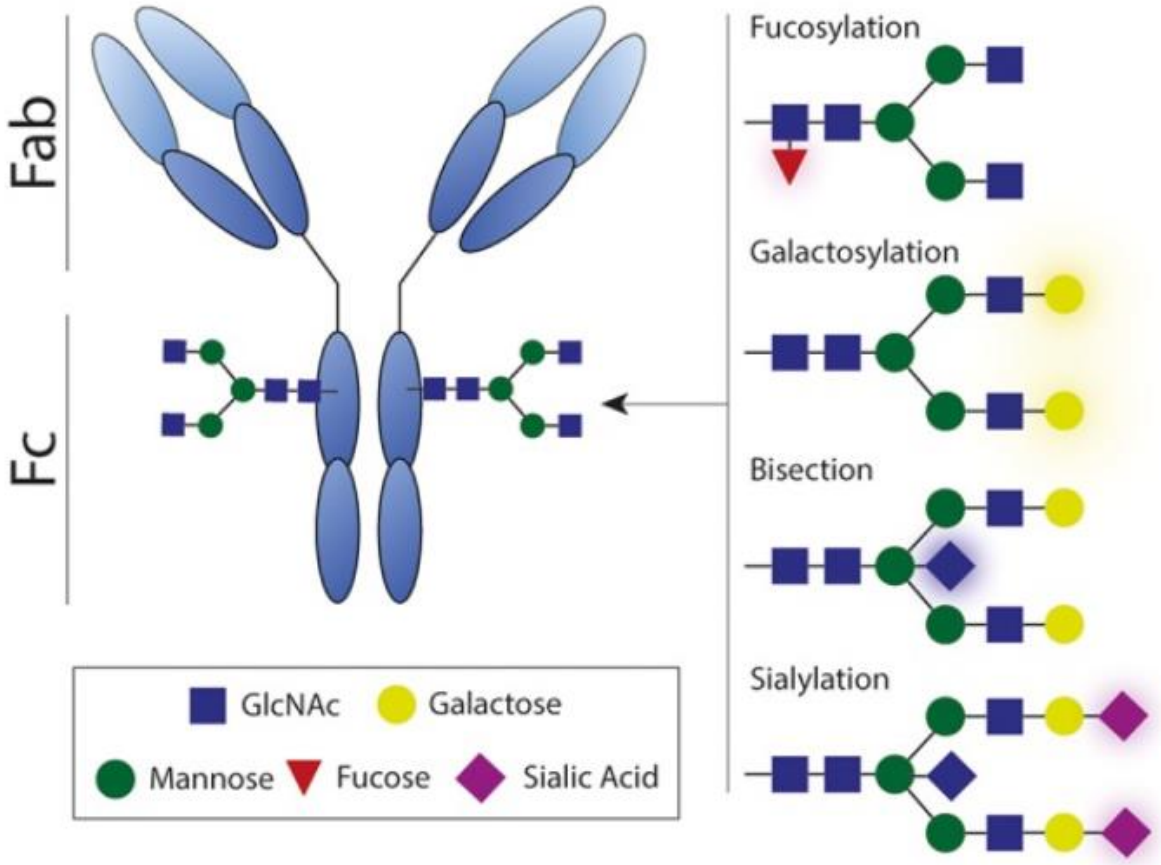


Figure 13: *Antibody with variable glycosylation in the Fc domain (from Van Erp et al., 2019).*

Commonly, therapeutic proteins that require glycosylation are produced in mammalian cell lines, such as Chinese Hamster Ovary (CHO), Human Embryonic Kidney (HEK) cells. Other hosts such as insect cell lines, yeast and plants are also used when the glycans do not need to be human-type. A lot of effort has been made in the last decade to glycoengineer these non-mammalian expression systems to produce homogenous and 'humanized' glycoproteins (van Landuyt et al., 2019; Hamilton et al., 2006; Harrison et al., 2006; Kallolimath et al., 2016; Dicker and Strasser, 2015, Margolin et al. 2020; Gupta and Shukla, 2018, Kightlinger et al., 2020) (Figure 14). This has been mainly achieved by modifying the biosynthetic machinery of the host (i.e. knocking out unwanted GT activities and/or introducing the missing GT activities) of or through metabolic engineering.

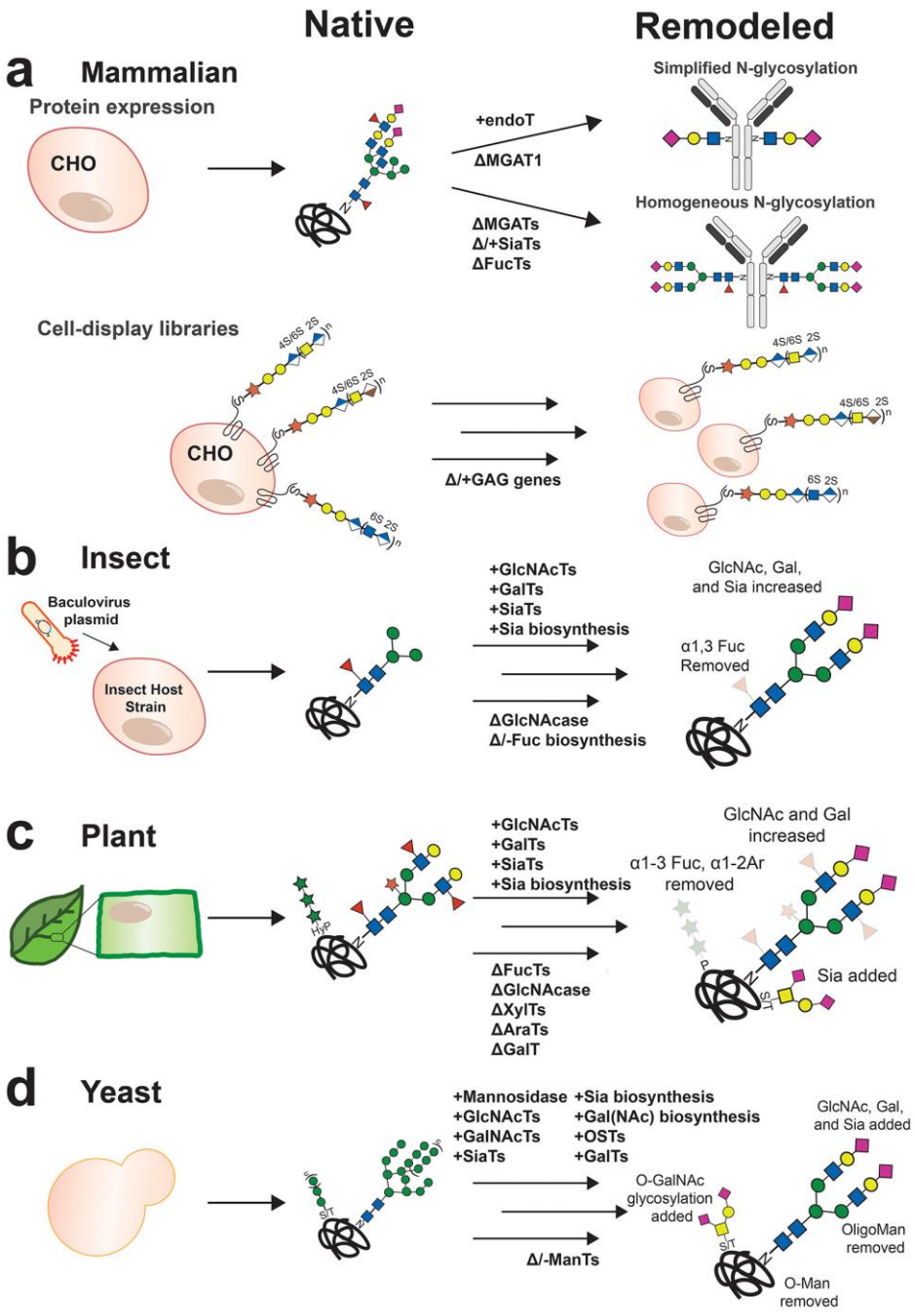


Figure 14: *Protein expression hosts (mammalian, insect, plant and yeast) and how synthetic glycosylation systems have been engineered by remodelling their naturally occurring glycosylation pathways (from Kightlinger et al., 2020).*

Moreover, there have been advances in the production of therapeutic proteins on bacterial expression platforms (Keys and Aebi, 2017; Tytgat et al., 2019; Du et al., 2019) and also in cell-free glycosylation systems, where the use of purified GTs or co-expression of GTs with the target molecule allow for glycosylation from *E. coli* extract (Kightlinger et al., 2019; Kightlingher et al., 2020) (Figure 15).

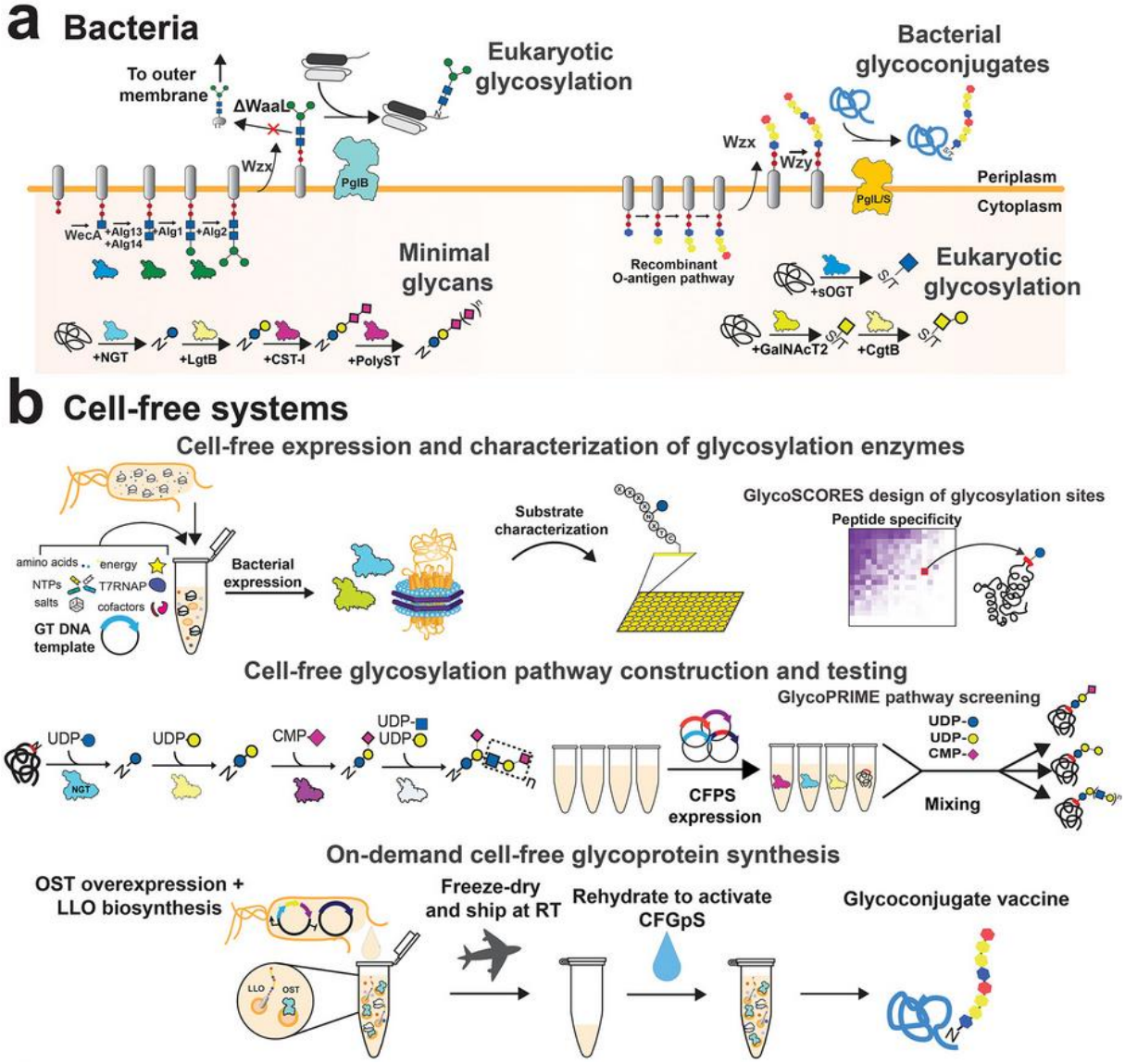


Figure 15: *Advances in bacterial glycosylation systems. a) Engineered eukaryotic glycosylation systems in E. coli bacteria via expression of eukaryotic GTs. b) Cell free expression platform utilising the bacterial lysate and GTs, purified or co-expressed, to construct a glycosylation pathway (from Kightlinger et al., 2020).*

A recent application of protein glycosylation engineered the monoclonal antibody, Trastuzumab, with enhanced N-glycosylation on the Fab domain, via insertion of glycosylation sites, which led to enhanced protein stability and resistance to protein aggregation (Cruz et al., 2021).

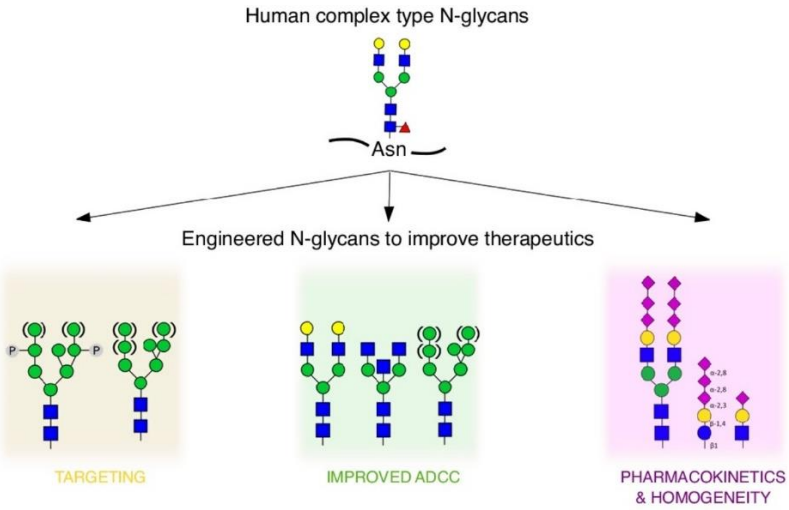


Figure 16: *Pharmaceutical effects of engineered N-glycans on therapeutic proteins. Specific glycosylation such as mannose-6-phosphate can improve cellular or subcellular targeting. Glycosylation can also improve Antibody-Dependent Cell Cytotoxicity (ADCC) and enhance pharmacokinetic properties, for example polysialylation reduces renal clearance and increases half-life (from Van Landuyt et al., 2019).*

Glycans can also be used for targeting therapeutic proteins to their site of action. For example, in enzyme replacement therapy (ERT) of lysosomal storage diseases, an enzyme with mannose-6-phosphate enhances its cellular uptake and targeting for lysosomal delivery (Kang et al., 2018). An example of this is Fabrazyme®, an ERT for Fabry disease, it contained mannose-6-phosphate from mammalian expression which targeted the enzyme to the lysosome (Prabakaran et al., 2012).

b) Small molecule glycosylation

As with proteins, glycosylation can improve chemical stability, solubility and polarity of drugs, increasing their concentrations in solution (Figure 17). The addition of the sugar moiety can facilitate the formation of hydrogen bonds that are vital for the specific recognition by their targets. Thus, in natural products that contain glycan moieties, the removal of the glycan can significantly diminish or kill bioactivity (He and Liu, 2002). For example, in the antimicrobial drug Clindamycin, the 3 hydroxyl groups in the sugar moiety form important hydrogen bonds with peptidyl transferase ligands that are important for its bioactivity (Schlünzen et al., 2001).

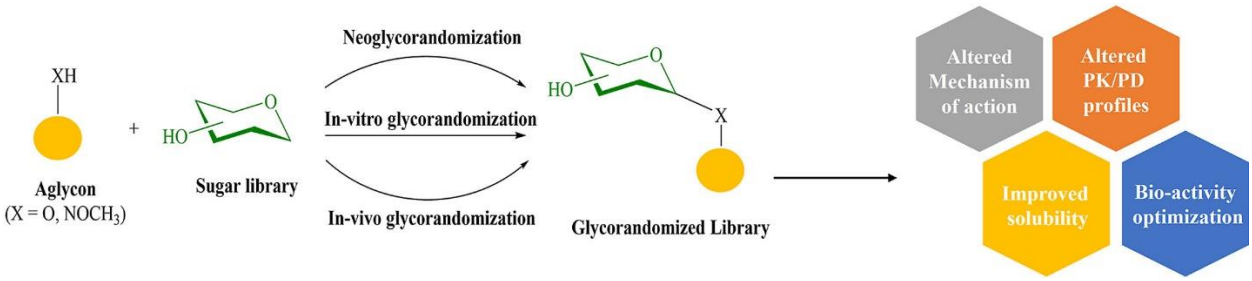


Figure 17: *Different pathways for glycorandomization and the effects of glycosylation on therapeutic drugs (from Goel et al., 2021)*

A significant use of GTs is in the area of drug discovery and drug development. A property of some GTs that produce natural products (usually small molecules) are their substrate promiscuity, either for the sugar donor or acceptor. This can be exploited in order to glycosylate a number of potential drugs with a variety of sugars. In a process called glycorandomization, small molecules, drug leads and existing drugs are rapidly diversified through the conjugation of glycan molecules. This process can lead to the discovery of new glycomolecule analogs with enhanced efficacy, specificity and pharmacokinetic/pharmacodynamic profiles (Griffith et al., 2005).

c) Glycosylation of membrane surfaces

An interesting application of recombinant GTs is in a technique called chemoenzymatic glycan modification, which has been used as a tool to edit glycan structures in a cellular environment (Hong et al., 2019: Hong et al., 2020). As with glycorandomization, the substrate promiscuity of certain GTs is exploited in order to add natural or unnatural sugar donors. In an approach called in situ glycan editing, recombinant GTs were used to incorporate monosaccharides to cell receptor glycans, as a way to probe their roles as modulators of cell signalling (Jiang et al., 2018) (Figure 18).

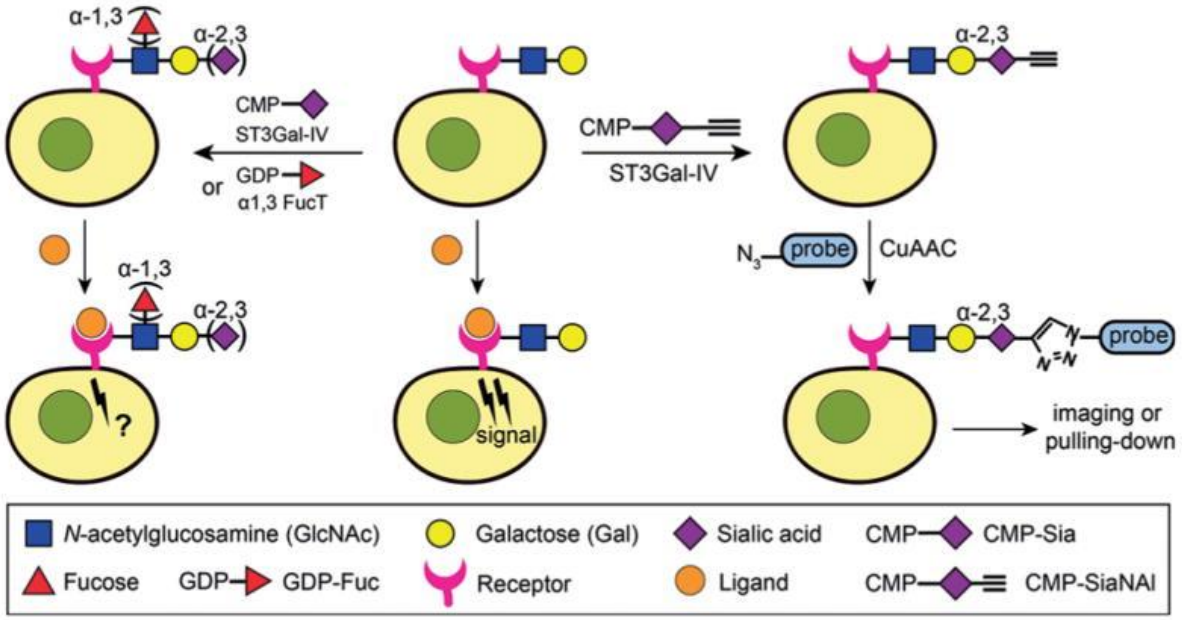


Figure 18: *In-situ glycan editing of a cell receptor glycan. Recombinant GTs and sugar donors incubated with the cell to add specific glycans to the cell surface. Promiscuous GTs were also used to add functionalised sugars for biorthogonal conjugation of a probe (from Jiang et al., 2018).*

Another example utilised multiple recombinant GTs, with relaxed sugar donor specificities, in order to develop a live cell-based assay for probing glycan-mediated influenza A virus infection (Hong et al., 2019).

Regarding the modification of synthetic membrane systems, there are not many examples in the literature. One example utilised β(1,4)-galactosyltransferase (β4Gal-T1) to glycosylate GUVs and LUVs composed of N-Acetylglucosaminolipids (Figure 19). This study highlighted the ability of some GTs to have increased activities in lipid micro-domains (Noble et al., 2012) (Figure 19).

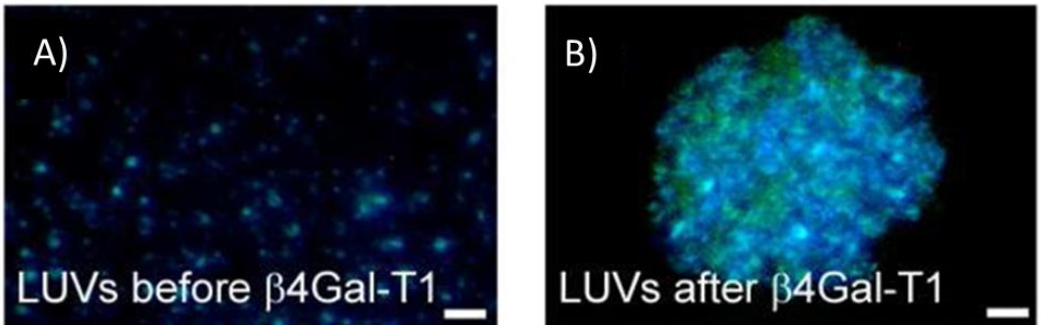


Figure 19: *Epi-fluoresence micrographs of LUVs incubated with fluorescently-labelled Erythrina cristagalli lectin (ECL). A) Dispersed vesicles before incubation with β4Gal-T1/UDP-Gal. B) Aggregated vesicles from lectin binding after β4Gal-T1/UDP-Gal incubation (from Noble et al., 2012).*

Another example utilised α-1,3-mannosyltransferase (GumH) to glycosylate a supported lipid bilayer and monitored real time glycosylation by QCM-D (Quartz-Crystal Microbalance with dissipation monitoring) (Figure 20). However, this study used immobilized alkanethiol self-assembled monolayers (SAM) as the inner leaflet attached to the solid support, while only the outer leaflet was a monolayer of glycolipids. Although not a true lipid bilayer, it nevertheless demonstrates that it is possible to glycosylate the surface of a lipid monolayer (Alves et al., 2011).

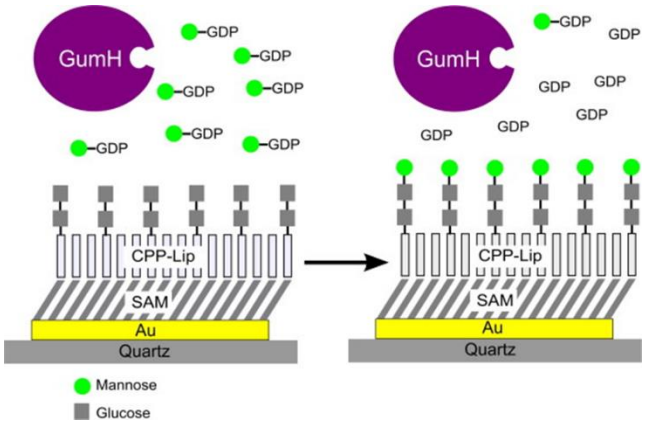


Figure 20: *Enzymatic glycosylation of a supported lipid bilayer by GumH (from Alves et al., 2011).*

These examples illustrate potential methods by which the activity of GTs on the surface of membranes can be characterized.

IV. Selected glycosyltransferases in the PhD study

1. The MonoGalactosylDiacylglycerol synthase (MGD1)

MGD1, from the plant *Arabidopsis thaliana*, is a GT28 family protein that transfers a galactose from UDP-*α*-galactose (UDP-Gal) to diacylglycerol (DAG) to form the galactolipid monogalactosyldiacylglycerol (MGDG) (Figure 21). Its role in the synthesis of the most abundant galactolipid in the thylakoid biosynthesis pathway makes it a very important enzyme in the formation of photosynthetic membranes (Jarvis et al., 2000).

MGD1 is nuclear-encoded and targeted to the inner envelope membrane of chloroplasts (Miège et al., 1999). Its primary sequence comprises 533 amino acids with a chloroplast transit peptide at its N-terminus (aa 1-106), followed by a short peptide region that is predicted to be disordered (aa 107-136) and a large catalytic domain (aa 137-153). MGD1 is a monotopic membrane protein that must bind for catalysis the water-soluble nucleotide sugar donor and the hydrophobic DAG acceptor. Its membrane binding properties are therefore crucial for optimal catalysis. Major advance was obtained in the production and purification of the catalytic domain of MGD1 expressed in *E. coli*. The recombinant protein is produced as a fully active and soluble monomeric form that retains its membrane binding properties (Rocha et al., 2013; Sarkis et al., 2014).

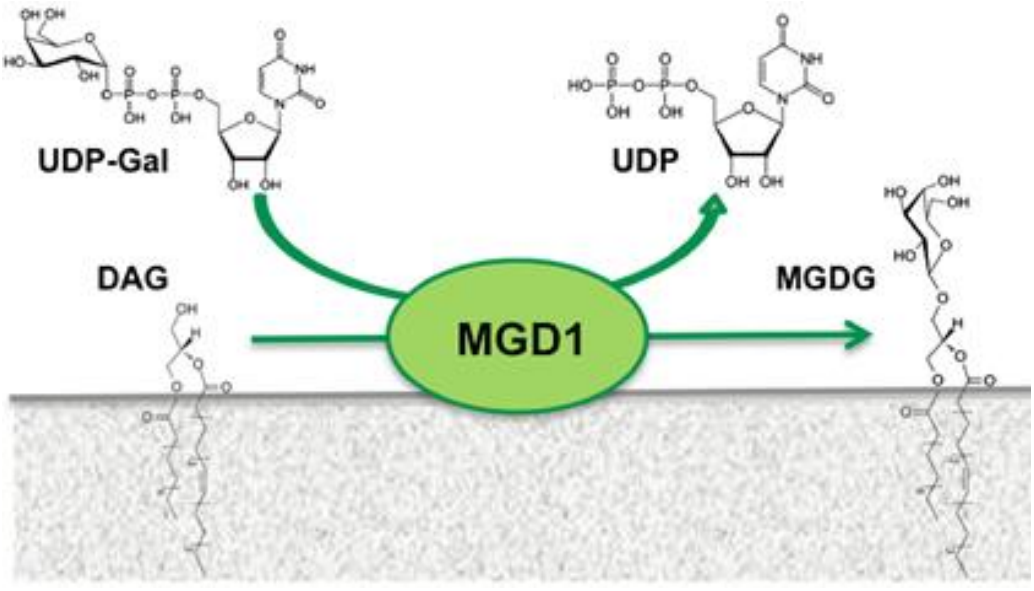


Figure 21: *Diagram of MGD1 activity (from Rocha et al., 2016b).*

a) Relevance in photosynthetic activity and membrane development

Non-phosphorus galactoglycerolipids, also called galactolipids, are the most abundant class of lipids on earth. The uncharged galactolipids, MGDG and digalactosyldiacylglycerol (DGDG), form up to 80% of the lipid content of thylakoid membranes (Table 1). The remaining lipids mainly include the anionic lipids phosphatidylglycerol (PG) and sulfoquinovosyldiacylglycerol (SQDG) (Block et al., 2007). The galactolipid composition has been remarkably conserved throughout evolution in photosynthetic organisms from cyanobacteria to plants (Petroutsos et al., 2014).

		MGDG	DGDG	SQDG	PG	PI	PC	PE	Others
Chloroplast	Thylakoids	53	27	7	7	2	0	0	4
	Inner envelope	49	30	5	8	1	6	0	1
	Outer envelope	17	29	6	10	5	32	0	1
Nongreen plastid		31.5	27.5	6	9	4.5	20	1	0.5
Synechocystis sp. PCC 6803		54	18	15	13	0	0	0	0

Table 1: *Membrane lipid composition of plant plastids and cyanobacteria (Kobayashi, 2016).*

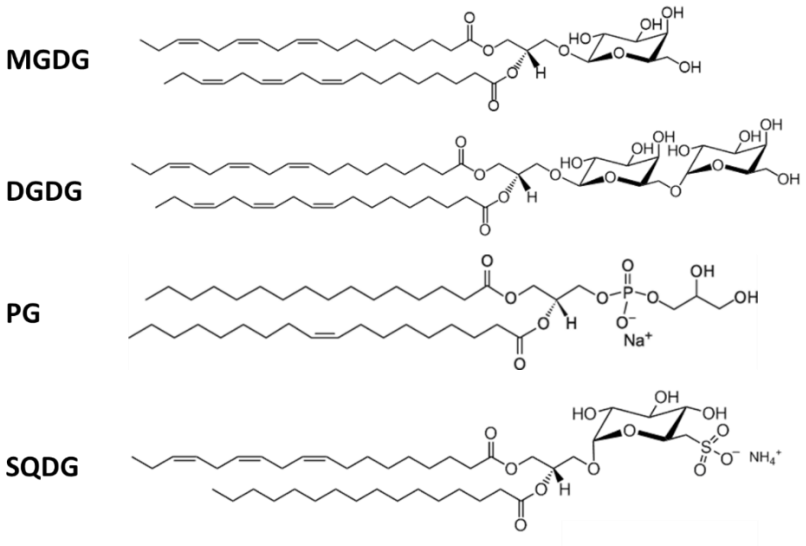


Figure 22: *The four major lipid constituents of plant photosynthetic membrane.*

In the thylakoid membrane, MGDG and DGDG play an essential role in photosynthesis by providing a lipid bilayer matrix and as structural components of protein photosystem complexes PSI and PSII (Fujii et al., 2014; Rocha et al., 2018). MGDG with its cone shape is a 'non-bilayer forming lipid' while DGDG, SQDG and PG are 'bilayer forming lipids' (Figure 23).

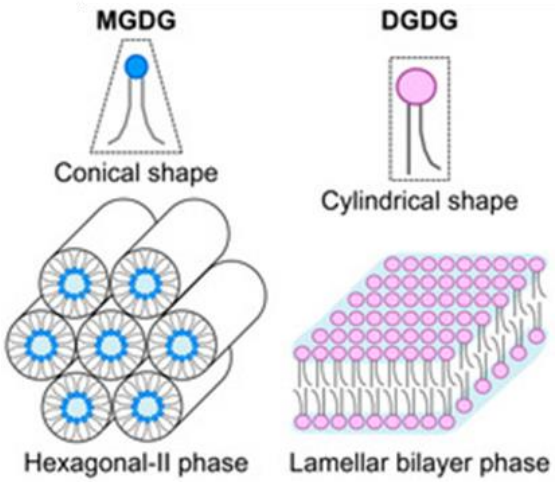


Figure 23: *Conceptual structure of MGDG and DGDG and their phase behaviours in aqueous mixtures (from Fujii et al., 2019).*

The ratio of MGDG/DGDG seems to be crucial for the stability and functional properties of photosynthetic membranes. Changes in this ratio may alter the membrane from hexagonal II (HII) to lamellar (Lα) phases, thus this ratio must be tightly regulated (Dörmann and Benning, 2002; Demé et al., 2014) (Figure 24).

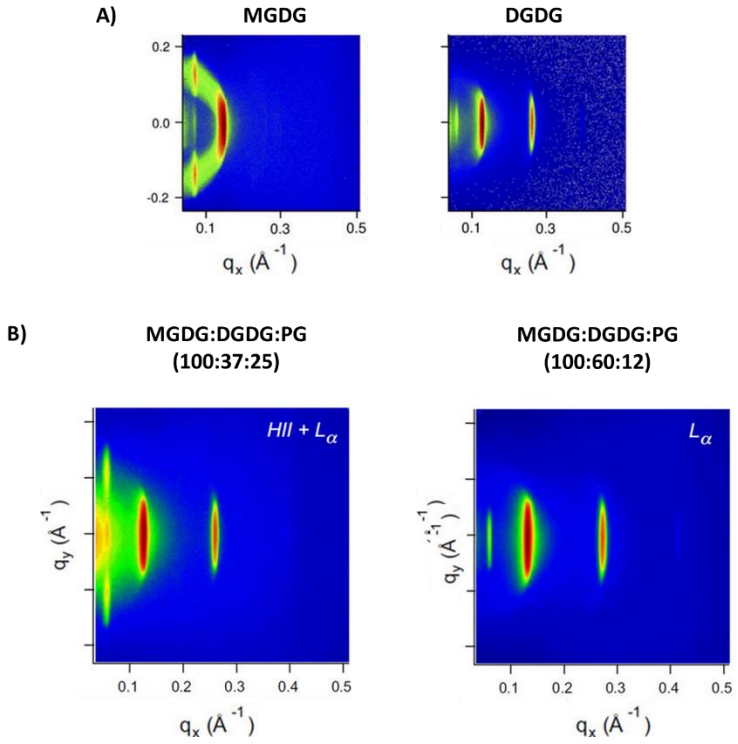


Figure 24: *Neutron diffraction images of lipid films consisting of major plant photosynthetic membranes. A) Neutron diffraction patterns of MGDG (HII) and DGDG (Lα). B) Lipid films of different lipid ratio. Demonstrated is the shift from HII to Lα phase via an increase in DGDG ratio (from Demé et al., 2014).*

The importance of galactolipids can be visually observed as deficiencies in MGDG have resulted in a largely altered chloroplast architecture with significant loss of thylakoid membranes and the formation of chloroplast invaginated inner-envelope membranes (Kobayashi et al., 2007; 2013). The loss of MGDG also impairs PSII activity and disrupts the formation of photosystem complexes such as the correct oligomerisation of the light-harvesting complex II (LHCII) and PSII (Fujii et al., 2014; Kobayashi et al., 2013; Schaller et al., 2011). Overall effects can be visualised in plant MGD1-knockout mutants that exhibit impaired cell development during embryogenesis, and can only grow in sucrose supplemented media, highlighting the importance of MGD1 and galactolipids (Fujii et al., 2019) (Figure 25).

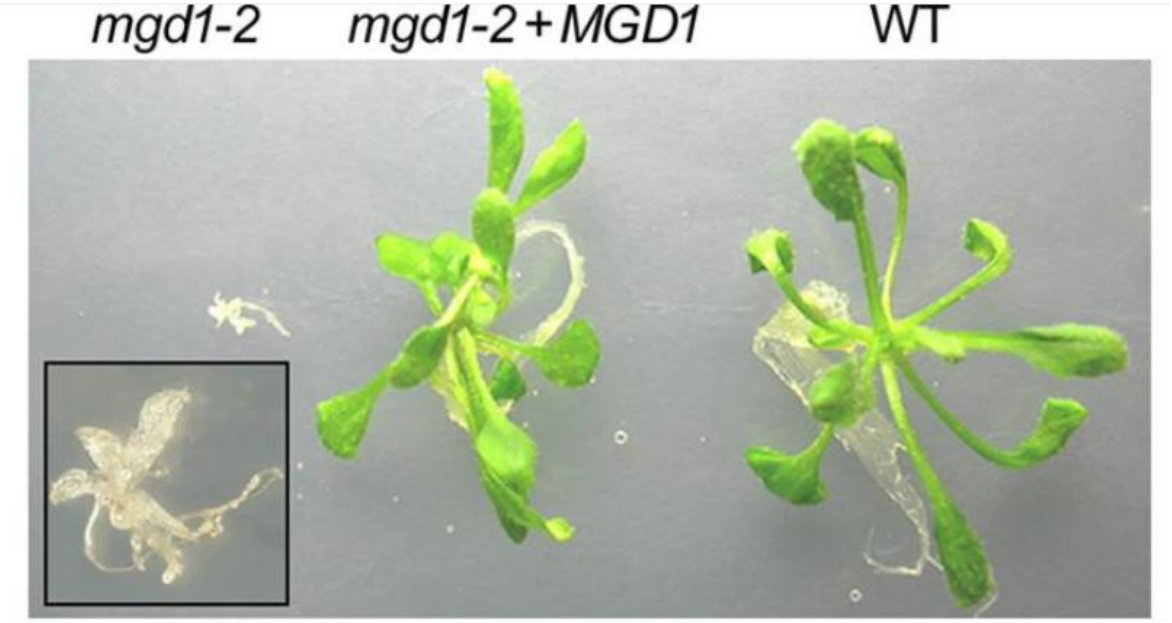


Figure 25: *Different varieties of Arabidopsis thaliana seedlings germinated and grown on sucrose containing media for 21 days. From left to right are: loss of function mgd1 mutant (mgd1-2), mgd1-2 mutant with exogenously introduced wild-type MGD1 cDNA and wild-type seedling (from Kobayashi et al., 2007).*

b) Membrane binding properties of MGD1

As mentioned above, in photosynthetic membranes, the majority of lipids are the galactolipids MGDG (~50%) and DGDG (~30%), and the MGDG/DGDG ratio must be tightly regulated. The reason for this was investigated by studying the influence of lipid constituents in MGD1 binding (Sarkis et al., 2014, Rocha et al., 2016b). In this study, binding of MGD1 to Langmuir monolayers consisting of essential plant lipids such as MGDG, DGDG, PG, and DAG were performed. Phosphatidic acid (PA) was also included in the study. MGD1 demonstrated high affinity for all monolayers tested except DGDG (Figure 26). As shown in Figure 26, a maximum insertion pressure (MIP) value above the protein interfacial tension (22 mN/m) and the estimated physiological surface pressure (30 mN/m), and a positive synergy factor (a) are indicative of a high affinity of the protein for the monolayer (Figure 26). Therefore, MGD1 demonstrated high affinity for all monolayers tested except for DGDG. The contrasted behaviour of MGD1 toward MGDG and DGDG is interesting. Indeed, MGDG, which is the reaction product, has a positive effect on MGD1 binding whereas DGDG has a negative effect

and tends to exclude MGD1. These results highlight the importance of the MGDG/DGDG ratio on MGD1 binding.

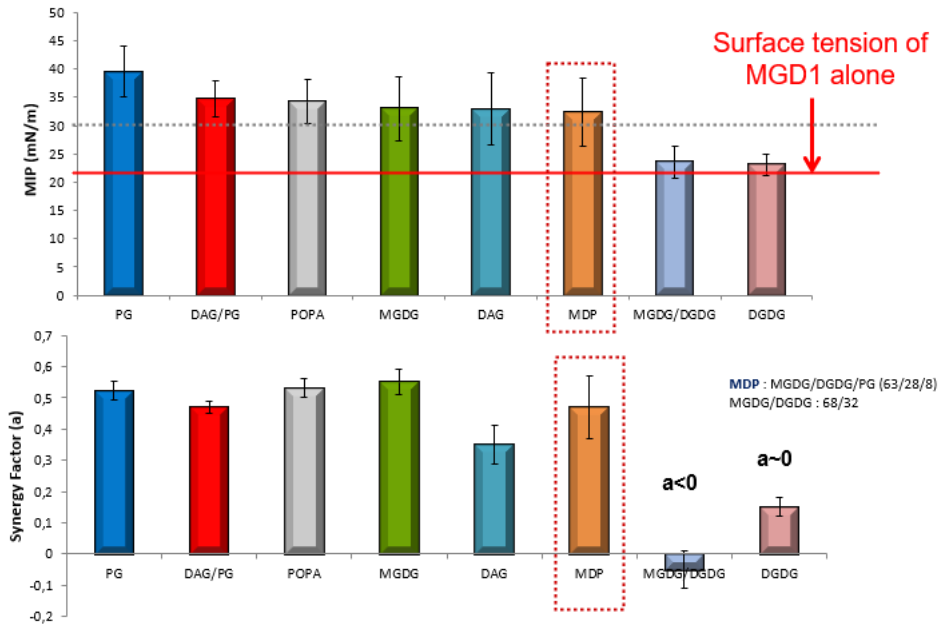


Figure 26: *Influence of the lipid nature on the binding of MGD1 to Langmuir monolayers. Maximum insertion pressure (MIP) and synergy factor (α). MDP is composed of MGDG:DGDG:PG in a 64/28/8 ratio. (Top) MIP values are determined from the increase in surface pressure upon injection of MGD1 under the monolayer and calculated at different initial surface pressures. The red line indicates the interfacial tension of MGD1 (22 mN/m). (Bottom) Synergy factor (α) of MGD1 for individual lipids. This provides information on the affinity of proteins for lipid monolayer. A positive* (a) *value is indicative of a favourable condition for protein binding whereas a negative* (a) *value is indicative of a repulsion phenomenon (from Sarkis et al., 2014).*

c) 3D structure of MGD1

The crystal structure of the catalytic domain of MDG1 (residues 137-533) in its apo-form and in complex with UDP has been recently solved and has given insight into its membrane binding properties and catalytic mechanism (Rocha et al., 2016). MGD1 is an inverting enzyme that

adopts a standard GT-B fold. It is comprised of two Rossmann-fold domains of similar size, aptly named the N- and C- domains, connected by a linker (Y319–P333) and stabilized by two long C-terminal α-helices (Figure 27). The catalytic site is located in the cleft between these two domains. The C-domain (residues K334–P506) contains the nucleotide-sugar donor binding sub-site whereas the N-domain (residues P140–V318 and E507–N524) is presumably involved in DAG binding. The closest structural homologs of MGD1 are bacterial GTs involved in the synthesis of peptidoglycan (MurG from family GT28), and the synthesis of the antibiotics calicheamycin (CalG1 and OleI from family GT1). A peculiarity of MGD1 is a large disordered region of almost 50 amino acids in the N-domain (residues 182-230), which is not visible in the crystal structure.

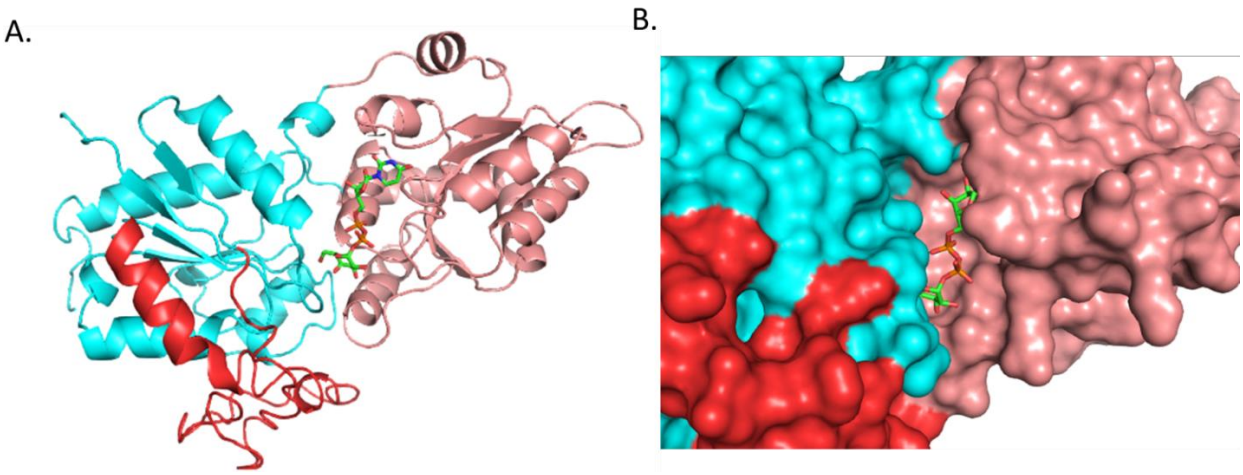


Figure 27: *Structure of MGD1 with modelled UDP-Gal and disordered loop region (Red) not seen in the published structures (Olga Makshakova, personal communication). Cartoon (A) and Surface (B) representation.*

d) UDP-Gal binding site

Although a complex of MGD1 with UDP-Gal has been obtained, only the electron density for UDP moiety is visible. This can be attributed to either flexibility/disorder of the Gal moiety, or hydrolysis of the donor. The UDP binds in a pocket at the interface between the two domains. There is marked difference in crystal structures between the apo-form (4WYI) and UDP-bound form (4X1T). A conformational change is observed in the UDP-bound form where a glycine-rich loop located between Cβ1 and Cα2 moves up 7Å. This closes the donor binding pocket and buries the nucleotide moiety while the sugar is expected to be exposed to solvent.

Computational docking experiments were used to model the Gal moiety in the MGD1 active site (Figure 28). The enzyme MurG, a N-acetylglucosaminyltransferase, has been chosen because of its similar sugar binding motif and near identical spatial positioning of key residues in their donor binding site, this is despite only having <20% overall sequence identity. UDP-Gal was modelled into a similar orientation as UDP-GlcNAc in MurG (Hu et al., 2003).

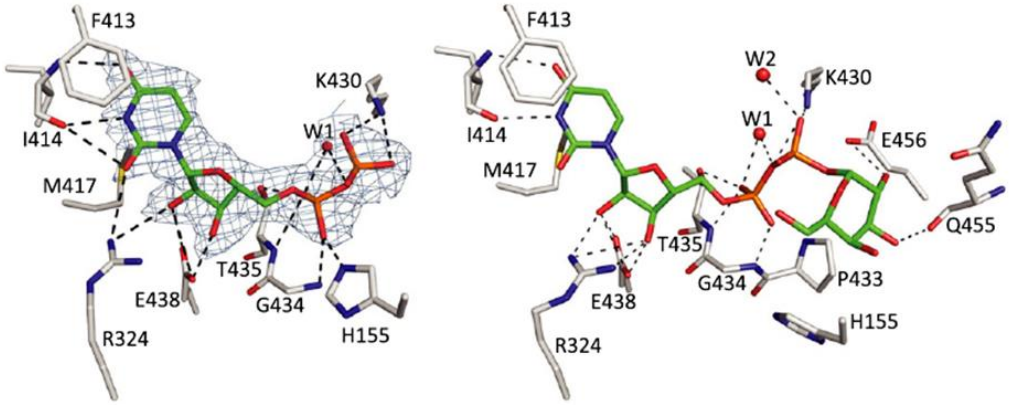


Figure 28: *UDP-Gal in the MGD1 active site. (Left) MGD1 with bound UDP. (Right) UDP-Gal modelled into the active site, based on the orientation of UDP-GlcNAc bound to MurG* (Rocha et al., 2016).

Residue F413 forms stacking interaction with uracil. The carbonyl backbone of I414 forms a hydrogen bond with uracil N3 while its main chain amide interacts with the O4 atom. O2 and O3 atoms of ribose make hydrogen bonding with R324 and E438. The main chain amide and hydroxyl group of T435, as well as G434, form interactions with α-phosphate. K430 hydrogen bonds to the β-phosphate. In the modeled Gal moiety, the O2 atom hydrogen bonds to the side chain of E456, while the main-chain carbonyl of Q455 interacts with O3.

Finally, the P433 residue seems to confer UDP-Gal specificity to MGD1 as the OH group at the C4 position of glucose is predicted to result in a steric clash with the proline (Rocha et al., 2016). There have been extensive mutation studies on the residues stated above which have resulted mostly in either complete loss or largely diminished activity (Table 2). In the context of the PhD, three residues will be particularly studied, namely P433, Q455 and E456, in our attempts to change the donor sugar specificity (UDP-Gal versus UDP-Glc).

MGD1 mutants	Enzyme activity (%)
Native	100
H155A	0
H155R	0
F413A	<2
K430A	0
K430R	0
T435A	75
P433A	<5
E438A	0
Q455A	15
E456A	0
E456N	0

Table 2: *List of MGD1 mutants studied that interact with UDP-Gal and their catalytic activity* (Rocha et al., 2016).

e) DAG binding site

As previously mentioned, the N-domain contains a large disordered region (residue W182-F230), also called the LOOP region. This region is believed to contain the predicted secondary structures Nβ2–Nα2 loop, the Nα2 helix, the Nβ3 strand and part of the Nα3 helix. The high quantity of basic and aromatic residues suggests that this region is involved in membrane anchoring and DAG binding. This is supported by mutational studies where MGD1 substitution mutants W182A, D184N, W188A have resulted in a ~90% decrease in enzyme activity (Botté et al., 2005).

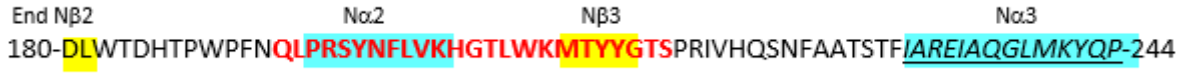


Figure 29: *Amino acid sequence of the disordered region in MGD1:UDP structure (4X1T). Residues in italics are visible in the apo-enzyme structure (4WYI). Residues in red were deleted in mutant MGD1Δ192-215 (from Rocha et al., 2016).*

In addition to this, a truncated form of MGD1 with a deletion in the disordered region corresponding to the Nα2 helix and Nβ3 strand (residues Q192–S215), named MGD1Δ192-215, was successfully produced with correct folding but it lacked activity. The Langmuir monolayer technique was used to investigate the DAG binding properties of MGD1Δ192-215. Results showed that the Q192-S215 deletion caused a 12-fold decrease in the rate of binding to DAG monolayers when compared to native MGD1, highlighting the importance of the disordered region in membrane binding (Rocha et al., 2016).

f) Lipid activators of MGD1

A highlight of MGD1 is the requirement of anionic lipids for catalytic activity. There is complete inactivity utilising only DAG and UDP-Gal. This was demonstrated as early as 2010 using phosphatidylglycerol (PG) and phosphatidic acid (PA) (Dubots et al., 2010). In this study, both PG and PA can activate MGD1 but they seem to proceed through different mechanisms, thus suggesting distinct binding sites. PG behaves as a classical activator of MGD1, whereas PA acts as an allosteric activator. In addition, PG and PA have a synergistic effect on MGD1. Kinetics data from the mutant P189A supports this assumption as the mutant lost activity in PG tests while retaining ~50% activity with PA (Nitenberg et al., 2020). Mutational data suggest that PG binds to a region close to the expected DAG binding site. SQDG, which is another important anionic lipid in chloroplast membrane can also activate MGD1 and it is expected to bind to the same binding site as PG (Nitenberg et al., 2020). Interestingly, PG and SQDG appear to be functionally redundant in plant chloroplasts. They can replace each other to maintain a constant content of negatively charged lipid (~15%), which seems to be required for correct functioning of photosynthetic membranes (Yu et al., 2002).

g) Proposed model for catalytic mechanism

Structural data has highlighted the role of H155 as the catalytic base that deprotonates the nucleophile group of the DAG acceptor but they did not reveal the role of anionic lipids in catalysis. A catalytic dyad (His-Asp/Glu) has been described in the closest structural homologs of MGD1 of family GT1, CalG1 and OleI (Bolam et al., 2007; Chang et al., 2011; Breton et al. 2012). Nevertheless, no equivalent acidic residue is present in MGD1 active site. In a recent study, using Langmuir membrane models, residues H155 and P189 were shown to be involved in PG binding (Nitenberg et al., 2020). This led the authors to propose a S_N2 type mechanism involving a PG-His catalytic dyad, where PG binding is stabilised by P189 and H155 (Figure 30). This acid-base dyad should facilitate the deprotonation of the OH group of DAG via H155. As far as we know, there is no example in the literature of a similar catalytic dyad involving an anionic lipid.

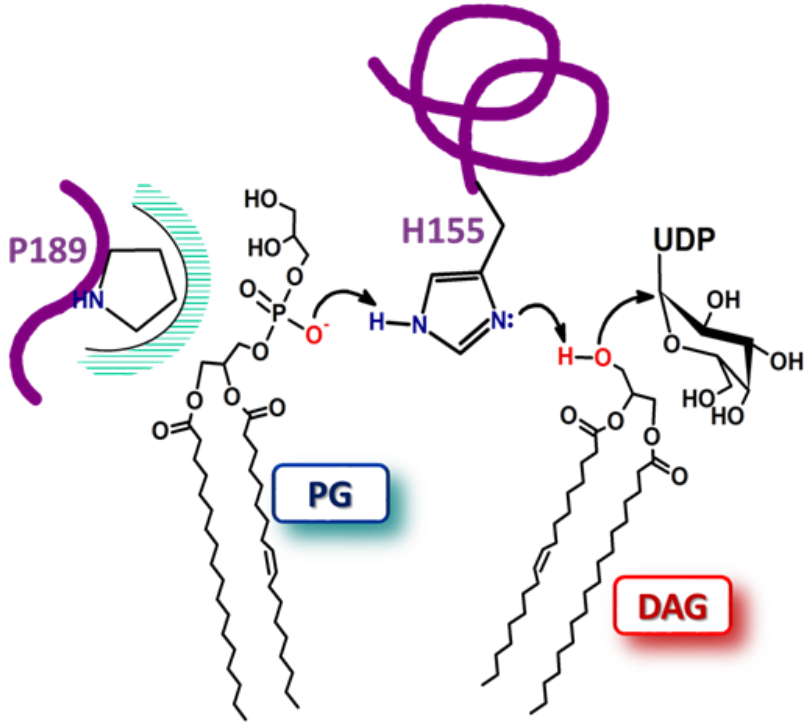


Figure 30: *Proposed mechanism of PG activation for MGD1* (Nitenberg et al., 2020).

Nevertheless, there is the issue of a mechanism utilising PA. The current hypothesis is that PA, having a smaller polar head, can bind to the active site in a different way while not interacting with P189 and still forming a hydrogen bond with H155. There is also the possibility of a second

regulatory binding site, which would clarify the allosteric properties of PA (Dubots et al., 2010). In the context of the PhD study, we will further explore potential binding sites in MGD1 for PA.

h) Protein–membrane interaction and capture of DAG and PG

The flux of DAG must be tightly controlled to meet the galactolipid demand during the massive and rapid expansion of thylakoid membranes upon light illumination. The DAG content in chloroplast membranes is very low (less than 1 mol%), thus suggesting that, once formed, DAG is rapidly converted to MGDG. One central question is to understand the mechanism of such an efficient recognition of DAG by MGD1. Interestingly, *in silico* analyses showed that MGD1 can induce a reorganisation of lipids by attracting DAG molecules to create an optimal platform for protein binding (Figure 31B). This effect was first observed using a model of bilayer formed of only DAG and PG (1:3 mol/mol) (Nitenberg et al., 2020). A model of MGD1 interacting with membrane has also been proposed. The molecular dynamics (MD) simulation also highlights the motion that happens in the protein, particularly in the C-domain, that brings the catalytic His155 residue close to membrane surface (Figure 31C).

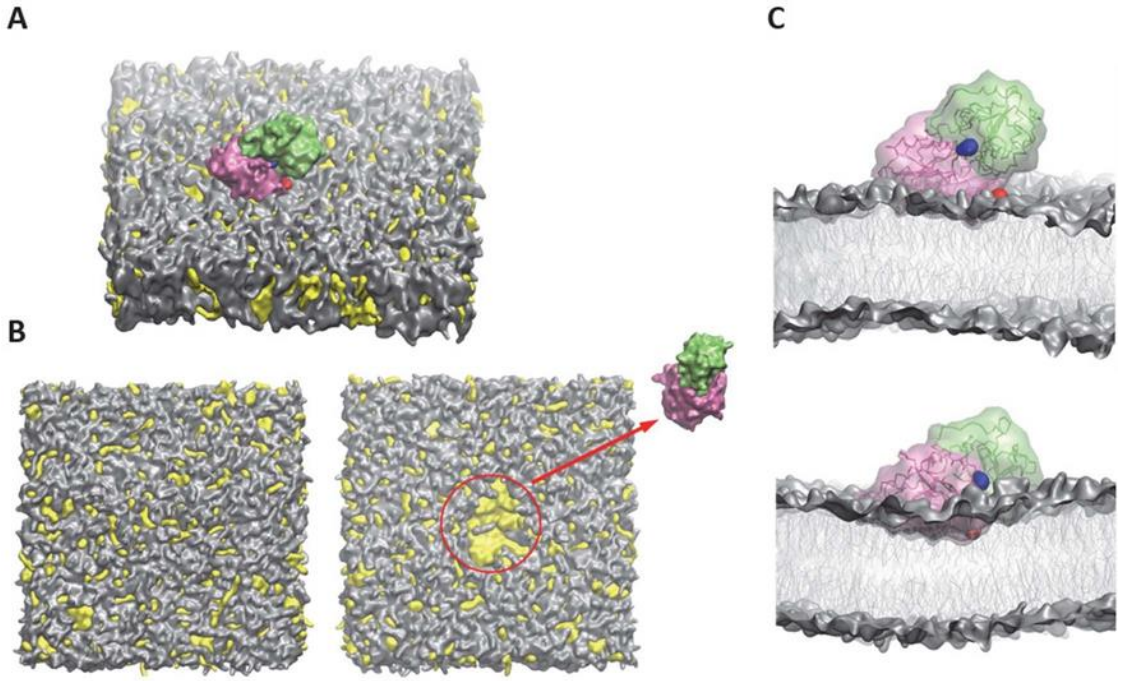


Figure 31: *(**A**) MD simulations (Coarse-Grain) in QuickSurf representation showing the interaction of MGD1 with a DAG:PG bilayer (1:3 mol/mol). Color coding: PG (grey), DAG (yellow), N-domain of MGD1 (magenta), C-domain of MGD1 (green). (**B**) Local clustering of DAG (red circle) in the area of MGD1 binding. (**C**) Internal motions in the MGD1 N- and C-domains bringing the catalytic residue, H155, near the membrane. P189 residue (red) and H155 residue (blue) (from Nitenberg et al., 2020).*

A similar computer simulation study carried out with a more complex membrane model, consisting in a mixture of representative lipids of the inner envelope membrane (MGDG/DGDG/PG/DAG; 40/30/25/5) came to the same conclusions (Makashakova et al., 2020). The protein tends to locate in the vicinity of DAG-PG clusters, thus facilitating the capture of both lipid ligands. Protein-lipid interactions occur primarily via residues of the N-domain, particularly those in the large “disordered” LOOP region that becomes ordered upon membrane binding (Figure 33).

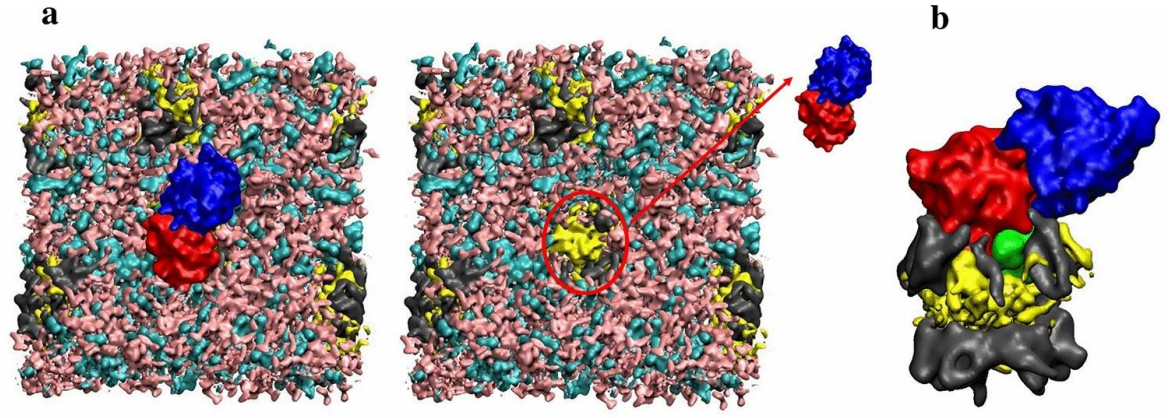


Figure 32: *MD simulations of MGD1 binding to a biomimetic membrane model. Lipid bilayer composed of MGDG(cyan)/DGDG(pink)/PG(grey)/DAG(yellow) in the molar ratio 40:30:25:5. a) Top view, showing the clustering of DAG and PG, upon MGD1 binding. B) MGD1 bound to the DAG/PG cluster. N-domain of MGD1 is in red, C-domain in blue, and the LOOP region in green (from Makshakova et al., 2020).*

2. Lipooligosaccharide α-1,4-galactosyltransferase C (LgtC)

Lipooligosaccharide α-1,4-galactosyltransferase C, more commonly known as LgtC, is a GT8 family protein from the Gram-negative bacteria *Neisseria meningitidis*. LgtC catalyzes a key step in the biosynthesis of lipo-oligosaccharide (LOS) by transferring a galactose residue from UDP-α-Gal, in a retaining Gal-α-1,4- linkage, to a terminal lactose acceptor of the LOS structure. LgtC is also able to use other glycolipid acceptors such as Lactosylceramide (LacCer) to produce Globotriaosylceramide (also named Gb3) (Figure 33) (Aldercreutz et al. 2010).

In the context of the PhD, LgtC was selected as a candidate to create artificial membrane surfaces displaying the Gb3 glycolipid.

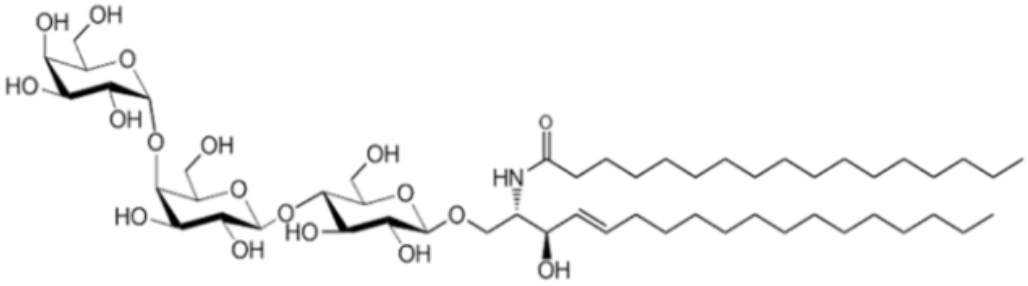


Figure 33: *Chemical structure of Gb3 (Galα-1,4-Galβ-1,4-Glcβ-ceramide).*

a) 3D structure of LgtC

LgtC has been extensively studied and structural data are available (Persson et al., 2001). This is one of the major reasons why we chose this enzyme as it opens the opportunity to engineer the enzyme for various applications. LgtC is a retaining enzyme that adopts the canonical GTA-fold and it is metal ion-dependent (Figure 34). It can be produced as a soluble enzyme in *E. coli* and purification protocols are available (Chan et al., 2012).

The crystal structure of LgtC was first solved in 2001 with a manganese ion, the acceptor analogue 4'-deoxylactose and in the absence/presence of the donor analogue UDP-2FGal (Persson et al., 2001). The catalytic domain (residues 1–247) has a standard GT-A fold containing the active site whereas a smaller helical C-terminal domain (residues 248–310) is hypothesised to mediate membrane attachment due to an abundance of basic, hydrophobic and aromatic residues.

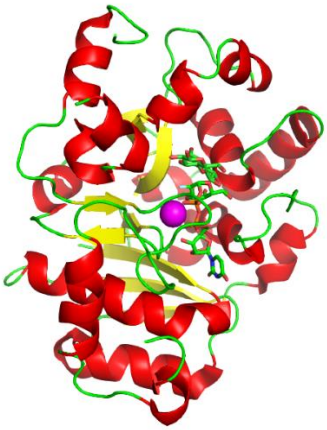


Figure 34: *Crystal structure of LgtC (PDB code 1GA8) in complex with manganese ion (purple), acceptor 4'-deoxylactose and donor analogue UDP-2-deoxy-2-fluorogalactose (UDP-2FGal) (*Persson *et al., 2001).*

b) LgtC as a target for drugs

In *Neisseria meningitidis* LgtC acts on lipooligosaccharides (LOSs) displayed on the cell surface of gram-negative bacteria. This results in the synthesis of glycan structures on LOSs that mimic human glycosphingolipids, such as Gb3, thereby allowing the bacteria to more easily evade the host immune system and causing infection such as meningitis (Moran et al., 1996). LgtC expression have also been linked to high-level serum resistance to non-typeable Haemophilus influenzae (NTHI) strains (Erwin et al., 2006). Thus, inhibition of LgtC and other GTs involved in the biosynthesis of LOSs have been explored as a strategy for anti-microbial drug discovery (Xu et al., 2017).

c) Gb3 glycolipid and Shiga toxin

The Gb3 antigen is studied in the context of glycolipid-mediated endocytosis of toxins, particularly the Shiga toxin. The Shiga toxin (Stx) is one of the most potent bacterial toxins known, which is produced by *Shigella dysenteriae* and enterohaemorrhagic *Escherichia coli* strains. Stx is composed of a toxic A-subunit and a receptor-binding B-subunit. The homopentameric B-subunit of Shiga toxin (StxB) specifically binds to Gb3 molecules (Melton-Celsa, 2014). Stx1B monomer was showed to have a K_D of 4 mM to the trisaccharide Gb3,

however, the apparent K_D for the Gb3 glycolipids was determined to be 4.2 nM. High affinity in Gb3 binding of StxB is achieved through avidity, each monomer has 3 binding sites with a total of 15 binding sites per holotoxin (Gallegos et al., 2012) (Figure 35).

Figure 35: *Structure of a mutated (W34A) Shiga-like Toxin B subunit in complex with Gb3 analogue (PDB code: 1D1I). Structure showed from the side view (A), top view (B) and bottom view (C).*

Gb3 also belongs to those glycolipids that are predominantly expressed in cancerous tissues, and is a biomarker in certain cancer types, and particularly, Gb3 is a possible target in colon cancer therapy (Falguières et al., 2008).

Gb3 is also used in tissue formation studies. Figure 36 displays an example of prototissue, which was formed from GUVs containing the glycosphingolipid receptor Gb3 and by cross-linking GUVs with the tetrameric lectin LecA from *Pseudomonas aeruginosa* that recognizes specifically Gb3.

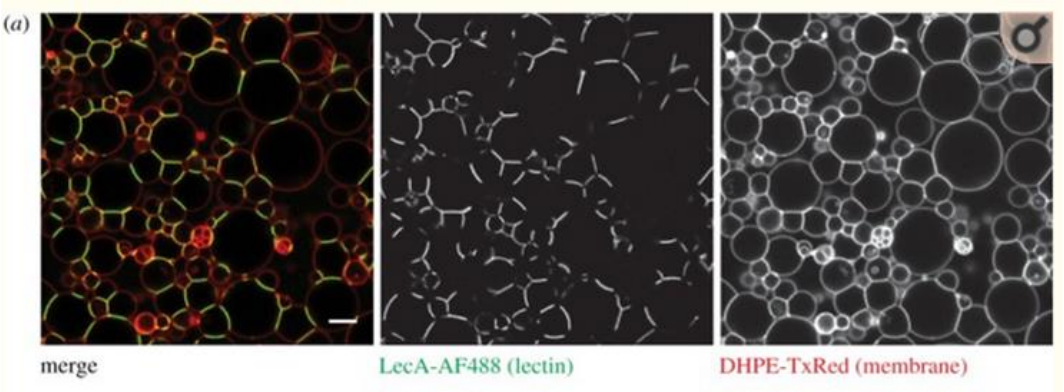


Figure 36: *Assembly of prototissues by lectin–glycan interactions. Confocal sections of an arbitrary plane of a prototissue made by cross-linking Gb3-containing GUVs (membrane marker: DHPE-TxRed, red colour) in the presence of Alexa488-labelled lectin LecA (200 nM, green colour). (Omidvar and Römer, 2019).*

V. Objectives of the PhD

1. Investigation of donor substrate specificity and mechanism of action of MGD1

As previously alluded to, most GTs currently being used for synthetic biology applications are those that exhibit some level of substrate promiscuity. This flexibility allows applications in the addition of non-native or novel sugars, as well as in the glycosylation of different target acceptors. However, not all GTs have such flexibility in recognizing substrates. There are GTs with very high specificity and therefore there is value in being able to engineer GTs with altered selectivity.

Thus, one of the main goals of the PhD project, is to increase our understanding of the sugar donor specificity and catalytic mechanism for glycosyltransferases. MGD1 is highly specific for its sugar donor (UDP-Gal) and is thus an ideal model for investigating its sugar donor specificity and mechanism of action. In addition, MGD1 has been extensively studied in the laboratory, and therefore many biochemical and structural data are available.

The main objectives are:

- Modify the sugar donor specificity from UDP-Gal to UDP-Glc
- Confirm the existence of a PG/His catalytic dyad for its mechanism of action
- Identify the allosteric site for the PA activator

We also explored the capacity of MGD1 (and also LgtC) to use modified nucleotide sugars carrying bio-orthogonal azido groups.

The approach to meet these objectives involves mutational studies coupled to molecular modelling and simulations to probe the effects of mutations on one or more residues of the protein.

2. Application of glycosyltransferases to glycosylate synthetic membrane surfaces

This PhD further explores the use of recombinant GTs to decorate synthetic membrane systems (GUVs and SLBs) with sugars. LgtC was chosen for its ability to produce the Gb3 antigen, the receptor for the lectin Shiga Toxin-B (StxB). LgtC has been widely studied with soluble acceptors, but not with lipid acceptors such as LacCer, and not in a membrane context (Lairson et al., 2004). In this PhD we explored the capability of LgtC to decorate lactosyl-ceramide-containing micelles, GUVs and SLBs, with galactose to form the Gb3 antigen. StxB was used as a reporter to monitor the Gb3 formation either by fluorescence microscopy (GUVs) or QCM-D (SLBs).

Altogether, results from this research will have applications in the use of GTs (native or engineered) to expand the toolbox for synthetic glycobiology.

Chapter 2

Materials and Methods

I. Materials

This chapter details the principles, protocols and materials employed for all experiments in Chapters 3 to 5. In addition, the computational modelling and simulations performed by collaborator, Olga Makshakova, are also described.

Molecular biology and protein expression

All primers used for mutagenesis and genes for RcMGlcD, RcMGD1 and LgtC were synthesised by Eurofins (Germany). **See appendix** for the list of primers. Vector pET-TEV was acquired from Nova lifetech Inc. and vector pET29b from Novagen/Sigma-Aldrich (Germany). Restriction enzymes: NdeI, NotI and XhoI were bought from New England Biolabs (France), and DpnI from Promega (France). PrimeSTAR Max DNA polymerase and DNA Ligation Mix/Mighty Mix were purchased from Takara (France). QIAquick DNA Extraction Kit was bought from Qiagen (France). Monarch DNA Gel Extraction Kit was bought from New England Biolabs (France). Taq'Ozyme Purple Mix 2 was purchased from Ozyme (France). NEB5α competent *E. coli* cells and BL21(DE3) competent *E. coli* cells were bought from New England Biolabs (France) and XL1 competent *E. coli* cells from Agilent Technologies (France). LB media and LB agar were purchased from Invitrogen. IPTG and Kanamycin were purchased from Euromedex (France).

Protein purification

Superdex 75 Increase 10/300 GL and HisTrap FF (1 ml and 5 ml) were purchased from Cytiva/Sigma-Aldrich. Benzonase nuclease and Imidazole were purchased from Sigma-Aldrich (Germany) and cOmplete protease inhibitor from Roche/Sigma-Aldrich.

Glycosyltransferase assays

UDP-Glo^{TM} glycosyltransferase assay kit and Ultra-pure UDP-Galactose were purchased from Promega (France). UDP-Gal for QCM-D experiments was purchased from Carbosynth (UK). TLC Silica gel 60 F254 was purchased from Supelco/Sigma-Aldrich (Germany), and Orcinol from Sigma-Aldrich. Shiga Toxin B-subunit (StxB) was provided by our synBIOcarb partners at enGenes Biotech (Vienna, Austria). Lectins RPL-Gal1 (*Pseudomonas aeruginosa*), RPL-Gal4 (*Photorhabdus asymbiotica*) were provided by GlycoSeLect (Ireland). Isolectin GS-IB4, from *Griffonia simplicifolia*, was purchased from Thermofisher Scientific.

Lipids

Monogalactosyldiacylglycerol (MGDG), Digalactosyldiacylglycerol (DGDG), C8 Lactosyl(ß) Ceramide (d18:1/8:0) (C8 LacCer), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate (16:0-18:1) (PA) and cholesterol were purchased from Avanti Polar Lipids (USA). L-α-phosphatidylglycerol (Egg yolk) (PG), 1,2-sn-dioleoylglycerol (DAG) and 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) were purchased from Sigma-Aldrich (Germany). Ceramide trihexosides (Gb3) was obtained from Matreya (USA).

Other chemicals not mentioned were of analytical or liquid chromatography grade. Unless stated otherwise, all procedures used MilliQ H_2O.

II. Expression and purification of recombinant proteins

1. Generation of protein constructs

a) MGD1

In a previous work, the MGD1 gene from *Arabidopsis thaliana* was cloned into a pET29b expression vector, via the NdeI and NotI restriction sites (Rocha et al., 2013). The MGD1 protein produced for experiments is the catalytic domain of wild-type MGD1 (cdMGD1), and is N-terminal truncated (residue 1-136). The deleted sequence corresponds to a chloroplast signal peptide sequence (1-107) followed by a short region predicted to be disordered and not necessary for activity. The gene construct also contains a non-cleavable C-terminal 6-histidine tag (His-Tag). It is transformed into *E. coli XL1* competent cells for plasmid amplification, and *E. coli* BL21 (DE3) for protein expression.

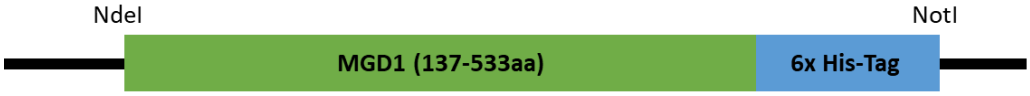


Figure 37: *Construct for wild-type MGD1.*

The pET29b-cdMGD1 plasmid codes for the wild-type MGD1 and is the DNA template used for subsequent mutagenesis procedures of MGD1.

b) MGD1 mutants

The mutagenesis procedure for MGD1 aimed to perform point mutations on the DNA sequence of MGD1. The MGD1 mutant, AQE, was created in a previous study which used the wild-type MGD1 plasmid as the template (Rocha et al., 2016). For MGD1 mutants, AEE, AEQ and GGG, we used the AQE mutant as the template for mutagenesis. For MGD1 mutants GGA and GGL, we used the GGG mutant as the template. All MGD1 mutants produced for **Chapter 4** used the wild-type MGD1 as the template. The primers used for mutagenesis were designed and then synthesised by Eurofins. **See annex** for the details on primers used.

For mutagenesis, the PCR reaction mix is assembled as described in the table below:

Reaction mixture for mutagenesis	
	Quantity
Template plasmid	20 ng
Sense primer (10 µM)	1 µl
Anti-sense primer (10 µM)	1 µl
Nuclease-free H_2O	Varies
PrimeSTAR Max DNA polymerase	25 µl
Final volume	50 µl

Table 3: *Composition of the reaction mix for mutagenesis.*

After assembly of the mutagenesis reaction mix, PCR was run on a Mastercycler gradient (Eppendorf), with the program described in Table 4.

Steps	**No. cycles**	**Temperature (°C)**	**Time**
1	1	95	2 min
2	30	95	20 sec
		55	20 sec
		72	15 sec/kb plasmid
3	1	72	3 min

Table 4: *PCR cycle for mutagenesis of MGD1.*

After PCR, the template DNA in the reaction mix was digested with the addition of 10 U DpnI restriction enzyme and was incubated for at least 1 hour at 37°. In parallel, 10 µl of PCR reaction was run on 1% agarose gel to check for amplification. After DpnI digestion, 2 µl of the PCR product was used to transform *E. coli* NEB5α competent cells by **heat shock: 45 seconds, 42°C, then 2 minutes on ice**. The transformed cells are incubated at shaking 180 rpm, 37°C for 1 hour before plating onto LB agar plates supplemented with 50 µg/ml kanamycin. Plates are incubated overnight at 37°C. Selected colonies are grown in 5 ml LB overnight at 37°C and plasmid is extracted using the QIAquick Gel Extraction Kit according to manufacturer instructions. The plasmid was sent for DNA sequencing. Upon confirmation of successful mutagenesis, the plasmid was transformed into expression strain *E. coli* BL21 (DE3) using the heat shock method described previously.

c) RcMGlcD and RcMGD1

RcMGlcD and RcMGD1 genes were synthesized by Eurofins. Both genes were codon optimised for *E. coli* expression. RcMGlcD is from *Roseiflexus castenholzii*, and the synthesised gene codes for the full-length protein (1-396). The plasmids containing the genes for RcMGlcD and RcMGD1 were transformed into *E. coli* NEB5α competent cells, and was grown overnight in 5 ml LB media to amplify the plasmid. The plasmid was extracted using the QIAquick Gel Extraction Kit. The genes and pET29b were digested with NdeI and NotI for 1 hour at 37°C. The typical reaction mix for restriction digest is described in Table 5:

Restriction digest reaction mix	
	Quantity
Plasmid	<2 µg
NdeI (10 µM)	1 µl
NotI (10 µM)	1 µl
Nuclease-free H_2O	Varies
CutSmart buffer	20 µl
Total volume	40 µl

Table 5: *Reaction mixture for restriction digest. Digestion is left to occur for 1 hour at room temperature.*

After digestion, the reaction mix was run on 1% agarose gel and the bands containing the digested RcMGlcD, RcMGD1 and pET29b were extracted using the Monarch DNA Gel Extraction Kit according to manufacturer instructions. Ligation of the RcMGlcD or RcMGD1 gene to pET29b was performed using the DNA Ligation Mix/Mighty Mix according to manufacturer's instructions (Table 6).

Ligation reaction mix	
	Quantity
Vector	Varies
Gene	Varies
Nuclease-free H_2O	Varies
DNA Ligation mix	5 µl
Total volume	10µl

Table 6: *Reaction mixture for DNA ligation. Vector and gene insert are added in a molar ratio of 1:3 and made to 5µl with nuclease-free H_2O before adding the DNA ligation mix. The reaction mix is incubated at 4 ^oC overnight or at room temperature for 20 minutes.*

Ligation created the pET29b-RcMGlcD and pET29b-RcMGD1 plasmids. The final constructs were transformed into *E. coli* NEB5α competent cells for plasmid amplification by heat shock (45 seconds, 42^{o}C, then 2 minutes on ice). The transformed cells are plated in a LB agar plate containing kanamycin (50 µg/ml) and incubated at 37^{o}C overnight. Kanamycin selects for colonies containing the pET29b vector. Colonies were then screened for the pET29b containing the RcMGlcD or RcMGD1 gene. A typical PCR mix for screening contained: *E. coli* colony, T7 sense and anti-sense primers (0.2 µM) and Taq'Ozyme Purple Mix 2 kit. The screening reaction was run on PCR with cycles according to Ozyme manufacturer instructions. Afterwards, the PCR reaction are run on 1% agarose gel and positive colonies are re-plated on LB agar plates. A positive colony was grown overnight in 5 ml LB to amplify the plasmid and was extracted for DNA sequencing. The pET29b-RcMGlcD and pET29b-RcMGD1 plasmids are transformed into *E. coli* BL21 (DE3) for protein expression.

d) CaMGD1

The CaMGD1 gene from *Caldilinea aerophila* was obtained from Dr Koichiro Awai from Shizuoka University, Japan. CaMGD1 was cloned, expressed and purified at CERMAV by Milène Nitenberg (Thesis, 2018). CaMGD1 is composed of 380 amino acids and the coding sequence was cloned into pET24a vector, which possessed kanamycin resistance and the T7 promoter, via the restriction enzymes NdeI and XhoI. The final construct expressed CaMGD1 with a non-cleavable His-Tag at the C-terminus.

e) LgtC

The gene of LgtC from *Neisseria meningitidis* was codon optimised for *E. coli* expression and synthesised with 25 amino acids truncated at the C-terminus. This region is believed to be involved in membrane association. Truncation was also performed to increase solubility and yields. Codons were optimised for *E. coli* expression and with the mutations C128S, C174S (Persson et al., 2001) and T273A which were reported to improve expression levels of the protein without compromising the enzyme activity (Chan et al., 2012). The pET-TEV plasmid was chosen as the expression vector for LgtC. The pET-TEV plasmid is a pET28a vector containing the Tobacco Etch Virus (TEV) cleavage site in-place of the original thrombin site.

The plasmid containing the LgtC gene was synthesised by Eurofins and transformed into *E. coli* XL1. Two small cultures containing the LgtC gene and pET-TEV were grown overnight to amplify the plasmids, and were extracted using the QIAquick Gel Extraction Kit. The LgtC gene and pET-TEV were digested with NdeI and XhoI for 1 hour at 37°C (Table 5). Digestion products were run on 1% agarose gel and the bands containing the digested LgtC gene and pET-TEV were extracted using the Monarch DNA Gel Extraction Kit. Ligation of the LgtC gene and pET-TEV (Table 6) was performed using the DNA Ligation Mix/Mighty Mix. This formed the LgtC-TEV construct which is ~6.3kb long and contains a N-terminal TEV protease site and His-Tag (Figure 38).

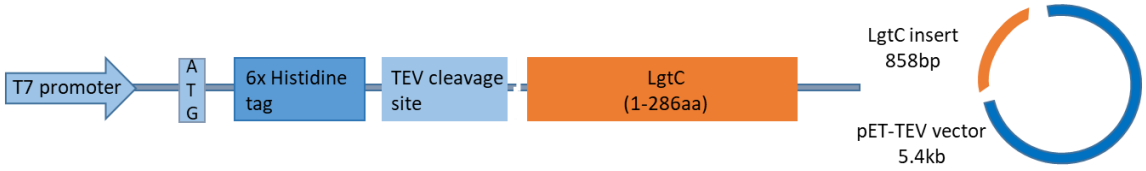


Figure 38: *Schematic for the expression construct of LgtC in pET-TEV vector (Lgtc-TEV).*

The construct was transformed into *E. coli* XL1 and plated in kanamycin supplemented (50 µg/ml) LB agar plates, and colonies left to grow overnight at 37°C. Colonies were then screened for the plasmid containing the LgtC gene. A PCR reaction contained a mix of *E. coli* colonies, T7 sense and anti-sense primers (0.2 µM) and Taq'Ozyme Purple Mix 2 kit. After PCR, the PCR mixes are run on 1% agarose gel and positive colonies are plated on LB agar plates. Selected positive colonies are grown overnight in 5 ml LB media and plasmid extracted for DNA sequencing. LgtC-TEV plasmid is transformed into *E. coli* BL21 (DE3).

2. Expression of recombinant proteins

Recombinant protein expression was performed on a large scale in 1 L LB medium (50-100 ml on a small scale) in baffled flasks with induction by isopropyl β- d-1-thiogalactopyranoside (IPTG). Encoded in the expression vector is the LacI gene, which codes for the lac repressor (LacI). The LacI binds to the lac operon, preventing RNA polymerase transcription and thus protein expression. IPTG binds to LacI removing it from the lac operon and thus protein expression is induced. Also included in the expression vector is the gene for antibiotic resistance (kanamycin). Thus, the medium contains the antibiotic to ensure the elimination of cells that do not acquire antibiotic resistance from the expression vector.

a) MGD1, MGD1 mutants, RcMGlcD and CaMGD1

Expression and purification of MGD1, MGD1 mutants, RcMGlcD and CaMGD1 follow the protocol established by Rocha et al. (2016a). A small starter culture of 20 ml LB containing *E. coli* BL21 (DE3) cells harbouring the MGD1 construct was grown overnight. A 1 L of LB medium supplemented with 50 µg/ml kanamycin was inoculated with the 20 ml starter culture. The culture was incubated at 37°C and 180 rpm. At optical density (OD_{600}) 0.6-0.8 the culture was induced with IPTG to a final concentration of 0.25 mM and incubated at 30°C for 4 hours. Cells were harvested by centrifugation at 7,000 x *g* at 4°C. Cell pellets were stored at -20°C until purification. 1 ml of culture was collected before IPTG induction and after expression, and was used to monitor protein expression in SDS-PAGE.

b) LgtC

The expression of LgtC was based on the procedure by Chan et al. (2012). A small starter culture of 20 ml LB medium containing *E. coli* BL21(DE3) cells harbouring the LgtC-pET-TEV construct was grown overnight. 1L of LB medium supplemented with 50 µg/ml kanamycin was prepared. The 1L LB medium was inoculated with 20 ml starter culture and incubated at 37^{o}C and 180 rpm. At OD_{600} of 0.6-0.8 the cells were induced with 0.5 mM IPTG and incubated overnight at 16^{o}C and 180 rpm for ~16 hours. Cells were harvested by centrifugation at 7,000 x *g* and 4^{o}C. Cell pellets were stored at -20^{o}C until purification. 1 ml of culture was collected before IPTG induction and after expression, and was used to monitor protein expression in SDS-PAGE.

3. Purification of recombinant proteins

Proteins are expressed in the cytoplasm of the bacteria. To release the proteins, cells were lysed by high pressure ell disruption. The resulting lysate was centrifuged at high speed (24,000 x g) and the target protein found in the supernatant. The supernatant was clarified through a 0.22 µm filter to remove any large particles. The expressed proteins contained a His-tag which allowed for immobilized metal affinity chromatography (IMAC) for the separation of the target proteins from other bacterial proteins present in the supernatant. The IMAC column contains immobilized nickel in its matrix. His-Tag binds to the nickel matrix, while other proteins pass through or bind with weaker affinity. The target protein can be released from the column with a gradient of imidazole. Due to their weak binding, non-specific bacterial proteins were washed away with a low concentration of imidazole (0-75 mM imidazole), while the target proteins were eluted with higher concentration of imidazole (Figure 39A). If cleavable, the protein was incubated with His-tagged-TEV protease to cleave the His-Tag from the protein of interest. The mixture was then further purified by IMAC. The TEV protease and His-Tag bound to the column, and the target protein were collected in the flow through.

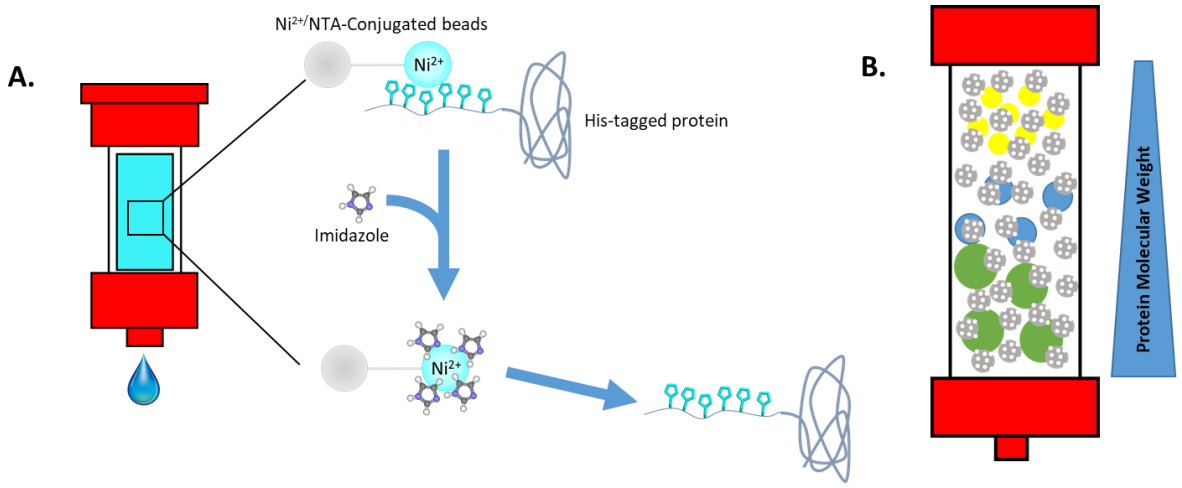


Figure 39: *Basic principles of (A) IMAC and (B) SEC. A) Proteins containing a His-tag bind to nickel ion. The addition of imidazole out competes for binding to nickel, releasing the His-tag protein. B) Porous matrix interacts with smaller proteins and, thus separates proteins by their molecular mass.*

A second step of purification involved size exclusion chromatography (SEC). A column containing a porous matrix separated proteins by their size. Smaller proteins interact with the porous matrix more strongly and therefore are retained in the column for a longer time. In practice, large proteins elute earlier from the column than small proteins (Figure 39B). By looking at the volume at which a protein elutes and comparing this to an established standard an approximate molecular weight of the protein can be obtained. This can be used as a quality control for identification of contaminant proteins or an indicator of oligomerisation of the target protein.

a) MGD1, MGD1 mutants, RcMGlcD and CaMGD1

For purification of MGD1, MGD1 mutants, RcMGlcD and CaMGD1, the following buffers were prepared:

- Lysis buffer: 25 mM HEPES, 500 mM NaCl, 1 M Urea, 10 mM imidazole, 1 mM DTT, Benzonase® nuclease (1 µl/10 ml of solution), cOmplete protease inhibitor (1 tablet/50 ml of solution), pH 7.5
- Binding buffer: 25 mM HEPES, 500 mM NaCl, 1 M Urea, 10 mM imidazole, 1 mM DTT, pH 7.5

- Elution buffer: 25 mM HEPES, 500 mM NaCl, 1 M Urea, 500 mM imidazole, 1 mM dithiothreitol (DTT), pH 7.5
- SEC buffer: 25 mM Bis-Tris propane, 150 mM NaCl, 5% (v/v) Glycerol, 1 mM tris(2-carboxyethyl)phosphine (TCEP), pH 6.5

Cell pellets from large scale expression were resuspended in lysis buffer (5 ml/g of cell pellet). The cells were disrupted using a Cell Disruptor CSL (Constant Systems, Ltd) at 1.8 kbar. The lysate was centrifuged at 24,000 x g for 30 min at 4°C. The supernatant was then clarified through a 0.22µm filter. The protein was first purified by IMAC using either the 1 ml or 5 ml HisTrap FF column and the binding/elution buffers. The chromatography system used for IMAC and SEC was performed on a NGC Quest 10 purification system (BioRad). The clarified lysate was loaded into the HisTrap FF column at 1 ml/min and the flow-through was collected. The HisTrap column was washed at 1 ml/min (or 2 ml/min for the 5 ml column) with 40 mM (8% elution buffer) and then 80mM (16% elution buffer) imidazole until Abs_{280nm} was stable. MGD1 was eluted from the column with an imidazole gradient (80-500 mM or 16-100% elution buffer). Fractions were pooled according to protein purity evaluated by SDS-PAGE and DLS, using Zetasizer Nano ZS (Malvern), if necessary.

Pooled fractions were concentrated to a maximum concentration of 10 mg/ml and injected into a Superdex 75 Increase 10/300 GL column. Proteins were purified through the SEC column by running the SEC buffer at 0.5 ml/min. Fractions were analysed by SDS-PAGE, and by Dynamic Light Scattering(DLS), using a Zetasizer Nano ZS (Malvern), if necessary. The appropriate fractions were pooled and concentrated up to 10 mg/ml using a Macrosep® concentrator (Pall) with a 30 kDa MWCO and stored at -20°C. Protein concentration was monitored using a NanoDrop™ ND-2000 spectrophotometer (Thermo Scientific). The Nanodrop parameters for MGD1 and MGD1 mutants are: 47900 $M^{-1}.cm^{-1}$ and 43.99 kDa. For RcMGlcD: 60,515 $M^{-1}.cm^{-1}$ and 42.50 kDa. For CaMGD1: 47,815 $M^{-1}.cm^{-1}$ and 41.00 kDa

b) LgtC

For LgtC purification, the following buffers were prepared:

- Lysis buffer: 20 mM Tris, 1 mM DTT, 500 mM NaCl, Benzonase nuclease, cOmplete protease inhibitor, pH 7.5

- Binding buffer: 20 mM Tris, 1 mM DTT, 500 mM NaCl, pH 7.5
- Elution buffer: 20 mM Tris, 1 mM DTT, 500 mM NaCl, 500 mM Imidazole, pH 7.5
- SEC buffer: 20 mM Tris, 5 mM TCEP, pH 8.5

Cell pellets from protein expression were resuspended in lysis buffer (5 ml/g cell pellet) and disrupted using a Cell Disruptor at 1.8 kbar. The lysate is centrifuged at 24,000 x g for 30 min at 4°C and the supernatant was clarified through a 0.22 µm filter before loading into a HisTrap FF column. The column was washed with binding/elution buffer at 25 mM (5%B) and 75 mM (15%B) imidazole each until the Abs_{280nm} became stable. LgtC was eluted with an imidazole gradient 75-500 mM (15-100%B) imidazole. The samples were collected in 1 ml fractions and analysed by SDS-PAGE for purity and DLS for homogeneity. The appropriate fractions were pooled and concentrated up to 10 mg/ml using a Macrosep® concentrator (Pall) with a 10kDa molecular weight cutoff (MWCO). The parameters for NanoDrop™ were 61,225 M^{-1} cm^{-1} and 34.85 kDa.

LgtC was injected into a Superdex 200 Increase 10/300 GL column equilibrated in SEC buffer. LgtC was purified by running SEC buffer through the column at 0.5 ml/min. The appropriate fractions were analysed by SDS-PAGE, pooled and concentrated up to 10 mg/ml. The purified protein was stored in aliquots at -20°C until further use.

4. Protein analysis methods

a) SDS-PAGE

A quality control for protein purification was performed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples were denatured with boiling, reducing agent (mercaptoethanol or DTT) and SDS, which is a strong anionic detergent. This treatment denatures the protein samples and SDS coats the proteins with a uniform negative charge. The protein samples are then run on a gel matrix with a determined polyacrylamide concentration. In an applied electric field, the negatively charged proteins migrate through the gel to the positive anode and are separated by their molecular weight. Since the proteins in the sample now have the same charge to mass ratio, there is no differential migration due

to charge. Proteins were thus separated by their molecular weight and this allowed for evaluation of the purity of the protein.

Described below is the typical protocol for SDS-PAGE analysis of fractions from purification of protein.

The appropriate amount of 5X SDS-PAGE Protein Loading Buffer (10% SDS, 500 mM DTT, 50% Glycerol, 500 mM Tris-HCL and 0.05% bromophenol blue) was added to protein samples to be analysed, typically 2 µl loading buffer to 10 µl sample. The samples were boiled at 100^{o}C for 5 minutes and sonicated in a bath sonicator (Bioblock Scientific) at 5kHz for 5 minutes. A 10% polyacrylamide gel was prepared and assembled in an electrophoresis cell (BioRad). The electrophoresis cell was submerged in SDS-PAGE running buffer: 25 mM Tris-HCl, 200 mM Glycine, 0.1% (w/v) SDS, pH 8.3. A protein ladder and protein samples were loaded in a 10% polyacrylamide gel. Migration of the proteins was started by applying a constant voltage of 130 V. When the dye has migrated near the bottom of the gel the electric current was stopped. The gel was extracted from the cell apparatus and proteins were stained with Coomassie blue overnight. The next day, the gels were analysed and recorded.

b) Western blot

Western blot is another analytical technique utilised in this PhD project. It is an immunoblot technique used for detection of specific proteins in a sample using antibodies. The protein samples undergo denaturation and size separation under SDS-PAGE. Blotting is the next process and works under the same principle as SDS-PAGE, where the natively charged proteins migrates to the anode. In blotting, the protein bands from the SDS-PAGE gel are transferred to a nitrocellulose membrane. Next is the blocking procedure, which coats the membrane surface in casein or bovine serum albumin (BSA) except for sample proteins. The blocking step reduces background signal and false positives. Finally, the membrane is probed with an antibody specific to the protein sample and visualised. In the case of the PhD, the antibody used is specific for proteins with a His-tag and is conjugated to horseradish peroxidase (Anti-His-HRP). Bands on the membrane are visualised by adding 3,3',5,5'-tetramethylbenzidine (TBM) staining reagent. TMB is oxidised by HRP in the presence of peroxide which results in a blue precipitate at the site of the His-Tag proteins. Western blotting was essential for detection of extremely low quantities of protein that could not be distinguished by SDS-PAGE.

Described below is the typical protocol for western blot analysis of fractions from purification of protein.

Protein samples underwent SDS-PAGE as previously described. However, a pre-stained protein ladder was run on the SDS-PAGE gel in order to visualise the molecular weight of bands on the nitrocellulose membrane. Afterwards, the SDS-PAGE gel was assembled for blotting on a nitrocellulose membrane (membrane and gel sandwiched by blotting paper) and placed on an electrophoresis cell. The cell was submerged with the blotting buffer: SDS-PAGE running buffer supplemented with 20% ethanol. Blotting is initiated by applying a constant voltage (100 V) for 1 hour. TBST buffer is prepared: 50 mM Tris, 250 mM NaCl, pH 7.6. Blocking solution was prepared, 30 ml: TBST buffer supplemented with 1% BSA. The membrane was assembled in the SNAP i.d. 2.0 Protein Detection System (Sigma) according to manufacturer's instructions. The membrane was blocked with the blocking solution for 1 min and solution drained away. Probing solution was prepared by mixing 2.5 µl Anti-His-HRP with 5 ml of TBST. Antibody binding to His-Tag proteins is started by completely covering the membrane surface with the probing solution and incubating for 15 minutes at room temperature. Afterwards, the membrane is washed with 4 cycles of 30 ml TBST. The His-Tag proteins were visualised by applying TBM staining reagent.

c) Dynamic light scattering

Dynamic Light Scattering (DLS) can be used as additional quality control to ensure purity, structural integrity and any multimeric formation of the protein. DLS works by the principle of light scattering by small particles in solution. Brownian motion of particles in solution results in intensity over time fluctuations of Rayleigh scattering. This fluctuation of intensity is measured, and by using mathematical tool, particles can be physically characterized. In regards to proteins in solution, the hydrodynamic radius of the protein is obtained and can be correlated to 'globular protein' molecular weight. Additional information that can be obtained from DLS are monodispersity and homogeneity of the protein sample.

DLS measurements were taken using the Zetasizer Nano ZS (Malvern) and analysed on the Zetasizer Nano software. For measurement of proteins, buffer composition is input into the

software in order to remove background signal from the buffer. Proteins were diluted to 0.1-1 mg/ml for measurements.

III. Enzyme activity assays

1. UDP-Glo™ assay – principle

UDP-Glo™ is a glycosyltransferase assay kit by Promega. The assay functions by means of a luciferase reaction which converts the UDP product into ATP, that generates luminescence measured as relative light units (RLU) (Figure 40). This luminescence can be correlated to UDP concentration using a standard curve of know UDP concentrations (Figure 41). GTs may have a small hydrolase activity that cleaves UDP from sugar. Thus, there must be a control (without the acceptor), the RLU value of this control is taken from the full enzyme reaction to obtain the true transferase activity. A control reaction with no donor or no enzyme also provides the background luminescence reading (no reaction), but is not necessary.

In a standard GT reaction, the enzyme reaction is stopped by mixing an equal amount of UDP detection reagent (UDR), which consists of UDP-Glo enzyme and Nucleotide detection buffer provided by the kit. Adding UDP also starts the bioluminescence process.

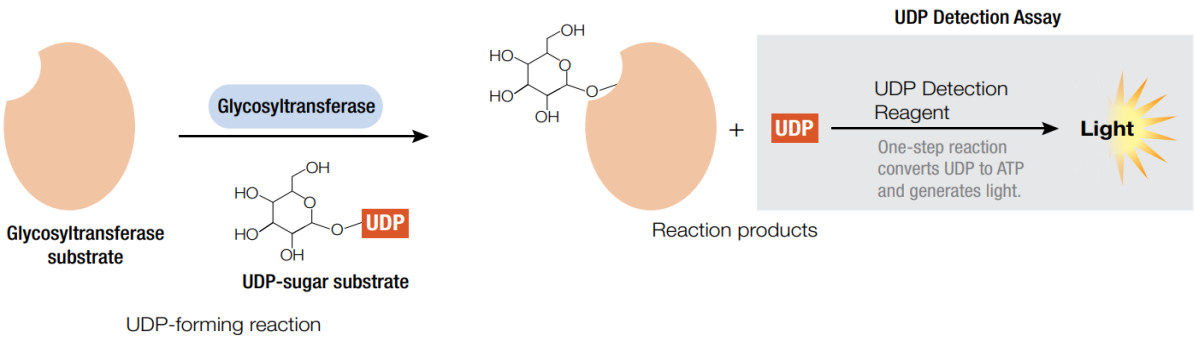


Figure 40: *Schematic of the enzymatic assay. Figure taken from the manufacturers protocol from Promega.*

The UDP standard graph is created from measuring the luminescence of known UDP concentrations. It is important to note, when performing the UDP-Glo assay, that the same homogenous UDP detection reagent mix must be used for both the UDP standard and the experiment. This is to ensure reliability when utilising the UDP standard graph for calculations.

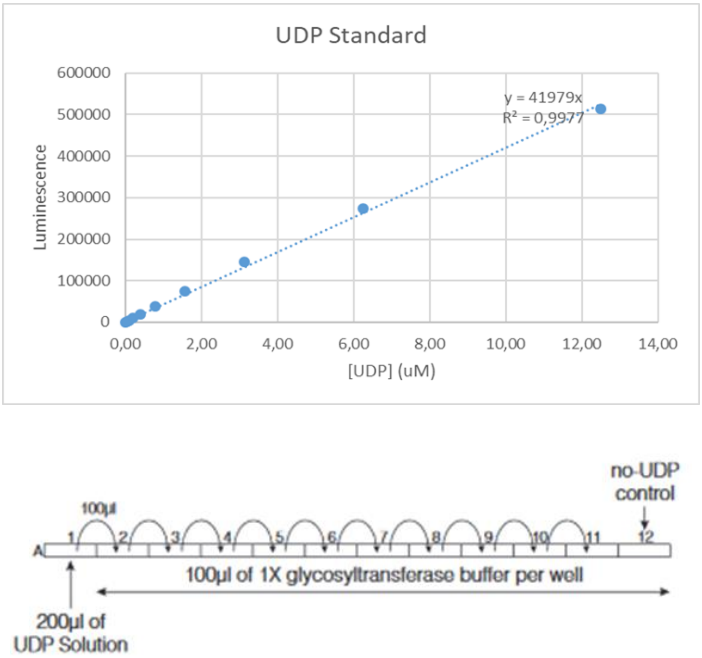


Figure 41: *A typical UDP standard graph and below the creation of the UDP standards via consecutive two-fold dilutions of a known UDP concentration.*

The enzymatic activity can be expressed in multiples ways. Activity can be expressed as the raw luminescence reading (RLU) of the reaction. This method does not require the creation of a UDP standard graph, but the data only useful in comparisons of enzyme reactions when all conditions are the same in all conditions (buffer, reaction time, substrates concentration and protein concentration) and performed in the same experimental batch. Using a UDP standard graph, the rate can be expressed as µM/min or as specific enzyme activity (nmole/min/mg).

2. Enzyme assay for MGD1, MGD1 mutants, RcMGlcD and CaMGD1

MGD1 enzyme activity was assayed using an adapted protocol based on Rocha et al., 2016. The MGD1 assay uses mixed micelles composed of the lipid acceptor diacylglycerol (DAG) and a lipid activator, phosphatidylglycerol (PG) or phosphatidic acid (PA).

The assay is carried out in mixed lipid/detergent micelles. In this case the constituents of mixed micelles are DAG, PG or PA and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS). The concentration of each lipid is expressed as the surface concentration of the lipid at the level of mixed micelle (mol-%). Each lipid is calculated as the mol-% of all constituents in the mixed micelle.

Composition of MGD1 reaction tubes			
	Full reaction (µl)	**No DAG control (µl)**	**No Donor control (µl)**
DAG 1 mg/ml	10	0	10
PA or PG 1 mM	4	4	4
Tp3X + CHAPS 18 mM	12	12	12
H_2O	17.5	17.5	22.5
MGD1 (Varies)	2	2	2
UDP-Gal (5 mM)	5	5	0
Total Volume	36.5	36.5	36.5

Table 7: *Composition of MGD1 reaction tubes. Quantity of protein and reaction time depend on the experiment. Tp3x: 75 mM Bis-Tris propane, 450 mM NaCl, 15% Glycerol, 3 mM TCEP. Detailed are the composition of a full MGD1 reaction, No DAG control and No UDP-Gal control.*

Using a Hamilton syringe, DAG and PG/PA (can be pre-made as a lipid mix) are placed in a glass tube and solvents evaporated under nitrogen (20 minutes) to form a lipid film. A master-mix consisting of Tp3x (CHAPS supplemented), water and MGD1 is created and added to the glass tubes to solubilise the lipid (Table 7). The tubes are sealed with caps and sonicated in an iced water bath sonicator (Bioblock Scientific) at 5 kHz for 10 minutes, with temperature maintained below 10^{o}C. This step forms the mixed micelles for the MGD1 reaction.

During this time the UDR is prepared according to manufacturer instructions. 25 µl UDR is added to a 96-well plate (Greiner Bio-One, Germany). After sonication to form micelles, the reaction can begin. The reaction is started by addition of UDP-Gal/UDP-Glc (5 mM) to the tubes and vortexing. Pipetting of UDP-Gal to each tube are performed in equal intervals (every 20 seconds). The final reaction mix consists of: 25 mM Bis-Tris propane, pH 6.5, 6 mM CHAPS, 150 mM NaCl, 5% glycerol, 1 mM TCEP, supplemented with 6.3 mol-% DAG, 1.5 mol-% PG/PA and 0.7mM UDP-Gal. In the assay, the MGD1 concentration reaction time can be varied. Enzyme assays are performed at 25°C. The reaction is stopped by taking 25 µl of the enzyme reaction and transferring it to the 96-well plate containing UDR. After all reactions have been stopped, the plate is covered in foil, shaken for 20 seconds and left to stand for 1 hour at room

temperature. Bioluminescence is measured on the SPARK® Multimode Microplate Reader (TECAN).

The protocol described above is the standard/traditional protocol used in **Chapter 3 and 4**. However, for adapting the protocol to GUVs, as described in **Chapter 5**, the assay protocol was altered slightly. During the sonication step to form micelles, MGD1 was not added. MGD1 was added separately after the sonication step. The change in the protocol was to mimic the LgtC reaction with GUVs (see Chapter 5), where GUVs are preformed and reaction initiated by adding LgtC and UDP-Gal. Aside from this, the protocol was followed as described.

3. Enzyme assay for LgtC

a) Establishing buffer conditions for a UDP-Glo assay

An enzymatic assay for LgtC using the UDP-Glo™ kit was developed in this work. Using lactose as the acceptor, two buffer conditions were tested:

- 20 mM Tris, 5 mM DTT, 5 mM $MnCl_2$, pH 8.5
- 20 mM HEPES, 5 mM DTT, 5 mM $MnCl_2$, pH 7.5

Both buffers were designed with consideration to the tolerable pH limits of the UDP-Glo™ kit, as described in the manufacturer's guidelines. HEPES buffer is based on previous activity studies with LgtC, which used neutral pH conditions (Persson et al., 2001). LgtC was reported to be very stable in the more basic Tris buffer condition, therefore this buffer condition was tested for its compatibility in an UDP-Glo assay (Chan et al., 2012).

For testing the HEPES and Tris buffer, LgtC reaction mix was composed of 25 ng LgtC (0.83 ng/µl), 62.5 mM Lactose, 0.5 mM UDP-Gal in either Tris or HEPES buffer. The enzyme reaction was left to run for 10 minutes at 25°C. Controls are reactions without lactose or UDP-Gal. After adding UDR, the plate was covered with aluminium foil, shaken for 30 seconds and left to stand for 1 hour at room temperature. Luminescence was read with a TECAN plate reader.

We also compared the LgtC activity at low lactose and LacCer concentrations (see manuscript in preparation in Chapter 5). DTT and $MnCl_2$ concentrations were lowered to 0.5 mM, to prevent the formation of precipitates on GUVs and SLBs. This was also done in lactose

reactions to allow comparison with LacCer reactions. The table below was used to assemble the reaction mixture for lactose activity (Table 8).

Composition of a LgtC reaction with low concentration lactose			
Reagent	**Full reaction (µl)**	**No lactose (µl)**	**No UDP-Gal (µl)**
Buffer 4x	25	25	25
LgtC 100ng/µl	10	10	10
Lactose 1 mM	10	0	10
Water	50	60	55
UDP-Gal 5mM	5	5	0
Final volume	100	100	100

Table 8: *Composition of a typical LgtC reaction with lactose using the HEPES buffer. Buffer 4x composition is: 80 mM HEPES, pH 7.5, 2 mM DTT, 2 mM $MnCl_2$. Conditions: 1 µg LgtC, 0.1 mM lactose and 0.25 mM UDP-Gal at 10-minute reaction time.*

b) Lactosylceramide assays

LacCer is formed into micelles for enzymatic assays. The critical micelle concentration of LacCer has been determined to be 150 nM (Ulrich-Bott and Wiegandt, 1984). This allowed for determination of the lower limit of LacCer concentration needed for micelle formation. The HEPES buffer tested previously was used as the basis for the conditions with LacCer assays. Due to formation of crystals as observed in GUVs, the buffer composition was modified with lower $MnCl_2$ and DTT.

C8 LacCer (Avanti) in Chloroform:Methanol (2:1) was added to a glass tube and dried under nitrogen for 20 minutes. The lipid was resolubilised in the reaction buffer: 20 mM HEPES buffer, pH 7.5, 0.5 mM $MnCl_2$, 0.5 mM DTT. Tubes were closed with caps and sonicated in an iced water bath for 10 minutes making sure the temperature did not exceed 10°C. LgtC was added to the tubes. The enzyme reaction is initiated by adding UDP-Gal (0.25 mM) (Table 9). Once the reaction was stopped, 25 µl of the reaction were transferred into a 96-well plate and mixed with 25 µl UDR. The plate was covered with aluminium foil, shaken for 30 seconds and

left to stand for 1 hour at room temperature. Luminescence was read with a TECAN plate reader.

Composition of a LgtC reaction with LacCer acceptor			
Reagent	**Full reaction (µl)**	**No LacCer (µl)**	**No UDP-Gal (µl)**
LacCer 1mM	Varies	0	Varies
Reaction buffer	25.5	25.5	27
LgtC (100 ng/µl)	3	3	3
UDP-Gal (5mM)	1.5	1.5	0
Final volume	30	30	30

Table 9: *Composition of a LgtC reaction with LacCer. Reaction buffer composition: 20 mM HEPES, 0.5 mM DTT, 0.5 mM $MnCl_2$, pH 7.5. Conditions: 300 ng LgtC, 0.25 mM UDP-Gal and a 10-minute reaction time. This reaction makes for 300 ng of LgtC but can be adjusted by changing the concentration of the LgtC, the volume added must remain the same. The same can be done for UDP-Gal if a different concentration is needed. Th reaction time can be adjusted freely.*

4. Thin layer chromatography

Thin layer chromatography (TLC) is a method of separation of compounds via their relative affinity towards the stationary phase (silica plate) and mobile phase (solvent). Mixture to be separated are spotted 1 cm on one side of the TLC plate. A small portion of the plate is submerged in a solvent of choice and allowed to migrate up the TLC plate. During this movement, compounds with higher affinity to the stationary phase move slower and therefore will have a lower position on the plate. After the solvent has almost reached the top of the plate, a staining reagent is applied to reveal the glycolipid product from the GT reaction.

GT reactions involving lipid acceptors were performed as previously described in UDP-Glo™ assays. On a Silica gel 60 F_{254} plate (Supelco), 5-10 µl of the reaction mixture were applied. The samples were transferred into a sealed container equilibrated with the eluent:

- MGD1 products: chloroform/n-hexane/2-propanol/tetrahydrofuran/water (25:50:40:0.5:1 by volume)

- LgtC products: chloroform/methanol/water (65:35:5 by volume)

After the solvent had migrated up to 1 cm from the top of the TLC plate, the plate was removed from the solvent and the solvent line was marked. Orcinol-sulfuric acid reagent was used for visualization of glycolipid products, which coloured purple when a heat gun is applied. The retention factor (Rf) value was calculated as the ratio between the distance travelled by the sample versus the distance travelled by the solvent front.

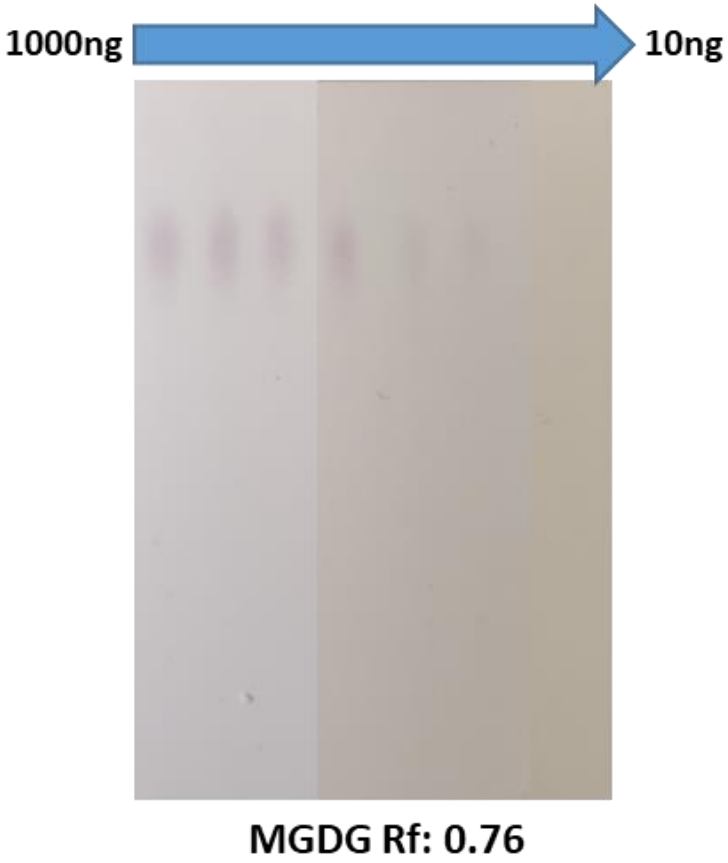


Figure 42: *Detection limit of orcinol-sulphuric acid staining for MGDG. The lowest MGDG amount that can be visually detected is 100 ng. Rf value for MGDG was calculated to be 0.76.*

IV. Applications of glycosyltransferases on synthetic membrane systems

1. Giant Unilamellar vesicles

a) Preparation of giant unilamellar vesicles (GUVs) - Principle

GUVs are useful as artificial cell models due to their size (>1 µm), which enable observation of membrane structure using light microscopy. GUVs are formed using the electroformation method. While the mechanism of electroformation is not fully understood, the general process involves the rehydration of a lipid film, swelling of the lipid film to form liposomes and fusion of adjacent liposomes due to mechanical stress (Li et al., 2016).

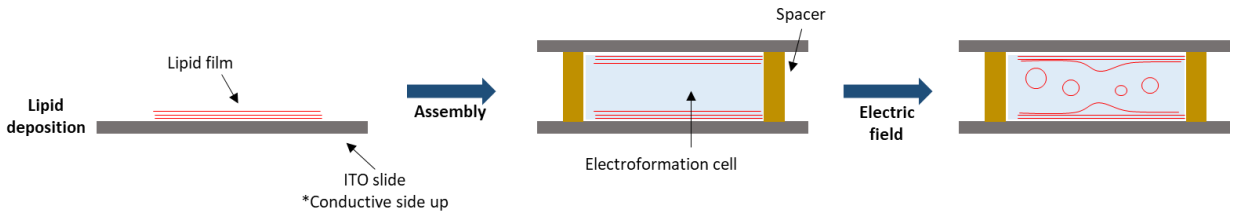


Figure 43: *General process of electroformation.*

The lipid solution is deposited on Indium-Tin-Oxide (ITO) covered slides and dried under vacuum in order to form a lipid film. Two ITO slides are assembled with a spacer (sigillum wax) between to form the electrofomation cell (lipid film on both slides of the cell). Sucrose or buffer solution is added to the cell and sealed with the spacer. Normally, sucrose is used during electroformation as the sucrose weight down the GUVs, making them sink in the chamber and immobilize them, which is ideal for microscopy. However, the osmolarity of the sucrose must match that of the buffer solution the GUVs will be subjected to. Otherwise, the difference in osmolarity will cause the GUVs to burst. GUVs are formed when an electric field is applied to the ITO slides. An alternating voltage (11 Hz, sine waveform, Vmax = 1.1 V) is applied for 2-3 hours. The slides are covered with aluminium foil throughout the whole electroformation to protect fluorescently labelled vesicles. A ramp generator is used to achieve a stepwise increase in voltage until Vmax is reached. This results in more homogenous results of electroformation (Figure 43).

For visualisation under confocal microscopy, observation chambers are built by attaching one or two glass cylinders to a round glass coverslip using UV-curable glue. The inside of the chamber is coated with β-casein for at least 30 minutes to prevent vesicle explosion upon contact with the chamber. Chambers are washed with ultrapure water and placed in a metal holder.

b) Tests with LgtC on GUVs

Lipid mixtures (total lipid concentration: 0.5 mg/ml) are prepared beforehand:

- Positive control: 64.7% DOPC, 30% Chol, 5% Gb3, 0.3% Atto-647 DOPE
- LacCer GUVs: 29,7% DOPC, 30% Chol, 40% LacCer, 0.3% Atto-647 DOPE

In the conductive side of two ITO slides, 15 µl of lipid were deposited. To ensure that all solvent was evaporated, the slides were placed under vacuum for at least 1 hour. Afterwards, the electroformation cell was assembled using sigillum wax as the spacer and tape to ensure it does not fall apart. Sucrose solution (318 $mOsm.l^{-1}$, ~300 mM) was added to the electroformation cell until it was full, sigillum wax was used to seal the cell, making sure there were no bubbles. Cable from the function generator was attached to each end of the ITO slide and an electric field was applied for 2-3 hours. After electroformation, the GUVs were transferred to an Eppendorf tube and covered with aluminium foil. GUVs can be stored at room temperature for 2 days.

For LgtC enzyme reaction, 30 µl of LacCer GUVs are added to a final 60 µl enzyme reaction mix. The final reaction mix consisted of: 20 mM HEPES, pH 7.4, 1 mM $MnCl_2$, 1 mM DTT, 100 ng/µl LgtC, 0.5 mM UDP-Gal and 40 mol-% LacCer GUVs (Table 10). Enzyme reaction was left to incubate for 1 hour. During this time, microscopy observation chambers were prepared.

To the observation chamber, 100 µl PBS (318 $mOsm.l^{-1}$) were added. Alexa Fluor 488 (Invitrogen) labelled Shiga toxin B (StxB-488) is prepared according to manufacturer's protocol. StxB-488 were added to the chamber, mixed with the PBS, to a final concentration of 500 nM. Once LgtC reaction was stopped, the treated GUVs are added, typically 5-10 µl, to the chamber. Images were acquired by confocal fluorescence microscopy (Nikon Eclipse Ti-E inverted microscope using a Nikon A1R confocal laser scanning system with laser lines: 405

nm, 488 nm, 561 nm, 640 nm; 60x oil immersion objective, NA = 1.49; Nikon Instruments, Inc.) and analysed using the corresponding NIS Elements software (NIS Elements Confocal 5.20, Nikon Instruments, Inc.) and ImageJ (Fiji win64, Open source).

Composition of LgtC reaction with GUVs	
	Volume (µl)
LacCer GUVs	30
4X Buffer	15
MQ Water	7.2
LgtC (800ng/µl)	7.5
UDP-Gal (100mM)	0.3
Total volume	60

Table 10: *Composition of LgtC reaction with GUVs. 4X Buffer composition: 80 mM HEPES, 4 mM DTT, 4 mM $MnCl_2$, pH 7.5.*

c) Tests with MGD1 on GUVs

Several lipid mixtures were prepared:

- MGDG+PG GUVs: 89.7 mol-% DOPC, 7.5 mol-% MGDG, 2.5 mol-% PG, 0.3 mol-% Atto-647 DOPE
- MGDG+DOPC GUVs: 92.2 mol-% DOPC, 7.5 mol-% MGDG, 0.3 mol-% Atto-647 DOPE
- DAG+PG GUVs: 89.7 mol-% DOPC, 7.5 mol-% DAG, 2.5 mol-% PG, 0.3 mol-% Atto-647 DOPE
- PG+DOPC GUVs: 97.2 mol-% DOPC, 2.5 mol-% PG, 0.3 mol-% Atto-647 DOPE
- DOPC GUVs: 69.7 mol-% DOPC, 30 mol-% Chol, 0.3 mol-% Atto-647 DOPE

All GUVs composed of the lipid mixtures described above were prepared using the same protocol previously described.

Detection of MGDG on GUVs using fluorescent lectins

MGDG+PG GUVs were formed. Lectins: RPL-Gal1, RPL-Gal4 and GSI-B4 were labelled with Alexa Fluor 488. 100 µl PBS (adjusted for osmolarity) was added to observation chambers. Fluorescently labelled lectins were added to the chambers at 200 nM. 5-10 µl of GUVs were added to the chambers. Chambers were visualised under fluorescence microscopy.

Binding of fluorescently-labelled MGD1 to GUVs

Wild-type MGD1 was labelled with Alexa Fluor 488 following the manufacturer's protocol. MGDG+PG, MGDG+DOPC, DAG+PG, PG+DOPC and DOPC GUVs were formed. GUVs (~20 µl) were added to chambers containing the MGD1-AF488 (200 nM) and fluorescence images were obtained as described previously.

2. Tests of LgtC on supported lipid bilayer using quartz crystal microbalance with dissipation (QCM-D)

a) QCM-D principle

QCM is an acoustic technology. It is a real-time, surface sensitive technique that detects mass changes on the surface of a sensor. In the centre of this technology is the oscillating unit, i.e. the sensor, which is a quartz crystal disk sandwiched between two electrodes. Piezoelectricity is a key phenomenon in QCM that relates the electrical and mechanical state of a material. Quartz is a piezoelectric material and its mechanical deformation gives rise to charge and vice versa. At an alternating voltage the quartz disk is excited to resonance and this is related to thickness (mass) of the disk, which can be quantified by the Sauerbrey equation. If the thickness changes so will the resonance frequency (f), and we can detect small changes in crystal mass by monitoring the changes in frequency (Δf).

An additional parameter that can be measured is the energy loss in the system, also called dissipation (D), hence the technique becomes QCM-D. Dissipation allows characterisation of how soft or viscoelastic is the layer on the crystal. Dissipation can give information if the film is rigid or not, and this is important whether quantification of mass using the Sauerbrey equation is possible (only possible in thin and rigid films). In hydrated systems, such as

membranes, the Sauerbrey equation cannot be used, but the *D* value is a valuable input in viscoelastic modelling and quantification of mass, thickness and viscoelastic properties of the layer.

Also important in QCM-D is the different harmonics, *n*, at which a QCM sensor can be excited to resonate. QCM crystals are excited in odd harmonics (*n* = 1 ,3, 5…), with the lowest n=1 called the fundamental and the others as overtones of the fundamental. Multiple overtones can give unique information about the system being studied and can contribute to viscoelastic modelling.

In this work, QCM-D will be used to monitor LgtC action on a SLB formed on the sensor. StxB will serve as a reporter for the detection of the reaction product, Gb3, on SLB. Throughout the experiments Δf and Δ*D* are monitored.

b) Tests with LgtC on supported lipid bilayer (SLB)

The lipid bilayer was formed by the adsorption of small unilamellar vesicles (SUVs), containing LacCer, to the sensor until equilibrium (i.e. the whole sensor surface was covered). LgtC and UDP-Gal were flowed to the system to initiate the GT reaction on the membrane. Afterwards, the system was washed with buffer to remove LgtC and UDP-Gal. Then StxB was flowed into the system and binding was monitored via the Δ*f*. Real-time glycosylation was also monitored and LgtC characteristics on a membrane was analysed.

Lipid mixtures with different compositions of LacCer or Gb3 were prepared.

- DOPC
- 1% LacCer and 99% DOPC
- 5% LacCer and 95% DOPC
- 20% LacCer and 80% DOPC
- 5% Gb3 and 95% DOPC
- 20% Gb3 and 80% DOPC
- 40% Gb3 and 60% DOPC

Silica-coated sensors (QSX303; Biolin Scientific, Västra Frölunda, Sweden)) (Biolin Scientific) and Q-Sense E4 system (Biolin Scientific), connected to a syringe pump (Legato; World Precision Instruments, Stevenage, UK), were used for QCM-D measurements. QSoft Analyzer software was used for data acquisition.

Small unilamellar vesicles (SUVs) containing DOPC and C8 LacCer are prepared. Lipids in chloroform are mixed at desired molar ratios (described above) at a total amount of 5 µmol, and dried under a stream of nitrogen gas, followed by drying in a vacuum desiccator for 2 h. The lipid mixtures were re-suspended in working buffer (10 mM HEPES, 150 mM NaCl, pH 7.4) at 2 $mg \cdot ml^{-1}$, and homogenized by five cycles of freezing, thawing and vortexing. To obtain SUVs, the lipid suspension was subjected to tip sonication in pulse mode (1 s on / 1 s off) for 15 min with refrigeration. The SUV suspension was then centrifuged at 12,100 × g for 10 min to remove titanium debris (shed from the sonicator tip). SUVs were aliquoted into 600 µl tubes, covered in nitrogen gas and stored at 4^{o}C until use.

QCM-D sensors were pre-treated in UV/ozone (30 min), placed in the flow chamber, and connected to the Q-Sense E4 apparatus. Following acquisition of a baseline in HBS buffer (10 mM HEPES, 150 mM NaCl, pH 7.4), SLBs were formed on the sensor surface by the flow of 50 µg/ml SUVs until frequency and dissipation were stable. SLBs were equilibrated first in **HBS** buffer, and then in reaction buffer (**RB**; 20 mM HEPES, 0.5 mM MnCl2, 4.4 mM Tris, 1.1 mM TCEP, pH 7.4), which was matched in content with the LgtC storage buffer to avoid solution effects on the QCM-D response upon starting the enzyme reaction. Enzyme reaction was initiated in RB containing 250 µM UDP-Gal and desired concentrations of LgtC (17.2 µM or 172 µM). To stop the LgtC reaction, SLBs were washed, first with RB and then with HBS for further analysis with StxB. A flow of 6.5 µM StxB in HBS was applied and StxB binding was observed. Δf was recorded at 10 minutes after applying StxB. Maintained throughout the experiment was a flow rate of 20 µl/min. Measurements were recorded and the software's Dfind and QSoft Analyzer was used for data analysis.

V. Molecular modelling and simulations

All molecular modelling and simulations in this PhD were performed by our collaborator, Olga Makshakova.

a) Docking methods used to investigate nucleotide sugar specificity (Gal vs Glc):

The simulations were done using local search docking protocol to keep the precise position of UPD moiety in the binding site, determined by X-ray. The orientation of the sugar was allowed to sample in the process of docking. The protein was considered as rigid but only the residues mutated were allowed to move. The calculations were performed in AutoDock4.2 using Gasteiger charges on atoms.

b) Molecular Dynamics (MD) simulations:

The calculations were performed using Gromacs engine and force fields Martini2.2 for Coarse-Grain simulations and Charmm36 for All-atom simulations. The protein was pre-oriented in respect to the membrane via calculating the free energy of its transfer from non-polar medium to water. Then the membrane with bound protein was placed into water surrounding and the charge was equilibrated by counter ions. An additional amount of Na and Cl ions was put to reach the ionic strength of 150 mM. The system was energy minimized and equilibrated in the nvt ensemble. The production phase was calculated in the npt ensemble at the pressure of 1 bar and temperature of 300K. The length of the production phase was 3 μs.

Chapter 3

Study of MGD1 sugar donor specificity

Galactose vs Glucose

I. Objective

As discussed earlier, the GTs used in most applications have promiscuous substrate specificity. However, not all GTs have such relaxed selectivity. Due to this, it is of importance to be able to understand how specificity is conferred. MGD1 is one example of GT highly selective for its donor and acceptor substrates. It catalyses the transfer of a galactose residue from UDP-Gal to its lipid acceptor diacylglycerol (DAG). MGD1 has been extensively studied at CERMAV and many biochemical and structural data are available for this GT which makes it an ideal model for understanding how sugar donor specificity is conferred. The understanding of how galactose specificity is conferred in MGD1 and which mutations are required to change this can be helpful in engineering other GTs with altered specificities.

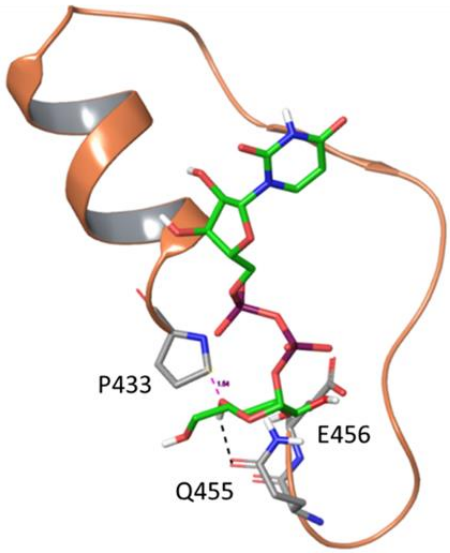


Figure 44: *Structure of the 'PQE' region conferring sugar donor specificity in MGD1. The substrate, UDP-Glc, was placed into the binding site of MGD1 (pdb code: 4X1T) by transferring the coordinates from the complex of Medicago truncatula UGT71G1 (pdb code: 2C1Z) to MGD1, whose structural region shown was superimposed by Ca-atoms.*

In this chapter we aimed to change the specificity of MGD1 from UDP-Gal to UDP-Glc. The main structural difference between these two monosaccharides is the orientation of the OH group at the C4, and so this objective seemed attainable. From previous structural and modelling studies, the residues P433, Q455, and E456, named the '**PQE'** motif, are predicted to be the key residues that confer galactose specificity in MGD1 (Figure 44). This is due to their role in interaction with the sugar moiety of the donor (Rocha et al., 2016b). In the galactose moiety, the Q455 side chain forms a hydrogen bond with O3 and

E456 side chain forms a hydrogen bond with O2. Looking at the crystal structure of MGD1 and from modelling of the galactose moiety into the active site, it is clear that P433 causes a steric clash with the O4 of glucose but not in galactose (Figure 28). Previously, a P433A mutant was produced in the lab which was designed to prevent this steric clash. This mutant lost the majority of activity for UDP-Gal (<5% of residual activity) but the mutation did not confer activity for UDP-Glc (Rocha et al., 2016b). Results from the P433A mutant suggest that perhaps additional mutations are required to confer glucosyltransferase (GlcT) activity to this mutant.

In order to rationalise possible mutations to test, we analysed the signature residues responsible for donor specificity in two families of GTs. Firstly, structural homologs of MGD1 in the GT1 family and which demonstrate glucosyltransferase (GlcT) activity. Secondly, GT28 family of proteins, which are the most related to MGD1 (>30% protein sequence similarity).

II. MGD1 mutants based on GT1 family glucosyltransferases

1. Structure and sequence analysis of structural homologs in GT1

When MGD1 was first crystallised, structural homologs were found in the GT1 family, even though there is very little sequence similarity (Rocha et al. 2016b). There are several fully characterised GlcT enzymes in the GT1 family with available crystal structures. One strategy for planning mutants was to take inspiration from sugar donor interactions in GT1.

In the GT1 family, there are three GT1 proteins (PDB codes: 2C1Z, 2ACW and 2PQ6), which are plant flavonoid GlcT enzymes (Offen et al. 2006; Shao et al., 2005; Li et al., 2007). We compared the equivalent 'PQE' region in the crystal structures of the three GT1 proteins and MGD1. As structural homologs we find remarkable similarity of the sugar donor binding region in the protein structures. The equivalent 'PQE' motif in the three GT1 proteins is 'W-D/E-Q'. Of note there is the inversion of the 'QE' into 'EQ' or 'DQ'. The exact reason for the inversion of residues is unknown as they still interact with the same hydroxyl groups in the sugar (O2 and O3).

A)

'PQE' sequence comparison – MGD1 vs GT1 Glucosyltransferases	
MGD1	GPGTIAEAMIRGLPIILNGYIAGQE
2C1Z	GWNSLWESVAGGVPLICRPFF-GDQ
2PQ6	GWNSTTESICAGVPMLCWPFFA-DQ
2ACW	GWNSILESMWFGVPILTWPIYA-EQ

B)

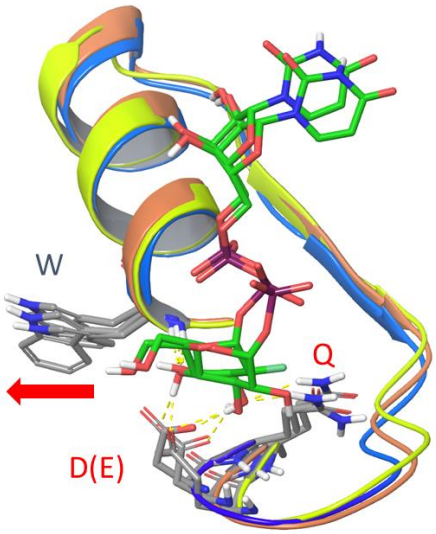


Figure 45: *(A) sequence comparison of the 'PQE' motif (highlighted in green) of MGD1 and the equivalent in three plant GT1 glucosyltransferases. Highlighted in blue is the highly conserved glycine in all four GTs. (B) structural superimposition of the 'PQE' motif in the three GlcTs from GT1: VvGT1 from Vitis Vinifera (PDB: 2C1Z, complex with UDP-2-Fluoro-Glc), UGT71G1 (2ACW, complex with UDP-Glc) and UGT85H2 (2PQ6, apo form) from Medicago Truncatula. The red arrow indicates the Trp side chain (considered as the equivalent P433 in MGD1) facing away from the sugar in the GT1 proteins. Blue structure is UGT71G1, orange is UGT85H2 and yellow is 2C1Z. Structural analysis by Olga Makshakova.*

Looking to mutate P433, in order to remove the proline ring steric clash, we look at the GT1 equivalent. In the GT1 structures, the P433 equivalent is a tryptophan residue. From the crystal structures of GT1 enzymes we observed that the side chain of tryptophan protrudes to the side of the helix, away from the sugar donor, and therefore does not cause a steric clash with the O4 of glucose (Figure 45). This might be the reason why UDP-Glc can be accommodated into the active site of the three GT1 enzymes. However, molecular modelling showed that in MGD1 there is no room to accommodate the large tryptophan side chain at this position, in contrast to GT1 enzymes (data not shown). Therefore, a P433W mutation was not considered.

Furthermore, we looked at a broader library of GT1 sequences, of various known functions including GlcTs and GalTs, in order to compare the equivalent 'PQE' motifs. We observed the most variation in the P433 position, where it is Trp in plant GlcTs, but can be Ser, Thr, Met or Ala in mammalian species. In GalTs, it is Cys, Trp, Leu or Gln. This is followed by the inversion of the acidic and neutral residues

('EQ' or 'DQ') in all the GlcT proteins. In GalT a His residue replaces the Gln residue (Figure 46). However, none of the examined sequences examined had a Pro at this position.

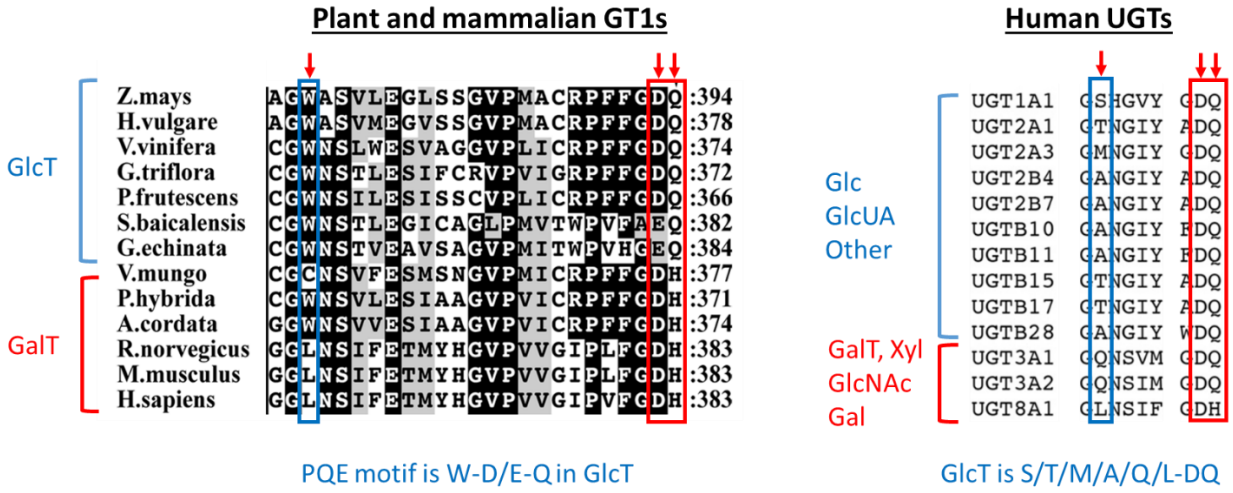


Figure 46: *Sequence alignment of glycosyltransferases from plant GT1 (Kubo et al., 2004) and human UGT (Nair et al., 2015). Red arrows indicate the positions that are the equivalent of the PQE motif in MGD1.*

Another important observation is a highly conserved Gly residue positioned before the P433 or equivalents. In all sequence analyses we have performed for GT1 family and MGD1 this remains conserved (Figures 48). While the role of this Gly is unknown, its conservation highlights its importance. However, in the context of altering sugar donor specificity this position was not relevant due to its conservation in GTs with other sugar donor specificities.

There are too many variations in the P433 position, with most of the potential candidate residues being unsuitable in MGD1 due to side chain clash with the sugar donor moiety. As a basis, to remove the proline ring and prevent the O4 steric clash, we maintained a P433A mutation. This was also ideal as the P433A mutant (thus named AQE) expressed and purified similarly to wild-type (WT) MGD1, highlighting the stability and folding of this mutant. For further mutations we looked at the Q455 and E456 positions. Based on the inversion of the neutral and acidic residues in the Q455 and E456 position as observed in GT1 glucosyltransferases, we produced two mutants: a double mutation P433A+455E (**AEE**) and a triple mutation P433A+Q455E+E456Q (**AEQ**).

2. Expression of wild-type MGD1 and AEE and AEQ mutants in *E. coli*

The protocol for MGD1 expression had been previously established in the laboratory (Rocha et al., 2013). A His-tagged version of the catalytic domain of MGD1 comprising residues 137 to 533 (named His6-MGD1Δ137) can be produced as a soluble form in *E. coli* cells (Figure 47A).

AEE and AEQ mutants were expressed in the same conditions as the wild-type MGD1: induction at OD_{600nm} ~0.6, by addition of IPTG (0.25 mM) for 4h at 30°C. The gel shows overexpression of MGD1 mutants with the majority soluble in the supernatant (Figure 47B). This is comparable to the expression of wild-type MGD1 and also indicates that the AEE and AEQ mutations does not perturb the folding of the protein.

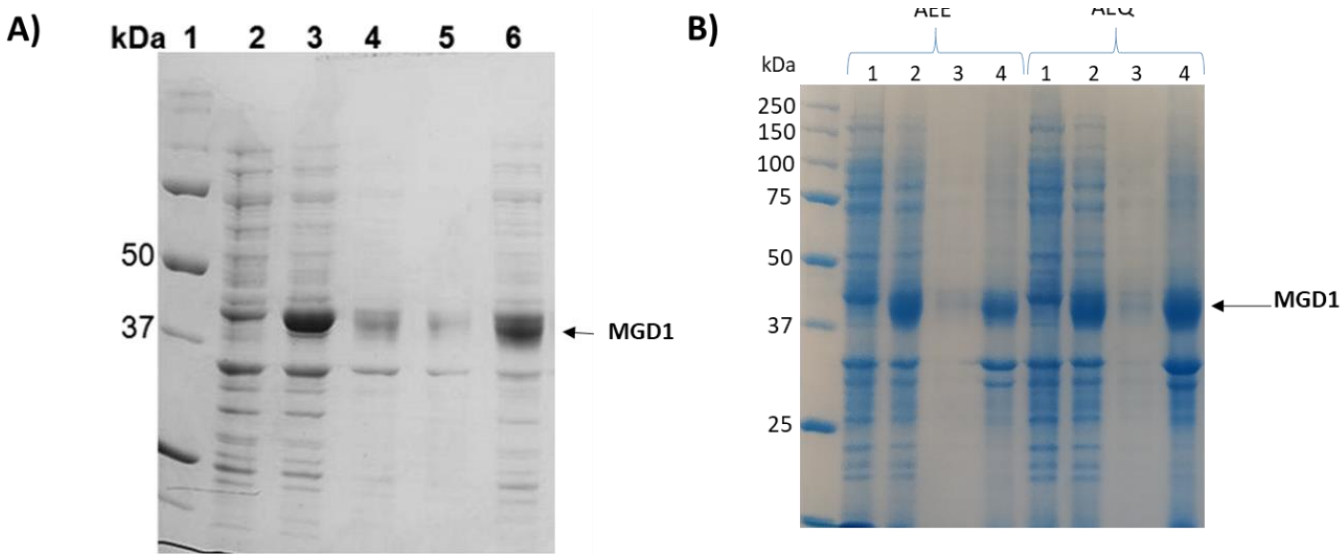


Figure 47: *Small scale expression trial of wild-type MGD1 and mutants AEE and AEQ. A) SDS-PAGE analysis of wild-type MGD1 expression in E. coli (BL21(DE3)) with lanes showing molecular ladder (1), total cell extract before (2) and after IPTG induction (3), insoluble fraction (4), membrane fraction (5) and soluble fraction (6) from induced cells (from Rocha et al., 2013). B) SDS-PAGE gel showing small scale expression of AEE and AEQ mutants. Samples taken are from total cell extract before (1) and after IPTG induction (2), insoluble fraction (3), and soluble fraction (4) from induced cells.*

3. Purification of wild-type MGD1 and AEE and AEQ mutants

The protocol of purification of MGD1 had been previously established (Rocha et al., 2013). It comprises of a two-step procedure: immobilized metal affinity chromatography (IMAC) using a HisTrap column and size-exclusion chromatography (SEC). MGD1 being a monotopic membrane protein, it has a strong tendency to precipitate during purification. The composition of buffers used in the purification scheme was optimized to avoid protein aggregation (Rocha et al., 2013). The same buffer was used for cell lysate and IMAC purification: 25 mM HEPES buffer at pH 7.5 (for optimal binding to the Ni^{2+}-resin), 0.5 M NaCl, 1 M urea, 1 mM DTT, 10 mM imidazole. The buffer used for SEC is: 25 mM Bis Tris pH 6.5, 150 mM NaCl, 1 mM TCEP, 5% glycerol. This buffer was optimal for activity and storage of the protein

Figure 48 illustrates what is classically obtained with wild-type MGD1 using this procedure.

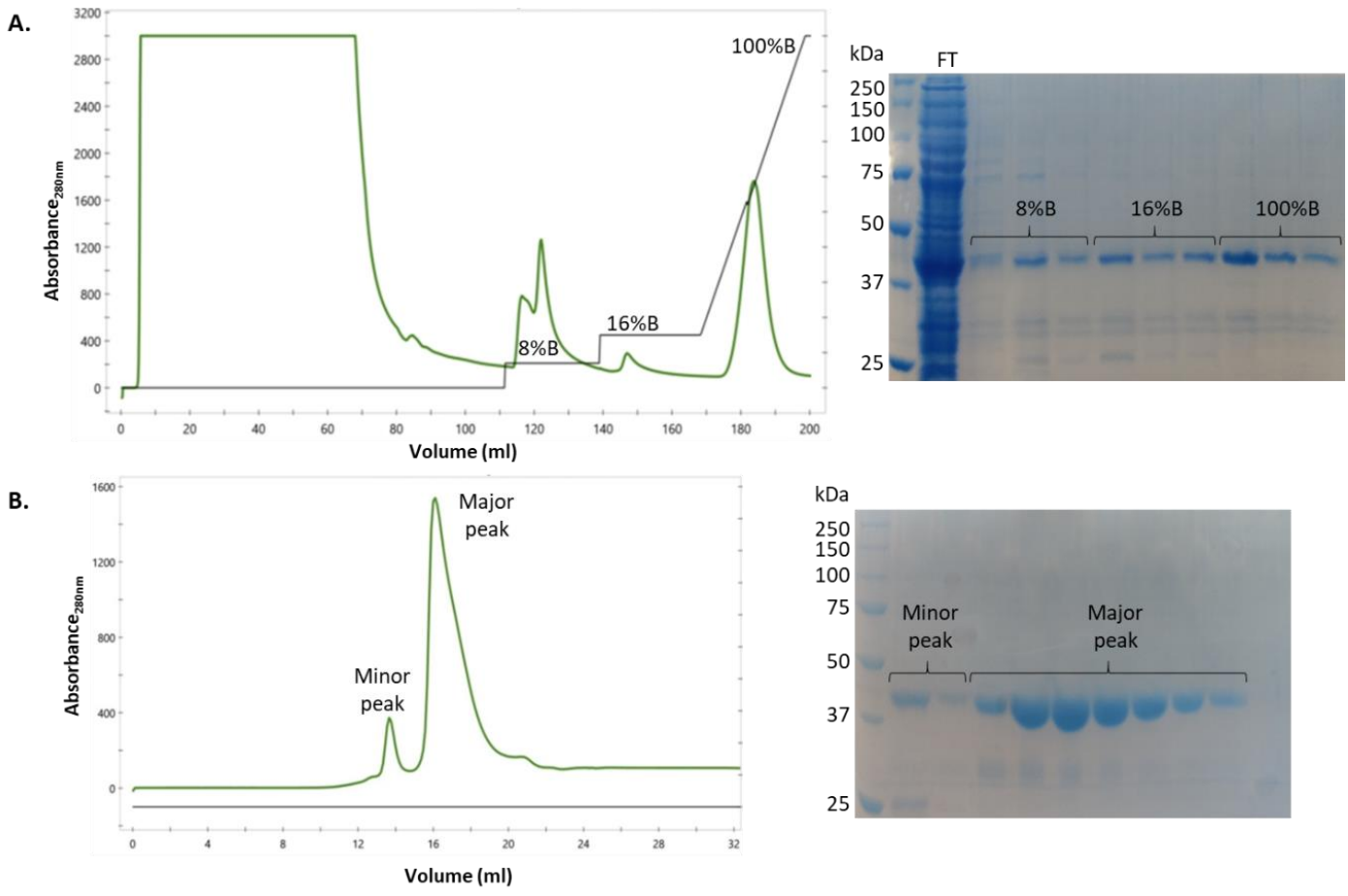


Figure 48: *IMAC (A) and SEC (B) purification of wild type MGD1 from a 4L culture. A) Included wells for SDS-PAGE analysis: flow through (FT), elution fractions from 8%, 16% and 16-100% imidazole. B) Included wells for SDS-PAGE analysis: elution fractions from the minor and major peaks in SEC.*

As shown in SDS-PAGE (Figure 48A), after a single step IMAC purification, the purity of protein was already satisfactory (estimated above 90-95%). The major peak observed in Figure 48B, after SEC, corresponded to a monomeric form of MGD1 (~45 kDa). The second step was necessary to remove a small fraction of dimeric MGD1 (minor peak). The protein was then safely concentrated (up to 10 mg/mL) and stored at -20°C until use. Typically, using this purification procedure, 5-10 mg of pure and homogenous protein from a 1L culture was produced.

The AEE and AQE mutants were purified using the same procedure. Similar to the purification of wild-type MGD1, the mutants were detected in all elution fractions but, judging from band intensity, the majority of protein eluted at 16-100% imidazole. However, it seems that AEE and AQE mutants were more poorly retained in the IMAC column compared to the wild-type. Analysis of the eluted fractions indicated that 16-100% imidazole fractions were deemed pure enough and did not require an additional SEC purification. Therefore, fractions were pooled and buffer exchanged into the storage buffer.

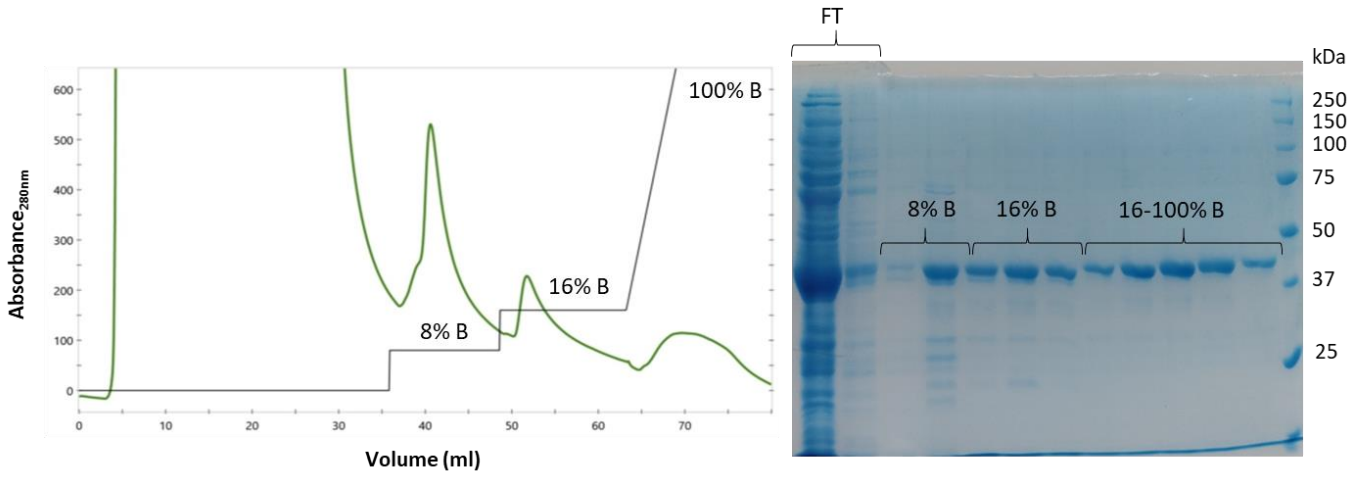


Figure 49: *IMAC purification of AEE mutant from 1L culture and the corresponding SDS-PAGE analysis of fractions from purification with focus on elution profile. Included wells: flow through (FT), elution fractions from 8%, 16% and 16-100% imidazole.*

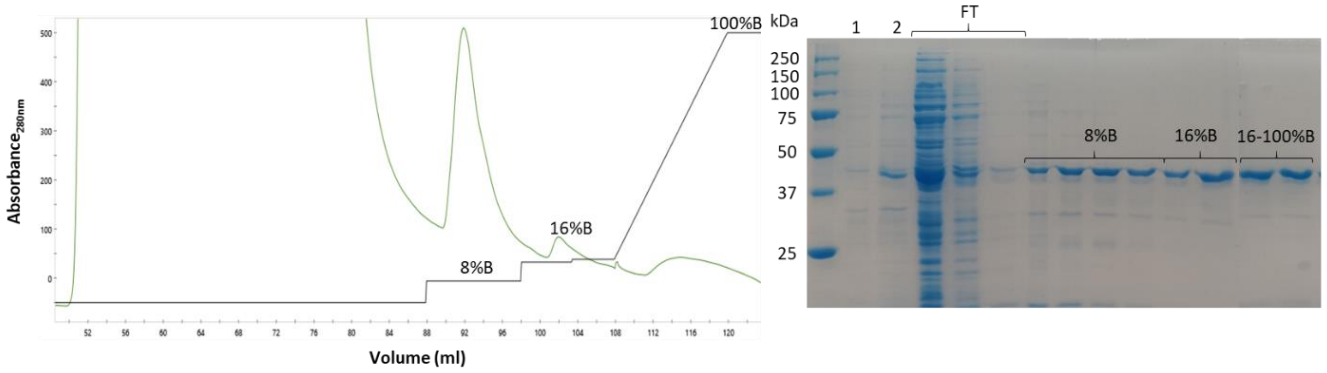


Figure 50: *IMAC purification of AEQ mutant from 1L culture and the corresponding SDS-PAGE analysis of fractions from purification with focus on elution profile. Included wells: total cell extract from before (1) and after (2) IPTG induction, flow through (FT), elution fractions from 8%, 16% and 16-100% imidazole.*

4. Enzyme assays of wild-type MGD1 and AEE and AEQ mutants

We used the highly sensitive Bioluminescent UDP-Glo™ GT assay from Promega, previously adapted in the laboratory to monitor MGD1 activity (Nitenberg et al., 2020). This enzyme test was developed to replace the classical radioactivity-based assay that was used previously (Rocha et al., 2013). The principle of the assay was to convert UDP, the product of the GT reaction, to ATP to generate light through a luciferase reaction. Briefly, the enzyme tests were done at 25°C in 25 mM Bis-Tris, pH 6.5, 150 mM NaCl, 1 mM TCEP or DTT, 5% glycerol, 6 mM CHAPS, DAG (6.3 mol-%), PG (1.5 mol-%), 700 µM UDP-Gal (or UDP-Glc), MGD1 or mutant, as described in the Materials and Methods. Control reactions were done in the absence of acceptor. The amounts of enzyme and the reaction time vary. For WT MGD1, typically the amount of MGD1 used was 0.14 ng/µL (5ng MGD1 per 36.5µL reaction), and a reaction time of 10 minutes. Using these reaction conditions, the specific activity of MGD1 in the presence of UDP-Gal and PG is estimated about 482 ± 21 nmol/min/mg protein. The WT enzyme showed no GlcT activity in these conditions.

Previously, the AQE mutant, where Pro433 has been replaced by Ala, was assayed which showed no GlcT activity and also reduced GalT to less than 5% activity (Rocha et al., 2016). AQE is the basis of the two other MGD1 mutants based on the GT1 family, AEE and AEQ. AQE mutant was again tested to confirm previous results. AQE mutant showed about 4% GalT activity compared to the WT enzyme and no GlcT activity (Table 11).

Activity of GT1 based MGD1 mutants		
	Activity (% of WT with UDP-Gal)	
	UDP-Gal	UDP-Glc
WT	**100**	0
AQE	4	0
AEE	<1	0
AEQ	0	0

Table 11: *Glycosyltransferase activity of MGD1 mutants AEE, AEQ and AQE in comparison to the Wild-Type MGD1 (WT) using the UDP-Glo assay.*

The AEE mutation further decreased GalT activity (less than 1% of WT). In the case of the AEQ mutant, GalT activity was completely abolished. Unfortunately, no GlcT activity was detected for AEE and AEQ mutants. This suggests that the mutations at positions 455 and 456 completely disrupted interactions with the galactose moiety of UDP-Gal, while bringing no positive interactions for accommodating a glucose instead.

III. MGD1 mutants based on the GT28 family

1. Sequence analysis of potential glucosyltransferases in GT28

MGD1 belongs to the GT28 family, which comprises to date only 3D structures for MGD1 and MurG, the latter being a GlcNAcT involved in peptidoglycan biosynthesis (Ha et al., 2000). However, sequences of bacterial glycolipid synthases are found in this family, and some of them have been described as GlcT.

RcMGlcD, from the photosynthetic bacterium *Roseiflexus castenholzii,* is described as a GT28 GlcT and has 33% sequence similarity to MGD1, making it a good basis for designing a GlcT activity to MGD1 (Yuzawa et al, 2011). Interestingly, there is very high sequence similarity in the 'PQE' region, with one key difference in RcMGlcD containing a 'GQE' motif instead (Table 12). Another interesting GT28 protein for comparison is CaMGD1 from *Caldilinea aerophila* which has 55% sequence similarity to MGD1 and is also activated by PA and PG (Nitenberg, 2018). CaMGD1 was described as possessing dual activity for both UDP-Gal and UDP-Glc with perhaps a preference for UDP-Glc due to a higher intensity of

monoglucosyldiacylglycerol (MGlcDG) bands observed in TLC (private communication – K. Awai). Sequence alignment shows a highly conserved protein sequence and that CaMGD1 also possesses the 'PQE' motif, thus questioning its dual activity (Table 12).

Sequence comparison of MGD1, RcMGlcD and CaMGD1		
Protein	**Sequence**	
MGD1	430-**K**AG**P**GTIA**E**	451-YIAG**QE**AG**N**
RcMGlcD	284-**K**AG**G**LSVS**E**	305-SAPG**QE**AG**N**
CaMGD1	286-**K**AG**P**GTIA**E**	307-FIPG**QE**EG**N**

Table 12: *Sequence comparison of MGD1, RcMGlcD and CaMGD1. Highlighted in blue is the 'PQE' motif or its equivalent in RcMGlcD and CaMGD1.*

While the capacity of RcMGlcD as a true GlcT has been demonstrated by TLC in a previous study, the authors used a crude extract of *E. coli* cells expressing the recombinant protein (Yuzawa et al, 2011). Similar to RcMGlcD studies, CaMGD1 also used *E. coli* crude extracts. The crude extract may contain other GTs that could influence the measured activities. To ascertain the true extent of the GlcT and/or GalT activity we expressed, purified and performed activity tests with RcMGlcD and CaMGD1.

2. Expression and purification of RcMGlcD and CaMGD1

The RcMGlcD gene was inserted into a pET29b vector and transformed into *E. coli* BL21 (DE3) cells for expression. Due to similar molecular weight and pI, RcMGlcD was expressed and purified using the same protocol as wild-type MGD1 (Figure 51). Small scale expression trials showed that RcMGlcD was largely insoluble (see appendix, Figure S4). However, a 1L expression of RcMGlcD was expressed and purified to see if we could obtain a small amount for activity test. Small bands on SDS-PAGE showed that RcMGlcD was mostly insoluble as only a small quantity of non-pure sample was obtained from affinity chromatography. Size exclusion chromatography was not attempted due to the low quantity of protein produced. Western blot confirmed the presence of His-tagged RcMGlcD in the sample. However, the presence of a second lower band (Figure 51C) suggeststhat the protein may have undergone proteolysis during expression, and this may explain the low yield. While we did not fully purify protein, the sample

is much purer than previous studies which only used crude extract for activity tests (Yuzawa et al, 2011). Due to time constraints no further attempts were made to optimise the expression and purification.

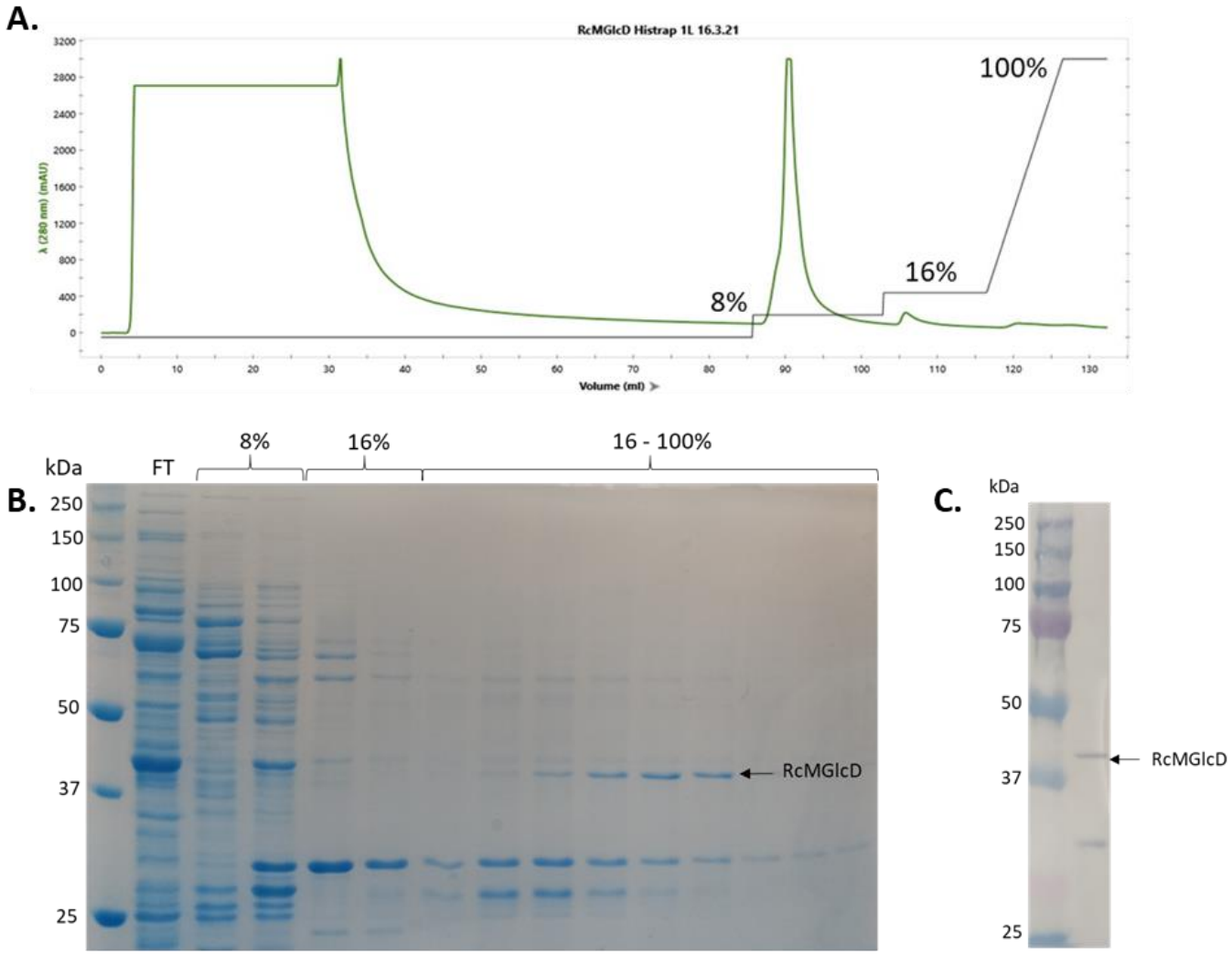


Figure 51: *Purification of RcMGlcD, SDS-PAGE and Western Blot analyses. A) IMAC purification. B) SDS-PAGE analysis of IMAC fractions. Included wells for SDS-PAGE analysis: flow through (FT), elution fractions from 8%, 16% and 16-100% imidazole. C) Western Blot of a pooled and concentrated sample from 16-100% imidazole fractions using anti-His antibody.*

The CaMGD1 gene was inserted into a pET29b vector and the protein was expressed and purified using the same method as MGD1 (Figure 52). The protein was successfully produced in *E. coli* and behave as MGD1. CaMGD1 has a molecular weight of 41 kDa which is close to wild-type MGD1 (44kDa), hence in SEC, CaMGD1 elutes in a similar elution profile to MGD1. The third peak contained pure monomeric form of CaMGD1 as it eluted at a similar volume to MGD1 (~17 ml). Purity was confirmed by SDS-PAGE

analysis (Figure 52B). CaMGD1 was also present in the first and second peak. The second peak is in a similar volume to that found for a dimeric form of MGD1 (Figure 48), which indicates the peak is a dimer of CaMGD1. Finally, the first peak is estimated to be >700 kDa, and may indicate aggregation of CaMGD1 with smaller proteins (~25 kDa) as they elute together with CaMGD1 in SEC.

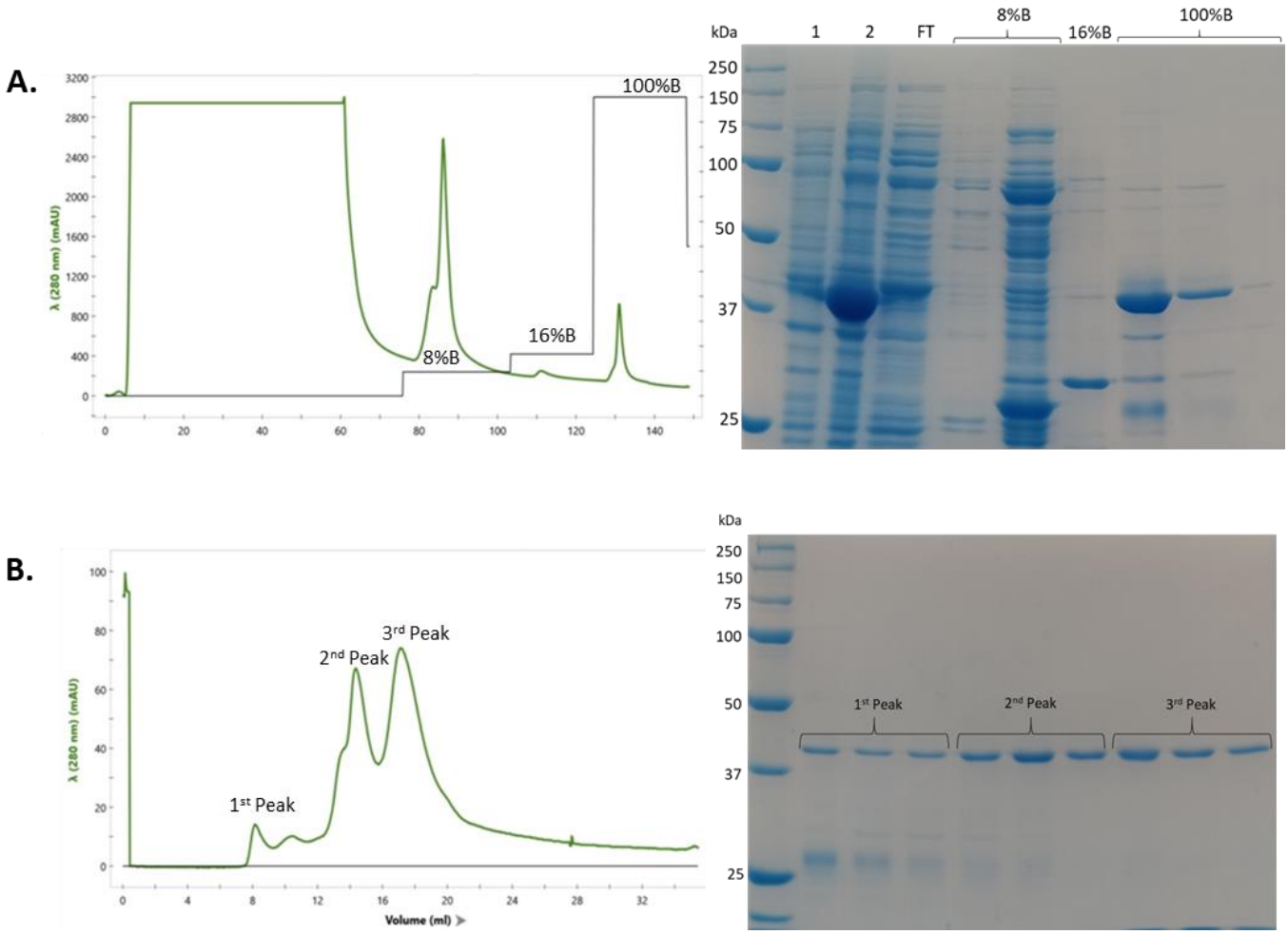


Figure 52: *Purification of CaMGD1 and SDS-PAGE analysis. A) IMAC purification. Included wells for SDS-PAGE analysis: total cell extract before (1) and after (2) IPTG induction and initial flow through (FT) from loading into the column. Elution fractions from 8%, 16% and 100% imidazole. B) SEC purification. Fractions from three different peaks analysed on SDS-PAGE.*

3. Activity assays of RcMGlcD and CaMGD1

Activity tests were performed using the UDP-Glo assay as described in the Materials and Methods. Results showed that RcMGlcD is a true GlcT protein but also possessed some residual GalT activity (Figure 53A). It was not possible to obtain a reliable specific activity as the sample tested was not pure enough. However, comparing bioluminescence readings of activity with UDP-Gal and UDP-Glc was enough to confirm the preferred sugar donor of RcMGlcD is UDP-Glc. To further confirm the GlcT activity, a 48-hour enzyme reaction was run on TLC to check for the formation of monoglucosyldiacylglycerol (MGlcDG). A spot migrating higher than MGDG was observed (Figure 53C). This is consistent with previous observations of migration of MGDG and MGlcDG on TLC (Yuzawa et al, 2011).

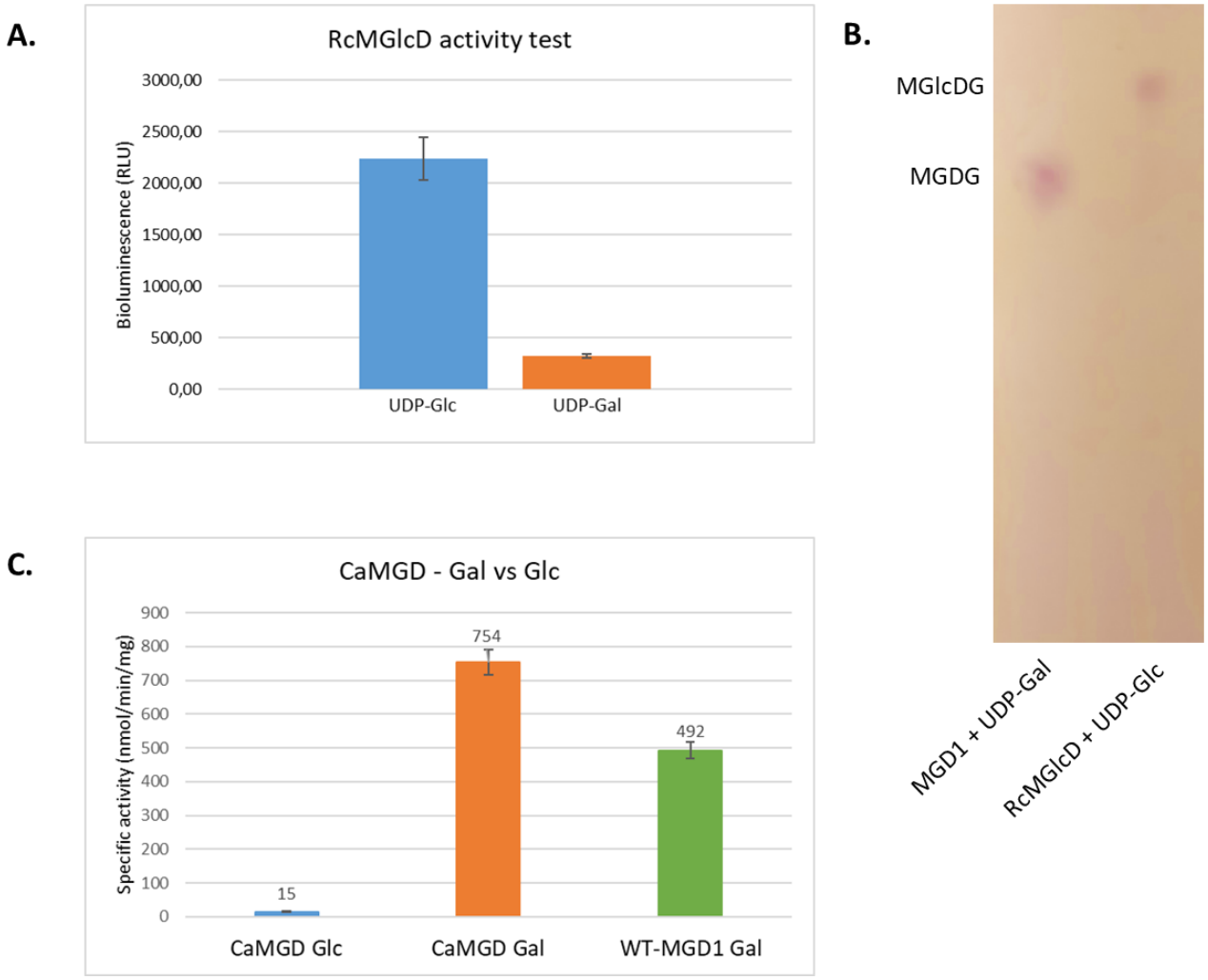


Figure 53: *Activity assays of RcMGlcD and CaMGD1. A) UDP-Glo assay comparing the activity of RcMGlcD using UDP-Gal and UDP-Glc as the sugar donor. B) TLC analysis of a 48-hour reaction of wild-type MGD1 with UDP-Gal (left lane) and RcMGlcD with UDP-Glc (right lane). C) Activity assay comparing CaMGD1 and MGD1. CaMGD1 was tested using UDP-Gal and UDP-Glc and wild-type MGD1 with UDP-Gal.*

CaMGD1 was tested for GalT and GlcT activity, and this was compared with wild-type MGD1. The data showed that CaMGD1 is a GalT with a rate slightly higher (~35%) to that of MGD1 (Figure 53B). Its specific activity is 754 ± 48 nmol/min/mg protein. Indeed, the GalT activity profile is close to that of MGD1, which maybe explained by the very high sequence similarity. In contrast, the GlcT activity of CaMGD1 is extremely low and only accounts for <2% of its GalT activity. Possibly, the role of P289 in steric blockade of glucose is much less pronounced in CaMGD1 allowing some GlcT activity. It may be interesting in a future project to mutate P289 of CaMGD1 to alanine or glycine and see if we can increase its GlcT activity further.

In conclusion, we have investigated two different GT28 family proteins with putative GlcT activity, RcMGlcD and CaMGD1. Of the two proteins, activity tests showed that CaMGD1 is predominantly a GalT protein, while RcMGlcD is a GlcT protein. Rational design of MGD1 mutants with GlcT activity was thus based on the RcMGlcD protein.

4. Rational design of MGD1 mutants based on RcMGlcD

Comparison of '**PQE**' motifs in MGD1 and RcMGlcD showed that the '**QE**' region is well conserved, including most of the surrounding residues (Table 13). However, we noted the presence of a Pro in position n-2 of the Gln residue. The environment of the '**P**' residue of PQE motif is interesting to consider: G**P**G in MGD1 and G**G**L in RcMGlcD. Considering that in MGD1, G**P**G is located at the end of an α-helix pointing towards the sugar moiety of nucleotide sugar (Figure 54), any change in this region should have an impact on the flexibility of this α-helix, and therefore on the binding to the nucleotide sugar.

Sequences of MGD1, RcMGlcD and MGD1 mutants		
Protein	**Sequence**	
MGD1	430-**K**AG**PG**TIA**E**	451-YIAG**QE**AG**N**
RcMGlcD	284-**K**AG**GL**SVS**E**	305-SAPG**QE**AG**N**
MGD1 mutants based on RcMGlcD		
GGG	430-**K**AG**GG**TIA**E**	451-YIAG**QE**AG**N**
GGA	430-**K**AG**GA**TIA**E**	451-YIAG**QE**AG**N**
GGL	430-**K**AG**GL**TIA**E**	451-YIAG**QE**AG**N**
RcMGD1	430-**K**AG**GLSVS**E	451-**SAP**G**QE**AG**N**

Table 13: *Sequence comparison between MGD1, RcMGlcD and mutants. Highlighted in blue is the 'PQE' motif important for donor specificity. In green is the G434 position (or equivalent in RcMGlcD) and in yellow are the other mutations introduced to the RcMGD1 mutant.*

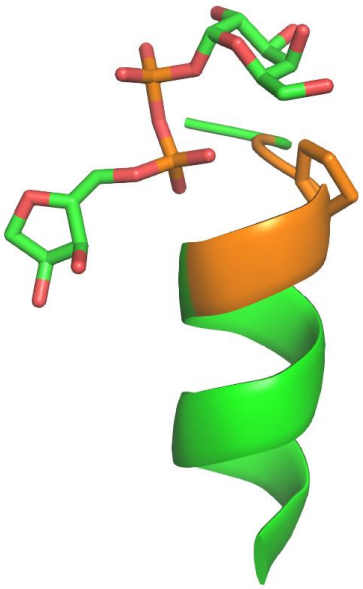


Figure 54: *MGD1 structure with UDP-Gal modelled into the active site. Coloured in orange in the α-helix are the positions of the 432-GPG sequence. Modelled by Olga Makshakova.*

Based on these observations, we designed four different MGD1 mutants; P433G (named GGG), P433G+G434A (GGA), P433G+G434L (GGL), and finally a mutant in which all residues of the 'PQE' motif of MGD1 were replaced by their equivalent in RcMGlcD, a mutant that was named RcMGD1 (Table 13).

The GGG mutant was designed based on the P433 residue being replaced by glycine, as seen in RcMGlcD (Table 13). This mutation generated a triple glycine sequence in the GGG mutant, which could be deleterious because it could provide too much flexibility in this region.

Mutants GGA and GGL were therefore designed to prevent a triple glycine sequence. A final mutant was designed, based on a complete transformation of the 'PQE' motif in order to match the residues observed in RcMGlcD (RcMGD1). Aside from the P433 and G434 positions of MGD1, there were six other positions that differed compared to RcMGlcD. Due to time constraints in testing each individual position, a mutant replacing all residues to that of RcMGlcD was made, called RcMGD1. In total, 8 amino acids were changed in RcMGD1.

5. Expression and purification of MGD1 mutants based on RcMGlcD

a) GGG mutant

The expression of GGG was considerably lower than the wild-type enzyme as we only observed a weak band after IPTG induction (Figure 55). To increase the purity of sample, the IMAC purification protocol was modified. The initial wash step at 8% imidazole was replaced by two wash steps done at higher imidazole concentrations, 20% and 40% Imidazole (100 mM and 200 mM). The goal was to remove more non-specific binding proteins that would remain in the nickel column.

As a result, some GGG mutant was wasted as a sizeable fraction of GGG eluted during the wash step (20%B), however, the fractions eluted from 40% and 100% imidazole (500mM) were deemed pure enough and did not need an additional SEC purification step. Fractions from 100% imidazole were pooled, concentrated and buffer exchanged into the storage buffer. This sample was in sufficient quantity to continue with activity tests.

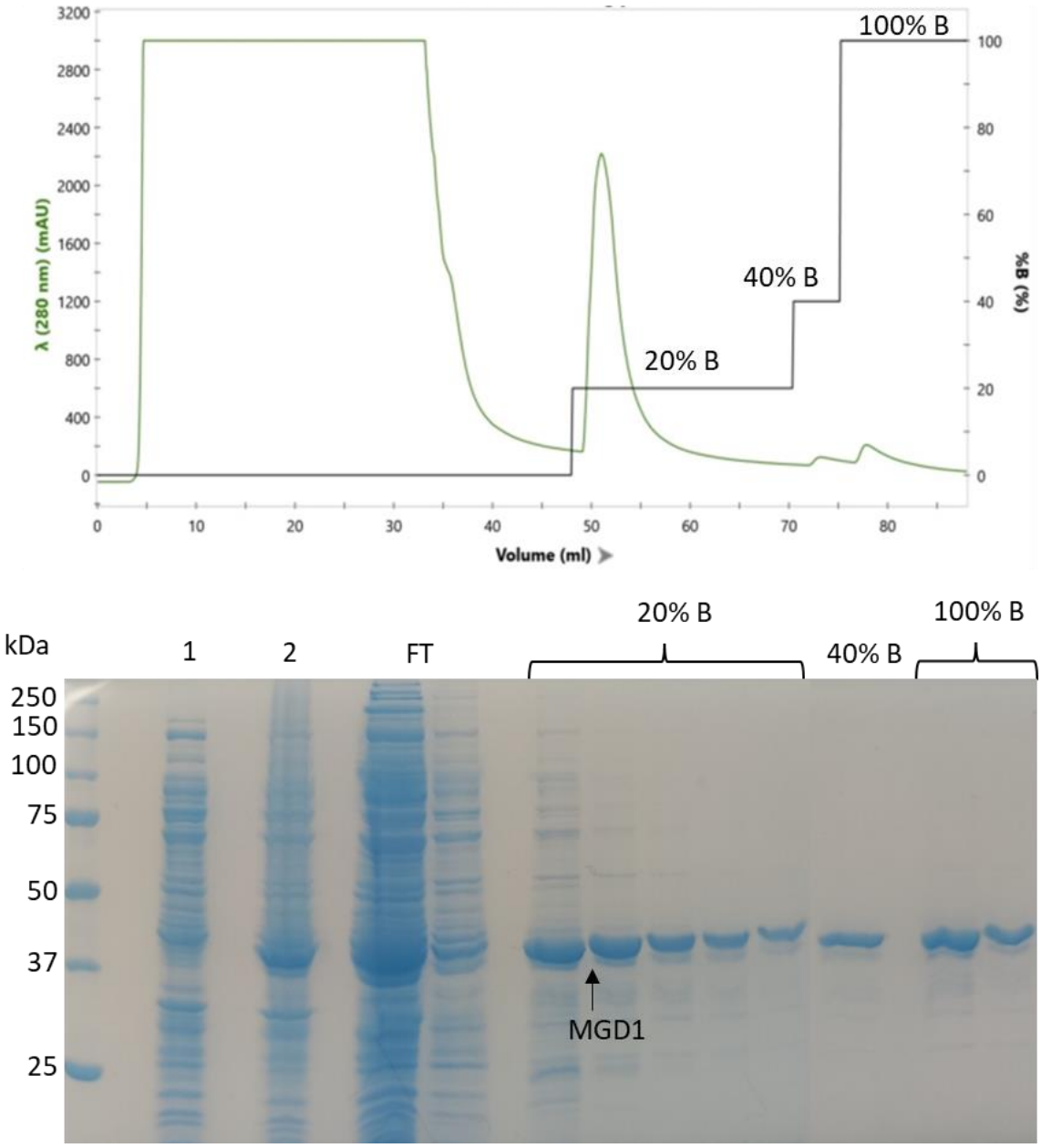


Figure 55: *IMAC purification of GGG mutant and SDS-PAGE analysis. Included wells for SDS-PAGE analysis: total cell extract before (1) and after (2) IPTG induction and initial flow through (FT). Elution fractions from 20%, 40% and 100% imidazole were also analysed.*

b) GGA mutant

The expression level of the GGA mutant was extremely low (Figure 56). This is surprising as the GGG mutant, which is hypothesized to be more unstable due to a triple glycine in the sequence, had a higher yield. To compensate for the low yield, a 4L culture was used for the purification of GGA mutant. We employed the same modified IMAC protocol used for the GGG mutant to increase purity but with an additional wash step at 15% imidazole (75 mM). Faint bands were observed from the 20% and 40% wash. As the yield was very low it was not ideal to do SEC purification. In the end, fractions from 40% and 100% were deemed pure and were concentrated and buffer exchanged to storage buffer. Because of the low yield, a western blot (using His-Tag antibody) was performed which confirmed the presence of GGA in the sample.

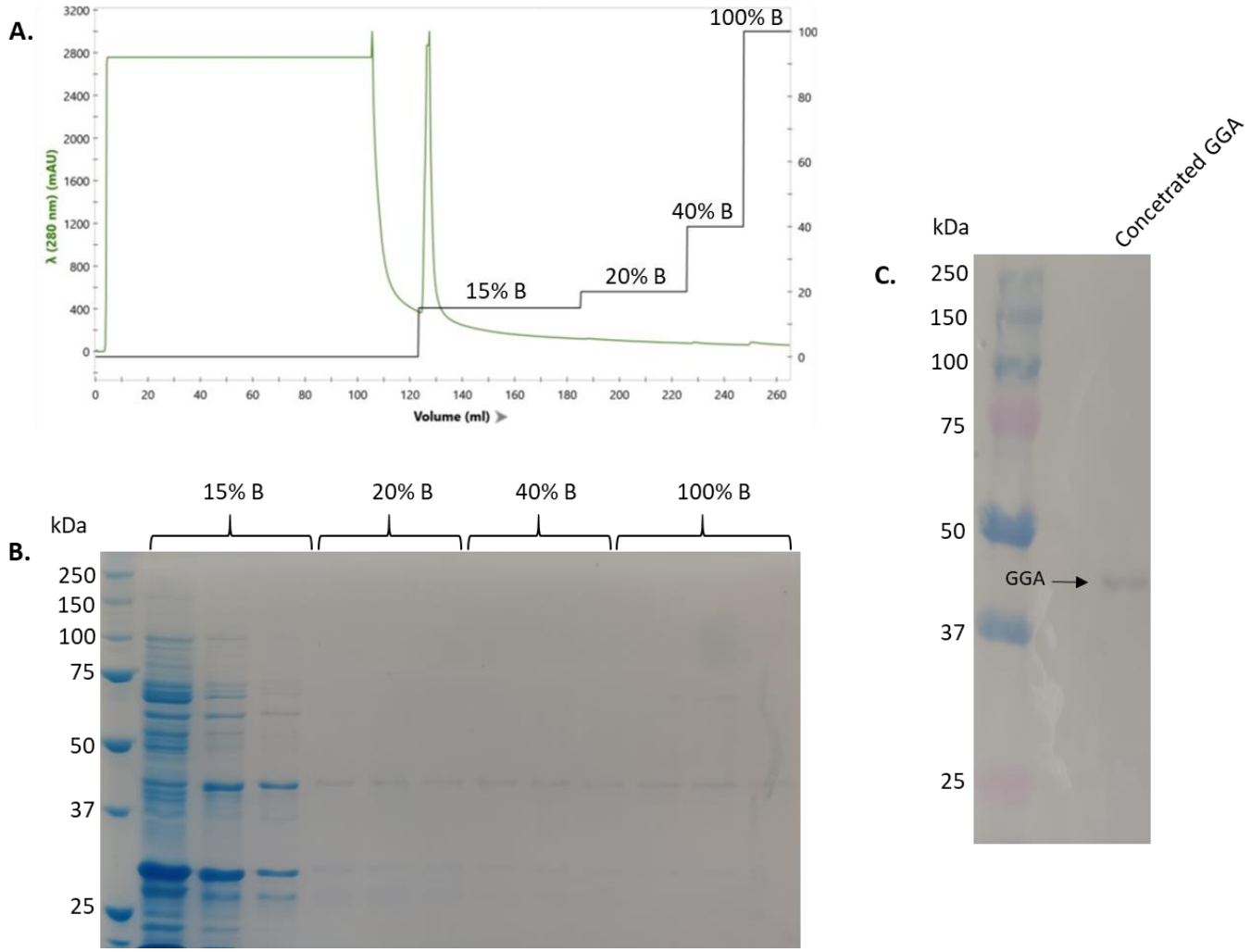


Figure 56: *IMAC purification of GGA mutant from a 4L culture with SDS-PAGE and Western Blot analyses. A) Chromatogram from IMAC purification. B) SDS-PAGE analysis from IMAC purification. C) Western blot of the concentrated sample from 40%B and 100%B fractions.*

c) GGL mutant

For the purification of the GGL mutant, we returned to the original protocol for MGD1 as this mutant had much higher expression levels. Compared to GGG and GGA mutants, this indicated that mutant GGL is definitely more stable and that its folding was not disturbed.

IMAC purification profile was comparable to that of the wild-type MGD1 (Figure 57). Elution yielded two major fractions containing the GGL mutant, obtained at 16% and 100% imidazole. The 100% fractions were pure samples, while 16% contained some low molecular weight protein contaminants. Fractions from 16%B and 100%B were concentrated and purified by SEC.

Both samples, from 16%B and 100%B, yielded pure protein after SEC (Figure 58). Major peaks from both fractions eluted at the same volume (~17 ml after injection to the SEC column) in SEC indicating that both fractions contained GGL mutant in monomeric form. Fractions from the major peak of both samples were pooled and concentrated. Overall, a total yield of 2 mg GGL was obtained from a 1L culture.

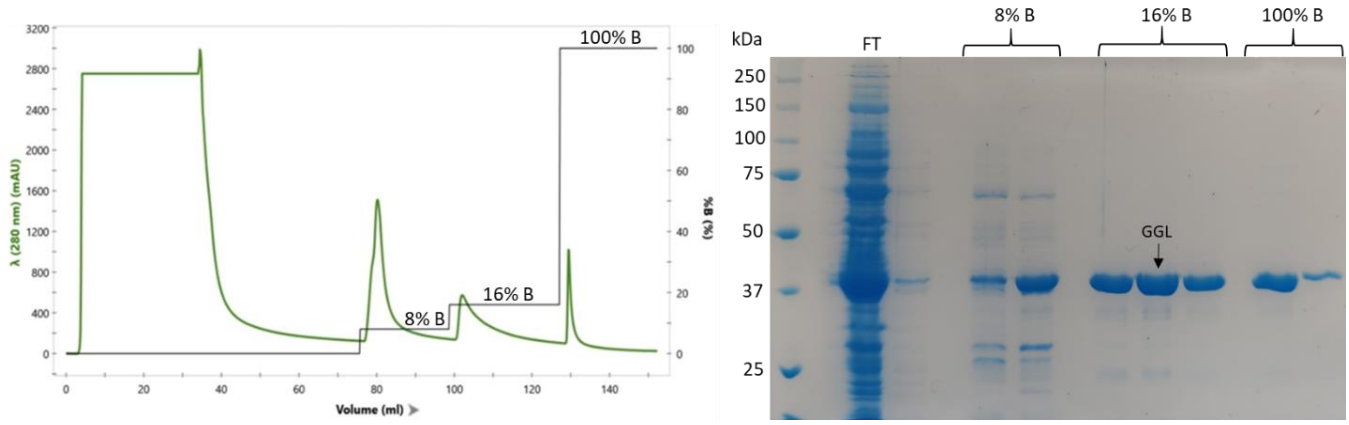


Figure 57: *Histidine tag purification of GGL mutant from a 1L culture. Included wells for SDS-PAGE analysis: flow through (FT) and elution fractions from 8%, 16% and 100% imidazole.*

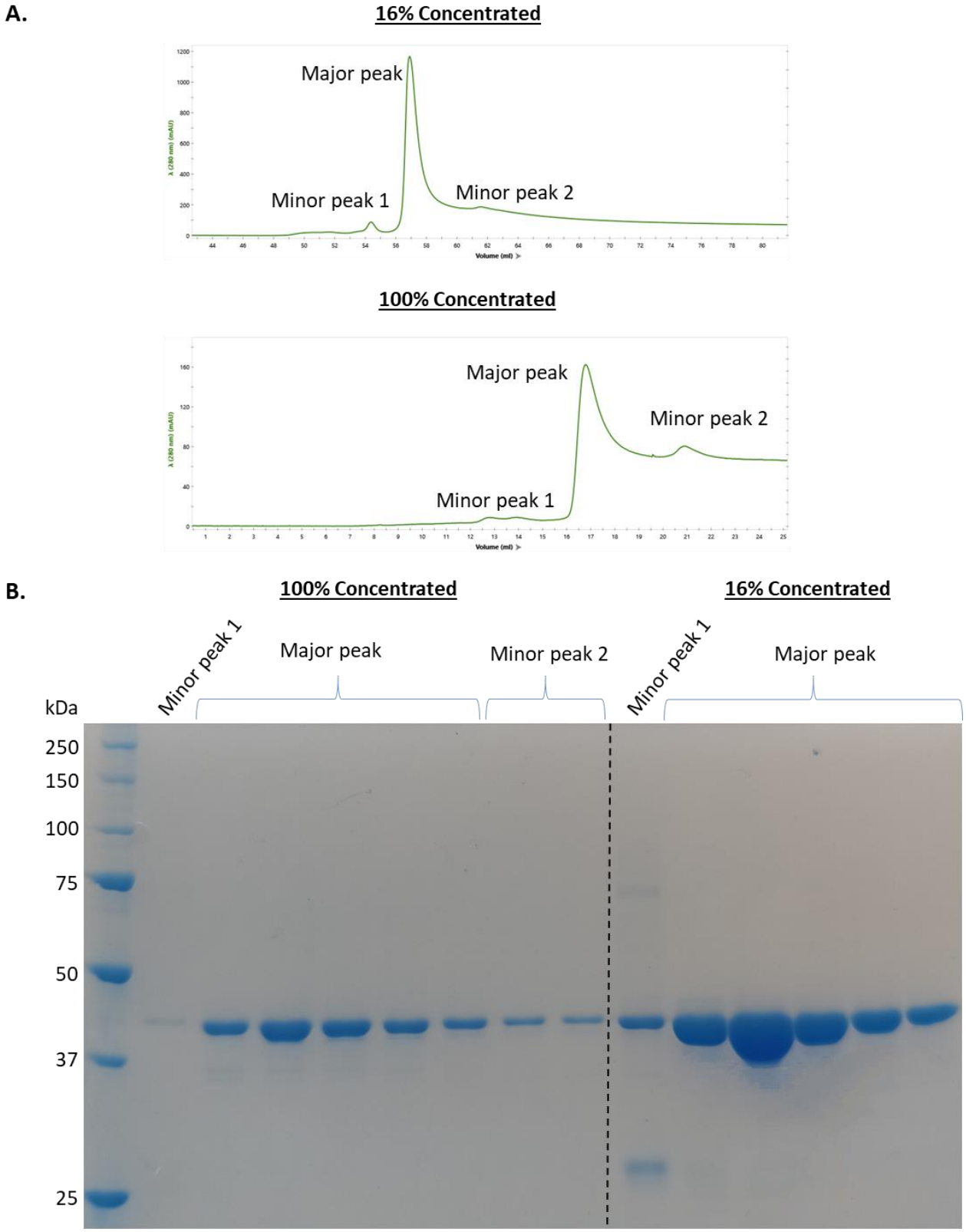


Figure 58: *SEC purification of the GGL mutant. A) Size exclusion purification of the concentrated samples, 16%B and 100%B, from IMAC purification. B) SDS-PAGE analysis of fractions from SEC purification.*

d) RcMGD1 mutant

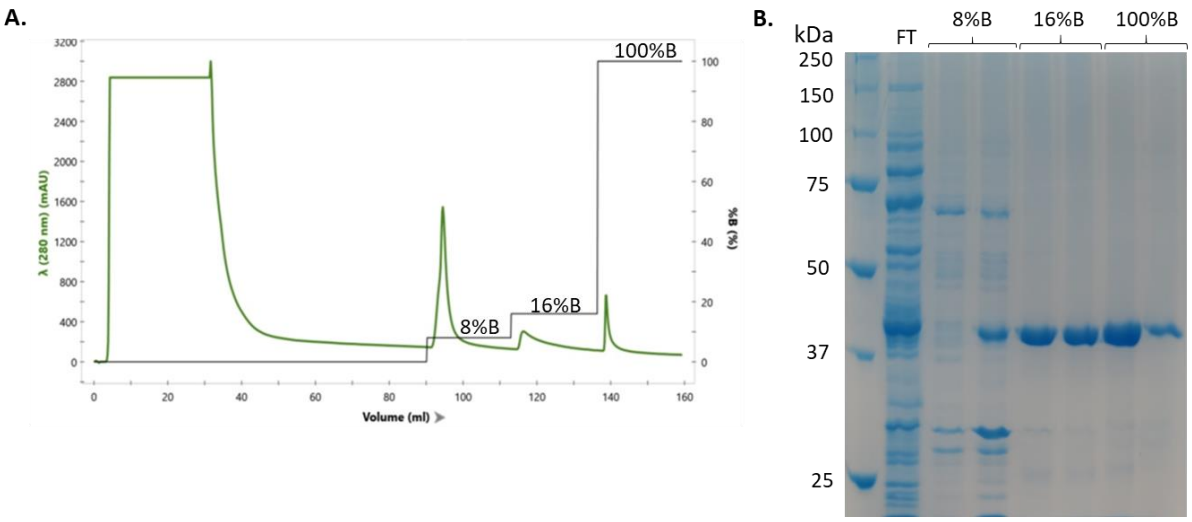


Figure 59: *IMAC purification of RcMGD1 mutant with SDS-PAGE analysis. A) Chromatogram from IMAC purification. B) SDS-PAGE analysis of IMAC purification. Included wells for SDS-PAGE analysis: flow through (FT) and elution fractions from 8%, 16% and 100% imidazole.*

RcMGD1 showed high expression levels comparable to the GGL mutant. Surprisingly, even though 8 residues were mutated, its folding was not perturbed.

The purification profile of RcMGD1 was similar to that of wild-type MGD1 and GGL mutant (Figure 59). Distinct bands indicated relatively pure RcMGD1 sample in the 16% and 100% imidazole fractions, thus size exclusion purification was not judged necessary. Fractions from 100% imidazole were concentrated, buffer exchanged to storage buffer, and used for subsequent activity tests.

6. Activity assays of MGD1 mutants based on RcMGlcD

a) GGG activity

Analysis of GGG mutant activity		
Protein	Activity (% of WT with UDP-Gal)	
	UDP-Gal	UDP-Glc
WT	100	0
GGG	1.13	~0

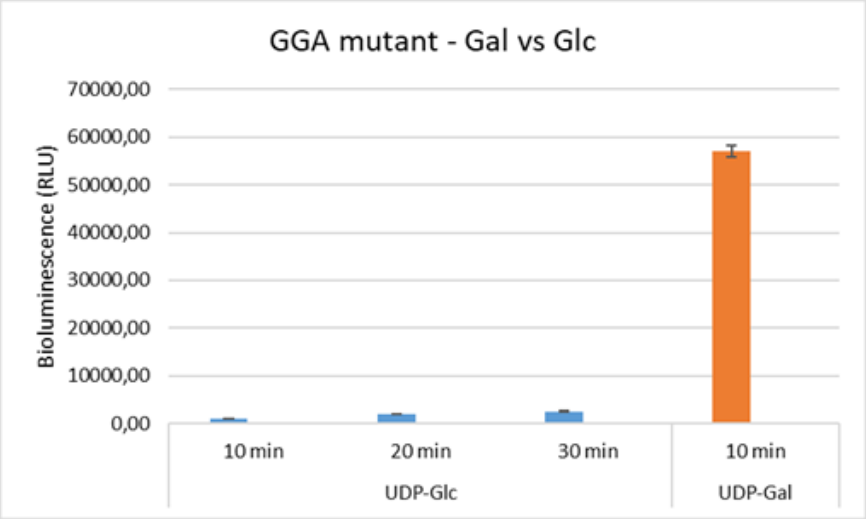
Table 14: *Activity of the GGG mutant with UDP-Gal and UDP-Glc (expressed as % of wild-type MGD1 activity).*

Initially, the GGG mutant (P433G) was produced to mimic the PQE motif of RcMGlcD. Since the AQE mutant (P433A) did not confer GlcT activity, one hypothesis was that the methyl side chain of alanine was still sufficient to block glucose, therefore a P433G mutation was interesting to test this hypothesis. The GGG mutant resulted in a more significant loss of activity compared to the AQE mutant (Table 14). The loss of activity is likely a result of the P433G mutation introducing a flexible triple sequence of glycine in the helix structure. The increased flexibility most likely disrupts the α-helix and thereby the active site resulting in the loss of activity.

b) GGA activity

RcMGlcD does not contain a triple glycine sequence in the PQE region. Instead in RcMGlcD there is a bulky aliphatic residue (Leu) adjacent to the glycine (Table 13). While retaining the P433G mutation, a second mutation in the adjacent G434 position was made, into alanine (GGA) or leucine (GGL), thereby preventing a triple glycine sequence. The activities of GGA and GGL were then assayed.

In Figure 60, activity assays showed that GGA displayed some GalT activity, and this was also confirmed by TLC (Figure 61B). Furthermore, the increase in RLU signal over time (10, 20 and 30-minutes) for UDP-Glc is indicative of GlcT activity, albeit extremely low, but no MGlcDG was detected by TLC after a 12h reaction (Figure 61B). Comparison of GalT and GlcT signal in a 10-minute reaction demonstrates that GGA is primarily a GalT enzyme.



		Bioluminescence (RLU)
UDP-Glc	10 min	1033
	20 min	1940
	30 min	2493
UDP-Gal	10 min	57088

Figure 60: *Activity test of the GGA mutant. UDP-Glo assay with GGA using UDP-Gal and UDP-Glc donor. Conditions: 100 ng GGA, reaction times at 10, 20 and 30-minute were tested for GlcT activity. A 10-minute reaction for UDP-Gal was used.*

c) GGL activity

This mutant is by far the most interesting of this series. On the one hand by its behavior in purification, close to that of the WT enzyme, and on the other hand, it was the first mutant that demonstrated a higher activity with UDP-Glc than with UDP-Gal (Figure 61A). As shown in this figure, the GGL mutant demonstrated both GalT and GlcT activities. The results show that the activity in the presence of UDP-Glc is about 1.7 times higher than that with UDP-Gal, thus making GGL more selective for UDP-Glc. The dual activity was further supported by TLC experiments, which demonstrated the formation of both MGDG and MGlcDG (Figure 61B).

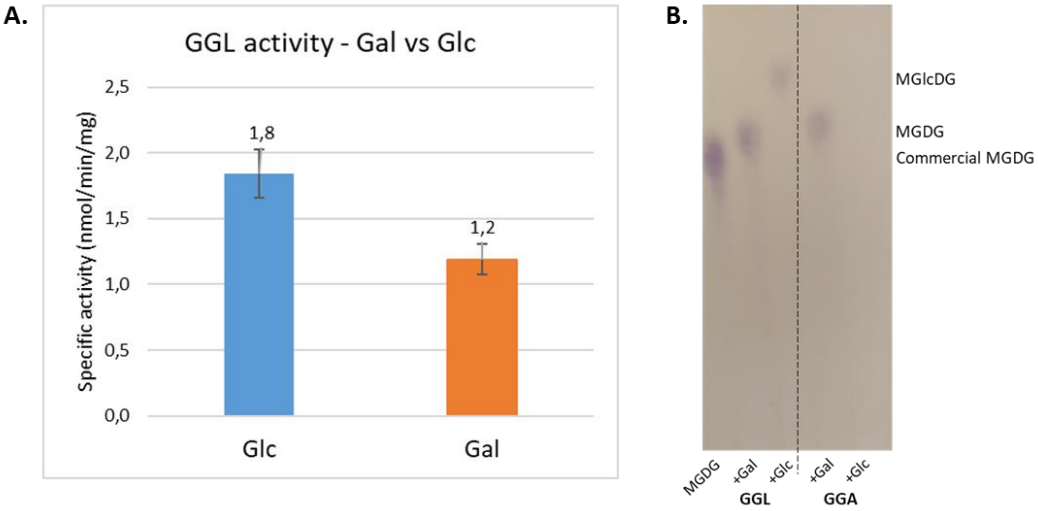


Figure 61: *Activity test of mutants GGA and GGL. A) UDP-Glo assay with GGL using UDP-Gal and UDP-Glc donor. C) TLC of a 12 hour-reaction of GGA and GGL with Gal and Glc donors.*

Although GGL was shown to be a bi-functional GT with GlcT and GalT activities, it must be stressed that the specific activity was very low compared to the WT MGD1. The values for GGL GlcT and GalT specific activities are 1.8 ± 0.2 nmol/min/mg and 1.2 ± 0.1 nmol/min/mg, respectively. As a reminder, the specific activity for MGD1 is about 482 nmol/min/mg.

d) RcMGD1 activity

RcMGD1 expressed very well highlighting that the mutation of 8 residues was compatible to the overall folding of MGD1. Bioluminescence signal was very low and not detectable in a 10 to 30-minute enzyme reaction. Therefore, a 24-hour reaction was attempted. In a 24-hour reaction, small GalT activity was observed while there is no GlcT activity (Figure 62A). However, a 24-hour reaction ran on TLC did not detect the presence of MGDG or MGlcDG (Figure 62B), indicating that the amount of product produced was below the sensitivity of the orcinol staining (limit of MGDG detection is ~100 ng). Taking into consideration the long reaction time used, the RcMGD1 mutant was considered inactive.

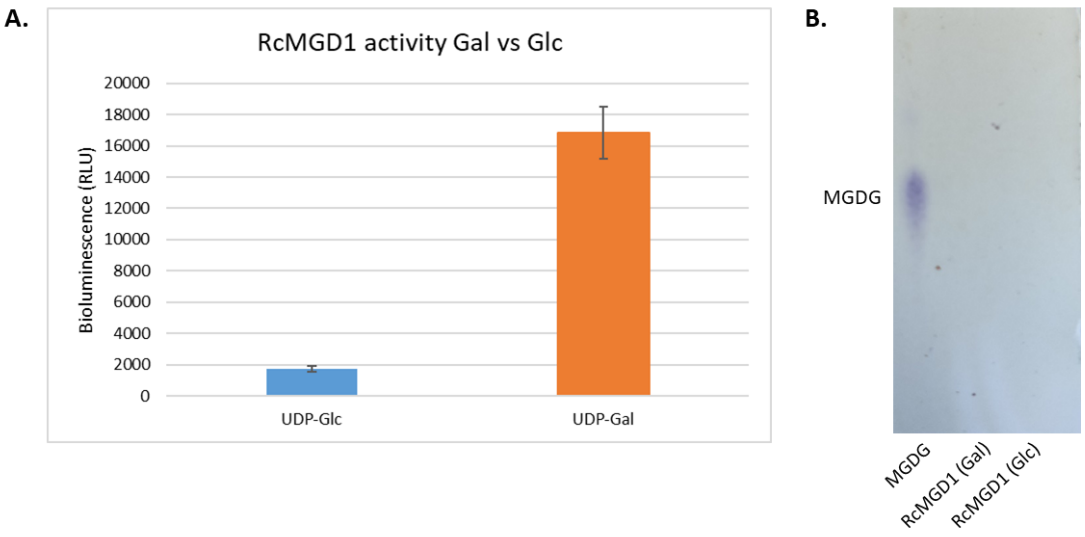


Figure 62: *RcMGD1 assay testing UDP-Gal and UDP-Glc activity in a 24-hour reaction.*

7. Discussion

a) Investigating the effects of mutations in the GGL mutant

The most promising mutant to confer GlcT activity was the GGL mutant (Figure 61). To investigate the mechanism by which GlcT activity was conferred, we looked into the effects of the P433G and G434L mutations in the protein structure and its interaction with the sugar donor.

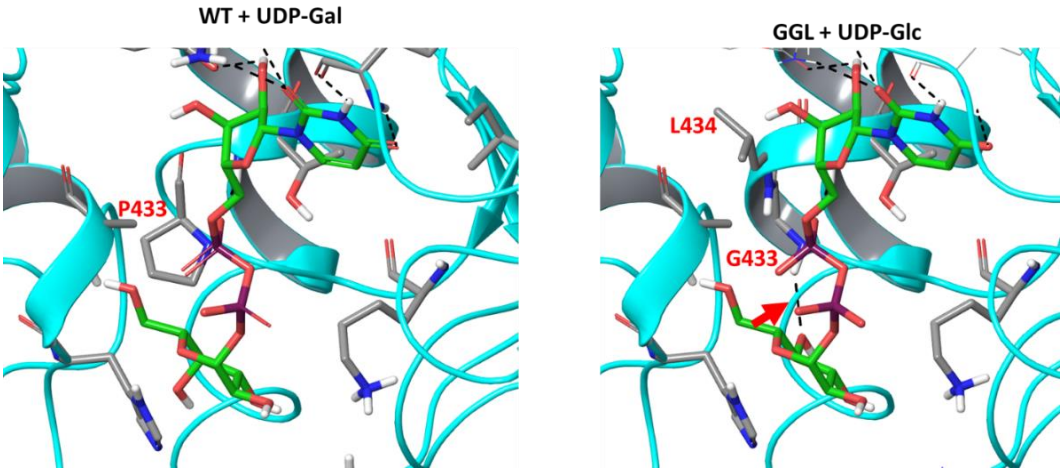


Figure 63: *Comparison of sugar donor binding of Wild-type MGD1 and GGL mutant. (Left) Wild-type MGD1 with UDP-Gal in the active site. (Right) GGL mutant with UDP-Glc modelled into the active site. The red arrow indicates the hydrogen bonding made by the glucose O4 and the MGD1 backbone (1.94 Å). Modelled by Olga Makshakova.*

From molecular modelling of the GGL mutant, we observed that the removal of the Proline-433 ring permitted glucose O4 to form a hydrogen bond, with a distance of 1.94 Å, to the α-helix backbone (Figure 63). This interaction was also observed in the previously analysed GT1 family GlcTs (Figure 45) and potentially in RcMGlcD. Therefore, the proline ring not only causes a steric clash, but also prevents hydrogen bonding of glucose O4 to the active site.

This information brought into question the lack of GlcT activity in the AQE mutant (P433A). One hypothesis is that the methyl group of alanine might still be close enough to block a glucose O4.

Provided the fold remains the same, it appears that the methyl side chain of A433 stays away from Glc, while the H-bond between O4 and the NH of backbone can form (Figure 64). Importantly, no conflict exists between the methyl group and surrounding protein residues. Maybe having Leu in the third position (G-A-L instead of G-A-G) could stabilize the helix (its dynamics) and help establish a stable H-bond.

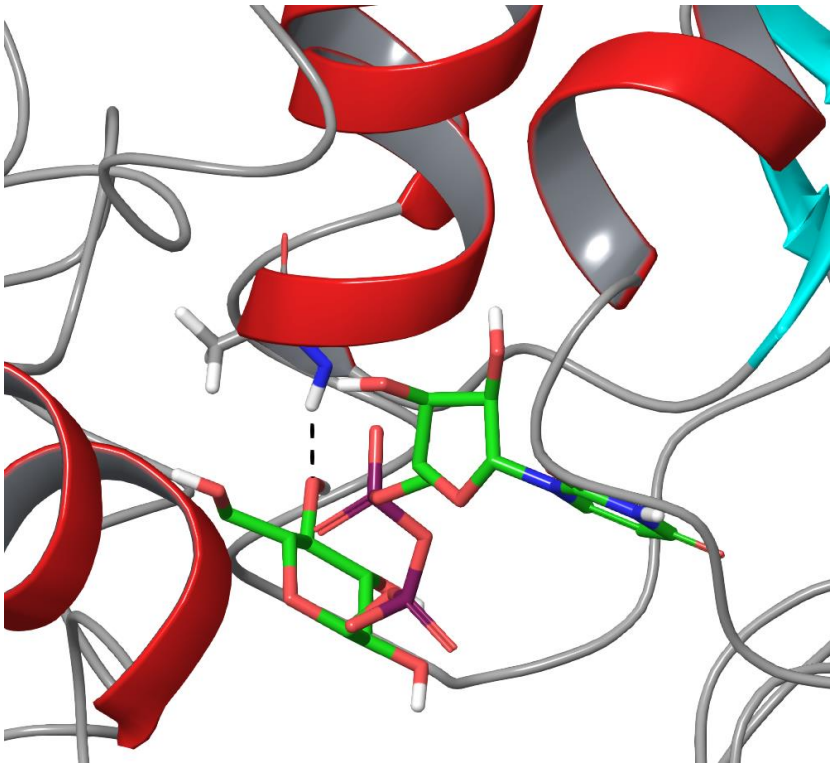


Figure 64: *Modelling of glucose O4 hydrogen bonding to an A433. Modelled by Olga Makshakova.*

From modelling of GGL, we observed that the L434 side chain fits into a pocket in the N-domain where it forms hydrophobic interactions with the surrounding residues: A157, T278, L321, P322 and V323 (Figure 65). These hydrophobic interactions likely restore the structure of the α-helix. The GGA mutant would lack these hydrophobic interactions and the α-helix structure would therefore not be as stable.

The integrity and stability of the α-helix is probably a key factor in protein folding, which could explain the expression problems encountered for GGG and GGA mutants. Thus, restoration of the α-helix in the GGL mutant also explains the similarity in expression and purification profile to the wild-type.

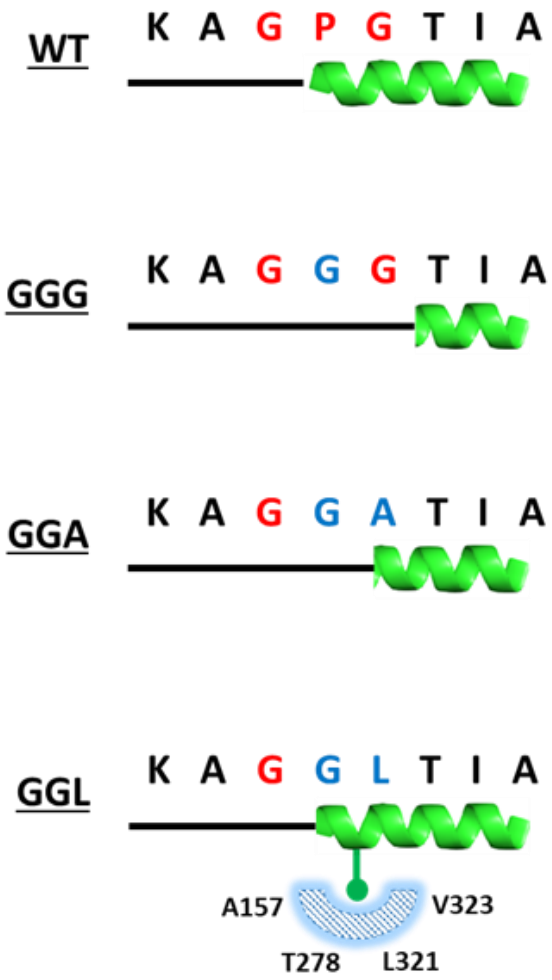


Figure 65: *Schematic representing the role of the 434 position in stabilisation of the protein.*

Along with the G432 residue, which is highly conserved in both GT1 and GT28 GTs (Figure 46), we also pointed to the importance of positions 433 and 434 in the donor specificity of MGD1. These three consecutive residues appear to be key for donor specificity in the GT28 family. We compared the 'GXX' region of the GlcT and GalT proteins we analysed. The known GalT enzymes (MGD1 and CaMGD1) have a 'Gly-Pro-Gly' sequence, where the proline plays the key role in discriminating for UDP-Gal, whilst the GlcT enzymes (RcMGlcD, GGL and possibly GGA mutants) have a 'Gly-Gly-Ala/Leu' sequence. This implies that a glycine in the second position and an aliphatic residue in the third position are required for GlcT activity. It is interesting to note that MurG, a bacterial GlcNAcT from family GT28, has a 'Gly-Ala-Leu'

sequence. Thus highlighting the importance of an aliphatic residue at the third position. The alanine in the second position brings into question the importance of glycine, and suggests that alanine could also be substituted in this position. A MGD1 mutant with the Gly-Ala-Leu (GAL) sequence could therefore be interesting to test as well.

Sequences of MGD1, RcMGlcD and MGD1 mutants	
Protein	**GalT activity**
MGD1	430-**K**A**GPG**TIA**E**
CaMGD1	286-**K**A**GP**GTIA**E**
	Glc activity
RcMGlcD	284-**K**A**GGL**SVS**E**
GGG	430-**K**A**GGG**TIA**E**
GGA	430-**K**A**GGA**TIA**E**
GGL	430-**K**A**GGL**TIA**E**
MurG (a GlcNAcT)	260-**R**S**GAL**TVS**E**

Table 15: *Comparison of the signature 'GXX' motif of MGD1, MGD1 mutants and other GT28 proteins. In yellow is the highly conserve first glycine. In blue is the second position. In green is the third position*.

In conclusion of this study on the enzymes of the GT28 family, the presence of a 'GPG' motif seems to be the signature of a GalT activity, whereas a 'GGX' motif (X being an aliphatic residue) would be the signature of a GlcT activity. To confirm our hypothesis, a GGI sequence is present in a processive β-GlcT (YpfP) from *Bacillus subtilis* (Jorasch et al., 1998). This may, therefore, represent predictive power for the many uncharacterized protein sequences in this family.

b) Future prospects

At this stage, we have generated a MGD1 mutant with dual activity (GlcT and GalT). It would be interesting to fully investigate the kinetic parameters (catalytic efficiencies, kcat and K_M) of GGL mutant for UDP-Gal and UDP-Glc. While the GGL mutant has met the aim of conferring GlcT activity, the specific activity is still very low (Figure 60). When compared to the wild-type GalT activity, GlcT activity is significantly lower. The aim of future work for this project will be to increase overall activity and

specificity to UDP-Glc. One strategy that can be employed is to use molecular simulations in order to predict possible mutations with enhanced binding to the UDP-Glc (Table 16).

	UDP-Gal	**UDP-Glc**
	Binding energy (kcal/mol)	**Binding energy (kcal/mol)**
Wild-type MGD1	-4.53	-4.09
GGL	**-5.11**	**-5.72**
Q455E	-5.32	-4.77
Q455D	-4.59	-5.17
GGL + Q455E	-4.81	-5.04
GGL + Q455D	**-4.89**	**-6.65**

Table 16: *Molecular simulations of sugar donor binding to MGD1 WT and mutant. The simulations were done using local search docking protocol, keeping the precise position of UDP moiety in the binding site, determined by X-ray. The orientation of the sugar was allowed to sample in the process of docking. The protein was considered as rigid but only the residues mutated were allowed to move. The calculations were performed in AutoDock4.2 using Gasteiger charges on atoms. Modelled by Olga Makshakova.*

In simulations, there is an increase in the binding energy of GGL for UDP-Glc (change from -4,09 to -5,72 kcal/mol) compared to wild-type MGD1, and this matches what is found in activity assays. Based on the GGL mutant, the binding energy from the addition of a third mutation was calculated. Of particular interest is a proposed 'GGL + Q455D' mutant. It is predicted that the additional Q455D mutation will decrease UDP-Gal binding and significantly increase UDP-Glc binding. We can therefore expect this mutant to be more specific for glucose. The Q455D mutation is interesting as this is the residue observed in GT1 GlcTs (Figure 46). Thus, it may still be useful to go back to take inspiration from the GT1 family. As the simulations predict, perhaps a MGD1 mutant inspired by both GT1 and GT28 enzymes will result in a more active GlcT enzyme.

In addition to the above suggestion, inspiration can also be taken from Akere et al (2020), who have used molecular dynamics simulation and deep learning analysis to improve the substrate recognition for the sugar donor of five UGT enzymes from *Arabidopsis thaliana* (Akere et al., 2020). Similar approaches were also made in a recent study by Zhang et al (2022), which used molecular docking and simulations to

identify regions (‘hot spots’) in the enzyme BsGT-1, that influence donor preferences. Although these studies dealt with GT1 family enzymes, the approaches could also be applied to MGD1 for identification of mutations that can further improve the UDP-Glc activity of the GGL mutant.

Chapter 4

Investigating MGD1 reaction mechanism:

Role of PA vs PG

I. Introduction

Understanding the catalytic mechanism of glycosyltransferases is essential for engineering GTs with altered activities. The general mechanisms for GT activity have been solved (Figures 11 and 12). However, some glycosyltransferases require an activator to be catalytically active, and thus their mechanism of action is more complex. For example, EryCIII, a GT involved in the biosynthesis of the antibiotic erythromycin D, is active only after forming a complex with the protein EryCII (Moncrieffe et al., 2012). In this study, the authors suggested that EryCII acts as a helper protein (or chaperone) that stabilizes the GT EryCIII, but it does not participate in catalysis. Other examples of this type, involving an auxiliary protein, exist in nature (Qasba et al., 2008; Wang et al. 2010).

In synBIOcarb, we were interested in GTs that can decorate lipid membrane surfaces and MGD1 was selected as a model enzyme. As previously mentioned, MGD1 is an example of a GT that needs an activator for its activity. The anionic lipids phosphatidic acid (PA), phosphatidic glycerol (PG) and sulfoquinovosyldiacylglycerol (SQDG) can fulfil this role of activator (Dubots et al., 2010; Nitenberg et al., 2020). Recent studies identified a possible PG binding site (Rocha et al., 2016b; Nitenberg et al., 2020). This led the authors to propose a novel reaction mechanism for MGD1 based on the existence of a PG-His catalytic dyad (Figure 30). However, kinetics data and mutagenesis studies indicated that PA and PG proceed through different mechanisms, thus suggesting distinct binding sites (Dubots et al., 2010; Rocha et al., 2016b; Nitenberg et al., 2020). PG which represents ~8-10% of total lipids of chloroplastic membranes is believed to be the 'natural' MGD1 activator in plant (Rocha et al., 2018). SQDG can also activate MGD1 and recent data suggested that it binds at the same site as PG. In addition, SQDG and PG are considered to be functionally redundant in plants (Yu et al., 2002). One question remains, which anionic lipid will be the preferred activator for MGD1 in a physiological context? The role of PA as an activator is less clear. It has been described as an allosteric activator of MGD1 but it is barely detectable in chloroplast membranes, thus rather suggesting a regulatory role in plants (Bastien et al., 2016, Rocha et al., 2018). It is clear that PA plays a central role in the metabolism of lipids of photosynthetic membrane (Dubots et al., 2012). A PA-induced conformational change of MGD1 may explain the observed allosteric activation.

In the context of the PhD study, we aimed to further investigate the reaction mechanism of MGD1 to tentatively confirm the existence of a PG/His catalytic dyad. By identifying the critical residues in MGD1 activation by PG and PA, we were also looking for potential PA-specific binding sites on MGD1 structure.

II. Strategy for testing the hypothesis of a PG-His catalytic dyad

In this section we wish to consolidate the hypothesis of a PG-His catalytic dyad. It must be stressed that an acid-base catalytic dyad (involving an Asp or Glu residue instead of PG) has been described in structural GT homologs of GT1 family (Rocha et al., 2016b). We therefore explored the possibility of introducing into MGD1 a mutation (in the form of Asp or Glu) that could replace the role of PG, thus creating an Asp/Glu-His catalytic. If this mutant was active in the absence of PG or PA, we would indirectly prove the presence of the PG-His dyad in wild-type MGD1.

A protein of particular interest in GT1 family is the MGDG synthase (MgdA) from the green sulfur bacterium, *Chlorobaculum tepidum.* Although from a different GT family, it is similar to MGD1 in structure. For example, it possesses a large loop ~50 aa that is proposed to also capture DAG as in MGD1. Most notably, MgdA does not need a lipid activator to be catalytically active (Masuda et al., 2011). In MgdA, the catalytic dyad consists of His93 and Asp179, by analogy to the 3D structures of GT1 family proteins CalG1 and CalG4 from *Micromonospora echinospora* (Chang et al., 2011). In structure predictions of MgdA, His93 and Asp179 are close enough, ~3.5 Å, to form the Asp-His catalytic dyad (Figure 66A). From structural comparison of MGD1 and MgdA, using the Phyre2 automatic fold recognition server (Kelley et al., 2015), Val250 is considered as the equivalent residue of Asp179 in MgdA (Figure 66B). However, the distance between His155 and Val250 in the crystal structure of MGD1 is ~6.5Å, which is too far to form an effective catalytic dyad. However, considering that MGD1 is a dynamic protein and that movements between the N- and C- domains have been evidenced upon lipid membrane binding (Nitenberg et al., 2020, Makshakova et al., 2020), these two residues may be able to make interactions. We therefore decided to produce two MGD1 mutants: V250E and V250D.

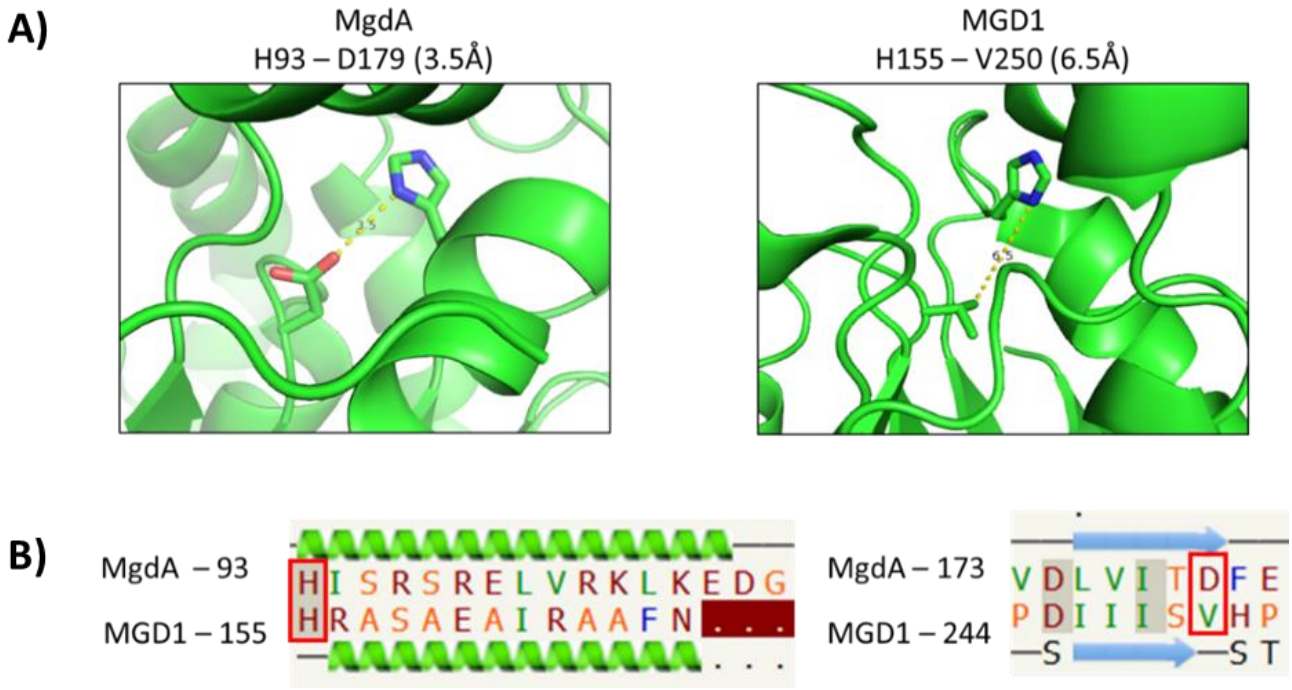


Figure 66: *Comparison of the residues forming the catalytic dyad. A) Structure of MGD1 and MgdA, displaying residues that would form a catalytic dyad and their distances. B) Alignment of MGD1 and MgdA by Phyre2, in the red boxes are the equivalent residues forming the catalytic dyad.*

Small scale expression trials with mutants V250D and V250E were performed and showed that the mutants are poorly produced, and in the insoluble fraction (see appendix, Figure S8). To check for any small quantity of soluble mutant, a 1L culture of each mutant, V250D and V250E, was prepared and IMAC purification was attempted. However, no protein was found from elution fractions (data not shown). The current hypothesis is that the polar and charged nature of Glu and Asp could destabilise the folding of the MDG1. This is because V250 is located in a β sheet in the N-domain where V250 and the surrounding residues are hydrophobic in nature. The V250 side chain forms hydrophobic interactions integral to MGD1 folding. This strategy was a bit simplistic but worth trying. See discussion at the end of chapter for a more thorough analysis.

III. Search for potential PA and/or PG binding sites

As previously described, an amino acid position (Pro189) in MGD1 that plays a role in PG activation was identified (Nitenberg et al., 2020). This was supported by the mutation of Pro189 into alanine, which yielded a mutant enzyme inactive with PG but which retained significant activity with PA, thus reinforcing the hypothesis of two distinct binding sites for PA and PG. The Pro189 is located in the LOOP region that is involved in DAG recognition.

If our hypothesis of a PG-His catalytic dyad is correct, this also suggests that PA can replace PG in the active site, at least in *in vitro* enzyme assays. Our current model is that PA, having a smaller polar head, can bind to the active site in a different way while not interacting with Pro189 but still forming a hydrogen bond with the catalytic base His155. There is also the possibility of a second regulatory binding site which would explain the allosteric properties of PA (Dubots et al., 2010). In the context of my PhD study, we will further explore potential binding sites in MGD1 for PA and/or PG.

1. Molecular dynamics (MD) simulations

For probing the PA and/or PG binding sites, we looked preferentially for positively charged amino acids as they may form strong interactions with the negatively charged phosphate group.

Molecular docking of PG and DAG to MGD1 has been previously attempted by our collaborator Olga Makshakova. Residues that interact preferably with PG or with PG and DAG are shown on Figure 67. Those that interact with DAG mostly occurred in the LOOP region, whereas residues that interact with PG are found in two regions, the LOOP region and the Nα4 helix. The binding site in Nα4 helix is mainly formed by positively charged residues, Arg260, Arg263, Ser264, Gly266, Leu268, Lys269. This MD study also revealed a potential role for a basic residue, **Arg156**, in establishing ionic interactions with the negatively charged polar head of PG (Makshakova et al., 2020). In the case that PA binds to the same site as PG, Arg156 could potentially interact with PA.

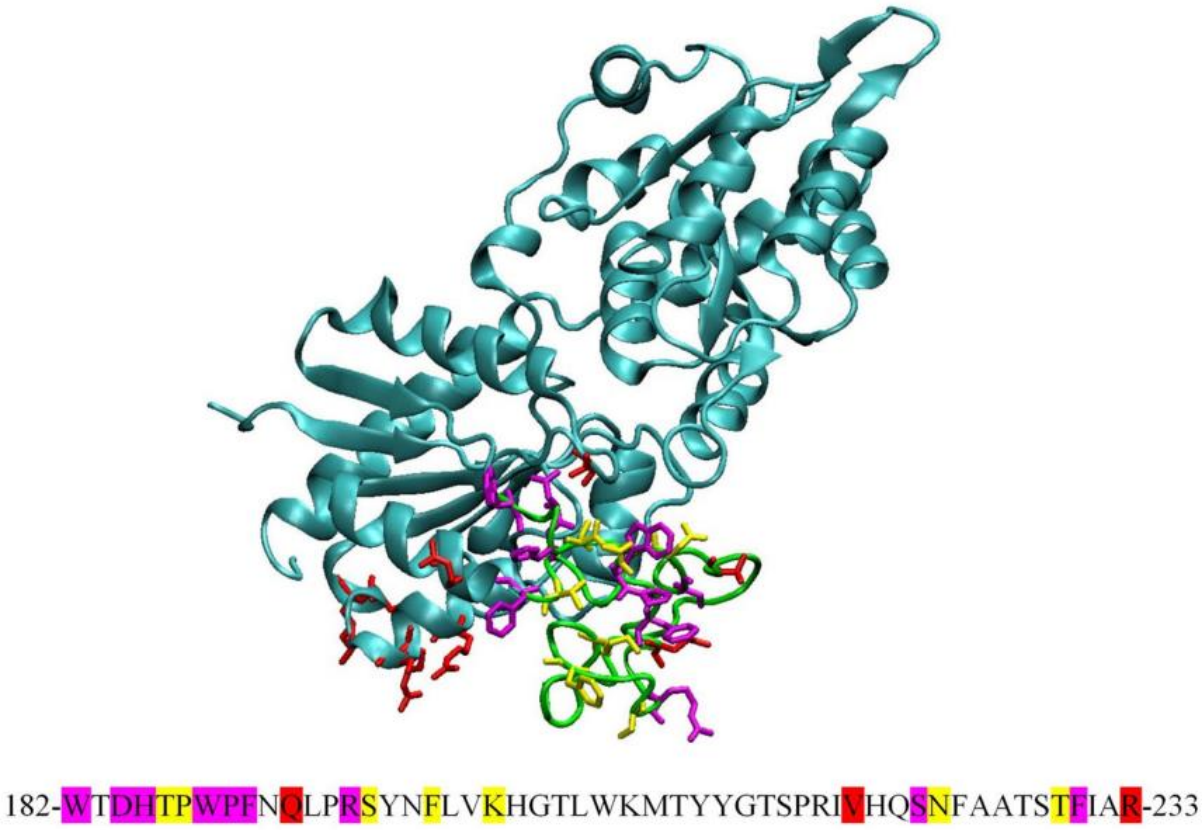


Figure 67: *MD simulations to identify potential PG and DAG binding sites in MGD1. LOOP region is in green and its sequence is indicated below the figure. Residues that interact preferably with PG or DAG are shown as sticks in red and yellow, respectively. Residues that interact with both PG and DAG are in magenta (from Makshakova et al., 2020).*

More recent computational study confirmed that the Nα4 helix is important for PA binding (Figure 68). Furthermore, the binding of PA was found more efficient than that of PG. Namely, the residence time of PA next to the residues Arg260, Arg263, Ser264 and Lys265 was about microseconds whilst that of PG counted for dozens of nanoseconds. The other site capable to tightly bind PA is organized by residues Arg195, Lys202, His203 and Lys208, which are located in the LOOP region. This was observed in MD simulations at the Coarse-Grained (CG) level of representation, which speed up the lipid mixing and enhance the sampling, as well as at the All-Atom (AA) level representation, which take into account the hydrogen bonds explicitly. The reader can refer to the publication from Makshakova et al. (2020) for details on CG and AA representations. The calculations were performed with a lipid composition that mimics the inner envelope membrane (IEM), namely MGDG/DGDG/PG/DAG/PA in the proportion

50/30/10/5/5, and with a simplified phospholipid/DAG composition, PG/PA/DAG in the proportion 60/15/25. The calculations were performed using Gromacs engine and force fields Martini2.2 for CG simulations and Charmm36 for AA simulations. We considered these results as a strong basis for the prediction of PA binding site to perform the site-directed mutagenesis experiments.

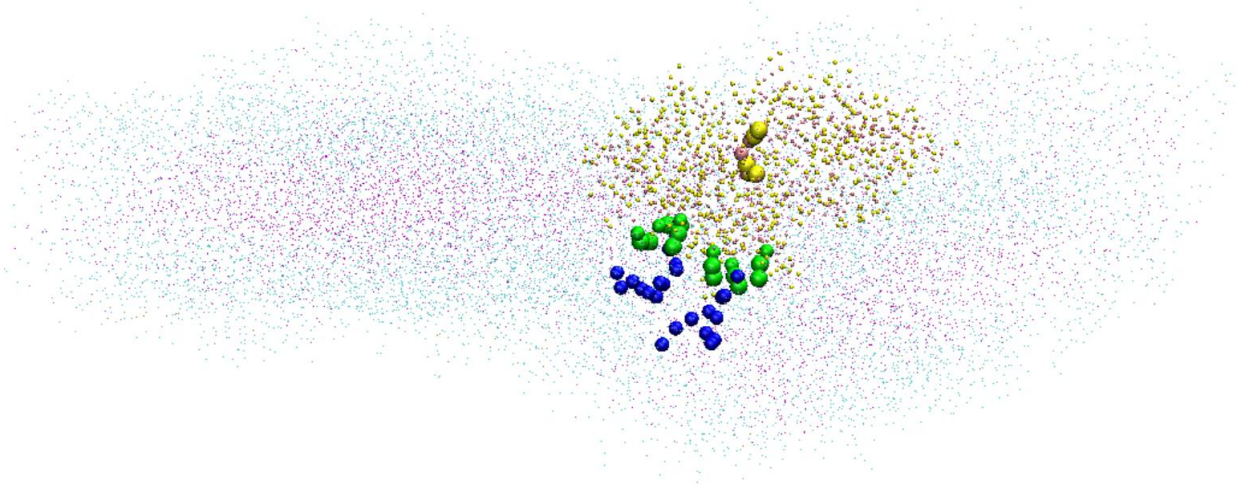


Figure 68: *The snapshot of Coarse-Grained (CG) MD trajectory illustrated the two PA molecules (in blue) tightly bound to MGD1 by (left) Arg260, Arg263, Ser264 and Lys265 and (right) Arg195, Lys202, His203 and Lys208, residues are shown in green. The membrane composed of MGDG/DGDG/PG/DAG/PA (50/30/10/5/5) is shown in cyan (polar heads) and magenta (hydrophobic core); protein is given in pink (backbone beads) and yellow (side chain beads), the catalytic His155 and Arg156 are shown as van der Waals beads. Results by Olga Makshakova.*

2. Sequence analysis

Another strategy for identifying residues that bind PA is to perform sequence analysis with MGD1 homologs. We turned to the two proteins of the GT28 family, which were studied in chapter 3, RcMGlcD from *Roseiflexus castenholzii*, and CaMGD1 from *Caldilinea aerophila*. Both proteins share close sequence and structural similarity to MGD1, and even contains a flexible LOOP region which is believed to capture DAG, just like in MGD1 (Rocha et al., 2016). These two proteins were produced (see Chapter 3) and were both confirmed to be activated by PA and PG (Figure 69). Another interesting protein, is MgdA from *Chlorobaculum tepidum*. Unlike MGD1, RcMGlcD and CaMGD1, MgdA does not require an activator.

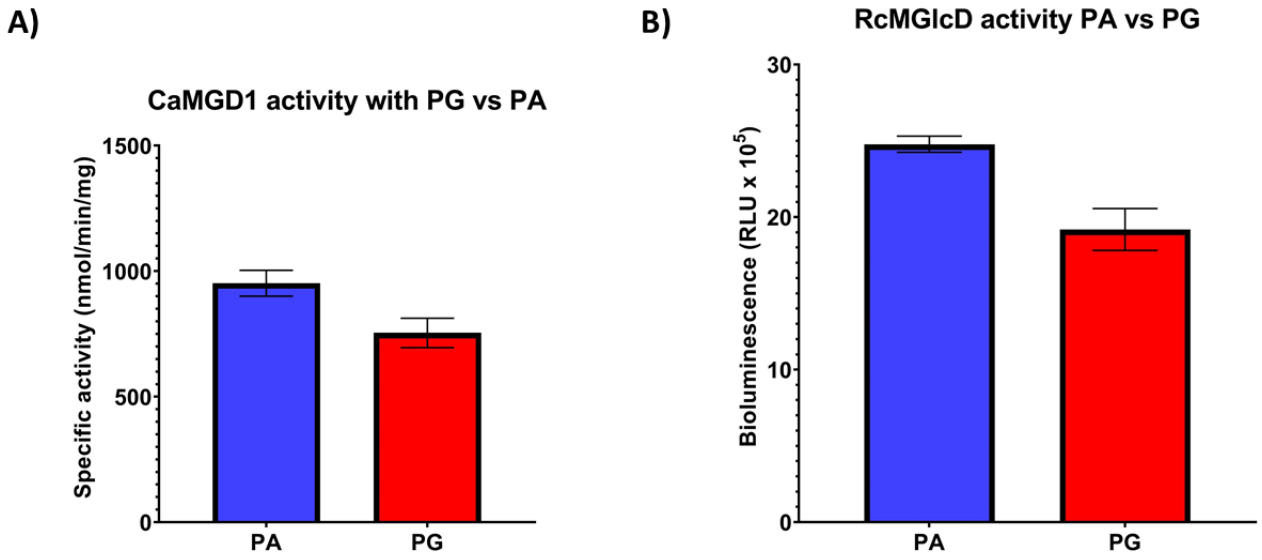


Figure 69: *Activity assay of CaMGD1 and RcMGlcD in the presence of PA or PG. UDP-Glo™ assay was used and followed the protocol for MGD1. Unless stated otherwise, conditions are as described for a standard MGD1 assay in the Methods and Material. A) Assay with CaMGD1. Conditions: 10 ng enzyme per reaction and 20-minute reaction time. B) Assay with RcMGlcD. Conditions: 100 ng enzyme per reaction and 10-minute reaction time.*

We performed sequence alignment of MGD1, RcMGlcD, CaMGD and MgdA, and looked more particularly at the basic amino acids at or near the proposed binding sites for PA: active site, LOOP region and the Nα4 helix. Furthermore, we focus on positive residues identified by molecular modelling and simulations (Part III.1). Ideally, sequence alignment of MGD1, RcMGlcD and CaMGD1 should reveal highly conserved basic residues that can direct us to possible PA binding sites. The absence of a basic residue at these positions in MgdA would be an additional clue.

Sequence alignment showed more conserved positive residues between MGD1 and CaMGD than with RcMGlcD (Figure 70). This correlates to the higher sequence similarity of CaMGD1 to MGD1 (CaMGD1 vs RcMGlcD / 55% vs 33%). MGD1 residues Arg195, Lys208, His251, His256 and His289, have an equivalent in CaMGD but not in RcMGlcD. MGD1 residue His203 has an equivalent in RcMGlcD but not in CaMGD. Of particular interest are MGD1 residues Arg156 and Lys265, which have an equivalent in both CaMGD1 and RcMGlcD.

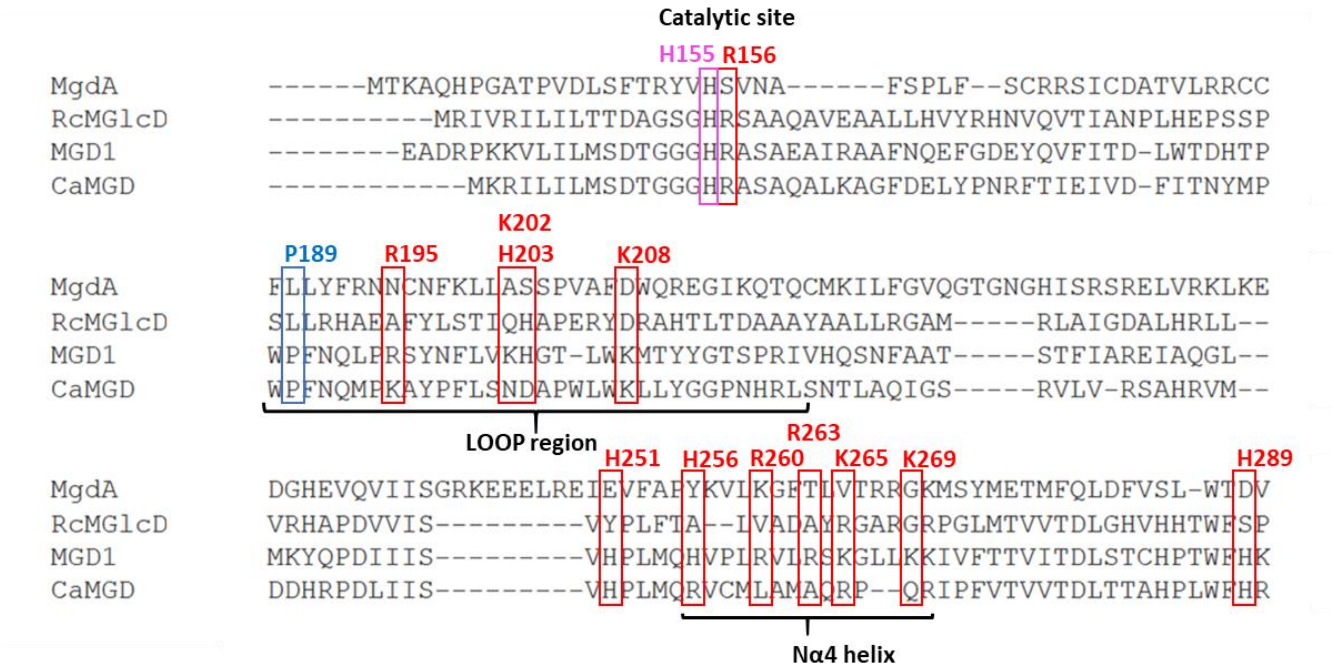


Figure 69: *Sequence alignment of MGD1, CaMGD1, RcMGlcD and MgdA in the three regions considered. In magenta is the catalytic residue His155. In blue is the vital residue in PG activation, Pro189. In red boxes are positive amino acids identified by computational work.*

From the basic residues identified in computer simulations and sequence analysis, we decided to produce the following MGD1 mutants:

- R156A and H251A for their proximity to the catalytic base His155
- R195A located in the LOOP region
- R260A and K265A located in the Nα4 helix
- H256A, H289A and the double mutant H256A-H289A. These two His residues are interesting because they are in close contact and interact in the MGD1 structure (see inset in Figure 71)

For each mutant generated, enzyme activity was assayed with PG and PA. A decrease in activity in both PA and PG may indicate similar binding sites. On the contrary, if one activity, PA or PG, is affected more, it can indicate an exclusive lipid-binding site.

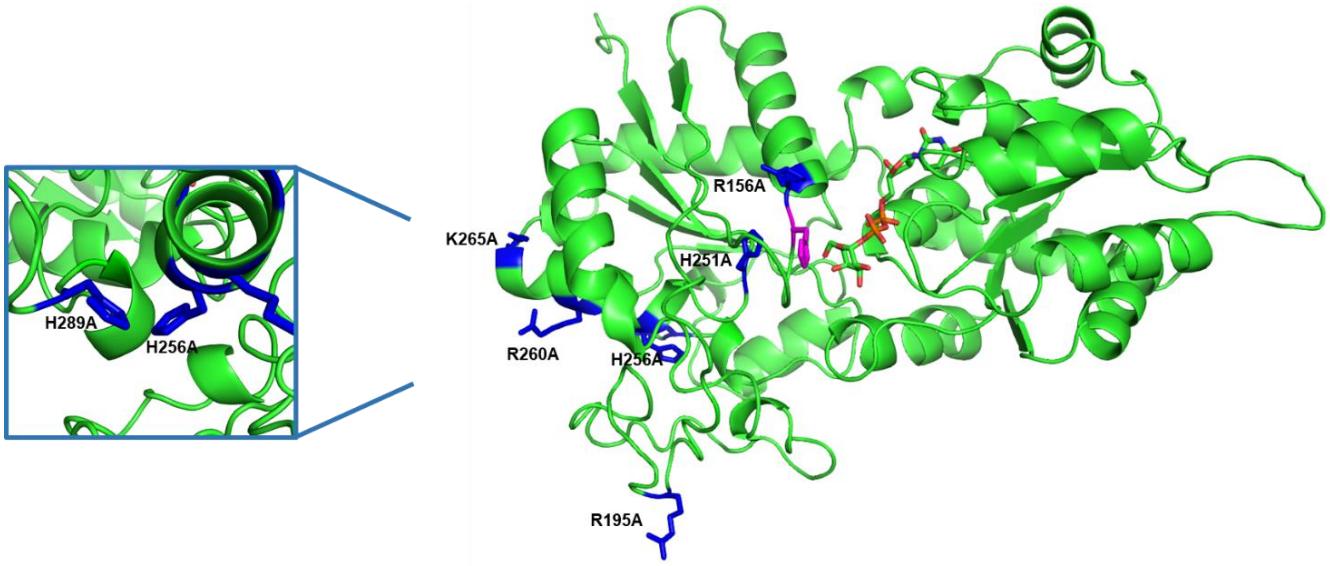


Figure 71: *MGD1 structure highlighting the targeted residues for mutation (Blue). In magenta is the catalytic residue H155. This is MGD1 with the LOOP region and UDP-Gal modelled into the structure by Olga Makshakova.*

IV. Production and activity of MGD1 mutants

1. Expression and purification of MGD1 mutants

All MGD1 mutants were expressed using the protocol established for wild-type MGD1 (Rocha et al., 2013). See appendix for small scale expression trials and all chromatograms from purification of MGD1 mutants produced for Chapter 4. R260A, K265A, and H289A had similar expression and purification profiles to Wild-type MGD1 (Table 17), indicating that these mutations do not perturb the folding of MGD1. The MGD1 mutants with lower expression levels and yields are R156A and H256A. Arg156 and His256 appear to have a more critical role in folding. Their mutation to alanine led to perturbation in folding and a higher proportion of the mutant was found in the insoluble fraction in expression trials. The mutants with the lowest expression level and yield are H251A and H256A+H289A. In the case of H251A, the folding was perturbed in a way that its binding to the IMAC column was much weaker and more protein eluted out in the 16%B (80mM Imidazole) wash step. The Arg195 residue, which is located in the flexible LOOP region seems to be vital for folding since its mutation into alanine resulted in an insoluble protein in expression trials, and attempted IMAC purification yielded no protein (see appendix).

Protein	Expression level
Wild-type MGD1	+++++
R156A	+++
R195A	-
H251A	+
H256A	+++
R260A	+++++
K265A	++++
H289A	++++
H256A+H289A	++

Table 17: *Expression levels of MGD1 mutants. Rating is based on the protein yield obtained from a 1L culture after IMAC purification and measurement of peak absorbance at 280 nm from 100%B elution step.*

2. Activity tests of MGD1 mutants

We used the UDP-Glo™ GT assay to monitor activity of MGD1 using the PA or PG activator. The enzyme tests were done at 25°C in 25 mM Bis-Tris, pH 6.5, 150 mM NaCl, 1 mM TCEP or DTT, 5% glycerol, 6 mM CHAPS, DAG (6.3 mol-%), PG or PA (1.5 mol-%), 700 µM UDP-Gal, MGD1 or mutant (5-15ng), as described in Materials and Methods. Comparison of mutant activity was made with the wild-type MGD1 under the same conditions.

When looking at the activities of the mutants, two factors must be taken into consideration. First, the role of the residue mutated on interaction with the activator. Second, the effect of mutation on protein folding. MGD1 is a dynamic protein. Changes in protein structure can have a severe effect on the dynamic movement of MGD1 and thus on the activity of the protein. Therefore, a perceived loss of activity may be caused by perturbation to the folding of MGD1 and does not necessarily mean that the residue plays a crucial role in activator binding. Additionally, as folding perturbations affect the whole protein, it would likely affect both PG and PA activity simultaneously.

In the light of activity assays, the mutants can be split into different categories: mutants affecting PG activity, mutants affecting PA activity, mutants affecting both PG/PA activity and mutants that do not affect activity (Table 18). We note that none of the mutations tested produced completely inactive enzymes.

Enzyme activity of PG/PA mutants			
	Activity (% Wild-type MGD1)		
	PG	PA	Evaluation
R156A	98	72	Not significant
H251A	4	33	Significant
H256A	31	98	Significant
P189A*	0	50	Significant
K265A	88	68	Not significant
H256A+H289A	10	14	Significant
R260A	100	97	Not significant
H289A	96	100	Not significant

Table 18: *Enzyme activity of MGD1 mutants in the presence of PA or PG. *Results from Nitenberg et al., 2020.*

We deemed that only mutations that reduced activity by >50% was significant and would likely play an important role in the catalytic mechanism. Of the 7 mutants assayed 4 of them were considered to have no significant effect on PA or PG activity: R156A, R260A, K265A and H289A. Based on computational work, Arg156 had been proposed to interact with the charged polar head of PG (Makshakova et al., 2020). Our result does not support this assumption. Also very surprising are the R260A and K265A mutants, as simulations predicted strong binding of PA to these residues, their mutation was expected to have a large effect on PA activity. This does not seem to be the case as neither PA activity nor PG activity were significantly affected.

Mutations H251A and H256A had significant effect on catalytic activity. These two residues are located close to the active site, with His256 being in the terminal of the Nα4 helix and His251 in close proximity to the catalytic residue His155 (as close as ~4.4 Å in some modelled structures). The mutation H251A significantly decreased activity for PG (~4% of WT) but the mutant retained some PA activity (33%) (Table 18). His251 was identified as one of the residues involved in the concerted fluctuations upon MGD1 membrane binding (Makshakova et al., 2020). Thus, the mutation could have affected the mechanical dynamics of MGD1 leading to the decreased activity observed for both PG and PA. Since a large % of PA activity remained, His251 plays a more critical role in PG mechanism. However, it must be taken into consideration that H251A showed low expression levels, indicating that the mutation also had a significant effect on the structure of MGD1 (Table 17). It is unknown to what extent activity was affected due to perturbation of folding or disruption in molecular dynamics specific to catalysis. Although the role of H251 in PA activation remains unclear, its proximity to the catalytic base H155 and a reduced PA activity for H251A are clues in favour of PA binding in the active site.

Results in Table 18 showed that the activity of mutant H256A is not affected in PA tests. However, its activity in PG tests is reduced to ~31% of WT activity (Table 18). As activity with PA was not affected, one can assume that the mutation did not perturb protein folding. These results suggest that His256 exclusively plays a role in PG catalytic mechanism. His256 is located in the Nα4 helix and is some distance away from the active site. Complementing the simulations from Makshakova et al. (2020), rather than binding directly to PG, His256 seems to be part of the downstream residues that pave the way for PG to the active site. This also supports the hypothesis that PG can bind both in the catalytic site and in the Nα4 helix.

Double mutant H256A+H289A is more complicated to interpret. Since the mutation H289A has no effect on activity, the double mutant is expected to behave similarly as H256A. Instead, a significant decrease in PG and PA activity is observed for H256A+H289A (~10-15% remaining) (Table 18). As expression levels for this mutant were very low, the overall decrease in activity is likely caused by perturbation of the protein structure.

V. Future work

We were unable to consolidate experimental proof for the PG-His catalytic dyad, due to the high instability that MGD1 V250D/E mutation confers. As observed in MGD1, Val250, is located at the terminus of a β strand and is involved in hydrophobic interactions essential for folding, and it is also further away from the catalytic His155. A proposed mutant could aim to introduce an Asp or Glu residue further into the loop and closer to the His155. Identified by computation simulation are some residues that can make contact with His155 and possibly form a catalytic dyad, the most interesting residues for mutation are His251, Thr278 and S158.

It is still inconclusive where PA binds to activate MGD1. However, there are many more residues identified from the computational work that need to be assayed, particularly in the Nα4 helix region there is Arg263 and Lys269, that are interesting targets. Another suggestion is a combination of multiple mutations. The combined interaction of the activator to multiple residues in the Nα4 helix may propagate the necessary conformational changes in MGD1 for allosteric regulation. The other promising PA binding site is in the LOOP region and residues: Lys202, His203 and Lys208 remain to be assayed.

Finally, the allosteric character of PA activation should also be assayed with mutants. This is because, while general activity with PA might not be affected, the dynamics that allows for allosteric regulation may be disrupted. The first study that discovered the allosteric effect by Dubots et al. (2010) assayed the MGD1 in mixed micelles containing PG and PA, 0.5 mol-% and 0.08 mol-% respectively. It would be interesting to see to what extent the allosteric effect is disrupted in MGD1 mutants, compared to wild-type MGD1.

Chapter 5

Application of glycosyltransferases on synthetic membrane systems

I. Introduction

As mentioned in chapter 1, GTs can be attractive tools for synthetic glycobiology. Advances have been made in the glycoengineering of complex glycosylation pathways to produce therapeutic glycosylated proteins (Van Landuyt et al., 2019; Hamilton et al., 2006; Harrison et al., 2006; Kallolimath et al., 2016; Dicker and Strasser, 2015, Margolin et al. 2020; Gupta and Shukla, 2018, Kightlinger et al., 2020). GTs are widely used in glycorandomization to generate a library of small glycosylated compounds with altered biological/therapeutic properties (Griffith et al., 2005; Goel et al., 2021).

GTs are also of interest for modifying the glycocalyx on live cells (Hong et al., 2020; Hong et al., 2019; Jiang et al., 2018). However, GTs used to decorate synthetic membranes such as giant unilamellar vesicles (GUVs) and supported lipid bilayers (SLBs) are sparse in literature. Two examples are reported in the literature. One study concerns the use of a β4-galactosyltransferase to decorate GUVs. The authors demonstrated that the formation of glycolipid microdomains enhances the reaction rate of the enzyme (Noble et al., 2012). This is reminiscent of the 'cluster glycoside effect', a term used to describe the higher affinity observed during interactions between multivalent glycans and some lectins (Müller et al., 2016; Dimick et al., 1999; Lundquist & Toone, 2002). This study showed that soluble GTs can decorate GUVs, and these membrane models are valuable tools for studying the behaviour of GTs in membrane environments such as lipid microdomains.

The second example concerns the use of a GT (GumH, a bacterial α3-mannosyltransferase) acting on a SLB (Alves et al., 2011). Glycosylation was monitored using a Quartz Crystal Microbalance with Dissipation (QCM-D). This study used an immobilized alkanethiol self-assembled monolayer (SAM) and a lipid monolayer, with the lipid outer leaflet exposed to the enzyme. Therefore, it is not a true lipid bilayer where both leaflets are lipids. These are the few examples of a recombinant GT decorating a lipid surface. Thus, we are still in the infancy regarding the application of GTs to glycosylate artificial membrane surfaces and what information can be derived with different models.

In the context of synBIOcarb, we were interested in applying GTs to decorate synthetic membrane systems. In this chapter, we investigated LgtC and MGD1, two very different galactosyltransferases in terms of fold, reaction mechanism and types of acceptors utilized. LgtC is a soluble GT-A fold protein that was used to glycosylate the lactosyl-ceramide (LacCer) acceptor substrate. On the other spectrum is the monotopic MGD1, a GT-B fold protein that transfers the first sugar into the lipid, DAG. We explored the applications of LgtC and MGD1 to decorate synthetic membrane models.

LgtC, from *Neisseria meningitidis*, is an α4-galactosyltransferase that participates in the biosynthesis of lipo-oligosaccharides (LOS) of the bacterial cell envelope. It catalyses the transfer of a galactose molecule in an α-1,4 position to the minimal glycan acceptor, lactose. LgtC is a well-studied enzyme with the crystal structure solved and kinetic parameters for lactose determined (Persson et al., 2001; Lairson et al., 2006). The interest of LgtC in the project lies in its ability to form the Gb3 antigen that is the receptor of the Shiga toxin B subunit (StxB). LgtC has been shown to glycosylate the glycolipid acceptor LacCer, which has been solubilized in solution with methyl-cyclodextrin (Adlercreutz et al., 2010). However, it has never been experimentally proven that LgtC can glycosylate LacCer included in a synthetic membrane, such as GUV and SLB.

In the PhD project, we explored the capability of LgtC to glycosylate micelles, GUVs, and SLBs containing LacCer.

All current activity assays with MGD1 were performed in mixed micelles. Thus, there was interest in expanding MGD1 applications to other membrane systems. Preliminary data on the MGD1 application on GUVs was also presented in this chapter.

Finally, LgtC and MGD1 were tested for their potential in chemoenzymatic applications. The capacity of LgtC and MGD1 to use UDP-Gal containing an azido functional group was assayed.

II. Expression and Purification of LgtC

The LgtC construct consist of 286 amino acids (1-286 aa) and was truncated of 25 amino acids in the C-terminus. This truncation is known to increase solubility and yields from purification, in addition the full-length protein was sensitive to proteolysis (Wakarchuk et al., 1998; Persson et al, 2001; Chan et al., 2012). Recombinant LgtC contained a N-terminal His-Tag with a cleavage site for TEV protease. Purification was comprised of a two-step procedure: immobilized metal affinity chromatography (IMAC) using a HisTrap column followed by size-exclusion chromatography (SEC). LgtC purification protocol was based on previous studies (Persson et al., 2001; Chan et al., 2012). IMAC purification used the same buffer for cell lysate and IMAC purification: 20 mM Tris, 1 mM DTT, 500 mM NaCl, 10 - 500 mM Imidazole, pH 7.5. Fractions containing LgtC were collected and further purified by SEC: 20 mM Tris, 5 mM TCEP, pH 8.5. The buffer used for SEC was optimized for protein stability and long-term storage of the protein at -20^{o}C. Due to steric blockage from LgtC, TEV protease was not able to cleave the His-Tag

(data not shown). The His-tag is located in the reverse-side from the active site and is predicted to not interfere with activity or binding to the membrane, thus the TEV cleavage was not necessary

IMAC purification showed that the binding of LgtC to the nickel column was weak due to the elution of LgtC at the wash steps (10 mM and 25 mM imidazole), indicated by large bands on SDS-PAGE (Figure 72A). However, a large portion remained bound, and fractions from 75 mM to 500 mM imidazole were deemed pure. These fractions were collected and further purified by SEC (Figure 72B). There were two minor peaks. The first minor peak seemed to be an oligomer of LgtC, and the second peak was a mixed oligomer consisting of LgtC and a contaminating protein between 50-75kDa. LgtC eluted as the major peak in SEC. The elution volume of the major peak was consistent with a monomeric form of LgtC. LgtC was very stable, and tests of stability showed that in 3 defrost cycles, 75% of LgtC was still soluble (see appendix, Figure S17).

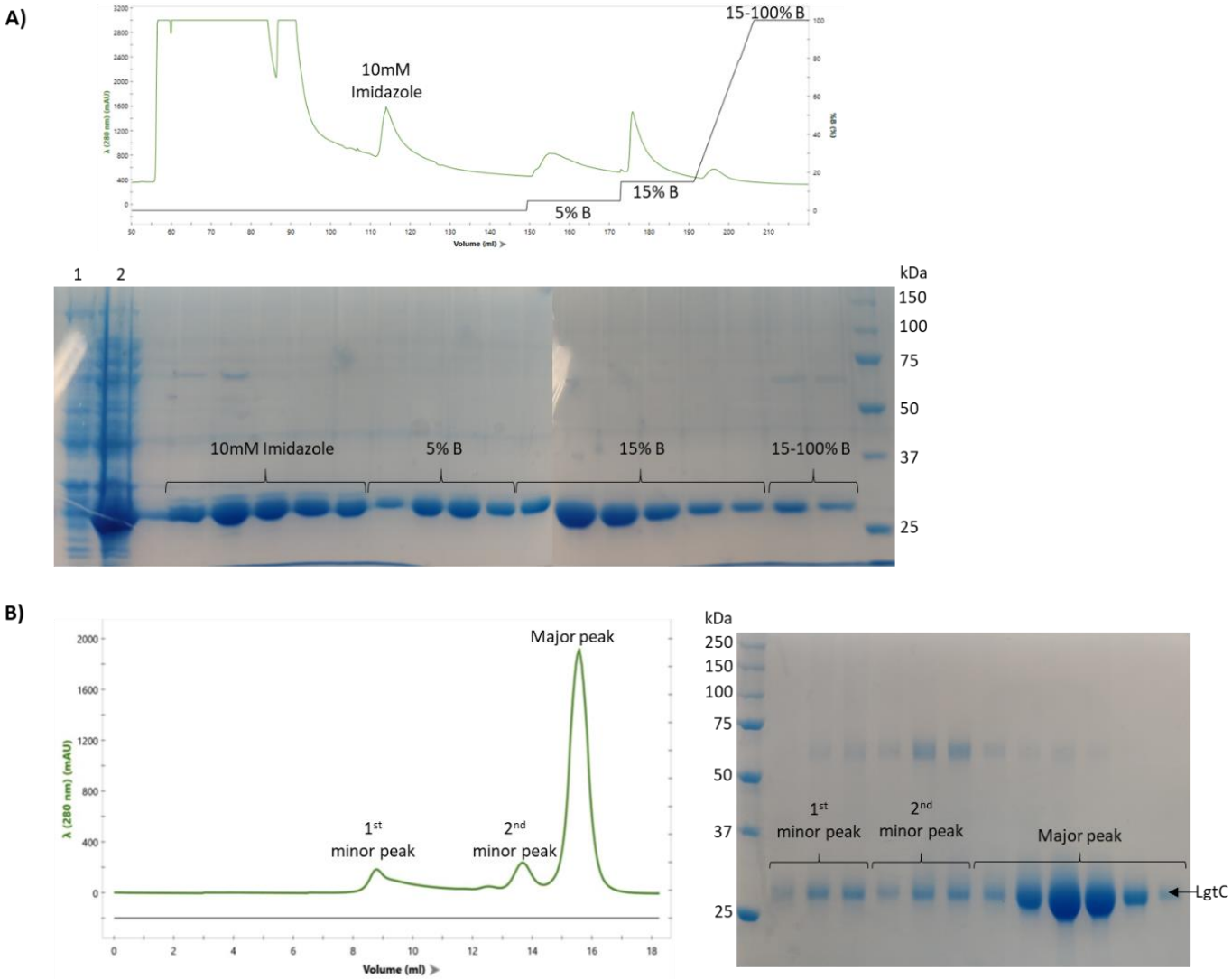


Figure 72: *A) IMAC purification of LgtC. SDS-PAGE analysis wells: total cell extract before (1) and after (2) IPTG induction, elution fractions from 10 mM, 25 mM (5%B), 75 mM (15%B) and 75-500 mM (15-100%B). B) SEC purification. SDS-PAGE analysis wells: 1st and 2nd minor peaks and the major peak from SEC. Yield from 1L culture was ~4 mg.*

III. Establishing experimental conditions for LgtC activity assays

The UDP-Glo™ assay from Promega was used to monitor LgtC activity. Buffer conditions that are compatible with the assay kit and show optimal enzyme activity were investigated. LgtC activity was first measured with the lactose acceptor. On the basis of data in the literature, two buffer conditions were tested: HEPES buffer at pH 7.5 and Tris buffer at pH 8.5. HEPES buffer was used previously for kinetic studies with LgtC (Persson et al., 2001). However, LgtC was reported to be very stable in solution with the more basic Tris buffer condition (Chan et al., 2012). Therefore, the Tris buffer was also tested for its compatibility in a UDP-Glo™ assay. In the Tris buffer condition, the conversion of manganese ions into permanganate at basic pH deprived the enzyme of manganese ions and thus lowered activity. As shown in Figure 73, the HEPES buffer was proven best for enzyme tests since it resulted in significantly higher enzyme activity at 4,668 ± 182 nmol.min^{-1}.mg^{-1}, this an 18-fold increase compared to the Tris buffer (254 ± 42 nmol.min^{-1}.mg^{-1}). Additionally, this experimental condition allows for the comparison of kinetics data with previous studies.

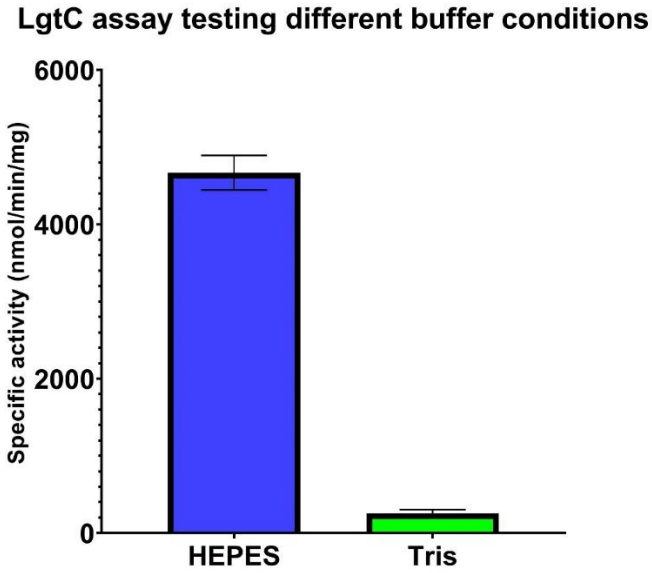


Figure 73: *LgtC assay testing two buffer conditions for optimal enzyme activity. HEPES buffer: 20 mM HEPES, 5 mM DTT, 5 mM $MnCl_2$, pH 7.5. Tris buffer: 20 mM Tris, 5 mM DTT, 5 mM $MnCl_2$, pH 8.5. Reaction conditions: 0.83 ng/µl LgtC (24 nM), 62.5 mM Lactose, 0.5 mM UDP-Gal, 10-minute reaction time.*

IV. Precipitation problems encountered in activity tests on GUVs

Initial trials of LgtC activity were performed on GUVs consisting of 29.7 mol-% DOPC, 30 mol-% Cholesterol, 40 mol-% LacCer and 0.3 mol-% Atto-647 DOPE, were formed. The LgtC reaction mixture was: 40 mol-% LacCer-containing GUVs, 20 mM HEPES, pH 7.5, 5 mM DTT, 5 mM $MnCl_2$, 100 ng/µl LgtC and 0.5 mM UDP-Gal. This reaction mix was incubated for 1 hour. After enzyme reaction, GUVs were incubated with 500 nM StxB-AF488 and visualised by confocal microscopy.

However, initial attempts at introducing LgtC to 40 mol-% LacCer GUVs resulted in the formation of precipitates, which look like crystals visible in confocal microscopy (Figure 74). The precipitates seem to be formed from aggregates of lipids and proteins. Specifically, clusters of StxB were observed in the same areas as the precipitant. We found that reducing the $MnCl_2$ and DTT concentrations from 5 mM to 1 mM stopped the formation of crystals. Precipitation was presumably caused by high concentrations of $MnCl_2$. The Mn^{2+} ions can form a complex with phospholipids and proteins, and this formed interfacial precipitates (Fullington and Hendrickson, 1966). For LgtC activity on GUVs, the buffer condition was changed to 20 mM HEPES, 1 mM $MnCl_2$, and 1 mM DTT at pH 7.5. This change in buffer conditions prevented the formation of precipitates and LgtC activity was observed on GUVs (manuscript in preparation).

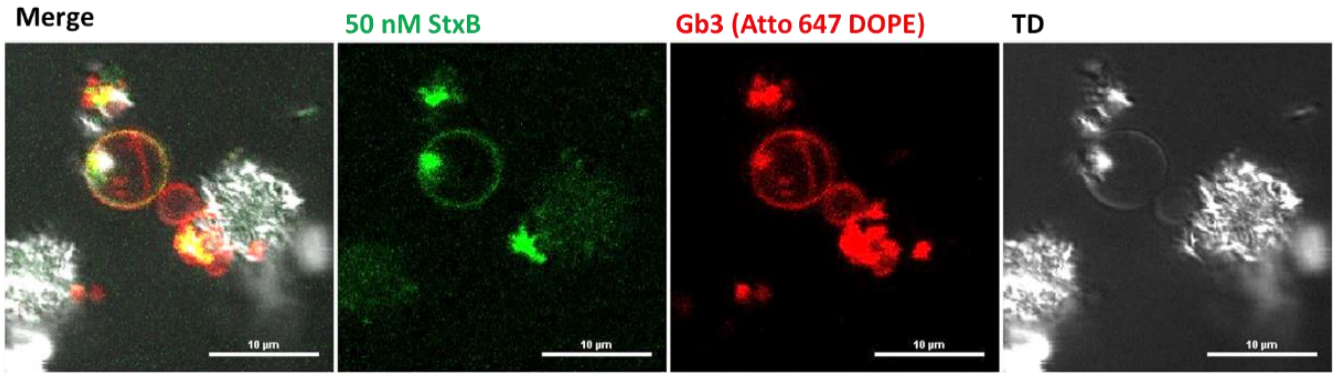


Figure 74: *Crystal formation upon addition of LgtC to LacCer-containing GUVs. Visualised fluorescence from StxB-Alexa fluor 488 (green) and A647-DOPE (red). Also visualised under confocal microscopy (TD). Finally, a merged image displays fluorescence image and from TD.*

V. Enzymatic glycosylation on synthetic membrane systems (manuscript in preparation)

Enzymatic glycosylation on synthetic membrane systems

Dylan Jabeguero[a]+, Lina Siukstaite[b]+, Chunyue Wang[c]+, Anna Mitrovic[b], Ralf P. Richter*[c], Winfried Römer*[b], Christelle Breton*[a]

[a] Centre de Recherches sur les Macromolécules Végétales (CERMAV) - CNRS
Université Grenoble AlpesGrenoble, France

[b] Faculty of Biology
Signalling Research Centres BIOSS and CIBSS
Freiburg Institute for Advanced Studies (FRIAS), Albert-Ludwigs-University Freiburg, Freiburg, Germany

[c] School of Biomedical Sciences, Faculty of Biological Sciences, School of Physics and Astronomy, Faculty of Engineering and Physical Sciences, Astbury Centre for Structural Molecular Biology, and Bragg Centre for Materials Research
University of Leeds
Leeds, United Kingdom

[+] These authors contributed equally to this work.

* Co-corresponding authors

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

Abstract: *Enzymatic glycosylation of lipids embedded in synthetic membrane systems has the potential to elucidate crucial information on glycosylation events occurring on the surface of membranes, yet hitherto has been sparsely explored. Here, we demonstrate that the soluble glycosyltransferase LgtC can glycosylate lactosylceramide to produce globotriaosylceramide (Gb3) on three distinct model membrane systems: micelles, supported lipid bilayers (SLBs) and giant unilamellar vesicles (GUVs). The formation of Gb3 in micelles was confirmed by thin layer chromatography. Binding of fluorescently-labelled B-subunit of Shiga toxin (StxB) to GUVs, observed by confocal microscopy, demonstrated glycosylation of a lipid bilayer membrane system. Quartz crystal microbalance with dissipation monitoring (QCM-D) demonstrated the transient binding of LgtC to the membrane during catalysis and showed the feasibility to monitor real-time glycosylation at the limits of sensitivity. LgtC showed decreasing catalytic efficiencies as we move from the soluble lactose to LacCer-containing micelles and then SLBs. The decreasing catalytic efficiency correlate to what would be expected in terms of substrate presentation. This work demonstrates the feasibility to utilize glycosyltransferases for the glycosylation of multiple types of membrane systems, with broader applications in identifying enzymatic behavior in physiological conditions.*

Glycosyltransferases (GTs) are responsible for the synthesis of membrane-bound glycoconjugates that form the glycocalyx, such as glycolipids and glycoproteins. These glycocalyx components play essential roles in many cellular events such as cell adhesion and signaling [1]. Synthetic membrane systems such as giant unilamellar vesicles (GUVs) and supported lipid bilayers (SLBs) have been utilized to study the underlying mechanisms underpinning adhesion and signaling processes occurring at or across the lipid membrane. For example, in conjunction with fluorescently labelled lipids and/or proteins, lipid-lipid or protein-lipids interactions on GUVs can be observed using confocal microscopy [2, 3, 14].

As tools in synthetic glycobiology, GTs have applications in decorating the surface of membranes. One study with lipid vesicles highlighted that the enzymatic activity of β(1,4)-galactosyltransferase was enhanced by an order of magnitude when the lipid substrate was presented in membrane microdomains

[4]. This example illustrates the capability of membrane model systems to characterize GT functions, and also highlights the sensitivity of GT activity to details of the membrane microenvironment. Another application utilized hybrid bilayers, consisting of a self-assembled alkanethiol monolayer and a lipid monolayer, in conjunction with QCM-D for real time monitoring of lipid mannosylation by α-1,3-mannosyltransferase (GumH) [5]. Micelles are classically used for membrane GTs acting on lipid substrates [6]. Being not planar, micelles can present the target lipidic substrate on a surface which could mimic substrate presentation on natural membranes. Overall, the utilization of GTs on synthetic membrane systems has potential for the characterization of glycosylation events that otherwise could not be perceived using classical *in-vitro* assays

In this report we utilized LgtC, a bacterial glycosyltransferase from *Neisseria meningitidis* involved in the synthesis of lipooligosaccharide. LgtC catalyzes the transfer of a galactose molecule from UDP-galactose to lactose-containing acceptors forming an α1,4 linkage. LgtC is also able to use glycolipid acceptors such as lactosylceramide (LacCer) to form globotriaosylceramide (Gb3, Galα1,4Galβ1,4Glcβ-Cer) [7]. LgtC has research value for its ability to synthesize Gb3, the product of which is the ligand of several lectins, e.g. the bacterial lectin Shiga-toxin B (StxB) from *Shigella dysenteriae* [15]. It also has application in synthetic chemistry owing to its acceptor substrate promiscuity [8]. LgtC is a well-studied enzyme, and was first produced as a soluble protein with a C-terminal truncation of 25 amino acids, which was used to determine its crystal structure [7]. The truncation was performed to increase solubility and production yields, while maintaining its activity [18]. Interestingly, the truncated region is proposed to play a role in membrane binding but this has not been demonstrated. Kinetic parameters for its minimal glycan acceptor, lactose, and many other acceptors have been determined [7, 8]. However, catalytic activity of LgtC using LacCer, the glycolipid acceptor, on membrane surfaces has not been investigated. This makes LgtC an ideal GT for the project, to demonstrate that soluble LgtC can be used to decorate a membrane surface and to understand the enzyme behavior in a membrane environment. In this work, we explore the capacity of LgtC to glycosylate LacCer in micelles, GUVs and SLBs. Specific binding of StxB to Gb3 was exploited for identification of the enzymatic product, and fluorescent-labelled StxB (StxB-AF488) was used for visualization of GUVs under confocal microscopy [9].

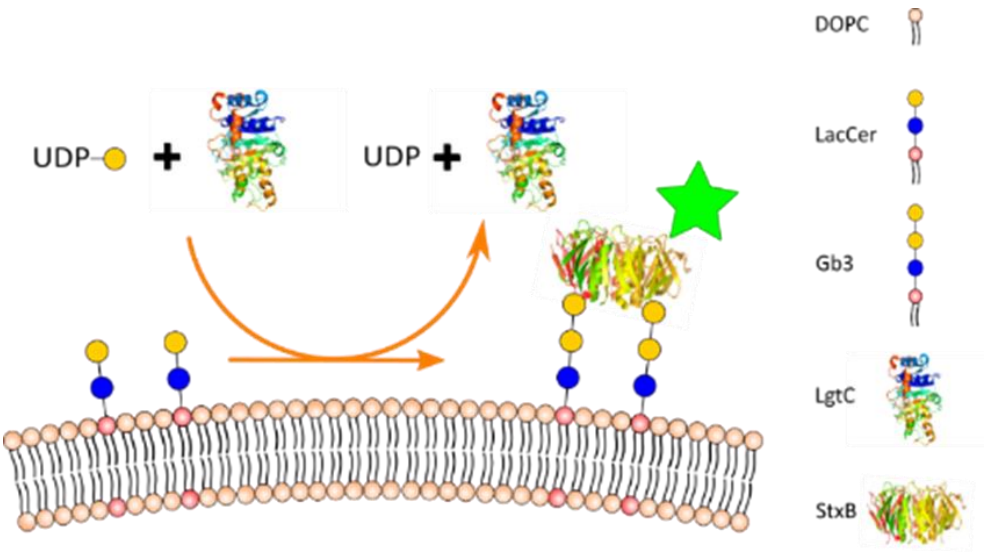


Figure 1. *Schematic representation of on-membrane glycosylation of LacCer by LgtC in a lipid bilayer. LgtC binds to UDP-Gal and then to LacCer in the lipid bilayer to initiate glycosylation turning LacCer to Gb3 by transferring a galactose molecule. LgtC then leaves the membrane along with the UDP product. Subsequently, StxB is introduced, which can be fluorescently labelled, and acts as a reporter molecule by specifically binding to Gb3. Sugars are represented as blue (glucose) and yellow (galactose) circles.*

Results and Discussion

LgtC activity on LacCer in micelles is less efficient than soluble acceptors. Recombinant LgtC with the 25 amino acid C-terminal truncation was produced and purified from *E. coli (BL21)* and enzymatic activity was first confirmed using lactose as the minimal acceptor (Figure 2A). A previous study validated the activity of LgtC with its glycolipid acceptor LacCer. However, LacCer was solubilised with methyl-β-cyclodextrin rather than presented in a lipid membrane environment [10]. In a first step to probe the impact of the membrane environment, we used a LacCer variant with a short C8 acyl chain, which is more soluble than naturally occurring LacCer, to form micelles. The resulting product is a C8 form of Gb3. Enzyme activity was performed with the UDP-Glo^{TM}Glycosyltransferase assay from Promega, which detects the UDP product formed from catalysis. As shown in Figure 2A, LgtC showed higher specific activity (2.6-fold) for the soluble acceptor lactose (307±5 nmole/min/mg protein) compared to LacCer (117±19 nmole/min/mg protein).

Thin layer chromatography analysis was performed as a confirmation of Gb3 formation during catalysis (Figure 2B). A TLC plate running on an eluent consisting of chloroform:methanol:water (65:35:5), can separate LacCer and Gb3. On TLC, the micelle enzyme reaction showed the presence of spots, for both LacCer substrate and Gb3 product indicating only partial conversion of LacCer to Gb3 after 12 h of enzyme reaction. This partial conversion suggests that, when present at a high enough concentration, the Gb3 product formed inhibits the enzyme by blocking its access to surrounding LacCer molecules in the micelles. Nevertheless, confirmation of activity for LacCer-containing micelles demonstrated the ability of soluble LgtC to glycosylate LacCer, without the aid of additives, and in a supramolecular lipid assembly. The next step of progression thus was tests of LgtC activity in synthetic lipid bilayer membrane systems.

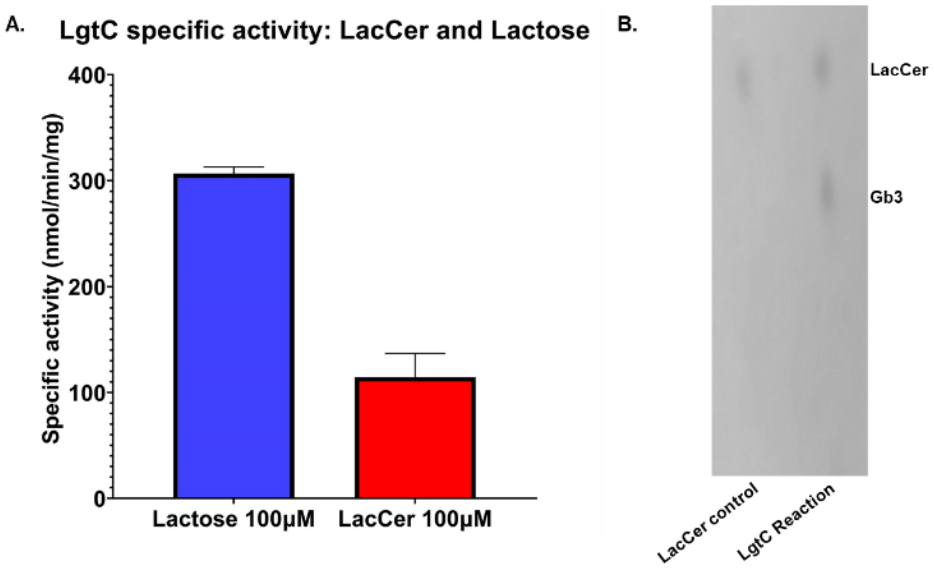


Figure 2: LgtC activity on LacCer micelles and soluble lactose. *A) Enzymatic assay comparing the specific activity with lactose or LacCer micelles. Conditions: 100 μM Lactose or LacCer, 0.25 mM UDP-Gal,10 ng/μl LgtC. B) TLC shows the formation of Gb3, enzyme reaction was run for 12 h to maximise the amount of Gb3. As a commercial C8-Gb3 is not available, only C8-LacCer was used as a negative control. C8-LacCer migrated with a Rf value of 0.61 and C8-Gb3 formed with a Rf value of 0.49.*

LgtC converts LacCer to Gb3 in GUVs. GUVs is a biomimetic model of cells composed of various membrane constituents, such as lipids and proteins. GUVs can be used to rebuild cellular processes for their study in a controlled environment. In the case of GTs, GUVs allow for qualitative evaluation of glycosylation reactions through confocal microscopy. GUVs were formed from a lipid solution containing LacCer (40 mol-%), DOPC (29.7 mol-%), cholesterol (30 mol-%) and Atto647N-DOPE (0.3 mol-%) using the electroformation procedure [16]. The GUVs formed were stable and varied in sizes ranging from 10 µm to 50 µm (Figure 3, red channel). The enzyme reaction consisted of 1 h incubation of the LacCer-containing GUVs with LgtC (100 ng/µl) and UDP-Gal (0.5 mM) at room temperature. Subsequently, GUVs were treated with 500 nM StxB-AF488 and visualized by confocal microscopy. The fluorescently labelled StxB (Figure 3, green channel) here reported the formation of Gb3. Indeed, StxB binding was observed on GUVs treated with LgtC, but not on untreated ones, demonstrating specificity of StxB binding and successful on-membrane conversion of LacCer into Gb3.

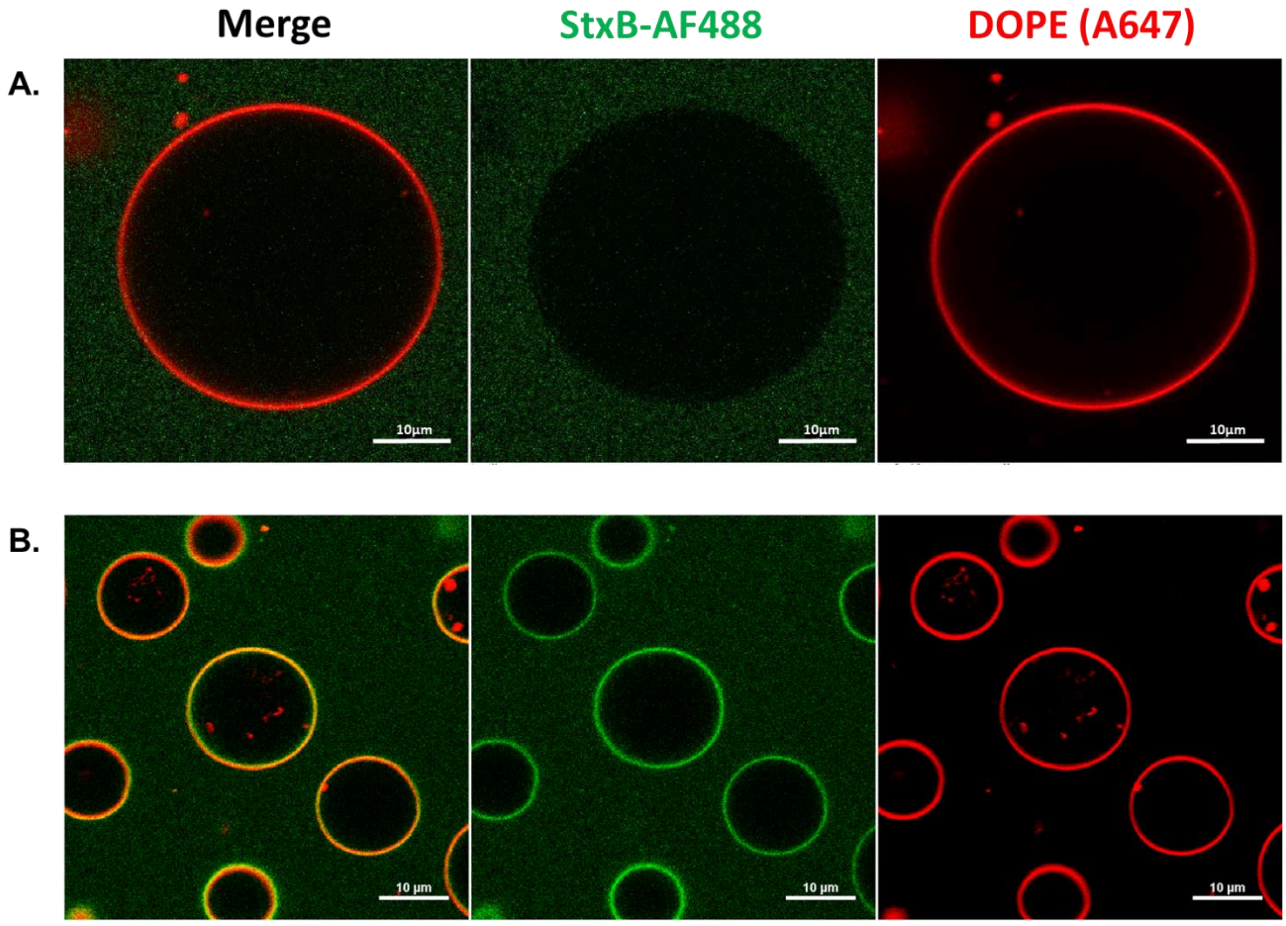


Figure 3: LgtC activity on LacCer in GUVs with StxB-AF488 binding to Gb3. *Confocal micrographs of LacCer-containing GUVs (visualised via Atto647N-DOPE tracer lipids; red channel) after enzyme reaction with LgtC and UDP-Gal, show StxB binding (visualized via Alexa488 label; green channel), demonstrating conversion of LacCer to Gb3. A) Negative control, with a focus on a single untreated 40 mol-% LacCer GUV. B) Zoomed out to multiple treated 40 mol-% GUVs. Reaction conditions: 20 mM HEPES, pH 7.4, 1 mM $MnCl_2$, 1 mM DTT, 100 ng/µl LgtC, 0.5 mM UDP-Gal and 40% LacCer GUVs, incubation time 1 h. Visualised by 500 nM StxB-AF488. Scale bar: 10 µm.*

Attempts were made at quantifying activity on GUVs by using the UDP-Glo assay. However, electroformation produced vesicles of varying size and quite likely also an uneven distribution of LacCer acceptors between vesicles, mirrored in uneven Gb3 production as seen from the varying levels of StxB staining of GUVs (Figure 3), thus impeding reliable quantification.

Monitoring enzymatic LacCer glycosylation in SLBs by QCM-D. To quantify LgtC activity in a lipid bilayer membrane system, we used SLBs instead. As illustrated in Figure 4A, small unilamellar vesicles (SUVs) were formed from a lipid solution composed of 5 mol-% LacCer and 95 mol-% DOPC, and SLBs were formed by the method of vesicle spreading on silica surfaces [17]. This method allows for robust control of membrane composition. QCM-D was used to monitor successful SLB formation and the subsequent enzyme reaction and reporting steps (Figure 4B-C). The enzyme reaction was initiated by exposure of the SLBs to a solution containing LgtC (600 ng/µl), UDP-Gal (0.25 mM) and $MnCl_2$ (0.5mM) under constant flow rate (20 µl/min). Afterwards, SLBs were washed with buffer and StxB binding was used as a reporter for Gb3 produced.

While the enzymatic reaction itself produced only a very small QCM-D response (*vide infra*), subsequent StxB binding clearly reported the presence of Gb3 after co-incubation of LacCer-containing SLBs with LgtC and UDP-Gal, indicating successful LacCer-to-Gb3 conversion in SLBs. At equilibrium, StxB bound to the enzyme-catalyzed LacCer surface with a frequency shift $\Delta f = 5.8$ Hz, and a dissipation shift $\Delta D = 0.05 \times 10^{-6}$, while there was no binding on surfaces in the absence of either LacCer, UDP-Gal or LgtC, demonstrating specificity of StxB binding to Gb3. Upon rinsing by buffer, the majority of bound StxB was removed from the surface, aligning with the weak interactions (Kd = 4mM) between StxB and Gb3 [13]. The pronounced yet reversible binding of StxB at µM concentrations is here likely due to multivalent recognition of several Gb3 molecules by the StxB pentamer harboring a total of 15 Gb3 binding sites.

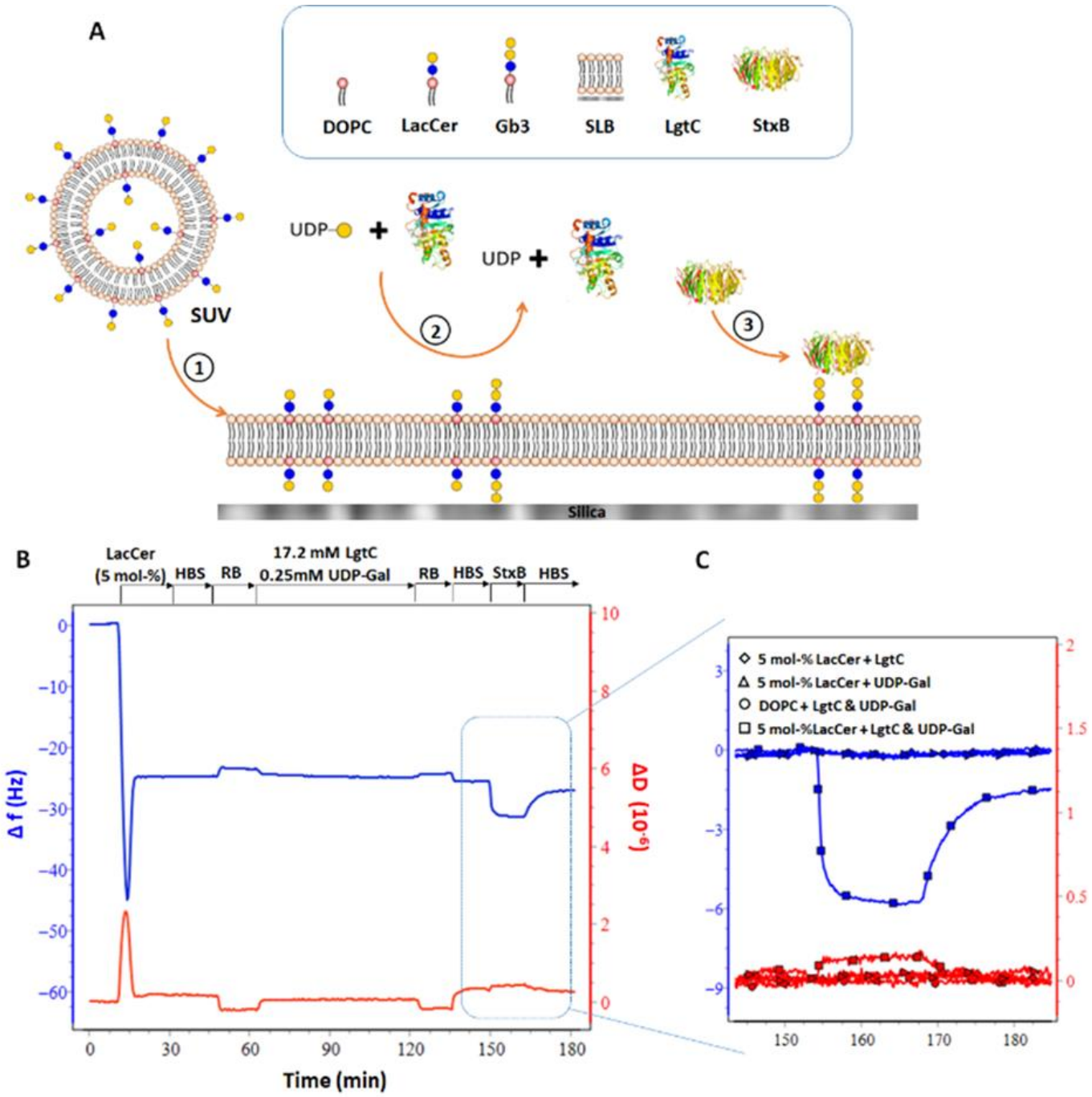


Figure 4: LgtC converts LacCer to Gb3 in supported lipid bilayers. *A) Schematic of the main steps of the assay: (1) supported lipid bilayer formation, (2) enzymatic conversion of LacCer into Gb3, (3) StxB binding reports on Gb3 content in the SLBs. B) Representative QCM-D data of the assay. A decrease in the frequency shift (Δf, blue line) indicates mass binding to the surface (including hydrodynamically coupled solvent). The dissipation shift (ΔD, red line) provides a measure for the softness of the surface-confined film. Conditions: HBS – HEPES buffered saline (10 mM HEPES, 150 mM NaCl, pH 7.4); RB – reference buffer (20 mM HEPES, 0.5 mM $MnCl_2$, 4.4 mM Tris, 1.1 mM TCEP, pH 7.4) ; SUV – 50 μg/mL; StxB – 6.5 μM. The start and the duration of incubation steps are indicated by the arrows above the graph. C) Zoom onto StxB binding step for the experiment shown in B, and controls of experiments that were lacking in panel B, either with LacCer, LgtC or UDP-Gal (lines with circles, triangles or diamonds respectively, as indicated), but otherwise identical.*

To shed more light on the reaction mechanism at the lipid membrane, the enzyme reaction was also performed under different conditions (Figure 5). Frequency shifts during the enzyme reaction remained minor even when the LacCer content in the SLB was increased 4-fold (to 20 mol-%) or LgtC concentration was increased 10-fold (to 17.2 µM). This indicates very little enzyme is bound to the surface at any time and a short-lived interaction (transient binding) between LgtC and LacCer is sufficient for an enzyme reaction.

Minor but significant decreases in the frequency (on the order of 1 Hz) were detectable during the enzyme reaction, demonstrating that QCM-D is sensitive to the enzymatic addition of a monosaccharide to LacCer. Indeed, the magnitude of the shifts (between -0.5 Hz and -2 Hz, depending in conditions; Figure 5) were comparable to $\Delta f \approx -3$ Hz predicted for a film thickness increase by 0.5 nm (i.e., equivalent to the size of a monosaccharide), and the slightly smaller values observed are attributed to incomplete coverage of the surface with Gb3.

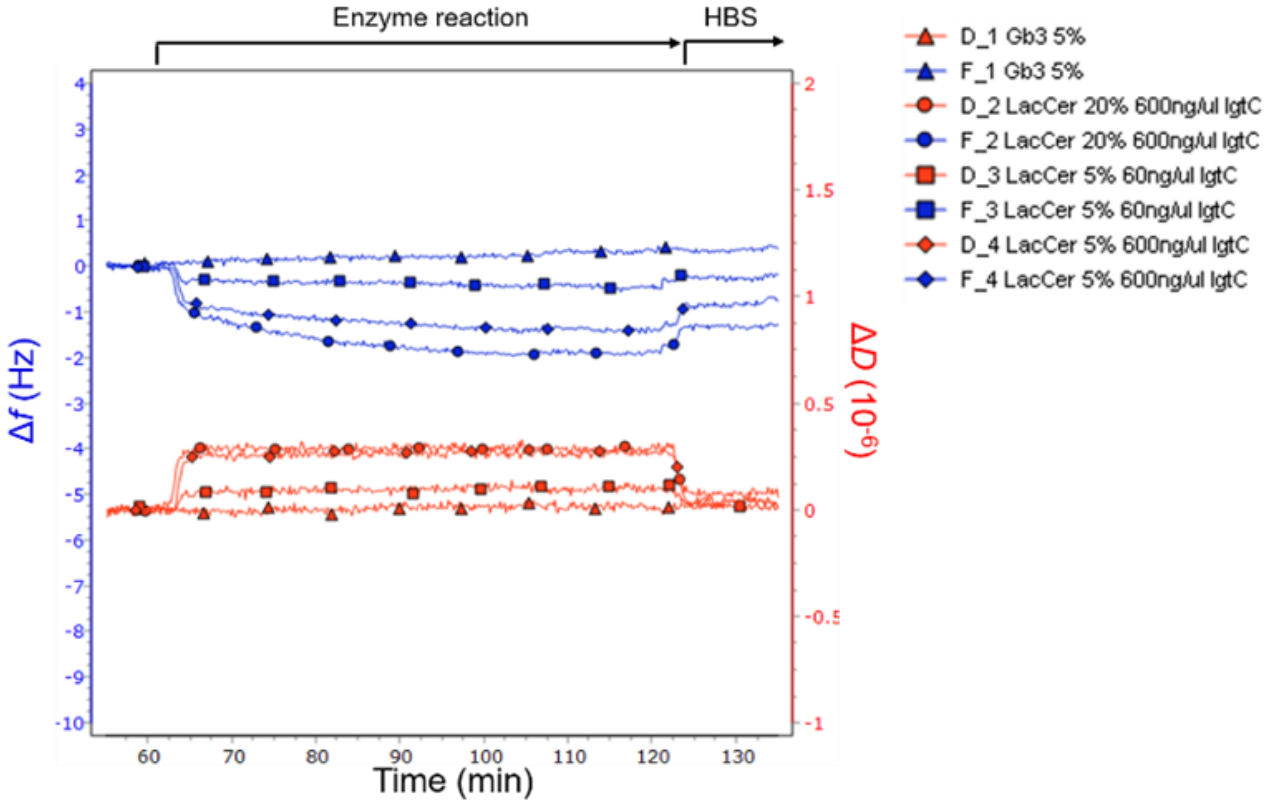


Figure 5: Direct monitoring of the enzyme reaction by QCM-D. *LgtC on LacCer containing SLBs at different reaction conditions. Conditions tested: 5 mol-% and 20 mol-% for LacCer, 1.72 µM and 17.2 µM for LgtC, reaction time - 1 h. Control: 5 mol-% Gb3, run with buffer for 1 hour.*

From the kinetics of the frequency response for 172 mM LgtC and 20 mol-% LacCer (Figure 5, line with circles), one can estimate that the reaction is completed within less than 1 h. Further analysis of StxB binding as a function of the reaction time corroborates this estimate and indicates half-maximal binding within approximately 20 min. From these numbers, a catalytic efficiency of ~24 $M^{-1}s^{-1}$ can be estimated for C8-LacCer in a lipid bilayer membrane context (see supporting information). Compared to the published data for lactose (4000 $M^{-1}s^{-1}$), this is substantially less efficient [8]. This value for LacCer must be taken only as a rough order of magnitude estimation due to two approximations, (i) the frequency shift is proportional to StxB surface density and (ii) the StxB surface density is proportional to Gb3 surface density. However, along with estimates of catalytic efficiency on micelles (735 $M^{-1}.s^{-1}$), which lies

between SLB and lactose, this would suggest reduced access to the substrate in the context of membranes (see supplementary data).

Future prospects. LgtC has a C-terminal region which is proposed to play a role in membrane interactions and activity as it is rich in basic residues suggesting interactions with phospholipids. It also contains several non-polar and aromatic residues that can form hydrophobic and electrostatic interactions with the membrane [7]. Therefore, while the truncated region may not have much effect on activity using soluble substrates such as lactose, it may play an important role on lipid membranes. For example, binding to the membrane, bringing the LgtC in close contact to the LacCer acceptor and presenting the lactosyl moiety in a more favorable orientation. In light of the low catalytic efficiency of LgtC in membranes (SLBs) and micelles, the C-terminal region may be vital for a good working activity of the enzyme in physiological conditions.

As previously demonstrated, glycosyltransferase activity can be magnified in the presence of lipid micro-domains formed from the lipid acceptor [4]. Evidence of LacCer lipid rafts is observed in immune cells, inflammation or apoptosis [11,12]. However, in regards to work on GUVs and SLBs, there is no evidence for the formation of micro-domains. This is due to the lack of sphingomyelin and low cholesterol in the SLBs and GUVs, which are essential in microdomain formation. As a future avenue of research, it may be of interest to investigate the catalytic activity of LgtC in the presence of LacCer micro-domains. Another avenue of research is the behavior of full-length LgtC in the presence of LacCer micro-domains, investigating the effects of the C-terminal region on activity and whether this can be applied to other GTs. Overall, increasing our understanding on the behavior of GTs in different membrane environment has value for engineering GTs. For example, a highly efficient GT in membranes will have applications in construction of glycocalyx components on vesicles.

Conclusion. We have demonstrated that soluble LgtC can be used to glycosylate LacCer in different membrane models, namely, micelles, GUVs and SLBs, to produce Gb3 in these models. Glycosylation was feasible but slowed down by an order of magnitude in micelles and two orders on lipid bilayer membranes, suggesting a reduced substrate access on lipid bilayers. This low efficiency should be considered and compensated for when glycosylating lipid membranes. Future work should aim to increase the efficiency of LgtC and other soluble GTs to make them more viable as synthetic glycobiology tools. As a tool for research purposes, the application of LgtC as an alternative method for producing Gb3 may be a cost effective option due to the relative difference in commercial value. On a wider scale this has implications on alternative methods for the synthesis of membrane systems, especially in scenarios where the glycolipid of interest is not commercially available. Moving a step forward, enzymatic modification of cells to create Gb3 is also a promising application that should be explored.

Acknowledgements

This work has received funding from synBIOcarb and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 814029. This work was supported by a German Research Foundation grant [Major Research Instrumentation (project number: 438033605)], the Excellence Initiative of the German Research Foundation [EXC 294 and EXC 2189], the Ministry of Science, Research and the Arts of Baden-Württemberg [Az: 33-7532.20], and the Freiburg Institute for Advanced Studies (FRIAS).

We thank Natalia Danielewicz and EnGenes Biotech for providing the StxB.

Keywords: GUVs • SLBs • Micelles • Gb3 • Lactosylceramide • LgtC • StxB • Glycosyltransferase •

References

[1] Kopitz, J., 2016. Lipid glycosylation: a primer for histochemists and cell biologists. *Histochemistry and Cell Biology*, 147(2), pp.175-198.

[2] Villringer, S., Madl, J., Sych, T., Manner, C., Imberty, A. and Römer, W., 2018. Lectin-mediated protocell crosslinking to mimic cell-cell junctions and adhesion. *Scientific Reports*, 8(1), 1932

[3] Schubert, T. and Römer, W., 2015. How synthetic membrane systems contribute to the understanding of lipid-driven endocytosis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(11), pp.2992-3005.

[4] Noble, G., Craven, F., Voglmeir, J., Šardzík, R., Flitsch, S. and Webb, S., 2012. Accelerated Enzymatic Galactosylation of N-Acetylglucosaminolipids in Lipid Microdomains. *Journal of the American Chemical Society*, 134(31), pp.13010-13017.

[5] Alves, C., Pedroso, M., de Moraes, M., Souza, D., Cass, Q. and Faria, R., 2011. Real-time investigation of mannosyltransferase function of a Xylella fastidiosa recombinant GumH protein using QCM-D. *Biochemical and Biophysical Research Communications*, 408(4), pp.571-575.

[6] Rocha, J., Audry, M., Pesce, G., Chazalet, V., Block, M., Maréchal, E. and Breton, C., 2013. Revisiting the expression and purification of MGD1, the major galactolipid synthase in Arabidopsis to establish a novel standard for biochemical and structural studies. *Biochimie*, 95(4), pp.700-708.

[7] Persson, K., Ly, H., Dieckelmann, M., Wakarchuk, W., Withers, S. and Strynadka, N. 2001. Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. *Nature Structural Biology*. 8(2), 166-175.

[8] Lairson, L., Watts, A., Wakarchuk, W. and Withers, S., 2006. Using substrate engineering to harness enzymatic promiscuity and expand biological catalysis. *Nature Chemical Biology*, 2(12), pp.724-728.

[9] Schubert, T., Sych, T., Madl, J., Xu, M., Omidvar, R., Patalag, L., Ries, A., Kettelhoit, K., Brandel, A., Mely, Y., Steinem, C., Werz, D., Thuenauer, R. and Römer, W. 2020. Differential recognition of lipid domains by two Gb3-binding lectins. Scientific Reports. 10(1), pp.1-12.

[10] Adlercreutz, D., Weadge, J., Petersen, B., Duus, J., Dovichi, N. and Palcic, M., 2010. Enzymatic synthesis of Gb3 and iGb3 ceramides. *Carbohydrate Research*, 345(10), pp.1384-1388.

[11] Bodas, M., Min, T. and Vij, N., 2015. Lactosylceramide-accumulation in lipid-rafts mediate aberrant-autophagy, inflammation and apoptosis in cigarette smoke induced emphysema. *Apoptosis*, 20(5), pp.725-739.

[12] Yu, W., Ying, J., Wang, X., Liu, X., Zhao, T., Yoon, S., Zheng, Q., Fang, Y., Yang, D. and Hua, F. 2021. The Involvement of Lactosylceramide in Central Nervous System Inflammation Related to Neurodegenerative Disease. *Frontiers in Aging Neuroscience*. 13.

[13] Gallegos, K., Conrady, D., Karve, S., Gunasekera, T., Herr, A. and Weiss, A. 2012. Shiga Toxin Binding to Glycolipids and Glycans. *PLoS ONE*. 7(2), e30368.

[14] Omidvar R and Römer W. 2019. Glycan-decorated protocells - novel features for rebuilding cellular processes. *Interface Focus*. 9(2): 20180084

[15] Siukstaite, L., Imberty, A. and Römer, W. 2021. Structural Diversities of Lectins Binding to the Glycosphingolipid Gb3. Frontiers in Molecular Biosciences. 8.

[16] Madl, J., Villringer, S. and Römer, W. 2016. Delving into Lipid-Driven Endocytic Mechanisms Using Biomimetic Membranes. Springer Protocols Handbooks.2, pp.17-36.

[17] Richter, R., Bérat, R. and Brisson, A. 2006. Formation of Solid-Supported Lipid Bilayers: An Integrated View. *Langmuir*. 22(8), 3497-3505.

[18] Wakarchuk, W.W., Cunningham, A., Watson, D.C. and Young, N.M. 1998. Role of paired basic residues in the expression of active recombinant galactosyltransferases from the bacterial pathogen Neisseria meningitidis. *Protein Engineering Design and Selection*. 11(4), pp.295–302.

Supporting Information

Experimental Procedures

Reagents

C8 lactosyl(β)-ceramide (LacCer), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), cholesterol and ganglioside GM3 from bovine milk were obtained from Avanti Polar Lipids (AL, USA). Atto 647N 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) was purchased from Sigma-Aldrich (Darmstadt, Germany). Ceramide trihexosides (Gb3) was obtained from Matreya (PA, USA). Shiga Toxin B-subunit (StxB) was kindly provided by enGenes Biotech (Vienna, Austria). UDP-Glo assay kit and ultra-pure UDP-Gal were bought from Promega (France). UDP-Gal for QCM-D analysis was purchased from Carbosynth (Compton, UK). HisTrap FF column was purchased from Cytiva (MA, USA). Superdex 200 Increase 10/300 GL was purchased from GE Healthcare (MA, USA). All other chemical reagents were of analytical or liquid chromatography grade.

LgtC Production

The gene of LgtC from *Neisseria meningitidis* was designed with 25 amino acids truncated at the C-terminus [1, 7]. Codons were optimised for *E. coli* expression and the mutations C128S, C174S [1] and T273A were introduced which were reported to improve expression levels of the protein without compromising the enzyme activity [2]. The gene was inserted into the pET-TEV vector using the restriction sites Ndel and Xhol, forming the LgtC-TEV construct. LgtC-TEV vector was transformed into the expression strain *E. coli BL21 (DE3).* Cell culture was grown in LB media to an optical density (OD_{600}) of 0.7-0.8 before induction with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Protein expression occurred overnight at 16^{o}C, with shaking at 180 rpm, before cells were harvested.

Cell pellets were disrupted using Cell Disruptor CSL (Constant Systems, Ltd) at 1.9 kbar and the cell lysate centrifuged at 24,000 × *g*, 4^{o}C, for 30 min. Supernatant was collected and clarified through a 0.22 μm filter and purified by 6-Histidine tag affinity chromatography using a nickel column. LgtC was further purified with size exclusion chromatography using a Superdex^{TM} 200 Increase 10/300 GL and stored in 20mM Tris, 5mM TCEP, pH 8.5 at -20C [2].

Enzyme activity assays

Measurement of activity - UDP-Glo assay

The UDP-Glo assay kit quantifies the amount of UDP product formed from a glycosyltransferase reaction. A white 96-well microplate (Ref. 655074, Greiner Bio-One) containing the nucleotide detection reagent was used to stop the glycosyltransferase reaction and initiate luminescence. Luminescence was read using the Spark microplate reader (TECAN). Enzymatic rate was determined using a UDP standard curve of known concentrations against luminescence, and activity obtained was expressed, depending on the data analysis, as relative light units (RLU), the specific enzyme activity (nmole/min/mg protein) or as a reaction rate (μM/min).

Lactose and lactosylceramide micelle activity

Lactose and C8 LacCer in micelle form were tested as acceptors for LgtC. Required amount of LacCer was dried in a glass tube under nitrogen and 30 µl buffer containing 20 mM HEPES, 1 mM $MnCl_2$, 1 mM DTT, pH 7.5 was added. To form micelles, the LacCer solution was sonicated at 35 kHz in an iced water bath sonicator (Bioblock Scientific) for 10 minutes, with temperature maintained below 20°C. LgtC was added to the reaction and the addition of UDP-Gal initiated the enzyme reaction. The enzymatic reaction contained 20 mM HEPES, 1 mM $MnCl_2$, 1 mM DTT, 250 µM UDP-galactose, varying lactosylceramide/lactose concentrations and 10 ng/µL (0.29 µM) LgtC at pH 7.5. The enzyme reaction was left to proceed for 10 minutes before the catalysis was stopped and the activity measured.

Thin layer chromatography

Thin layer chromatography (TLC) was used to detect the formation of Gb3 by chemical staining [3]. An enzyme reaction (10 µl) as described previously was spotted on a TLC Silica gel 60 F_{254} (Supelco) and ran using a chloroform:methanol:water (65:35:5) eluent. A sulphuric acid-Orcinol reagent was used to stain the Gb3.

Giant unilamellar vesicles (GUVs)

The GUVs were formed by a classical electroformation protocol as previously described [4]. In brief, solutions of lipid at a concentration of 0.5 mg/mL, composed of DOPC, cholesterol, Atto 647N DOPE and the desired glycosphingolipid, LacCer C8 or Gb3, at a molar ratio of 29.7:30:0.3:40 mol-% or 64.7:30:0.3:5 mol-%, respectively. Solutions were prepared in chloroform and spread on indium tin oxide (ITO) covered glass slides. To remove residual solvent, the slides were incubated under vacuum for at least one hour. A chamber was assembled with two slides, filled with 318 $mOsm.L^{-1}$ sucrose solution as formation buffer and an AC electrical field with a voltage of 1 V mm-1 was applied (to the chamber) for 2.5 h at room temperature. LgtC reaction on 40 mol-% LacCer GUVs was initiated in: 20 mM HEPES, pH 7.4, 1 mM $MnCl_2$, 1 mM DTT, 100 ng/µl LgtC, 0.5 mM UDP-Gal and incubated for 1 hour. GUVs were observed in hand-built chambers using 318 mOsm L-1 PBS as imaging buffer. Images were acquired by confocal fluorescence microscopy (Nikon Eclipse Ti-E inverted microscope using a Nikon A1R confocal laser scanning system with laser lines: 405 nm, 488 nm, 561 nm, 640 nm; 60x oil immersion objective, NA = 1.49; Nikon Instruments, Inc.) and analysed using the corresponding NIS Elements software (NIS Elements Confocal 5.20, Nikon Instruments, Inc.) and ImageJ (Fiji win64, Open source).

Quartz crystal microbalance with dissipation (QCM-D)

QCM-D measurements on silica-coated sensors (QSX303; Biolin Scientific, Västra Frölunda, Sweden) were performed with a Q-Sense E4 system (Biolin Scientific) equipped with 4 independent flow modules, connected to a syringe pump (Legato; World Precision Instruments, Stevenage, UK) to deliver a fluid flow of 20 $\mu L \cdot min^{-1}$. The working temperature was set to 24 °C. Changes in resonance frequency (Δf_i) and dissipation (ΔD_i) were acquired from six overtones (i = 3, 5, 7, 9, 11 and 13, corresponding to resonance frequencies of $f_i \approx$ 5, 15, 25, 35, 45, 55 and 65 MHz). Results from overtone i = 3 are presented unless stated otherwise, and frequency shifts are presented normalised by the overtone number ($\Delta f = \Delta f_i/i$). All other overtones provided qualitatively similar data.

Small unilamellar vesicles (SUVs) containing DOPC and C8 LacCer were prepared as previously described, with modifications. Briefly, lipids in chloroform were mixed at desired molar ratios (100:0, 95:5 and 80:20) at a total amount of 5 µmol, and dried under a stream of nitrogen gas followed by drying in a vacuum desiccator for 2 h. The lipid mixtures were re-suspended in working buffer at 2 $mg \cdot mL^{-1}$, and homogenised by five cycles of freezing, thawing and vortexing. To obtain SUVs, the lipid suspension was subjected to tip sonication in pulse mode (1 s on / 1 s off) for 15 min with refrigeration. The SUV suspension was then centrifuged at 12,100 × *g* for 10 min to remove titanium debris (shed from the sonicator tip), and stored at 4°C under nitrogen gas until use.

QCM-D sensors were pre-treated in UV/ozone (30 min), placed in the flow chamber, and connected to the Q-Sense E4 apparatus. Following acquisition of a baseline in HBS buffer (10 mM HEPES, 150 mM NaCl, pH 7.4), SLBs were formed on the sensor surface by the flow of 50 µg/ml liposomes until frequency and dissipation were stable. SLBs were equilibrated first in HBS buffer, and then in reaction buffer (RB; 20 mM HEPES, 0.5 mM $MnCl_2$, 4.4 mM Tris, 1.1 mM TCEP, pH 7.4), which was matched in content with the LgtC storage to avoid solution effects on the QCM-D response upon starting the enzyme reaction. Enzyme reaction was initiated in RB containing 250 µM UDP-Gal and desired concentrations of LgtC (1.72 µM or 17.2 µM). SLBs were washed, first with RB to stop the enzyme reaction, and then with HBS for further analysis. A flow of 6.5 µM StxB in HBS was applied and StxB binding was observed.

Supporting Results and Discussion

Varying concentration of UDP-Gal

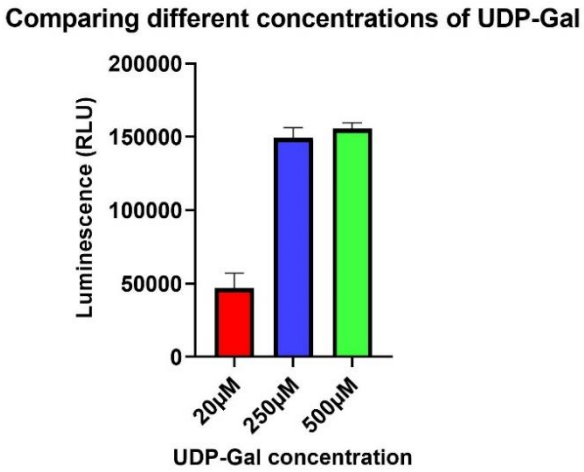


Figure S1: *Comparing enzyme activity at different UDP-Galactose concentrations. Conditions: 10 ng/µL (0.29 µM) LgtC and 0.1mM LacCer was maintained.*

As previously described, the K_m of LgtC for UDP-Gal was 18 µM [1]. However, this K_m was measured using lactose as the acceptor, which may change with LacCer as the acceptor. The activity of LgtC at differing UDP-Gal concentrations was measured in order to ascertain future reactions with LacCer were under saturating UDP-Gal concentrations. Signal from 20 µM is approximately a third compared to the other two concentrations. This suggests that UDP-Gal

K_m may be slightly higher when using a LacCer acceptor. Figure S1 demonstrates that from 250 μM to 500 μM there is no change in the activity, indicating that reactions at this range are at saturating conditions of UDP-Gal. For efficient usage of UDP-Gal, a concentration of 250 μM was used in further experiments.

Varying concentrations of LacCer in micelles

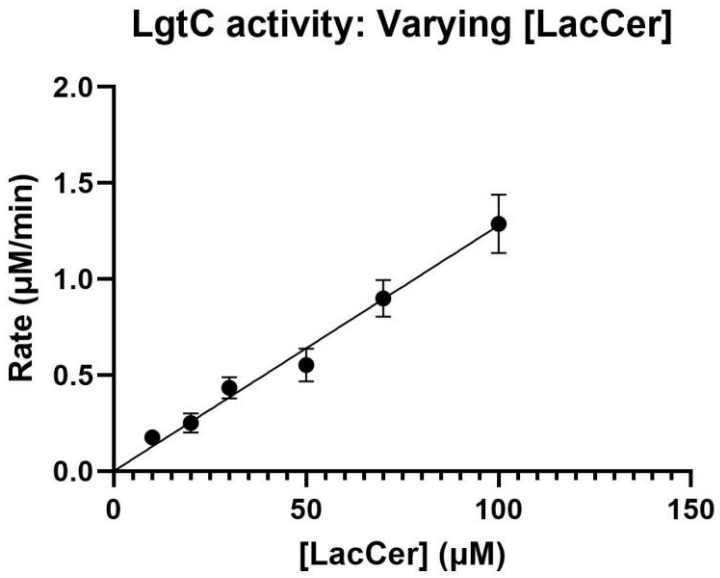


Figure S2: *LgtC at varying LacCer micelle concentrations. Conditions: 10 ng/μL LgtC and 250 μM UDP-Gal. R^2 = 0.95. Data was analysed with GraphPad Prism.*

A critical micelle concentration (CMC) of 150 nM for mixed LacCer (mixture of variable length acyl chain species) was identified [5]. While the CMC of C8 LacCer would vary from the reported value for mixed LacCer, this information nevertheless allows for estimation of a lower bound for the formation of LacCer micelles. At ~200 μM LacCer, precipitation was observed. This is in line with the maximum solubility concentration observed with other short chain ceramides at 100 μM [7]. However, this limited the LacCer range at which LgtC activity could be measured. As a result, saturating concentrations of LacCer for the LgtC reaction was not reached and a full Michaelis-Menten curve could not be achieved. Nevertheless, the data shows a linear rate at the measured LacCer concentrations. This implies that values measured are in the linear rate phase and therefore at $[S] \ll K\text{m}$, the analytical region for the analysis of enzyme substrates. In this condition, the Michaelis-Menten equation simplifies to:

$$v_0 \approx \frac{Vmax[S]}{Km} = \frac{Kcat[E][S]}{Km}$$

[Eq. 1]

The best fit results in a straight line equation of $Y = 0.01279X$. Where:

$$\frac{Kcat[E]}{Km} = 0.01279$$

[Eq. 2]

Solving for catalytic efficiency results in:

$$\frac{Kcat}{Km} = \frac{0.01279}{[E]}$$

[Eq. 3]

Upon input of total enzyme concentration (0.29 µM), a catalytic efficiency of 735 $M^{-1}.s^{-1}$ is obtained. This is an order of magnitude lower than that published for lactose (4000 $M^{-1}.s^{-1}$), meaning that LgtC is less efficient for LacCer micelles.

Approximating catalytic activity on supported lipid bilayers with QCM-D

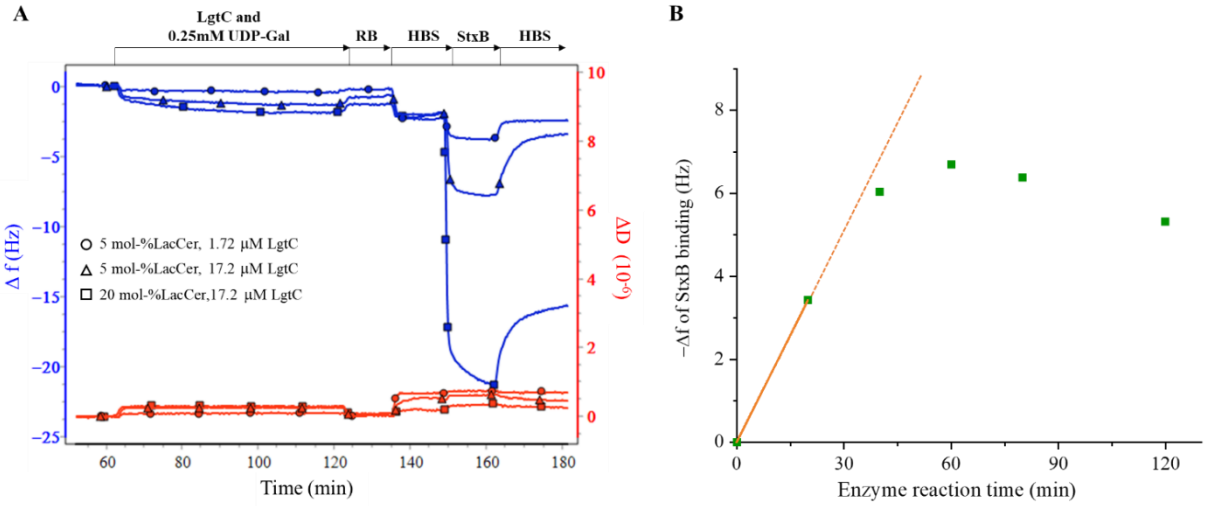


Figure S3. Parameters determining the reaction rate. *A) Representative QCM-D data of the enzyme reaction under different conditions (lines with symbols as indicated), and subsequent reporting of Gb3 by StxB binding. The net frequency changes after the enzyme reaction step were minor, ranging from -0.3 Hz (5mol-% LacCer with 1.72µM LgtC) to -0.8 Hz (5mol-% LacCer with 17.2µM LgtC) to -1.0 Hz (20mol-% LacCer with 17.2µM LgtC), and the net dissipation shifts were negligible. The frequency shifts at equilibrium of StxB binding were -1.5 Hz, -5.8 Hz and -18.8 Hz, respectively.* ***B)*** *Plot of -Δf for equilibrium StxB binding (at 6.5 µM StxB; a measure of the Gb3 content in the SLB) vs. enzyme reaction time. Conditions: 5 mol-% LacCer, 17.2 µM LgtC and 0.25 mM UDP-Gal. The reaction is complete within approximately 60 minutes, leading to -Δf = 6.5 ± 0.5 Hz (n = 3). The slope of the orange line (0.17 Hz/min) is a measure of the reaction rate.*

Figure S3A shows that LacCer turnover increases with enzyme and LacCer concentration. This indicates that the reaction is only limited by LacCer concentration as it has been proven that we are already working under saturating UDP-Gal concentration (Figure S1). From varying the enzyme reaction time, we observed a total conversion of LacCer, in the outer leaflet, within 1hour (Figure S3B). A simple model suggests a catalytic efficiency of approximately 24 M^{-1} s^{-1}, and that the reaction (under the conditions probed) is limited by the affinity between LacCer and LgtC (but not by the affinity between UDP-Gal and LgtC). The model is clarified below:

The below derivation takes into account the particulars of enzymatic reactions at surfaces *via* the Langmuir isotherm. We consider the enzymatic reaction:

$$\mathrm{E} + \mathrm{S} \rightleftarrows \mathrm{ES} \rightarrow P$$

[Eq. 4]

In our case, the enzyme E is free to diffuse in the solution phase (and thus expressed in molar concentration), but the substrate S, the enzyme-substrate complex ES and the product P are confined to the surface (and thus expressed in molar surface density).

By definition, the enzyme converts the bound substrate (ES) into a product through first order rate constant (also called turnover number) k_{cat}

$$\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t} = v = k_{\mathrm{cat}}[\mathrm{ES}]$$

[Eq. 5]

Assuming instantaneous chemical equilibrium at the surface, due very low interaction with the LacCer SLB (Figure 4), the surface density of ES is given by the Langmuir isotherm

$$[\mathrm{ES}] = [\mathrm{S}]_{\mathrm{tot}} \frac{[\mathrm{E}]}{K_{\mathrm{d}} + [\mathrm{E}]}$$

[Eq. 6]

where K_{d} is the equilibrium dissociation constant of ES, and $[\mathrm{S}]_{\mathrm{tot}}$ the total surface density of available substrate. In the limit of small enzyme concentrations ($[\mathrm{E}] \ll K_{\mathrm{d}}$) we can approximate $K_{\mathrm{d}} + [\mathrm{E}] \approx K_{\mathrm{d}}$. Eq. 5 and 6 then combine to:

$$v = \frac{k_{\mathrm{cat}}}{K_{\mathrm{d}}} [\mathrm{E}][\mathrm{S}]_{\mathrm{tot}}$$

[Eq. 7]

In our essay (constant flow), the enzyme concentration in solution is maintained constant and equal to the enzyme concentration injected in the system

$$v = \frac{k_{\mathrm{cat}}}{K_{\mathrm{d}}} [\mathrm{E}_0][\mathrm{S}]_{\mathrm{tot}}$$

[Eq. 8]

Near the start of the reaction, the total substrate concentration $[\mathrm{S}]_{\mathrm{tot}}$ is close to the initial substrate concentration $[\mathrm{S}_0]$, and thus the initial reaction velocity is

$$v = \frac{k_{\mathrm{cat}}}{K_{\mathrm{d}}} [\mathrm{E}_0][S_0]$$

[Eq. 9]

By measuring v_0 for known total enzyme and substrate concentrations, we can thus readily calculate the catalytic efficiency

$$\frac{k_{\mathrm{cat}}}{K_{\mathrm{d}}} = \frac{v_0}{[\mathrm{E}_0][\mathrm{S}_0]}$$

[Eq. 10]

In our experiment with LgtC and LacCer, K_d is estimated to be in the mM range (literature indicates $K_\mathrm{M} = 40$ mM for IgtC and lactose, and the K_d is likely to be similar or higher for LacCer). The enzyme concentration is $[\mathrm{E}_0] = 17.2\ \mu\mathrm{M}$. The assumption of low enzyme concentration thus is clearly fulfilled and Eq. 7 is validated.

From Figure S3B, we find half-maximal frequency shift in StxB binding after about 20 min of enzyme reaction. If we make the crude approximations that (i) the frequency shift is proportional to StxB surface density and (ii) the StxB surface density is proportional to product (Gb3) surface density, then this amounts to $v_0 \approx \frac{[\mathrm{S}_0]}{2}/20\ \mathrm{min}$ and $\frac{k_\mathrm{cat}}{K_\mathrm{d}} \approx \frac{1}{17.2\ \mu\mathrm{M} \times 2400\ \mathrm{s}} = 24\ \mathrm{M}^{-1}\mathrm{s}^{-1}$.

Assumptions (i) and (ii) above are rather crude, and the catalytic efficiency $\frac{k_\mathrm{cat}}{K_\mathrm{d}} \approx 24\ \mathrm{M}^{-1}\mathrm{s}^{-1}$ thus should be considered a rough 'order-of-magnitude' estimate.

In the above analysis, we have neglected that LgtC has two substrates. The above model can be extended if needed to include the effect of UDP-Gal on the reaction rate. However, in the assay conditions described above, the concentration of UDP-Gal (250μM) is >10-fold the K_m of LgtC for UDP-Gal (18μM). and we can therefore assume that LgtC is essentially saturated with UDP-Gal throughout the reaction process. This was evidenced experimentally, where the enzymatic rate didn't really change when the concentration of UDP-Gal increased from 250μM to 500μM (Figure S1). In summary, the above analysis provides an order-of-magnitude estimate of the catalytic efficiency of LgtC ($\frac{k_\mathrm{cat}}{K_\mathrm{d}} \approx 24\ \mathrm{M}^{-1}\mathrm{s}^{-1}$).

Comparison of catalytic efficiency of different acceptors

Comparing the catalytic efficiency from lactose, and LacCer-containing micelles and SLBs, we see a pattern: Lactose > LacCer-Micelle > LacCer-SLB. It correlates well with what would be expected in terms of substrate access. Therefore, a possible reason for the drop in catalytic efficiency is the constraint in substrate presentation as we transition from a free soluble substrate, to a surface presenting substrate. As micelles have much more freedom of movement in solution, it is logical that they will have a higher catalytic efficiency compared to SLBs, where the LacCer is constrained to the planar membrane. As a final perspective, this drop in activity at membrane interface brings into question the role of the truncated C-terminal of LgtC, which could play a role in membrane association and activity.

References

[1] K. Persson, H. Ly, M. Dieckelmann, W. Wakarchuk, S. Withers, N. Strynadka, *Nature Structural Biology* 2001, *8*, 166-175.

[2] P. H. W. Chan, S. Weissbach, M. Okon, S. G. Withers, L. P. McIntosh, *Biochemistry* 2012, 51, 41, 8278–8292

[3] D. Adlercreutz, J. Weadge, B. Petersen, J. Duus, N. Dovichi, M. Palcic, *Carbohydrate Research* 2010, 345, 1384-1388.

[4] J. Madl, S. Villringer, W. Römer, *Springer Protocols Handbooks* 2016, 17-36.

[5] B. Ulrich-Bott, H. Wiegandt, *Journal of Lipid Research* 1984, 25, 1233-1245.

[6] J. Sot, F. Goñi, A. Alonso, *Biochimica et Biophysica Acta (BBA) - Biomembranes* 2005, 1711, 12-19.

[7] W. Wakarchuk, A. Cunningham, D. Watson, N. Young, Protein Engineering Design and Selection 1998, 11, 295-302.

VI. LgtC activity on SLB with no flow rate – preliminary tests

1. Introduction

After LgtC reactions on SLBs were performed, we realised that a large amount of LgtC and UDP-Gal substrate were wasted. Briefly, we investigated the possibility to make the LgtC reaction on SLBs more cost effective.

Enzymatic activity monitored using QCM-D was typically run under a constant flow rate. This was ideal as the solution was being refreshed, meaning that substrate concentration remained constant. The advantage of this is that UDP-Gal is not depleted and free UDP in solution is washed away, therefore, reaction rate are kept at an optimal level. However, this approach results in large amounts of wasted LgtC and UDP-Gal. While the costs of nucleotide sugars have steadily decreased, they remain expensive, especially for certain donors such as GDP-fucose. In addition, the constant flow may 'push away' the enzyme from the membrane, thus slowing catalysis. Thus, we also investigated whether the flow rate affected the enzymatic activity.

In this section, we performed LgtC catalysis on LacCer-SLBs under zero flow rate (still chamber) and compared this to a reaction under constant flow rate. For the two methods, we monitored the frequency shift (Δf) by QCM-D and compared the Δf from StxB binding. We aimed to identify whether the flow rate affects catalysis and whether it is more cost-effective to perform a still chamber reaction.

2. Still chamber LgtC reaction

The still chamber experiments followed the same method as described in the manuscript with minor changes described briefly below. SLBs containing 5 mol-% or 20 mol-% LacCer were formed. For each LacCer concentration two enzyme concentrations were tested 60 ng/µl and 600 ng/µl. This results in a total of 4 enzyme reactions (Figure 75). Once SLBs are formed, HBS and RB buffers are used to equilibrate the SLBs. To start the still chamber reaction, the reaction solution was introduced into the SLB chambers at a flow rate of 20 µl/min for 10 minutes. The reaction solution contained: 20 mM HEPES, pH 7.4, 0.5 mM MnCl2, 0.25 mM

UDP-Gal and 60 or 600 ng/µl LgtC. This was to ensure that all the chambers contained the same starting amount of substrates before flow rate was stopped. After 10 minutes, the flow rate was stopped and the enzyme reaction was allowed to continue a further 50 minutes. This was a total reaction time of 1 hour (10 minutes on flow rate and 50 minutes without flow). Afterwards, enzyme reaction was stopped by washing away LgtC and UDP-Gal with RB and HBS buffer. After the enzyme reaction, Gb3 production was monitored by introducing SxtB (6.5 µM / 50 µg/ml) to the chamber at 20 µl/min and the change in frequency (Δf) from StxB binding to the SLBs was recorded 10 minutes after StxB first was flowed into the chamber.

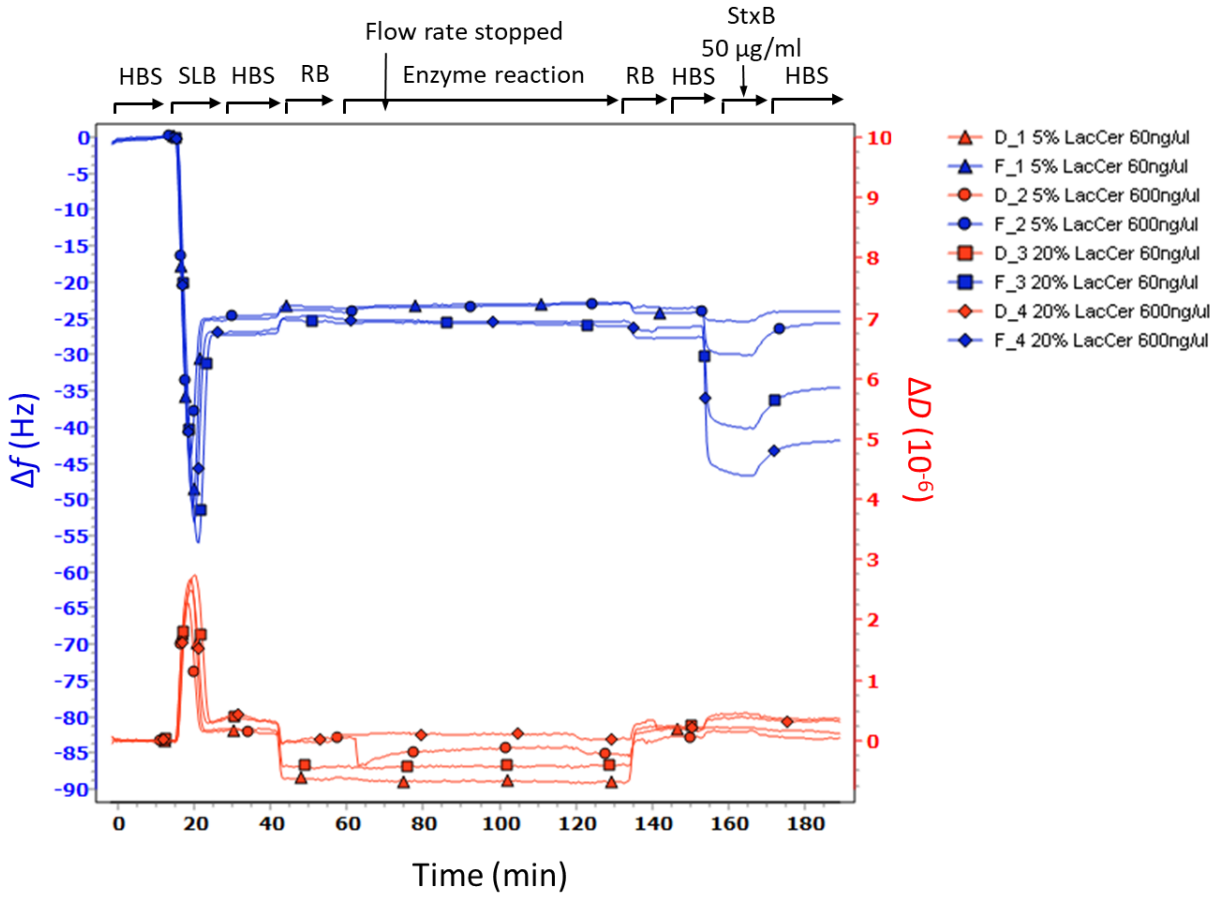


Figure 75: *Overview of the entire 'Still chamber' reaction experiment.*

3. Comparison of StxB binding – Flow rate vs Still Chamber

There is not much difference in the Δf from StxB binding when running a 'Still chamber' or a 'Flow rate' (Table 19). This indicated that a similar amount of Gb3 was synthesized in both methods. Therefore, a still chamber is a more cost-effective method. The large concentration of UDP-Gal can explain the similarity between the two methods. Reactions were run at 0.25 mM UDP-Gal, more than 10-fold the Km value for UDP-Gal (18 µM) (Persson et al., 2001). Likely not enough UDP-Gal was depleted throughout the 1-hour reaction to affect the rate of reaction.

As similar Δf were obtained from both methods, it also means that the flow rate did not affect the rate of reaction of LgtC. The flow of solution within the reaction chamber is far away from the surface of the SLB, thus any force applied to the SLB from the flow is minimal.

A constant flow rate allows us to maintain concentrations of substrates and proteins constant in a reaction. This is ideal for real-time monitoring of catalysis. However, if the measurement of the end product is the only goal, i.e. StxB binding to Gb3, then a 'Still chamber' is a better option.

Comparison of StxB binding between a Still chamber or Flow rate reaction		
	Frequency shift - Δf (Hz)	
	Still chamber	**Flow rate**
5% LacCer (60ng/µl)	-1.06	-1.54
5% LacCer (600ng/µl)	-6.36	-5.77
20% LacCer (60ng/µl)	-12.21	-14.57
20% LacCer (600ng/µl)	-20.31	-18.77

Table 19: *Comparison between 'Still chamber' and' Flow rate' reaction by StxB binding. Δf values taken at the 10-minute time point after flowing StxB (6.5µM / 50 µg/ml) to the chamber at 20 µl/min.*

VII. MGD1 activity on GUVs – Preliminary results

1. Introduction

There are still many more questions to be answered in regards to the behaviour of MDG1 on membranes that can be investigated.

In this section we took the opportunity to demonstrate that MGD1 can have activity on GUVs. All activity tests involving MGD1 that have been performed to date were with mixed micelle composed of DAG and PG/PA. While micelles are ideal for use in activity assays, there are several differences compared to liposomes such as GUVs, LUVs or SUVs. For example, structural presentation of the lipids, the dynamic movement of lipids on the membrane and even the curvature. The binding of MGD1 towards Langmuir monolayers consisting of photosynthetic membrane components such as: PG, PA, DAG, MGDG, DGDG and combinations of them have been previously described (Sarkis et al., 2014; Rocha et al., 2016; Nitenberg et al., 2020).

In this work we aimed to demonstrate MGD1 activity on the surface of GUVs. To achieve this goal, we explored the possibility to detect the product reaction MGDG formed in GUVS, using a fluorescently labelled lectin, which can then be visualised on confocal microscopy. Also, we explored the application of fluorescently labelled MGD1 (MGD1-AF488) by testing its binding to GUVs made of different lipid compositions.

This work was performed as part of the secondment in the BIOSS laboratory (University of Freiburg, Germany), with guidance from fellow synBIOcarb colleague Lina Siukstaite and Professor Winfried Römer.

2. Detection of MGDG using various lectins

We synthesized GUVs using the electroformation procedure described in chapter 2. The GUVs contained the lipids MGDG 7.5 mol-%, PG 2.5 mol-%, DOPC 89.7 mol-% and Atto-647 DOPE 0.3 mol-%. This composition was chosen as it was optimal for formation of abundant and stable GUVs. MGDG at 30 mol-% and PG 10 mol-% were also tested but resulted in very small and unstable GUVs (see appendix). MGDG was included as it is the product from MGD1

reaction. DAG was not present for these preliminary assays but PG was included as it is required for MGD1 reaction. Fluorescently labelled DOPE (Atto-647 DOPE) was included in small amounts to visualise the GUVs without lectin binding.

DOPC is a bilayer forming lipid that facilitates the formation of stable GUVs. Additionally, DOPC is a neutral lipid that MGD1 does not bind (Dubots et al., 2010).

We tested different lectins for their ability to bind to MGDG and thus act as a reporter for MGD1 activity (Figure 76). The galactose moiety in MGDG is a β-Gal. Therefore, the three lectins chosen for this experiment were specific for terminal β-Gal. RPL-Gal1 (*Pseudomonas aeruginosa*), RPL-Gal4 (*Photorhabdus asymbiotica*) and GS-IB4 isolectin (*Griffonia simplicifolia*). GUVs were incubated with 200 nM of each fluorescently-labelled lectin and instantly visualised. Under fluorescence microscopy, none of the lectins could bind to MGDG-containing GUVs. Perhaps as the galactose is the first sugar in the lipid, it is too close to the surface of the membrane. In this case, the membrane itself can block binding via steric hindrance.

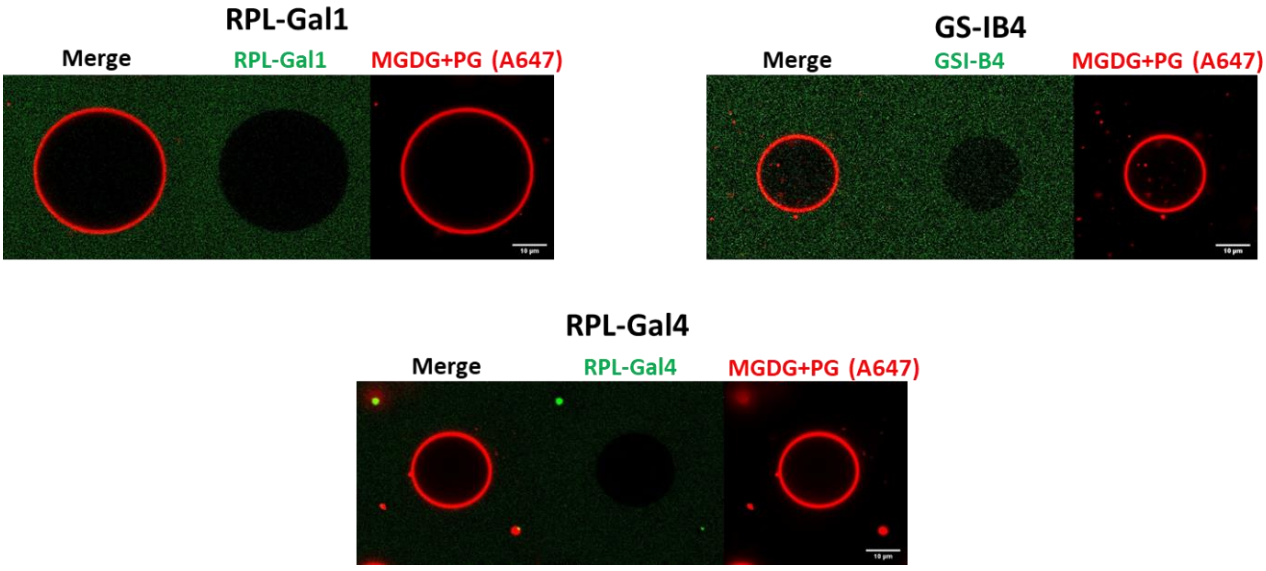


Figure 76: *Detection assay of MGDG on the surface of GUVs using fluorescent lectins. Lectins tested are: RPL-Gal1, RPL-Gal4 and GS-IB4 isolectin fluorescently labelled with Alexa Fluor 488 (AF488). GUV composition: DOPC (89.7%), MGDG (7.5%), PG (2.5%), Atto 647 DOPE (0.3%). All GUVs are incubated in 200 nM of lectin. 10 µm scale bar.*

We could not find a lectin to bind MGDG in the short time frame of the secondment and thus could not proceed with activity tests of MGD1 on GUVs. With more time, other solutions could have been considered. Thin layer chromatography can be used as an indicator of MGDG production in GUVs. However, the GUVs would have to be formed without sucrose due to potential co-migration of sucrose with the lipids on TLC plate. MGD1 enzyme reaction can be initiated in GUVs as described for LgtC and then run on TLC, similarly to what is performed in this Chapter.

A UDP-Glo™ assay could also be performed by substituting micelles with GUVs. While it may be possible to detect a signal from UDP formation, the obtained signal would probably not be reproducible. This is due to the large range of size of the GUVs synthesized (Figure 76b) and also the inner leaflet of the membrane is inaccessible for MGD1. Thus there is no control on the amount of DAG that is accessible for enzyme reaction and the signal would vary from assay to assay. Therefore, kinetic parameters cannot be obtained from GUVs, using the UDP-Glo™ assay. However, the assay can still be used as an indicator of MGD1 activity in GUVs from simple detection of UDP.

3. Fluorescently labelled MGD1 binding to GUVs

In order to mimic conditions required for MGD1 activity, DAG+PG GUVs were formed, which contained: DAG 7.5 mol-%, PG 2.5 mol-%, DOPC 89.7 mol-% and Atto-647 DOPE 0.3 mol-%. Fluorescently labelled MGD1 (MGD1-AF488) was incubated with GUVs at 200 nM and, under fluorescence microscopy, was shown to bind strongly to GUVs containing DAG and PG (Figure 77).

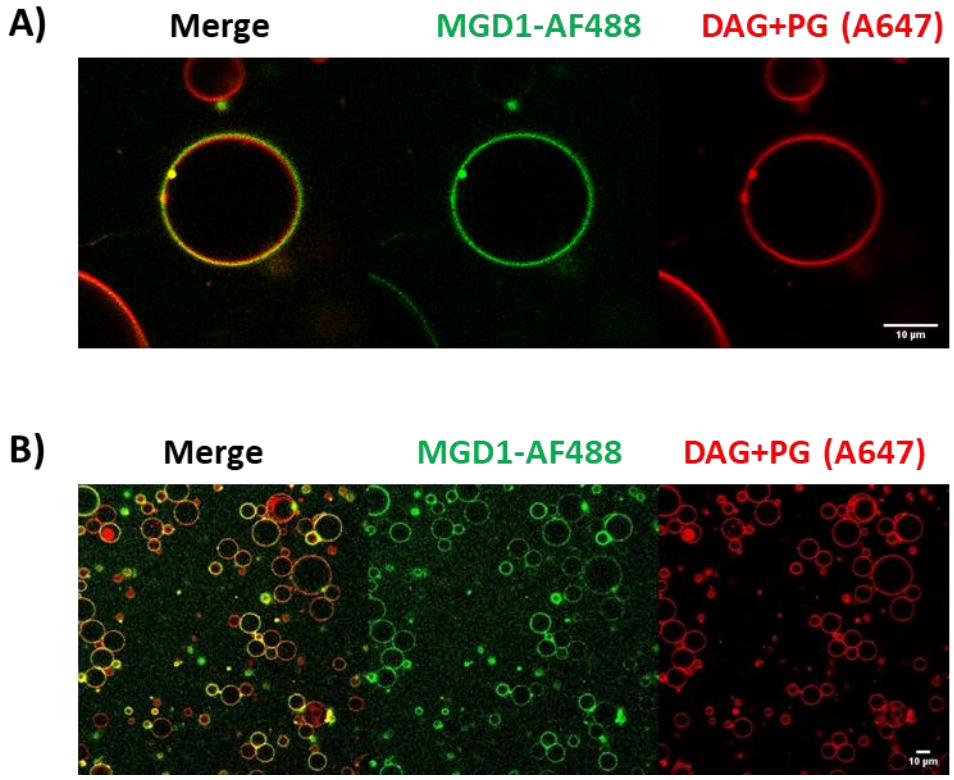


Figure 77: *DAG+PG GUVs incubated with fluorescently labelled (AF488) MGD1. A) Focus on a single GUV. B) Zoomed out view on multiple GUVs. GUV composition: DOPC (89.7%), DAG (7.5%), PG (2.5%), Atto-647 DOPE (0.3%). GUVs were incubated in 200nM fluorescently labelled MGD1. 10 µm scale bar.*

Furthermore, we investigated the binding preference of MGD1-AF488 for the individual lipids., We formed different compositions of GUVs that only contained one of the essential MGD1 lipids and DOPC. The GUVs formed were: PG+DOPC, MGDG+DOPC or DOPC (Figure 77). DOPC GUVs were previously formed in the lab and contained cholesterol, which can facilitate GUV formation and increase the average size of GUVs (Karal et al., 2022). Cholesterol was not included in the other GUVs in order to maximise the amount of DOPC to facilitate the formation of bilayers. This was necessary as DAG and MGDG are hexagonal II phase and discouraged the formation of bilayers.

Very weak fluorescence intensity correlates to weak binding. We observed weak binding to PG+DOPC and MGDG+DOPC GUVs (Figure 78A-B). There is no binding to DOPC GUVs (Figure 77C). Results from this indicated that the strong binding of MGD1-AF488 to the DAG+PG GUVs

was solely due to the presence of DAG. This was further proven when DAG GUVs consisting of: 92.2 mol-% DOPC, 7.5 mol-% DAG, 0.3 mol-% Atto-647 DOPE, showed similar binding to DAG+PG GUVs.

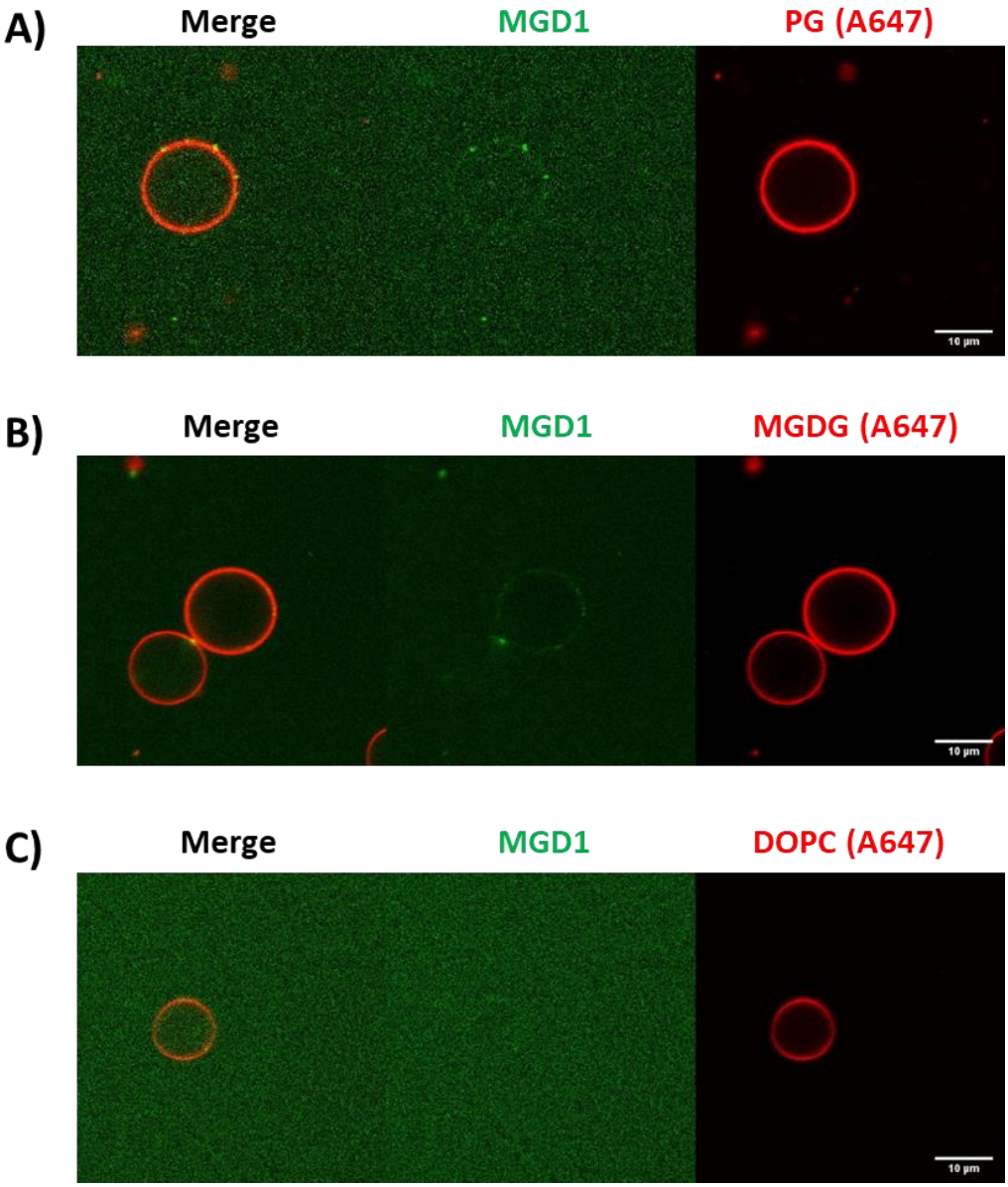


Figure 78: *Assay of MGD1-AF488 binding to PG, MGDG and DOPC GUVs. A) PG GUVs. Lipid composition: DOPC (97.2%), PG (2.5%), Atto-647 DOPE (0.3%). B) DOPC (92.2%), MGDG (7.5%), Atto-647 DOPE (0.3%). C) DOPC (69.7%), Chol (30%), Atto-647 DOPE (0.3%). GUVs were incubated in 200 nM Fluorescently labelled MGD1. 10µm scale bar.*

Based on previous binding studies of MGD1 to Langmuir monolayers, we expected MGD1-AF488 to also bind MGDG and PG similarly to DAG (Sarkis et al., 2014; Nitenberg et al., 2020). However, this was not the case in our GUV experiments as we saw very weak binding to some

MGDG and PG GUVs, but not to others. Certainly, the two experiments are hardly comparable due to the different methods used for GUVs and Langmuir monolayers. However, it hints at a lower binding for MGDG and PG. Indeed, it has been shown that the rate of bind to pure DAG monolayer is 10 fold more rapid than for MGDG or PG (Rocha et al., 2016b). From qualitative analysis of fluorescence intensity, MGD1-AF488 seems to have favourable binding to DAG but not to MGDG. It was not comparable for PG as we tested 2.5 mol-% PG GUVs, thus 7.5 mol-% PG GUVs should also be tested to compare binding with MGDG and DAG.

In the scenario where MGD1-AF488 binding to PG was weak or compromised, fluorescence labelling could be the cause. Several lysine residues have been identified as potential PG interacting residues in molecular simulations (Makshakova et al., 2020). Fluorescent labelling occurs by NHS/amine conjugation via lysine residues. It is possible that the fluorescence labelling has conjugated to lysine residues involved in PG binding (Figure 79). For example, it could prevent crucial interactions between lysine and PG, potentially K269 and K270 from the Nα4 helix. DAG binding remains strong because other residues (not lysine) that interact with DAG were not affected, such as T151, W182, D184, H185, W188, P189, F190, R195, S222 and F230 (Makshakova et al., 2020).

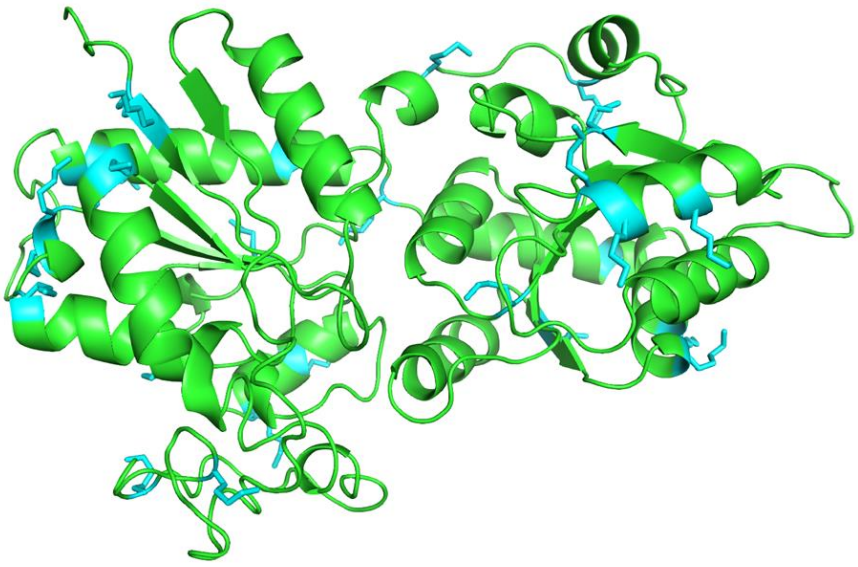


Figure 79: *MGD1 structure highlighting all lysine residues (cyan).*

Surprisingly, in this experiment was the very low binding for MGDG GUVs. The chloroplast inner envelope membrane (IEM) is composed of around 50 mol-% MGDG, 30 mol-% DGDG and the rest (<10%) consists of PG, PC, SQDG, PI and DAG. Langmuir monolayer studies find

strong interaction towards MGDG monolayers (Sarkis et al., 2014; Rocha et al., 2016), and this is not evident in GUVs. A potential reason for this is MGDG at 7.5 mol-% is not high enough in GUVs. MGD1 binding may improve with increased MGDG. Thus, 50 mol-% MGDG GUV should be tested. However, this was difficult to test as the quality and size of GUVs deteriorated as we increased the MGDG mol-% content (see appendix, Figure S18). This is due to increased hexagonal II (HII) (MGDG) to lamellar (Lα) (DOPC) lipid ratio, and this not ideal conditions for the formation of GUVs.

4. Adapting MGD1 activity on GUVs to UDP-Glo™ assay

There was an interest to adapt the MGD1 test using the UDP-Glo™ assay to use GUVs instead of mixed micelles. In glycosylation of GUVs by GTs, the enzyme activity is initiated by adding enzyme and sugar donor to pre-formed GUVs, this is demonstrated in the work presented on LgtC (see manuscript in preparation). In MGD1 activity assays, micelles were formed by sonication of a solution containing the lipids and MGD1, and subsequently enzyme reaction was initiated by the addition of UDP-Gal.

As the first step in adapting the MGD1 assay to the use of GUVs, we modified the traditional method of the MGD1 assay (Figure 80). The new method formed the micelles without the presence of MGD1. Then enzyme reaction was initiated by adding MGD1 and then UDP-Gal. Activity characteristics of MGD1 with different concentrations of PG and PA were analysed using the modified method.

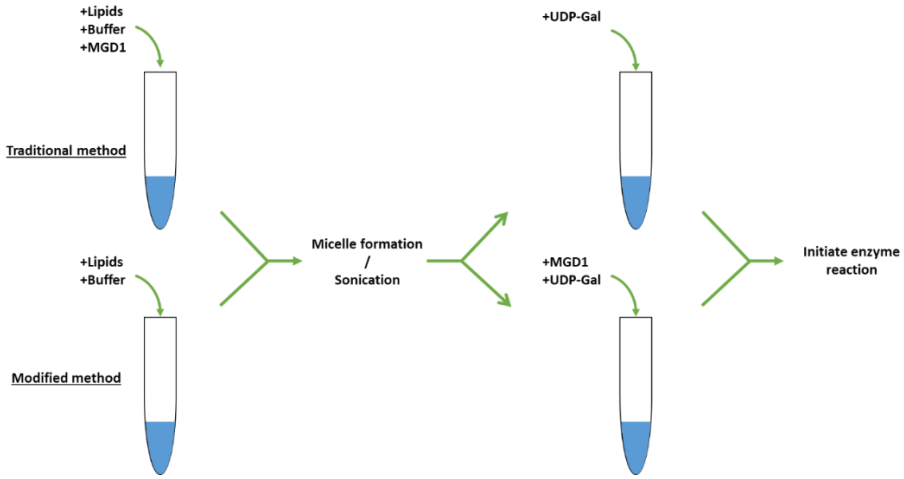


Figure 80: *Modification to the traditional activity assay. MGD1 is added after formation of micelles. The enzyme reaction is initiated by the addition of UDP-Gal. Otherwise; all steps are the same as the traditional method.*

a) Altered characteristics for PG and PA activators with the different methods

We performed the traditional and modified MGD1 assay and compared their activity. PG or PA activators at 1.5 mol-% was used, which is the standard concentration established by Rocha et al. (2013). From comparison, it was observed that MGD1 activity with PG is much higher than with PA in the modified method (Figure 81B). This is reversed in the traditional method, where PA activator displays higher activity (~2 fold more) (Figure 81A) (Rocha et al., 2013).

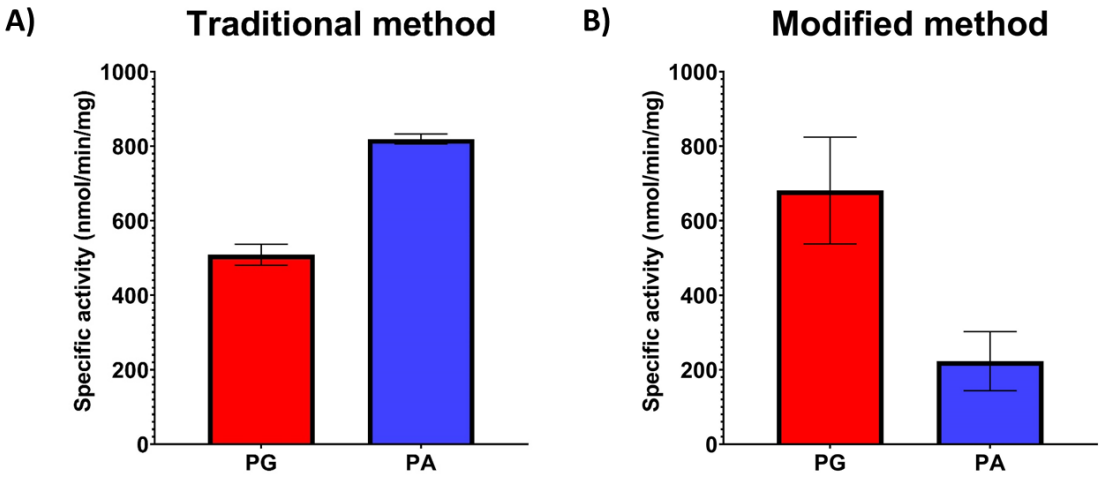


Figure 81: *Comparison of MGD1 activity using the traditional (A) and modified method (B). MGD1 reaction is as described in Chapter 2. Variable MGD1 reaction conditions: 5 ng (0.14*

ng/µl) MGD1, 10-minute reaction time. Activity was tested using PG and PA activator at 1.5 mol-%.

b) No synergistic effect in the modified protocol

In the presence of both PG and PA, there is an observed synergistic effect on MGD1 activity rate (Dubots et al., 2010). Following the discovery of this change in PG and PA characteristics, we wanted to know if the modified method altered the synergistic effect. We performed the assay with micelles containing PG and PA, at 0.15 or 1.5 mol-% (Figure 82). Activity assays showed an increased rate going from 0.15% PG/PA to 1.5% PG/PA. This is consistent with findings from Dubots et al. (2010). The key observation in this experiment indicates that, in the presence of equal amounts of PG and PA, there is no additive effect. Also, in the presence of both activators, MGD1 seems to favour PA activation.

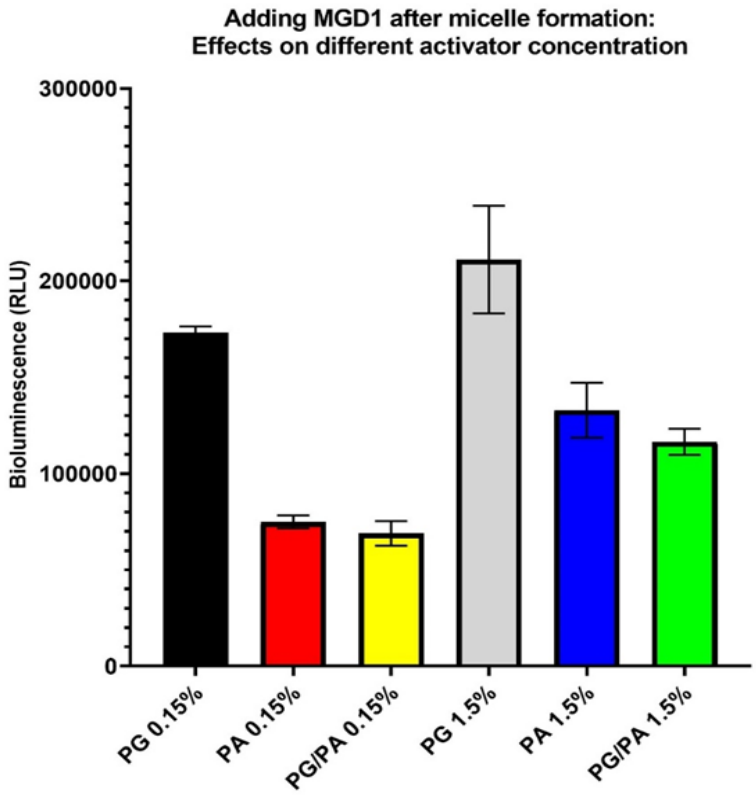


Figure 82: *Activity of wild-type MGD1 using different PG/PA concentrations (mol-%). Assay used the modified protocol. Activity represented as bioluminescence (RLU). Conditions: 15-minute reaction, 0.14 ng/µl MGD1 (5 ng/reaction tube), 50 µM UDP-Gal.*

In the paper by Dubots et al. (2010) a sigmoidal curve (initial rate vs [UDP-Gal]) was observed at 0.15 mol-% PA, which is characteristic of allosteric regulation. This changed at 1.5 mol-% PA, where rate curve becomes hyperbolic. From that study, it was expected that a synergistic effect would be present at 0.15 mol-% but not in 1.5 mol-%, however, this was not the case in our results. It is unknown exactly why there is no additive effect. A possible reason is that PA only displays this additive effect in very low concentrations compared to PG. Indeed, previous work displayed synergistic effect at a 0.08 mol-% / 0.5 mol-% (PA/PG) ratio (Dubots et al., 2010).

Structural differences to the micelles, dictated by the method used, could be a factor that could explain the differences observed in MGD1 activity. When calculating the components of the micelle, only DAG, PG/PA and CHAPS are considered. However, as MGD1 binds strongly to the lipid components of the micelle, MGD1 should also be considered in the final structure of the micelle.

Adding MGD1 during micelle formation could lead to differently structured micelles compared to the modified method. For example, MGD1 becomes more buried in the structure of the micelle and the more hydrophobic regions (such as the LOOP region) can be more involved in the micelle core. The modified method would have MGD1 that is less incorporated in the micelle structure. In the end, the change in protocol could result in MGD1 binding micelles differently and this can influence its activity greatly.

c) Future work

This work has revealed a different behaviour for MGD1 resulting from one change in the protocol. It is worth investigating the change in activity profile because the way MGD1 binds to the micelles might be different in the two methods. It was interesting to note that MGD1 seems to favour PA activity when both PG and PA are present in the same quantity, this had never been tested. It would be interesting to assay a range of PA mol-%, while maintaining a constant PG mol-%. This should be tested with both protocols and could reveal whether PA only acts as an allosteric activator at very low quantities compared to PG. Another change caused by the protocol is that MGD1 is not subjected to sonication. While unlikely, it is possible

that sonication disturbs the folding slightly. The effect of sonication on MGD1 should also be tested.

VIII. Testing the capacity of MGD1 and LgtC to transfer an azido-galactose

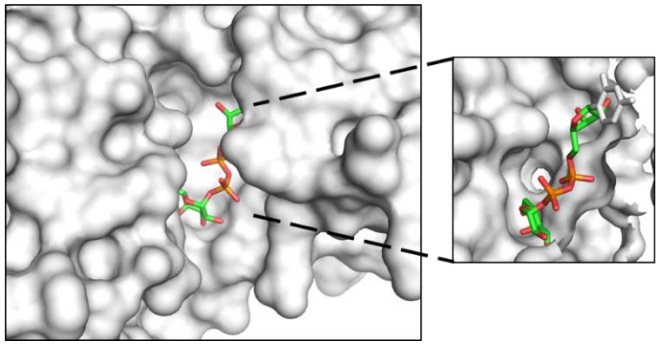
The use of functionalized sugar donors such as azido sugars can give additional opportunities through click chemistry, further expanding the variety of uses of GTs in synthetic glycobiology. For this study, we assayed the ability of LgtC and MGD1 to use UDP-6-azido-galactose (UDP-GalAz) as a sugar donor.

Activity tests showed that LgtC could not use the UDP-GalAz as a sugar donor (Figure 83A). While the mode of action of LgtC, a retaining GT, is not yet fully understood, the crystal structure of LgtC with UDP-2-deoxy-2-fluoro-galactose (UPF) shows a more closed and tightly bound conformation around the UPF. Thus, there is likely no space for the azido group in the active site (Figure 83C). In contrast, we observed that MGD1 could use the modified nucleotide sugar but with lower efficiency (~16%) (Figure 83A). Furthermore, looking at the X-ray structure of MGD1, we observed that the active site region for the sugar donor is more open to the solvent and contains pockets where the azido group may be able to fit (Figure 83B).

A.

MGD1 and LgtC activity assay: UDP-Gal vs UDP-GalAz		
	Bioluminescence (RLU x 10^{-3})	
	UDP-Gal	UDP-GalAz
LgtC	1219 ± 58	2 ± 1
MGD1	407 ± 23	64 ± 3

B.



C.

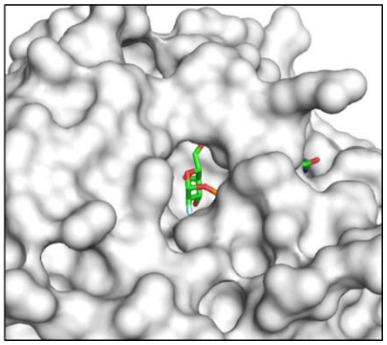


Figure 83: *Activity assay testing the capacity of MGD1 and LgtC to use UDP-GalAz. A) Assay comparing MGD1 and LgtC activity with UDP-Gal and UDP-GalAz. Activity expressed as RLU. LgtC reaction: 62.5 mM Lactose, 0.5 mM UDP-Gal, 0.4 ng/µl LgtC, 10-minute reaction. MGD1 reaction: 6.3 mol-% DAG, 1.5 mol-% PG and 0.7 mM UDP-Gal, 0.14 ng/µl MGD1, 30-minute reaction time. B) MGD1 structure with UDP-Gal modelled in the active site and a close up into an open pocket where the azide group can be accommodated Model provided by Olga Makshakova. C) LgtC structure in complex with donor analogue, UDP-2-deoxy-2-fluoro-galactose (PDB: 1GA8).*

In order to expand the variety of uses for LgtC, it must be mutated to accept UDP-GalAz. We predict that the Q187 side chain may be responsible for blocking the azido group from looking at the crystal structure. Therefore, a mutation of Q187 to alanine may be able to open space in the active site for an azide group and allow the transferase reaction to occur.

General Conclusions and Perspectives

In the field of synthetic glycobiology, glycosyltransferases (GTs) have emerged as powerful tools for building or editing glycan components. Indeed, the most widely used application of GTs are the development of therapeutic drugs. However, there is a need to explore the application of GTs to decorate membrane models. Synthetic membranes such as giant unilamellar vesicles (GUVs) and supported lipid bilayers (SLBs) have arisen as excellent models to uncover the cellular events that occur from the interaction of carbohydrate binding proteins to components of the glycocalyx. In this context, GTs are attractive tools to build or modify components of the glycocalyx. They also offer the possibility to add functionalised sugars to serve for biorthogonal chemistry. Within the scope of the synBIOcarb consortium, there is an interest to explore the potential of GTs as synthetic biology tools and their application to decorate membrane surfaces. Thus, the work presented in this thesis consists of two main parts:

1. Engineering of a model glycosyltransferase
2. Application of GTs to glycosylate synthetic membrane systems

Part 1

The first part aimed to investigate the catalytic mechanism of the model protein, MGD1 from *Arabidopsis thaliana*, a GT working on the inner chloroplast membrane. In the context of decorating membrane surfaces, it was of interest to investigate the mechanism by which a GT can add the first sugar on a lipid. This part of the project focused on two aspects of MGD1:

- Sugar donor specificity – altering from UDP-Gal to UDP-Glc
- Mechanism of activation of MGD1 by anionic lipids

The first objective aimed to change the sugar donor specificity of MGD1 from UDP-Gal to UDP-Glc. This was performed by mutational studies. For the project we used a structure- and sequence-guided approach for designing MGD1 mutants with GlcT activity. Initial MGD1 mutants were based on structural and sequence analysis of GT1 homologs of MGD1. Three mutants were produced based on the GT1 family (AQE, AEE and AEQ), however, they did not confer GlcT activity. There was a library of other mutants to try, however, the project based

on GT1 proteins was halted as we found a more promising strategy based on GT28 enzymes. In the GT28 family, RcMGlcD and CaMGD1, were promising candidates for our design of MGD1 mutants. RcMGlcD and CaMGD1, were produced and characterized in order to ascertain their true donor specificities. Assays proved that RcMGlcD is a GlcT, and CaMGD1 a GalT. As a result, MGD1 mutants based on RcMGlcD sequence were produced. In the end, a two-point mutation was required to confer some GlcT activity to MGD1. The most promising mutant was the GGL mutant which demonstrated the highest GlcT activity. Investigating the effect of the two mutations by molecular modelling, we found that the Pro433 ring was not only causing a steric clash but was also preventing an important hydrogen bonding between the glucose O4 and the protein backbone. An additional aliphatic residue in an adjacent position was also found to be important for GlcT activity by forming side chain hydrophobic interactions, which has the effect of stabilizing an α-helix. We performed sequence analysis of characterized GT28 proteins and hypothesized that the GalT activity is conferred by a 'GPG' motif, while GlcT activity is a 'GGX' motif with an aliphatic residue in the third position. This finding can be used for the prediction of activities of uncharacterized GT28 proteins.

The aim of the project was to alter the donor specificity of MGD1 from UDP-Gal to UDP-Glc. This goal was partially met by the GGL mutant, however, much work can still be done to improve the activity and selectivity of the GGL mutant. Possible strategies include using molecular simulations to guide further mutants by predicting binding interactions of UDP-Glc. This will help identify "hot spots", which can then be explored through alanine scanning and/or saturation mutagenesis. For expansion of synthetic biology tools, the work done, while only focused on galactose and glucose specificity, can still be used as guides in engineering other GTs with altered specificity.

Outside the scope of synthetic glycobiology, the GGL mutant and the information discovered regarding specificity can be utilized to study plant biology and photosynthetic membrane synthesis. Chloroplast membranes are primarily made of galactolipids (MGDG) synthesised by MGD1 and this lipid composition is preserved throughout evolution from bacteria to plants. It would be interesting to introduce a MGD1 mutant such as GGL into a plant and observe the effect on photosynthetic membrane development.

A second objective of the PhD project was to investigate another aspect of the MGD1 catalytic mechanism, the activation of MGD1 by phosphatidylglycerol (PG) and phosphatidic acid (PA).

While an activation mechanism for PG has been proposed for MGD1, involving a PG/His catalytic dyad, the binding site and allosteric activation by PA was still not understood. In this work, we attempted to gain more information on the binding site of the allosteric activator PA. We have probed multiple residues that target three regions of the protein as possible binding sites for PA and/or PG. The first region is the catalytic site, located in the cleft between the N- and C- domains of MGD1, and is the region where PG is expected to bind to form the PG/His catalytic dyad. The second site is the Nα4 helix, located in the N-domain of MGD1, of which there was strong prediction from computational work. The third site probed for PA binding was the LOOP region. The residues targeted for mutations were primarily basic amino acids (Histidine, Lysine, or Arginine) likely to form strong interactions with the negatively charged phosphate group of PA. Residues chosen were guided by computational work performed by Olga Makshakova and by sequence comparison to GT28 proteins CaMGD, RcMGlcD and GT1 homolog MgdA. We produced MGD1 mutants R156A, R195A, H251A, H256A, R260A, K265A, H289A, and H256A+H289A. However, the majority of these mutants did not have significant effect on activity with either PA or PG. The mutants H251A and H256A showed significantly decreased activity, however, this was mostly linked to PG activity and not PA activity. While H251A demonstrated significant loss of PA activity, 30% PA activity remained, thus indicating it is not a key residue of PA binding site. Thus, the binding site for PA remains unsolved. More mutants targeting the Nα4 helix, LOOP region, and catalytic site have to be assayed to gain more information.

Part 2

In parallel to the engineering work performed on MGD1, we also explored the application of glycosyltransferases to glycosylate synthetic membrane systems.

Recombinant LgtC was produced, and activity was assayed in micelles and synthetic membranes (GUVs and SLBs) that contained the LacCer acceptor. LgtC activity was demonstrated in GUVs by StxB binding. This showed for the first time that LgtC can glycosylate a membrane-bound LacCer. StxB binding on treated SLBs further evidenced this. Also discovered by QCM-D was the transient binding of LgtC to the membrane, which only occurs during catalysis. Finally, catalytic efficiency was estimated in micelles (735 $M^{-1}.s^{-1}$) and SLBs

(24 $M^{-1}.s^{-1}$). The low efficiency compared to lactose seems to be correlated with reduced access to the acceptor on membranes. In light of the weak membrane binding and low catalytic efficiency, there is now a question of whether the full-length LgtC will behave similarly. The recombinant LgtC used in this project was truncated by 25 amino acids in the C-terminus, to improve solubility and yield. However, the truncated region is predicted to associate with the membrane. Therefore, it will be interesting to test the membrane binding and catalytic efficiency of a full-length LgtC.

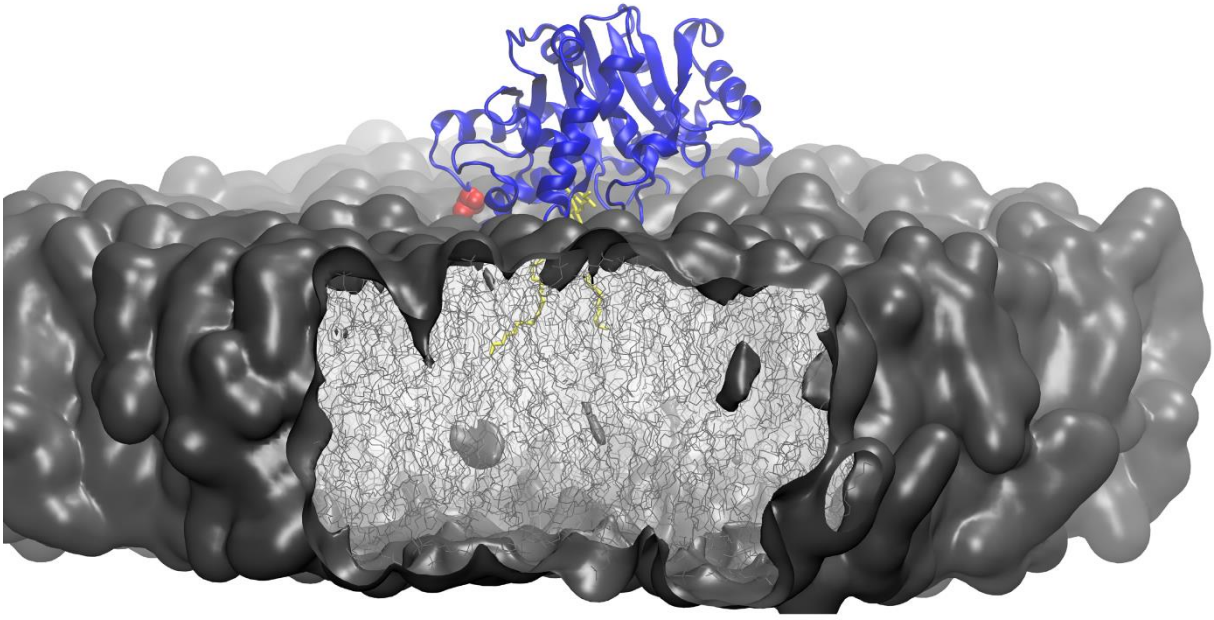


Figure 84: *Lipid bilayer with LgtC bound to LacCer. Shown are LgtC (Blue), its C-terminus (Red) and LacCer (Yellow). Bilayer composed of 5 mol-% LacCer and 95 mol-% DOPC. Modelling work by Olga Makshakova.*

On the other hand, we undertook preliminary studies for wild-type MGD1 working on lipid membranes. Established in this work are ideal lipid compositions for the formation of GUVs containing the lipids MGDG, PG, and DAG. For reporting of MGDG formation, three lectins specific for βGal were tested but could not bind MGDG. Also demonstrated in this work is a fluorescently-labelled MGD1 (MGD1-AF488), which showed strong binding to DAG but not to the other lipids MGDG and PG. The selective binding of MGD1-AF488 means it has a potential application to be further explored. Also undertaken was the initial work to adapt a UDP-Glo™ assay that uses GUVs. This involved the modification of the classical MGD1 activity assay. The addition of MGD1 after the formation of micelles resulted in a change in the activity profile of MGD1 with its activators (PG and PA). The reason for this change is unknown and has opened

questions about the way MGD1 binds to various membranes. Due to a lack of time, we could not complete the experiments with the MGD1 on a GUV model, but the work has laid the foundations for future experiments to continue the research.

Finally, we evaluated the application of LgtC and MGD1 as tools in chemoenzymatic synthesis by testing their capacity to use UDP-GalAz. Activity assays showed that LgtC could not use UDP-GalAz. However, MGD1 can use UDP-GalAz with 16% efficiency. In addition, previous studies have shown that LgtC has high substrate promiscuity for the acceptor. Therefore, augmenting the sugar donor promiscuity further expands the potential of LgtC as a synthetic glycobiology tool.

REFERENCES

Adlercreutz, D., Weadge, J.T., Petersen, B.O., Duus, J.O., Dovichi, N.J. and Palcic, M.M. 2010. Enzymatic synthesis of Gb3 and iGb3 ceramides. *Carbohydrate Research*. 345(10), pp.1384–1388.

Aerts, J.M., Groener, J.E., Kuiper, S., Donker-Koopman, W.E., Strijland, A., Ottenhoff, R., Van Roomen, C., Mirzaian, M., Wijburg, F.A., Linthorst, G.E., Vedder, A.C., Rombach, S.M., Cox-Brinkman, J., Somerharju, P., Boot, R.G., Hollak, C.E., Brady, R.O. and Poorthuis, B.J. 2008. Elevated globotriaosylsphingosine is a hallmark of Fabry disease. *Proceedings of the National Academy of Sciences of the United States of America.* 105(8), pp.2812-2817.

Akere, A., Chen, S.H., Liu, X., Chen, Y., Dantu, S.C., Pandini, A., Bhowmik, D. and Haider, S. 2020. Structure-based enzyme engineering improves donor-substrate recognition of Arabidopsis thaliana glycosyltransferases. *Biochemical Journal*. 477(15), pp.2791–2805.

Albesa-Jové, D., Giganti, D., Jackson, M., Alzari, P.M. and Guerin, M.E. 2014. Structure–function relationships of membrane-associated GT-B glycosyltransferases. *Glycobiology*. 24(2), pp.108-124

Alves, C.A., Pedroso, M.M., de Moraes, M.C., Souza, D.H.F., Cass, Q.B. and Faria, R.C. 2011. Real-time investigation of mannosyltransferase function of a Xylella fastidiosa recombinant GumH protein using QCM-D. *Biochemical and Biophysical Research Communications*. 408(4), pp.571–575.

Bastien, O., Botella, C., Chevalier, F., Block, M.A., Jouhet, J., Breton, C., Girard-Egrot, A. and Maréchal, E. 2016. New Insights on Thylakoid Biogenesis in Plant Cells. *International Review of Cell and Molecular Biology*. 323, pp.1–30.

Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M.E., Alvarez, R., Bryan, M.C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D.J., Skehel, J.J., Van Die, I., Burton, D.R., Wilson, I.A., Cummings, R., Bovin, N., Wong, C.H. and Paulson, J.C. 2004. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proceedings*

of the National Academy of Sciences of the United States of America. 101(49), pp.17033-17038

Block, M.A., Douce, R., Joyard, J. and Rolland, N. 2007. Chloroplast envelope membranes: A dynamic interface between plastids and the cytosol. *Photosynthesis Research*. 92(2), pp.225–244.

Bohl, H., Bai, L, and Li, H. 2021. Recent progress in structural studies on the GT-C superfamily of protein glycosyltransferases. *Subcellular Biochemistry*. 96, pp.259-271

Bolam, D.N., Roberts, S., Proctor, M.R., Turkenburg, J.P., Dodson, E.J., Martinez-Fleites, C., Yang, M., Davis, B.G., Davies, G.J. and Gilbert, H.J. 2007. The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proceedings of the National Academy of Sciences of the United States of America*. 104(13), pp.5336-5341

Botté, C., Jeanneau, C., Snajdrova, L., Bestien, O., Imberty, A., Breton, C. and Maréchal, E. 2005. Molecular modeling and site-directed mutagenesis of plant chloroplast monogalactosyldiacylglycerol synthase reveal critical residues for activity. *Journal of Biological Chemistry*. 280(41), pp.34691–34701.

Breton, C., Fournel-Gigleux, S. and Palcic, M.M. 2012. Recent structures, evolution and mechanisms of glycosyltransferases. *Current Opinion in Structural Biology*. 22(5), pp.540–549.

Breton, C., Mucha, J. and Jeanneau, C. 2001. Structural and functional features of glycosyltransferases. *Biochimie*. 83(8), pp.713-718.

Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J. and Imberty, A. 2006. Structures and mechanisms of glycosyltransferases. *Glycobiology*. 16(2), pp.29R-37R.

Castillo-Acosta, V.M., Balzarini, J. and González-Pacanowska, D. 2017. Surface Glycans: A Therapeutic Opportunity for Kinetoplastid Diseases. *Trends in Parasitology*. 33(10), pp.775–787.

Chan, P.H.W., Cheung, A.H., Okon, M., Chen, H.M., Withers, S.G. and McIntosh, L.P. 2013. Investigating the structural dynamics of α-1,4-galactosyltransferase C from Neisseria meningitidis by nuclear magnetic resonance spectroscopy. *Biochemistry*. 52(2), pp.320–332.

Chan, P.H.W., Weissbach, S., Okon, M., Withers, S.G. and McIntosh, L.P. 2012. Nuclear magnetic resonance spectral assignments of α-1,4- galactosyltransferase lgtc from neisseria meningitidis: Substrate binding and multiple conformational states. *Biochemistry*. 51(41), pp.8278–8292.

Chang, A., Singh, S., Helmich, K.E., Goff, R.D., Bingman, C.A., Thorson, J.S. and Phillips, G.N. 2011. Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regiospecificity. *Proceedings of the National Academy of Sciences of the United States of America.* 108(43), pp.17649–17654.

Chaubard, J.L., Krishnamurthy, C., Yi, W., Smith, D.F. and Hsieh-Wilson, L.C. 2012. Chemoenzymatic Probes for Detecting and Imaging Fucose-α(1-2)-galactose Glycan Biomarkers. *Journal of the American Chemical Society*. 134(10), pp.4489-4492

Chiu, C.P.C., Lairson, L.L., Gilbert, M., Wakarchuk, W.W., Withers, S.G. and Strynadka, N.C.J. 2007. Structural analysis of the α-2,3-sialyltransferase Cst-I from Campylobacter jejuni in apo and substrate-analogue bound forms. *Biochemistry*. 46(24), pp.7196–7204.

Coutinho, P.M., Deleury, E., Davies, G.J. and Henrissat, B. 2003. An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology*. 328(2), pp.307–317.

Critcher, M., O'Leary, T. and Huang, M.L. 2021. Glycoengineering: Scratching the surface. *Biochemical Journal*. 478(4), pp.703–719.

Cruz, E., Sifniotis, V., Sumer-Bayraktar, Z., Reslan, M., Wilkinson-White, L., Cordwell, S. and Kayser, V. 2021. Glycan profile analysis of engineered trastuzumab with rationally added glycosylation sequons presents significantly increased glycan complexity. *Pharmaceutics*. 13(11), 1747.

Demé, B., Cataye, C., Block, M.A., Maréchal, E. and Jouhet, J. 2014. Contribution of galactoglycerolipids to the 3-dimensional architecture of thylakoids. *FASEB Journal*. 28(8), pp.3373–3383.

Dicker, M. and Strasser, R. 2015. Using glyco-engineering to produce therapeutic proteins. *Expert opinion on biological therapy*. 15(10), p.1501-1516

Diniz, F., Coelho, P., Duarte, H.O., Sarmento, B., Reis, C.A. and Gomes, J. 2022. Glycans as Targets for Drug Delivery in Cancer. *Cancers*. 14(4). 991.

Dörmann, P. and Benning, C. 2002. Galactolipids rule in seed plants. *Trends in Plant Science*. 7(3), pp.112–118.

Du, T., Buenbrazo, N., Kell, L., Rahmani, S., Sim, L., Withers, S.G., DeFrees, S. and Wakarchuk, W. 2019. A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans. *Cell Chemical Biology*. 26(2), pp.203-212.

Dubots, E., Audry, M., Yamaryo, Y., Bastien, O., Ohta, H., Breton, C., Maréchal, E. and Block, M.A. 2010. Activation of the Chloroplast Monogalactosyldiacylglycerol Synthase MGD1 by Phosphatidic Acid and Phosphatidylglycerol. *The Journal of Biological Chemistry*. 285(9), pp.6003-6011

Dubots, E., Botté, C., Boudière, L., Yamaryo-Botté, Y., Jouhet, J., Maréchal, E. and Block, M.A. 2012. Role of phosphatidic acid in plant galactolipid synthesis. *Biochimie*. 94(1), pp.86–93.

Bastien, O., Botella, C., Chevalier, F., Block, M.A., Jouhet, J., Breton, C., Girard-Egrot, A. and Maréchal, E. 2016. New Insights on Thylakoid Biogenesis in Plant Cells*. International Review of Cell and Molecular Biology*. 323, pp.1–30.

Erwin, A.L., Allen, S., Ho, D.K., Bonthius, P.J., Jarisch, J., Nelson, K.L., Tsao, D.L., Unrath, W.C.T., Watson, M.E., Gibson, B.W., Apicella, M.A. and Smith, A.L. 2006. Role of lgtC in resistance of nontypeable Haemophilus influenzae strain R2866 to human serum. *Infection and Immunity*. 74(11), pp.6226–6235.

Falguières, T., Maak, M., Von Weyhern, C., Sarr, M., Sastre, X., Poupon, M.F., Robine, S., Johannes, L. and Janssen, K.P. 2008. Human colorectal tumors and metastases express Gb3 and can be targeted by an intestinal pathogen-based delivery tool. *Molecular Cancer Therapeutics*. 7(8), pp.2498–2508.

Formosa-Dague, C., Castelain, M., Martin-Yken, H., Dunker, K., Dague, E. and Sletmoen, M. 2018. The Role of Glycans in Bacterial Adhesion to Mucosal Surfaces: How Can Single-Molecule Techniques Advance Our Understanding? *Microorganisms*. 6(2), 39.

Fujii, S., Kobayashi, K., Nakamura, Y. and Wada, H. 2014. Inducible Knockdown of Monogalactosyldiacylglycerol Synthase1 Reveals Roles of Galactolipids in Organelle Differentiation in Arabidopsis Cotyledons. *Plant Physiology*. 166(3), pp.1436–1449.

Fujii, S., Wada, H. and Kobayashi, K. 2019. Role of Galactolipids in Plastid Differentiation Before and After Light Exposure. Plants. 8(10), 357.

Fullington, J.G. and Hendrickson, H.S. 1966. Phospholipid-metal complexes. Interaction of triphosphoinositide- and phosphatidylserine-metal complexes with ethylenediamine, polyaminoacids, and protein. *Journal of Biological Chemistry*. 241(17), pp.4098-4100

Gallegos, K.M., Conrady, D.G., Karve, S.S., Gunasekera, T.S., Herr, A.B. and Weiss, A.A. 2012. Shiga Toxin Binding to Glycolipids and Glycans. *PLoS ONE*. 7(2). e30368.

Geyer, P.E., Maak, M., Nitsche, U., Perl, M., Novotny, A., Slotta-Huspenina, J., Dransart, E., Holtorf, A., Johannes, L. and Janssen, K.P. 2016. Gastric Adenocarcinomas Express the Glycosphingolipid Gb3/CD77: Targeting of Gastric Cancer Cells with Shiga Toxin B-Subunit. *Molecular cancer therapeutics*. 15(5), pp.1008–1017.

Goel, B., Tripathi, N., Mukherjee, D. and Jain, S.K. 2021. Glycorandomization: A promising diversification strategy for the drug development. *European Journal of Medicinal Chemistry*. 213, 113156.

Gómez, H., Polyak, I., Thiel, W., Lluch, J.M. and Masgrau, L. 2012. Retaining glycosyltransferase mechanism studied by QM/MM methods: Lipopolysaccharyl-α-1,4-

galactosyltransferase C transfers α-galactose via an oxocarbenium ion-like transition state. *Journal of the American Chemical Society*. 134(10), pp.4743–4752.

Griffin, M.E. and Hsieh-Wilson, L.C. 2016. Glycan Engineering for Cell and Developmental Biology. *Cell Chemical Biology*. 23(1), pp.108–121.

Griffith, B.R., Langenhan, J.M. and Thorson, J.S. 2005. 'Sweetening' natural products via glycorandomization. *Current Opinion in Biotechnology*. 16(6), pp.622-630

Guerin, M.E., Kordulakova, J., Schaeffer, F., Svetlikova, Z., Buschiazzo, A., Giganti, D., Gicquel, B., Mikusova, K., Jackson, M. and Alzari, P.M. 2007. Molecular Recognition and Interfacial Catalysis by the Essential Phosphatidylinositol Mannosyltransferase PimA from Mycobacteria. *Journal of Biological Chemistry*. 282(28), pp.20705–20714.

Gupta, S.K. and Shukla, P. 2018. Glycosylation control technologies for recombinant therapeutic proteins. *Applied Microbiology and Biotechnology*. 102(24), pp.10457–10468.

Ha, S., Walker, D., Shi, Y. and Walker, S. 2000. The 1.9 Å crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Science*. 9(6), pp.1045–1052.

Haga, Y. and Ueda, K. 2022. Glycosylation in cancer: its application as a biomarker and recent advances of analytical techniques. *Glycoconjugate Journal 2022 39:2*. 39(2), pp.303–313.

Haga, Y. and Ueda, K. 2022. Glycosylation in cancer: its application as a biomarker and recent advances of analytical techniques. *Glycoconjugate Journal*. 39(2), pp.303–313.

Hamilton, S.R., Davidson, R.C., Sethuraman, N., Nett, J.H., Jiang, Y., Rios, S., Bobrowicz, P., Stadheim, T.A., Li, H., Choi, B.K., Hopkins, D., Wischnewski, H., Roser, J., Mitchell, T., Strawbridge, R.R., Hoopes, J., Wildt, S. and Gerngross, T.U. 2006. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science*. 313(5792), pp.1441–1443.

Harrison, R.L. and Jarvis, D.L. 2006. Protein N-Glycosylation in the Baculovirus–Insect Cell Expression System and Engineering of Insect Cells to Produce "Mammalianized" Recombinant Glycoproteins. *Advances in Virus Research*. 68, pp.159–191.

He, X.M. and Liu, H.W. 2002. Formation of unusual sugars: mechanistic studies and biosynthetic applications. *Annual review of biochemistry*. 71, pp.701–754.

Henrissat, B., Sulzenbacher, G. and Bourne, Y. 2008. Glycosyltransferases, glycoside hydrolases: surprise, surprise! *Current Opinion in Structural Biology*, 18(5), pp527-533.

Hong, S., Grande, G., Yu, C., Chapla, D.G., Reigh, N., Yang, J.Y., Yang, Y., Izumori, K., Moremen, K.W., Xie, J. and Wu, P. 2020. HFUT1-Based Live-Cell Assay to Profile α1-2-Fucoside-Enhanced Influenza Virus A Infection. *ACS Chemical Biology*. 15(4), pp.819–823.

Hong, S., Shi, Y., Wu, N.C., Grande, G., Douthit, L., Wang, H., Zhou, W., Sharpless, K.B., Wilson, I.A., Xie, J. and Wu, P. 2019. Bacterial glycosyltransferase-mediated cell-surface chemoenzymatic glycan modification. *Nature Communications*. 10(1), 1799.

Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M. and Walker, S. 2003. Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proceedings of the National Academy of Sciences of the United States of America*. 100(3), pp.845–849.

Huang, M.L., Tota, E.M., Verespy, S. and Godula, K. 2018. Glycocalyx scaffolding to control cell surface glycan displays. *Current protocols in chemical biology*. 10(2), e40.

Hudak, J.E. and Bertozzi, C.R. 2014. Glycotherapy: New advances inspire a reemergence of glycans in medicine. *Chemistry and Biology*. 21(1), pp.16–37.

Imberty, A. and Varrot, A. 2008. Microbial recognition of human cell surface glycoconjugates. *Current Opinion in Structural Biology*. 18(5), pp.567–576.

Jarvis, P., Dörmann, P., Peto, C.A., Lutes, J., Benning, C. and Chory, J. 2000. Galactolipid deficiency and abnormal chloroplast development in the Arabidopsis MGD synthase 1

mutant. *Proceedings of the National Academy of Sciences of the United States of America*. 97(14), pp.8175–8179.

Jiang, H., López-Aguilar, A., Meng, L., Gao, Z., Liu, Y., Tian, X., Yu, G., Ovryn, B., Moremen, K.W. and Wu, P. 2018. Modulating Cell-Surface Receptor Signaling and Ion Channel Functions by In Situ Glycan Editing. *Angewandte Chemie*. 57(4), p.967-971

Jorasch, P., Wolter, F.P., Zähringer, U. and Heinz, E. 1998. A UDP glucosyltransferase from Bacillus subtilis successively transfers up to four glucose residues to 1,2-diacylglycerol: expression of ypfP in Escherichia coli and structural analysis of its reaction products. *Molecular Microbiology*. 29(2), pp.419–430.

Kailemia, M.J., Park, D. and Lebrilla, C.B. 2017. Glycans and Glycoproteins as Specific Biomarkers for Cancer. *Analytical and bioanalytical chemistry*. 409(2), pp.395-410.

Kallolimath, S., Castilho, A., Strasser, R., Grünwald-Gruber, C., Altmann, F., Strubl, S., Galuska, C.E., Zlatina, K., Galuska, S.P., Werner, S., Thiesler, H., Werneburg, S., Hildebrandt, H., Gerardy-Schahn, R. and Steinkellner, H. 2016. Engineering of complex protein sialylation in plants. *Proceedings of the National Academy of Sciences of the United States of America.* 113(34), pp.9498–9503.

Karal, M.A.S., Mokta, N.A., Levadny, V., Belaya, M., Ahmed, M., Ahamed, M.K. and Ahammed, S. 2022. Effects of cholesterol on the size distribution and bending modulus of lipid vesicles. *PLOS ONE*. 17(1), e0263119.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J.E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*. 10(6), pp.845–858.

Keys, T.G. and Aebi, M. 2017. Engineering protein glycosylation in prokaryotes. *Current Opinion in Systems Biology*. 5, pp.23–31.

Kightlinger, W., Duncker, K.E., Ramesh, A., Thames, A.H., Natarajan, A., Stark, J.C., Yang, A., Lin, L., Mrksich, M., DeLisa, M.P. and Jewett, M.C. 2019. A cell-free biosynthesis

platform for modular construction of protein glycosylation pathways. *Nature Communications*. 10(1), 5404.

Kightlinger, W., Warfel, K.F., Delisa, M.P. and Jewett, M.C. 2020. Synthetic Glycobiology: Parts, Systems, and Applications. *ACS Synthetic Biology*. 9(7), pp.1534–1562.

Kobayashi, K. 2016. Role of membrane glycerolipids in photosynthesis, thylakoid biogenesis and chloroplast development. *Journal of Plant Research*. 129(4), pp.565–580.

Kobayashi, K., Kondo, M., Fukuda, H., Nishimura, M. and Ohta, H. 2007. Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis, photosynthesis, and embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 104(43), pp.17216–17221.

Kobayashi, K., Narise, T., Sonoike, K., Hashimoto, H., Sato, N., Kondo, M., Nishimura, M., Sato, M., Toyooka, K., Sugimoto, K., Wada, H., Masuda, T. and Ohta, H. 2013. Role of galactolipid biosynthesis in coordinated development of photosynthetic complexes and thylakoid membranes during chloroplast biogenesis in Arabidopsis. *Plant Journal*. 73(2), pp.250–261.

Kovbasnjuk, O., Mourtazina, R., Baibakov, B., Wang, T., Elowsky, C., Choti, M.A., Kane, A. and Donowitz, M. 2005. The glycosphingolipid globotriaosylceramide in the metastatic transformation of colon cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 102(52), pp.19087–19093.

Kubo, A., Arai, Y., Nagashima, S. and Yoshikawa, T. 2004. Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. *Archives of Biochemistry and Biophysics*. 429(2), pp.198–203.

Lairson, L.L., Henrissat, B., Davies, G.J. and Withers, S.G. 2008. Glycosyl transferases: Structures, functions, and mechanisms. *Annual Review of Biochemistry*. 77, pp.521–555.

Lairson, L.L., Watts, A.G., Wakarchuk, W.W. and Withers, S.G. 2006. Using substrate engineering to harness enzymatic promiscuity and expand biological catalysis. *Nature Chemical Biology*. **2**(12), pp.724–728.

Lairson, L.L., Chiu, C.P.C., Ly, H.D., He, S., Wakarchuk, W.W., Strynadka, N.C.J. and Withers, S.G. 2004. Intermediate trapping on a mutant retaining α-galactosyltransferase identifies an unexpected aspartate residue. *Journal of Biological Chemistry*. 279(27), pp.28339–28344.

Laughlin, S.T., Baskin, J.M., Amacher, S.L. and Bertozzi, C.R. 2008. In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science*. 320(5876), pp.664–667.

Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P. and Walker, S. 2011. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature*, 469(7331), pp.564-567.

Lee, S.S., Hong, S.Y., Errey, J.C., Izumi, A., Davies, G.J. and Davis, B.G. 2011. Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. *Nature Chemical Biology***.** 7(9), pp.631–638.

Li, N., Zhang, W., Lin, L., Shah, S.N.A., Li, Y. and Lin, J.M. 2019. Nongenetically Encoded and Erasable Imaging Strategy for Receptor-Specific Glycans on Live Cells. *Analytical Chemistry*. 91(4), pp.2600–2604.

Li, Q., Wang, X., Ma, S., Zhang, Y. and Han, X. 2016. Electroformation of giant unilamellar vesicles in saline solution. *Colloids and Surfaces B: Biointerfaces*. 147, pp.368–375.

Li, Y., Teneberg, S., Thapa, P., Bendelac, A., Levery, S.B. and Zhou, D. 2008. Sensitive detection of isoglobo and globo series tetraglycosylceramides in human thymus by ion trap mass spectrometry. *Glycobiology*. 18(2), pp.158–165.

Li, Z., Han, K., Pak, J.E., Satkunarajah, M., Zhou, D. and Rini, J.M. 2017. Recognition of EGF-like domains by the Notch-modifying O-fucosyltransferase POFUT1. *Nature Chemical Biology*. 13(7), pp.757–763.

Lira-Navarrete, E., Valero-González, J., Villanueva, R., Martínez-Júlvez, M., Tejero, T., Merino, P., Panjikar, S. and Hurtado-Guerrero, R. 2011. Structural Insights into the Mechanism of Protein O-Fucosylation. *PLoS ONE*. 6(9), e25365.

Lin, L.Y.C., Rakic, B., Chiu, C.P.C., Lameignere, E., Wakarchuk, W.W., Withers, S.G. and Strynadka, N.C.J. 2011. Structure and mechanism of the lipooligosaccharide sialyltransferase from Neisseria meningitidis. *Journal of Biological Chemistry*. 286(43), pp.37237–37248.

Magdalou, J., Fournel-Gogleux, S. and Ouzzine, M. 2010. Insights on membrane topology and structure/function of UDP-glucuronosyltransferases. *Drug Metabolism Reviews*. 42(1). pp.159-166.

Mahal, L.K., Yarema, K.J. and Bertozzi, C.R. 1997. Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science*. 276(5315), pp.1125–1128.

Makshakova, O., Breton, C. and Perez, S. 2020. Unraveling the complex enzymatic machinery making a key galactolipid in chloroplast membrane: a multiscale computer simulation. *Scientific Reports*. 10(1), pp.1–15.

Margolin, E., Crispin, M., Meyers, A., Chapman, R. and Rybicki, E.P. 2020. A Roadmap for the Molecular Farming of Viral Glycoprotein Vaccines: Engineering Glycosylation and Glycosylation-Directed Folding. *Frontiers in Plant Science*. 11. 609207.

Masuda, S., Harada, J., Yokono, M., Yuzawa, Y., Shimojima, M., Murofushi, K., Tanaka, H., Masuda, H., Murakawa, M., Haraguchi, T., Kondo, M., Nishimura, M., Yuasa, H., Noguchi, M., Oh-Oka, H., Tanaka, A., Tamiaki, H. and Ohta, H. 2011. A Monogalactosyldiacylglycerol Synthase Found in the Green Sulfur Bacterium Chlorobaculum tepidum Reveals Important Roles for Galactolipids in Photosynthesis. *The Plant Cell*. 23(7), pp.2644–2658.

Mattner, J., DeBord, K.L., Ismail, N., Goff, R.D., Cantu, C., Zhou, D., Saint-Mezard, P., Wang, V., Gao, Y., Yin, N., Hoebe, K., Schneewind, O., Walker, D., Beutler, B., Teyton, L., Savage, P.B. and Bendelac, A. 2005. Exogenous and endogenous glycolipid antigens activate NKT cells during microbial infections. *Nature*. 434(7032), pp.525–529.

Maverakis, E., Kim, K., Shimoda, M., Gershwin, M.E., Patel, F., Wilken, R., Raychaudhuri, S., Ruhaak, L.R. and Lebrilla, C.B. 2015. Glycans in the immune system and The Altered

Glycan Theory of Autoimmunity: A critical review. *Journal of Autoimmunity*. 57, pp.1–13.

McMorran, B.J., McCarthy, F.E., Gibbs, E.M., Pang, M., Marshall, J.L., Nairn, A. V., Moremen, K.W., Crosbie-Watson, R.H. and Baum, L.G. 2016. Differentiation-related glycan epitopes identify discrete domains of the muscle glycocalyx. *Glycobiology*. 26(10), pp.1120–1132.

Melton-Celsa, A.R. 2014. Shiga Toxin (Stx) Classification, Structure, and Function. *Microbiology Spectrum*. 2(4). EHEC-0024

Miège, C., Maréchal, E., Shimojima, M., Awai, K., Block, M. A., Ohta, H., et al. 1999. Biochemical and topological properties of type A MGDG synthase, a spinach chloroplast envelope enzyme catalyzing the synthesis of both prokaryotic and eukaryotic MGDG. *European Journal of Biochemistry*. 265, pp.990–1001.

Migliorini, E., Thakar, D., Sadir, R., Pleiner, T., Baleux, F., Lortat-Jacob, H., Coche-Guerente, L. and Richter, R.P. 2014. Well-defined biomimetic surfaces to characterize glycosaminoglycan-mediated interactions on the molecular, supramolecular and cellular levels. *Biomaterials*. 35(32), pp.8903–8915.

Mitra, N., Sinha, S., Ramya, T.N.C. and Surolia, A. 2006. N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends in biochemical sciences*. 31(3), pp.156–163.

Möckl, L. 2020. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. *Frontiers in Cell and Developmental Biology*. 8(253), pp.1–14.

Moncrieffe, M.C., Fernandez, M.J., Spiteller, D., Matsumura, H., Gay, N.J., Luisi, B.F. and Leadlay, P.F. 2012. Structure of the Glycosyltransferase EryCIII in Complex with its Activating P450 Homologue EryCII. *Journal of Molecular Biology*. 415–20(1), pp.92–101.

Moran, A.P., Prendergast, M.M. and Appelmelk, B.J. 1996. Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease. *FEMS Immunology & Medical Microbiology*. 16(2), pp.105–115.

Moremen, K.W. and Haltiwanger, R.S. 2019. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nature Chemical Biology*. 15(9), pp.853–864.

Mulivor, A.W. and Lipowsky, H.H. 2002. Role of glycocalyx in leukocyte-endothelial cell adhesion. American journal of physiology. *Heart and circulatory physiology*. 283(4), pp.H1282–H1291.

Nair, P.C., Meech, R., Mackenzie, P.I., McKinnon, R.A. and Miners, J.O. 2015. Insights into the UDP-sugar selectivities of human UDP-glycosyltransferases (UGT): A molecular modeling perspective. *Drug Metabolism Reviews*. 47(3), pp.335–345.

Narimatsu, Y., Joshi, H.J., Nason, R., Van Coillie, J., Karlsson, R., Sun, L., Ye, Z., Chen, Y.H., Schjoldager, K.T., Steentoft, C., Furukawa, S., Bensing, B.A., Sullam, P.M., Thompson, A.J., Paulson, J.C., Büll, C., Adema, G.J., Mandel, U., Hansen, L., Bennett, E.P., Varki, A., Vakhrushev, S.Y., Yang, Z. and Clausen, H. 2019. An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Molecular Cell.* 75(2), pp.394–407.

Ng, B.G. and Freeze, H.H. 2018. Perspectives on Glycosylation and its Congenital Disorders. *Trends in genetics*. 34(6), pp.466–476.

Ni, L., Sun, M., Yu, H., Chokhawala, H., Chen, X. and Fisher, A.J. 2006. Cytidine 5'-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from Pasteurella multocida. *Biochemistry*. 45(7), pp.2139–2148.

Nitenberg, M., Makshakova, O., Rocha, J., Perez, S., Maréchal, E., Block, M.A., Girard-Egrot, A. and Breton, C. 2020. Mechanism of activation of plant monogalactosyldiacylglycerol synthase 1 (MGD1) by phosphatidylglycerol. *Glycobiology*. 30(6), pp.396–406.

Nitenberg, M. 2018. Biosynthèse des galactolipids de plantes: étude structurale et fonctionnelle de la MGDG synthase 1 (MGD1) d' Arabidopsis thaliana. Thesis. *Université Grenoble Alpes*

Noble, G.T., Craven, F.L., Voglmeir, J., Šardzík, R., Flitsch, S.L. and Webb, S.J. 2012. Accelerated enzymatic galactosylation of N -acetylglucosaminolipids in lipid microdomains. *Journal of the American Chemical Society*. 134(31), pp.13010–13017.

Oikawa, A., Lund, C.H., Sakuragi, Y. and SQchekker, H.V. 2013. Golgi-localized enzyme complexes for plant cell wall biosynthesis. *Trends in Plant Science*. 18(1), pp.49-58.

Omidvar, R. and Römer, W. 2019. Glycan-decorated protocells: Novel features for rebuilding cellular processes. *Interface Focus*. 9(2). 20180084.

Ovchinnikova, O.G., Mallette, E., Koizumi, A., Lowary, T.L., Kimber, M.S. and Whitfield, C. 2016. Structural Insights into the Mechanism of Protein O-Fucosylation. *Proceedings of the National Academy of Sciences of the United States of America*. 113(22), pp.E3120–E3129.

Oyelaran, O., Li, Q., Farnsworth, D. and Gildersleeve, J.C. 2009. Microarrays with varying carbohydrate density reveal distinct subpopulations of serum antibodies. *Journal of Proteome Research*. 8(7), pp.3529–3538.

Pak, J.E., Satkunarajah, M., Seetharaman, J. and Rini, J.M. 2011. Structural and mechanistic characterization of leukocyte-type core 2 β1,6-N-acetylglucosaminyltransferase: A metal-ion-independent GT-A glycosyltransferase. *Journal of Molecular Biology*. 414(5), pp.798–811.

Parera Pera, N., Branderhorst, H.M., Kooij, R., Maierhofer, C., Van Der Kaaden, M., Liskamp, R.M.J., Wittmann, V., Ruijtenbeek, R. and Pieters, R.J. 2010. Rapid screening of lectins for multivalency effects with a glycodendrimer microarray. *ChemBioChem*. **11**(13), pp.1896–1904.

Patnaik, S.K. and Stanley, P. 2006. Lectin-Resistant CHO Glycosylation Mutants. *Methods in Enzymology*. 416, pp.159–182.

Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G. and Strynadka, N.C.J. 2001. Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. *Nature Structural Biology*. 8(2), pp.166–175.

Petroutsos, D., Amiar, S., Abida, H., Dolch, L.J., Bastien, O., Rébeillé, F., Jouhet, J., Falconet, D., Block, M.A., McFadden, G.I., Bowler, C., Botté, C. and Maréchal, E. 2014. Evolution of galactoglycerolipid biosynthetic pathways – From cyanobacteria to primary plastids and from primary to secondary plastids. *Progress in Lipid Research*. 54(1), pp.68–85.

Pinho, S.S. and Reis, C.A. 2015. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*. 15(9), pp.540–555.

Prabakaran, T., Nielsen, R., Satchell, S.C., Mathieson, P.W., Feldt-Rasmussen, U., Sørensen, S.S. and Christensen, E.I. 2012. Mannose 6-Phosphate Receptor and Sortilin Mediated Endocytosis of α-Galactosidase A in Kidney Endothelial Cells. *PLOS ONE*. 7(6), e39975.

Pulsipher, A., Griffin, M.E., Stone, S.E. and Hsieh-Wilson, L.C. 2015. Long-Lived Glycan Engineering to Direct Stem Cell Fate. *Angewandte Chemie*. 54(5), pp.1466–1470.

Purcell, S.C. and Godula, K. 2019. Synthetic glycoscapes: addressing the structural and functional complexity of the glycocalyx. *Interface Focus*. 9(2), 20180080.

Qasba, P.K. and Ramakrishnan, B. 2007. Letter to the Glyco-Forum: Catalytic domains of glycosyltransferases with 'add-on' domains. *Glycobiology*. 17(5), pp.7G-9G.

Qasba, P.K., Ramakrishnan, B. and Boeggeman, E. 2008. Structure and Function of β-1,4-Galactosyltransferase. *Current drug targets*. 9(4), pp.292-309

Qiu, H., Gong, J., Butaye, P., Lu, G., Huang, K., Zhu, G., Zhang, J., Hathcock, T., Cheng, D. and Wang, C. 2018. CRISPR/Cas9/sgRNA-mediated targeted gene modification confirms the cause-effect relationship between gyrA mutation and quinolone resistance in Escherichia coli. *FEMS Microbiology Letters*. 365(13). fny127.

Ribeiro, J.P., Villringer, S., Goyard, D., Coche-Guerente, L., Höferlin, M., Renaudet, O., Römer, W. and Imberty, A. 2018. Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells. *Chemical Science*. 9(39), pp.7634–7641.

Richter, R.P., Bérat, R. and Brisson, A.R. 2006. Formation of solid-supported lipid bilayers: An integrated view. *Langmuir*. 22(8), pp.3497–3505.

Rocha, J., Audry, M., Pesce, G., Chazalet, V., Block, M.A., Maréchal, E. and Breton, C. 2013. Revisiting the expression and purification of MGD1, the major galactolipid synthase in Arabidopsis to establish a novel standard for biochemical and structural studies. *Biochimie*. 95(4), pp.700–708.

Rocha, J., Chazalet, V. and Breton, C. 2016a. Expression, Purification and Crystallization of Recombinant Arabidopsis Monogalactosyldiacylglycerol Synthase (MGD1). *Bio-Protocol*. 6(24). e2064.

Rocha, J., Nitenberg, M., Girard-Egrot, A., Jouhet, J., Maréchal, E., Block, M.A. and Breton, C. 2018. Do galactolipid synthases play a key role in the biogenesis of chloroplast membranes of higher plants? *Frontiers in Plant Science*. 9, 126.

Rocha, J., Sarkis, J., Thomas, A., Pitou, L., Radzimanowski, J., Audry, M., Chazalet, V., de Sanctis, D., Palcic, M.M., Block, M.A., Girard-Egrot, A., Maréchal, E. and Breton, C. 2016b. Structural insights and membrane binding properties of MGD1, the major galactolipid synthase in plants. *The Plant journal: for cell and molecular biology*. 85(5), pp.622–633.

Sarkis, J., Rocha, J., Maniti, O., Jouhet, J., Vié, V., Block, M.A., Breton, C., Maréchal, E. and Girard-Egrot, A. 2014. The influence of lipids on MGD1 membrane binding highlights novel mechanisms for galactolipid biosynthesis regulation in chloroplasts. *The FASEB Journal*. 28(7), pp.3114–3123.

Saxon, E. and Bertozzi, C.R. 2000. Cell Surface Engineering by a Modified Staudinger Reaction. *Science*. 287(5460), pp.2007–2010.

Schaller, S., Latowski, D., Jemioła-Rzemińska, M., Dawood, A., Wilhelm, C., Strzałka, K. and Goss, R. 2011. Regulation of LHCII aggregation by different thylakoid membrane lipids. *Biochimica et Biophysica Acta - Bioenergetics*. 1807(3), pp.326–335.

Schlünzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A. and Franceschi, F. 2001. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*. 413(6858), pp.814–821.

Šnajdrová, L., Kulhánek, P., Imberty, A. and Koča, J. 2004. Molecular dynamics simulations of glycosyltransferase LgtC. *Carbohydrate Research*. 339(5), pp.995–1006.

Sola, R.J. and Griebenow, K. 2009. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *Journal of pharmaceutical sciences*. 98(4), pp.1223–1245.

Stimmer, L., Dehay, S., Nemati, F., Massonnet, G., Richon, S., Decaudin, D., Klijanienko, J. and Johannes, L. 2014. Human breast cancer and lymph node metastases express Gb3 and can be targeted by STxB-vectorized chemotherapeutic compounds. *BMC Cancer*. 14(1). 916.

Tang, T.M.S., Cardella, D., Lander, A.J., Li, X., Escudero, J.S., Tsai, Y.H. and Luk, L.Y.P. 2020. Use of an asparaginyl endopeptidase for chemo-enzymatic peptide and protein labeling. *Chemical Science*. 11(23), pp.5881–5888.

Taujale, R., Zhou, Z., Yeung, W., Moremen, K.W., Li, S. and Kannan, N. 2021. Mapping the glycosyltransferase fold landscape using interpretable deep learning. *Nature Communications*. 12(1). 5656.

Taylor, S.G., McKenzie, I.F.C. and Sandrin, M.S. 2003. Characterization of rat α(1,3)galactosyltransferase: Evidence for two independent genes encoding glycosyltransferases that synthesize Galα(1,3)Gal by two separate glycosylation pathways. *Glycobiology*. 13(5), pp.327–337.

Tytgat, H.L.P., Lin, C. wei, Levasseur, M.D., Tomek, M.B., Rutschmann, C., Mock, J., Liebscher, N., Terasaka, N., Azuma, Y., Wetter, M., Bachmann, M.F., Hilvert, D., Aebi, M. and Keys,

T.G. 2019. Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. *Nature Communications*. 10(1). 5403.

Ulrich-Bott, B. and Wiegandt, H. 1984. Micellar properties of glycosphingolipids in aqueous media. *Journal of Lipid Research*. 25(11). pp.1233-1245

Van Breedam, W., Pöhlmann, S., Favoreel, H.W., de Groot, R.J. and Nauwynck, H.J. 2014. Bitter-sweet symphony: glycan–lectin interactions in virus biology. *FEMS Microbiology Reviews*. 38(4), pp.598–632.

Van Erp, E.A., Luytjes, W., Ferwerda, G. and Van Kasteren, P.B. 2019. Fc-Mediated Antibody Effector Functions During Respiratory Syncytial Virus Infection and Disease. *Frontiers in Immunology*. 10(MAR), 548.

Van Landuyt, L., Lonigro, C., Meuris, L. and Callewaert, N. 2019. Customized protein glycosylation to improve biopharmaceutical function and targeting. *Current Opinion in Biotechnology*. 60, pp.17–28.

Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L. and Seeberger, P.H. 2017. *Essentials of glycobiology, third edition*.

Villringer, S., Madl, J., Sych, T., Manner, C., Imberty, A. and Römer, W. 2018. Lectin-mediated protocell crosslinking to mimic cell-cell junctions and adhesion. *Scientific Reports.* 8(1). 1932.

Wakarchuk, W.W., Cunningham, A., Watson, D.C. and Young, N.M. 1998. Role of paired basic residues in the expression of active recombinant galactosyltransferases from the bacterial pathogen Neisseria meningitidis. Protein Engineering Design and Selection. 11(4), pp.295–302.

Wang, Y., Ju, T., Ding, X., Xia, B., Wang, W., Xia, L., He, M. and Cummings, R.D. 2010. Cosmc is an essential chaperone for correct protein O-glycosylation. *Proceedings of the National Academy of Sciences of the United States of America*. 107(20), pp.9228–9233.

Wang, L.X., Tong, X., Li, C., Giddens, J.P. and Li, T. 2019. Glycoengineering of antibodies for modulating functions. Annual Review of Biochemistry. 88, pp.433–459.

Wheeler, K.M., Cárcamo-Oyarce, G., Turner, B.S., Dellos-Nolan, S., Co, J.Y., Lehoux, S., Cummings, R.D., Wozniak, D.J. and Ribbeck, K. 2019. Mucin glycans attenuate the virulence of Pseudomonas aeruginosa in infection. *Nature microbiology*. 4(12), pp.2146–2154.

Wu, Y., Sha, Q., Du, J., Wang, C., Zhang, L., Liu, B.F., Lin, Y. and Liu, X. 2018. Determination of N-glycans by high performance liquid chromatography using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate as the glycosylamine labeling reagent. *Journal of Chromatography A*. 1535, pp.114–122.

Xu, Y., Smith, R., Vivoli, M., Ema, M., Goos, N., Gehrke, S., Harmer, N.J. and Wagner, G.K. 2017. Covalent inhibitors of LgtC: A blueprint for the discovery of non-substrate-like inhibitors for bacterial glycosyltransferases. *Bioorganic & Medicinal Chemistry*. 25(12), pp.3182–3194.

Yamamoto, F.I., Clausen, H., White, T., Marken, J. and Hakomori, S.I. 1990. Molecular genetic basis of the histo-blood group ABO system. *Nature*. 345(6272), pp.229–233.

Yu, B., Xu, C. and Benning, C. 2002. Arabidopsis disrupted in SQD2 gene encoding sulfolipid synthase is impaired in phosphate-limited growth. *Proceedings of the National Academy of Sciences of the United States of America.* 99, pp.5732–5737.

Yuzawa, Y., Nishihara, H., Haraguchi, T., Masuda, S., Shimojima, M., Shimoyama, A., Yuasa, H., Okada, N. and Ohta, H. 2012. Phylogeny of galactolipid synthase homologs together with their enzymatic analyses revealed a possible origin and divergence time for photosynthetic membrane biogenesis. *DNA Research*. 19(1), pp.91–102.

Zeng, Y. 2017. Endothelial glycocalyx as a critical signalling platform integrating the extracellular haemodynamic forces and chemical signalling. *Journal of Cellular and Molecular Medicine*. 21(8), pp.1457–1462.

Appendix

MGD1 mutant primer designs

P433G (GGG) 5’ – GC ATC ATA ACA AAG GCT GGT GGA GGC ACT ATA GCT GAA GC – 3’ 5’ – GC TTC AGC TAT AGT GCC TCC ACC AGC CTT TGT TAT GAT GC – 3’ Translation: IITKAGGGTIAE
P433G + G434A (GGA) 5’ – CT GC ATC ATA ACA AAG GCT GGT GGT GCG ACT ATA GCT GAA GC – 3’ 5’ – GC TTC AGC TAT AGT CGC ACC ACC AGC CTT TGT TAT GAT GC AG – 3’ Translation: IITKAGGATIAE
P433G + G434L (GGL) 5’ – C ATA ACA AAG GCT GGT GGA CTG ACT ATA GCT GAA GCT ATG ATA AG – 3’ 5’ – CT TAT CAT AGC TTC AGC TAT AGT CAG TCC ACC AGC CTT TGT TAT G– 3’ Translation: ITKAGGLTIAEAMI
P433A + Q455E (AEE) 5’ – CGG TAC ATT CCC TGC CTC TTC ACC AGC GAT GTA CCC – 3’ 5’ – GGG TAC ATC GCT GGT GAA GAG GCA GGG AAT GTA CCG – 3’ Translation: GYIAGEEAGNVP
P433A + Q455E + E456Q (AEQ) 5’ – CGG TAC ATT CCC TGC CTG TTC ACC AGC GAT GTA CCC – 3’ 5’ – GGG TAC ATC GCT GGT GAA CAG GCA GGG AAT GTA CCG – 3’ Translation: GYIAGEQAGNVP
R156A 5’ – GAT ACT GGT GGT GGT CAC GCG GCT TCT GCG GAA GCC – 3’ 5’ – GGC TTC CGC AGA AGC CGC GTG ACC ACC ACC AGT ATC – 3’ Translation: DTGGGHAASAEA
R195A 5’ – C CAG AAA ATT ATA GCT CGC GGG AAG CTG GTT GAA CGG– 3’ 5’ – C CAG AAA ATT ATA GCT CGC GGG AAG CTG GTT GAA CGG– 3’ Translation: PFNQLPASYNFL
H251A 5’ – GAT ATC ATA ATC AGT GTT GCG CCG TTG ATG CAA CAT GTC CCA C – 3’ 5’ – G TGG GAC ATG TTG CAT CAA CGG CGC AAC ACT GAT TAT GAT ATC – 3’ Translation: DIIISVAPLMQHVP
H256A 5’ – GTT CAT CCG TTG ATG CAA GCG GTC CCA CTT CGT GTC C – 3’ 5’ – G GAC ACG AAG TGG GAC CGC TTG CAT CAA CGG ATG AAC – 3’ Translation: VHPLMQAVPLRV
R260A 5’ – G ATG CAA CAT GTC CCA CTT GCG GTC CTA AGA TCA AAA GGC TTG – 3’ 5’ – CAA GCC TTT TGA TCT TAG GAC CGC AAG TGG GAC ATG TTG CAT C – 3’ Translation: MQHVPLAVLRSKGL
K265A 5’ – CCA CTT CGT GTC CTA AGA TCA GCA GGC TTG CTT AAG AAA ATC GTG – 3’ 5’ – CAC GAT TTT CTT AAG CAA GCC TGC TGA TCT TAG GAC ACG AAG TGG – 3’ Translation: PLRVLRSAGLLKKIV
H289A 5’ – GC CAT CCT ACA TGG TTT GCT AAG CTT GTG ACA AGA TGT TAT TG – 3’ 5’ – CA ATA ACA TCT TGT CAC AAG CTT AGC AAA CCA TGT AGG ATG GC – 3’ Translation: HPTWFAKLVTRCY
V250D 5’ – CCA GAT ATC ATA ATC AGT GAT CAT CCG TTG ATG CAA CAT GTC – 3’ 5’ – GAC ATG TTG CAT CAA CGG ATG ATC ACT GAT TAT GAT ATC TGG – 3’

Translation: PDIIISDHPLMQHV
V250E 5’ – CCA GAT ATC ATA ATC AGT GAA CAT CCG TTG ATG CAA CAT GTC C – 3’ 5’ – G GAC ATG TTG CAT CAA CGG ATG TTC ACT GAT TAT GAT ATC TGG – 3’ Translation: PDIIISEHPLMQHV

S1: *Primer designs for MGD1 mutants. In red are the mutated codons and the corresponding amino acid after mutation.*

Chapter 3: Study of MGD1 sugar donor specificity: Galactose vs Glucose

Expression Trials for Mutants GGG, GGA, GGL, AEE, AEQ and RcMGlcD

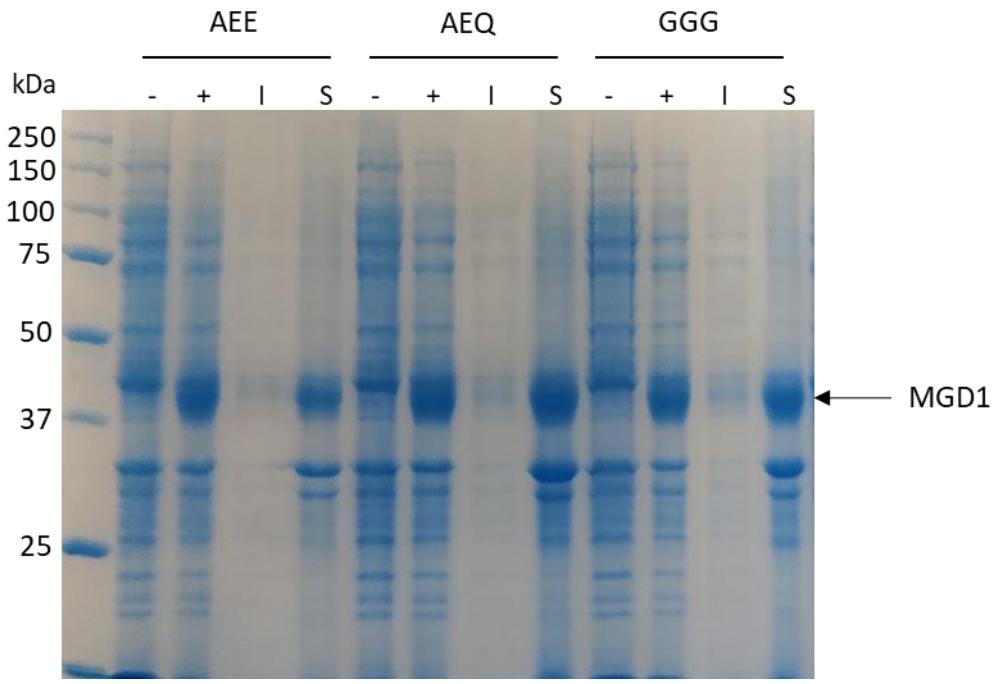


Figure S2: *Small scale expression trial for the MGD1 mutants AEE, AEQ and GGG. SDS-PAGE gel (A) and Western blot (B). Samples taken are total cell extract before (-) and after (+) IPTG induction, soluble (S) and insoluble (I) fractions from induced cells.*

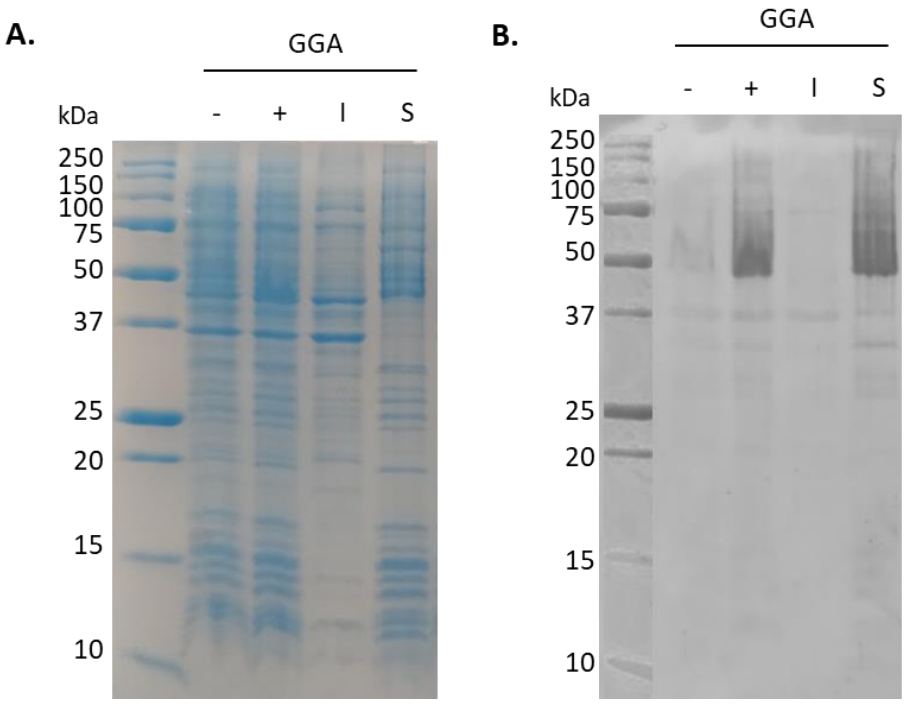


Figure S3: *Small scale expression trial for the MGD1 mutants GGA. SDS-PAGE gel (A) and Western blot (B). Samples taken are total cell extract before (-) and after (+) IPTG induction, soluble (S) and insoluble (I) fractions from induced cells.*

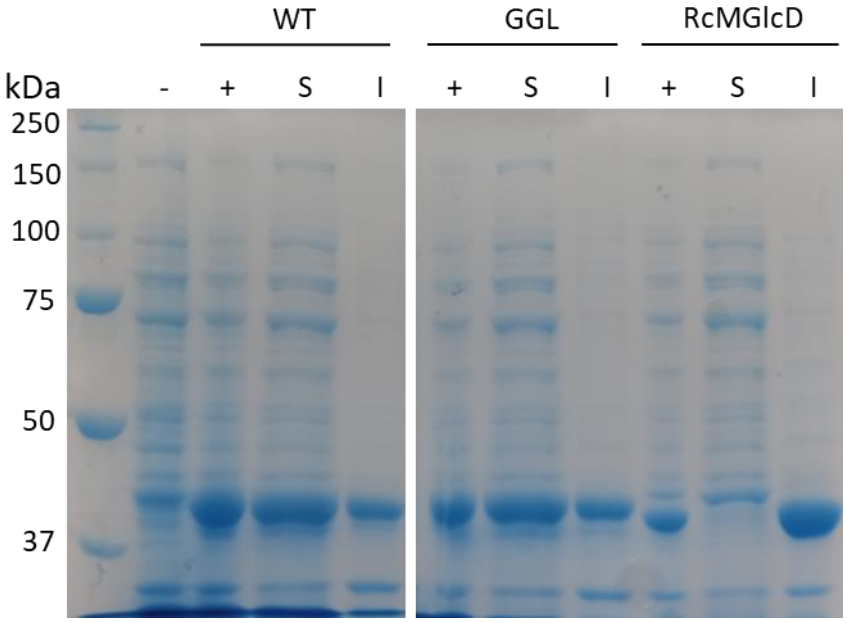


Figure S4: *Small scale expression trial for the wild-type (WT), GGL and RcMGlcD mutants. SDS-PAGE gel (A) and Western blot (B). Samples taken are total cell extract before (-) and after (+) IPTG induction, soluble (S) and insoluble (I) fractions from induced cells.*

Chapter 4: Investigating MGD1 reaction mechanism: Role of PA vs PG

1. Expression trials

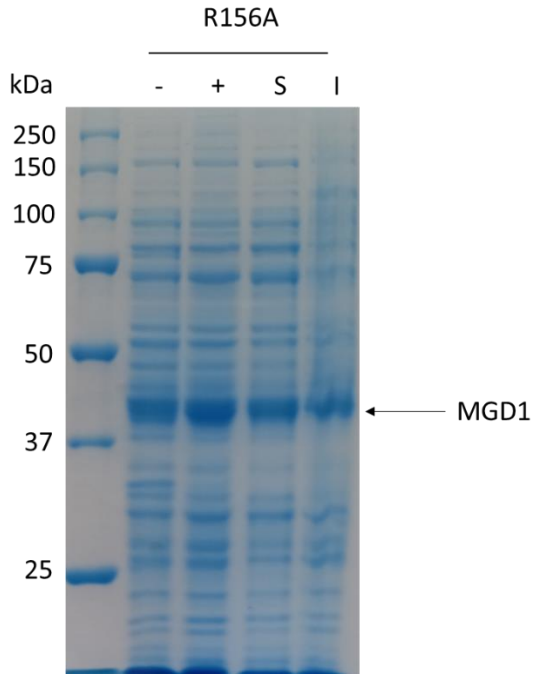


Figure S5: *Small scale expression trial for the MGD1 mutant R156A. SDS-PAGE gel (A) and Western blot (B). Samples taken are total cell extract before (-) and after (+) IPTG induction, soluble (S) and insoluble (I) fractions from induced cells.*

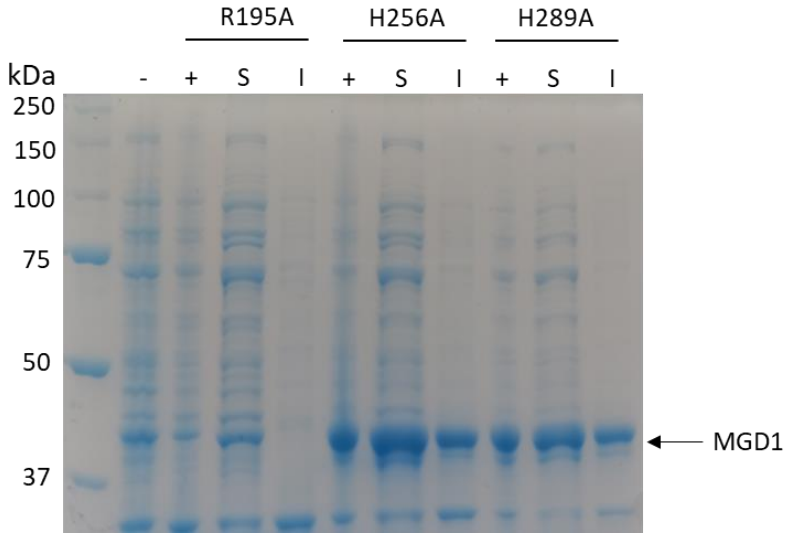


Figure S6: *Small scale expression trial for the MGD1 mutants R195A, H256A and H289A. SDS-PAGE gel (A) and Western blot (B). Samples taken are total cell extract before (-) and after (+) IPTG induction, soluble (S) and insoluble (I) fractions from induced cells.*

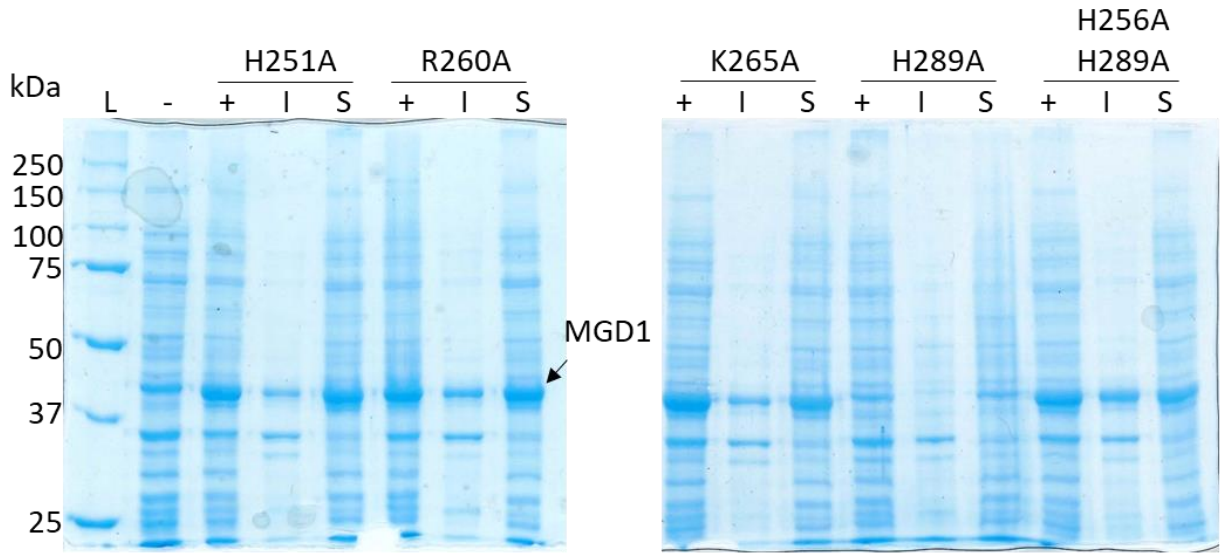


Figure S7: *Small scale expression trial for the MGD1 mutants H251A, R260A, K265A and H256A+H289A. SDS Page 10%, loading OD=100. Wells: (L) Ladder Precision Plus Protein standards (Biorad), (-) Total fraction before induction, (+) Total fraction after induction, (I) Insoluble fraction (pellet after 19 000g, 30min), (S) Soluble fraction (supernatant after 19 000g, 30min).*

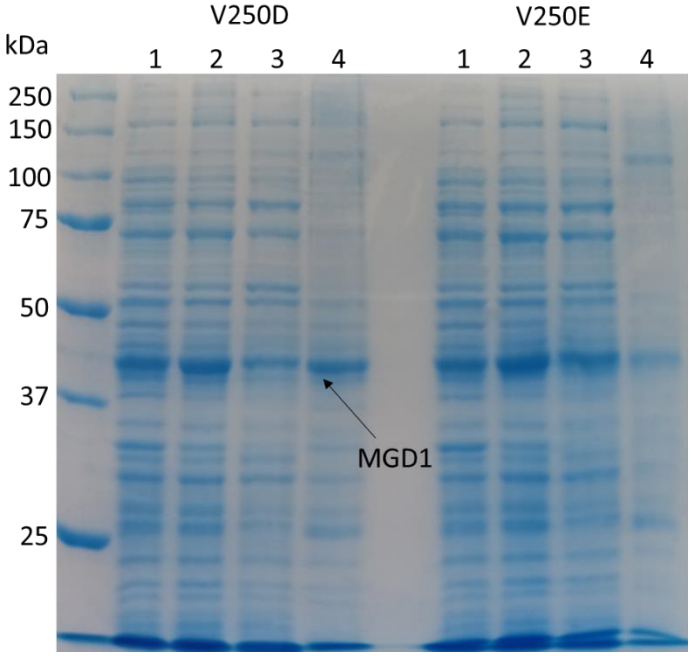


Figure S8: *Small scale expression trials of MGD1 mutants V250D and V250E. SDS-PAGE wells are: total cell fraction before (1) and after (2) IPTG induction, soluble fraction (3) and insoluble fraction.*

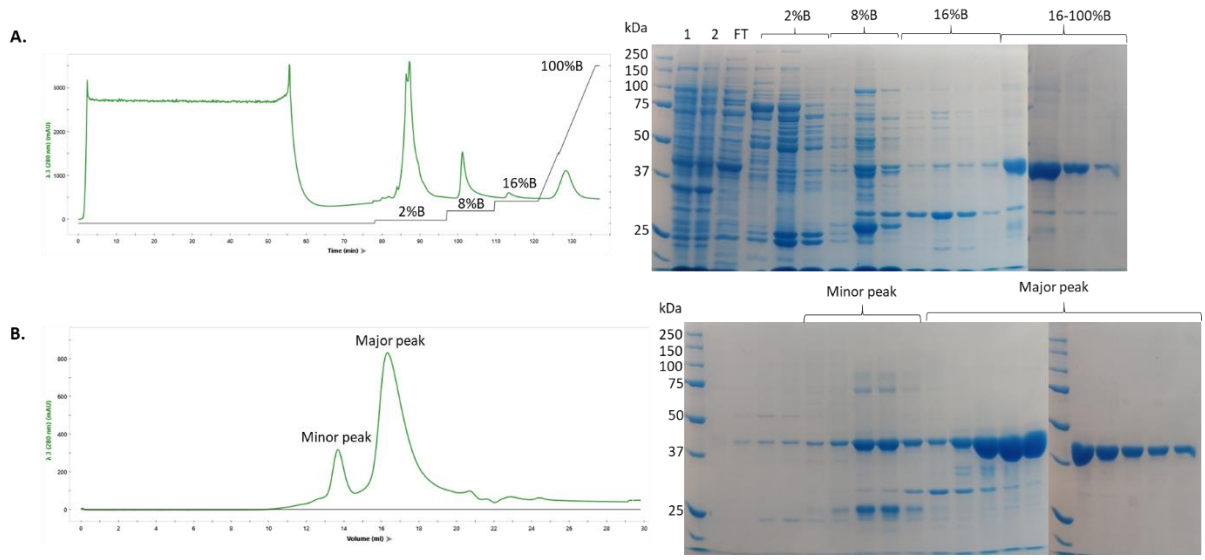


Figure S9: *IMAC (A) and SEC (B) purification of R156A mutant from a 1L culture. A) Wells from SDP-PAGE analysis: Total cell extract before (1) and after (2) IPTG induction, flow through (FT), elution fractions from 2%, 8%, 16% and 100% imidazole concentration. B) SDS-PAGE analysis from the minor and major peaks.*

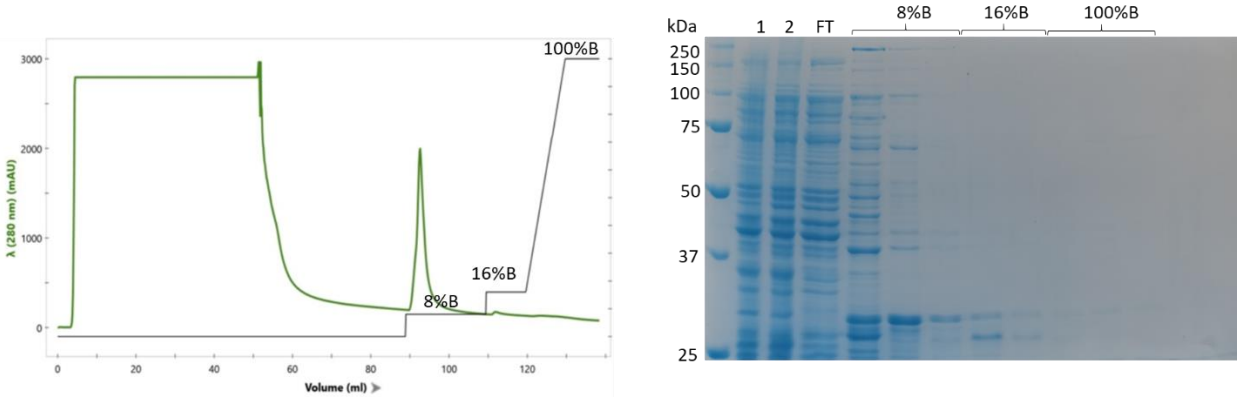


Figure S10: *IMAC purification of R195A. Wells from SDP-PAGE analysis: Total cell extract before (1) and after (2) IPTG induction, flow through (FT), elution fractions from 8%, 16% and 100% imidazole concentration.*

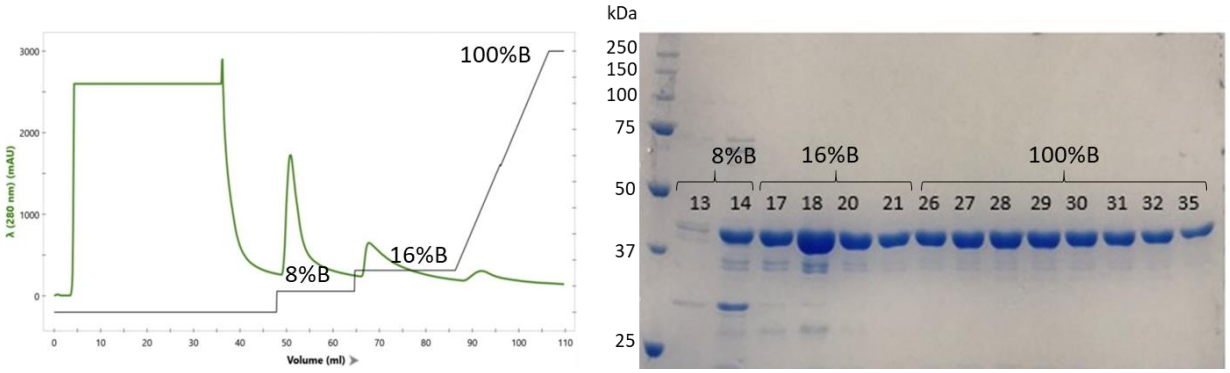


Figure S11: *IMAC purification of H251. Wells from SDP-PAGE analysis: Elution fractions from 8%, 16% and 100% imidazole concentration.*

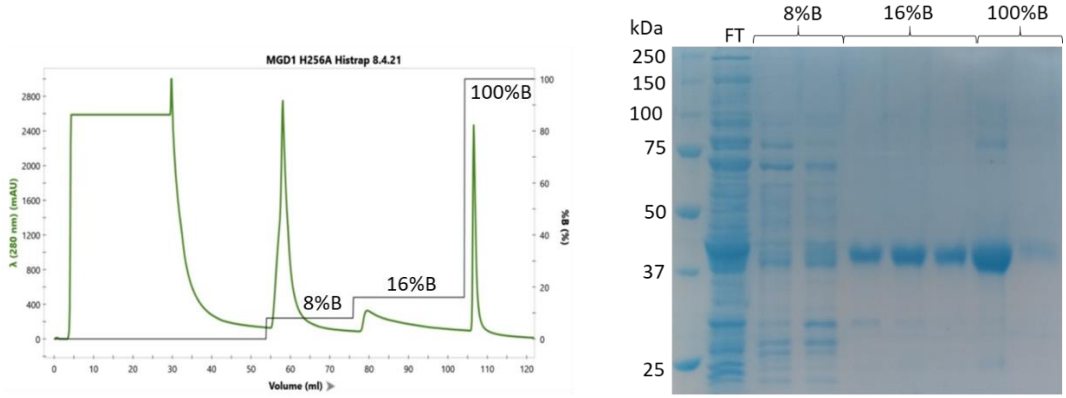


Figure S12: *IMAC purification of H256A. Wells from SDP-PAGE analysis: Flow through (FT), elution fractions from 8%, 16% and 100% imidazole concentration.*

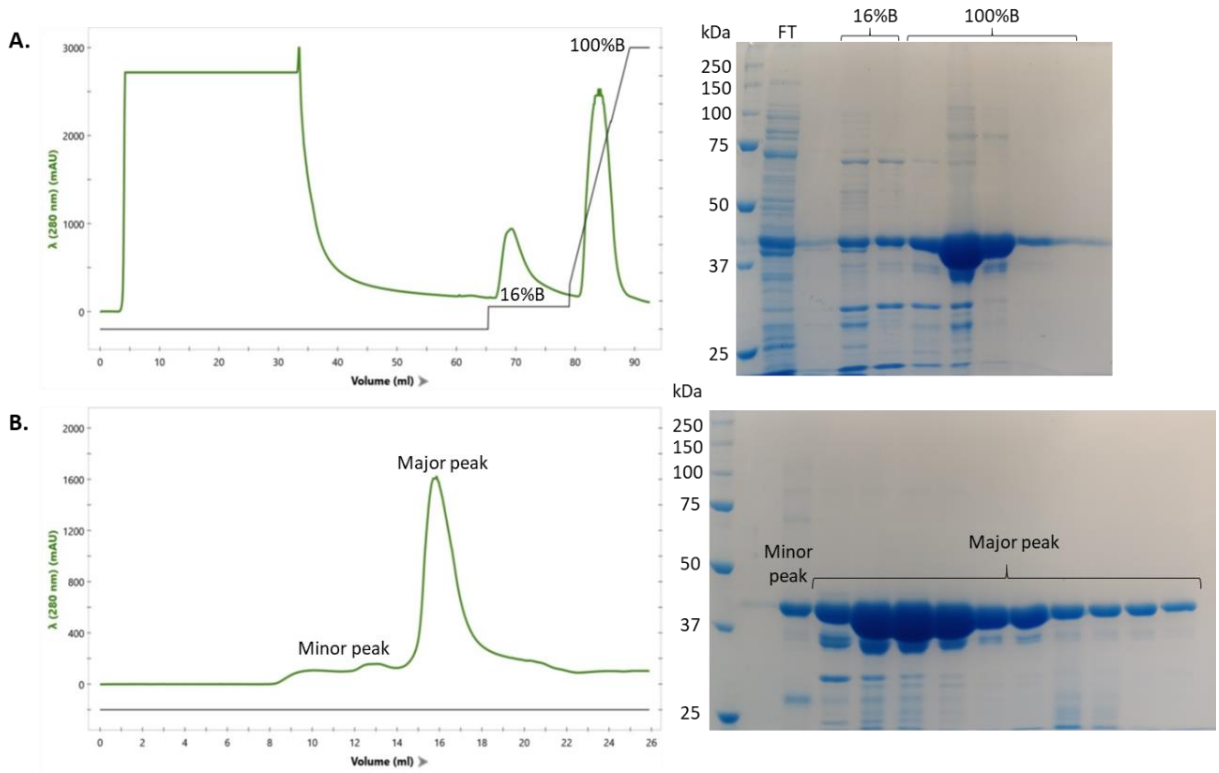


Figure S13: *IMAC (A) and SEC (B) purification of R260A mutant from a 1L culture. A) Wells from SDP-PAGE analysis: Flow through (FT), elution fractions from 16% and 100% imidazole concentration. B) SDS-PAGE analysis from the minor and major peaks.*

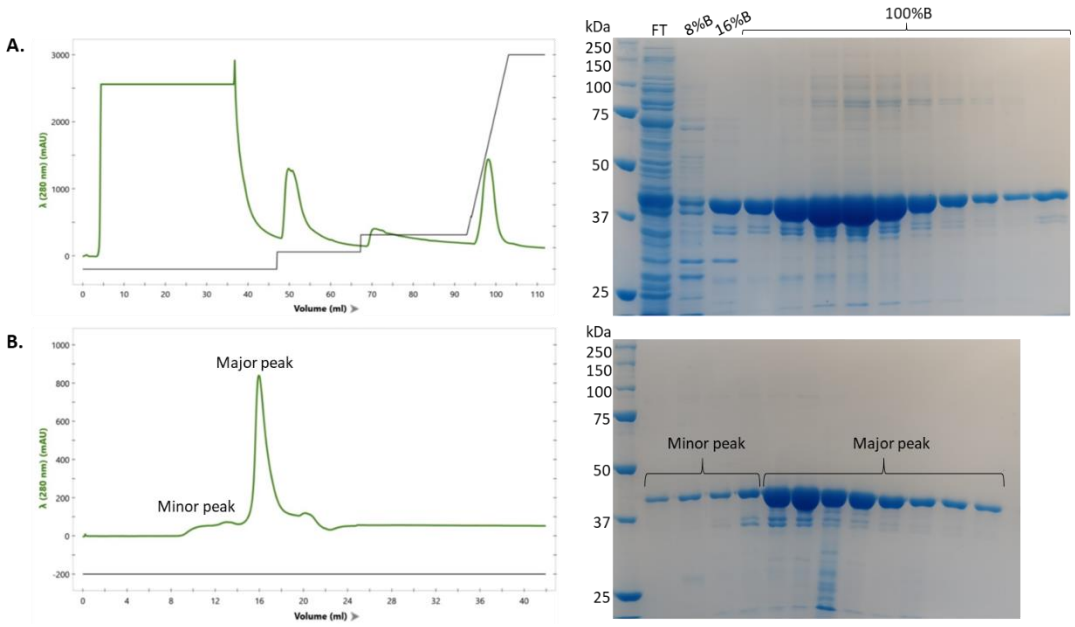


Figure S14: *IMAC (A) and SEC (B) purification of K265A mutant from a 1L culture. A) Wells from SDP-PAGE analysis: Flow through (FT), elution fractions from 8%, 16% and 100% imidazole concentration. B) SDS-PAGE analysis from the minor and major peaks.*

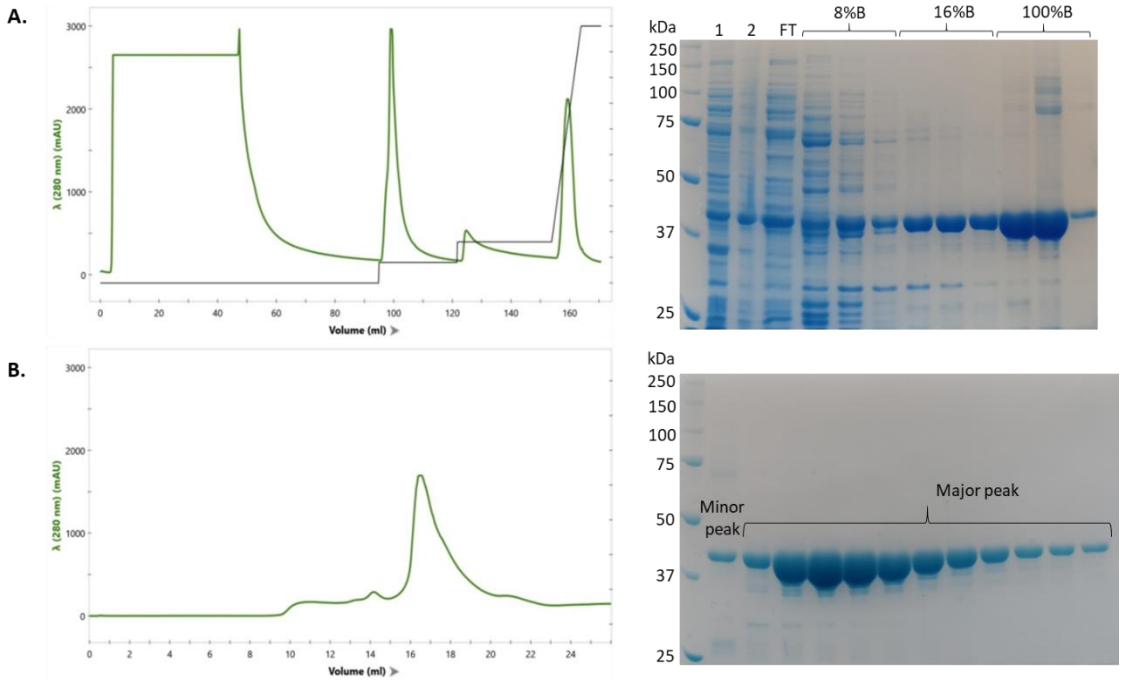


Figure S15: *IMAC (A) and SEC (B) purification of H289A mutant from a 1L culture. A) Wells from SDP-PAGE analysis: Total cell extract before (1) and after (2) IPTG induction, flow through (FT), elution fractions from 8%, 16% and 100% imidazole concentration. B) SDS-PAGE analysis from the minor and major peaks.*

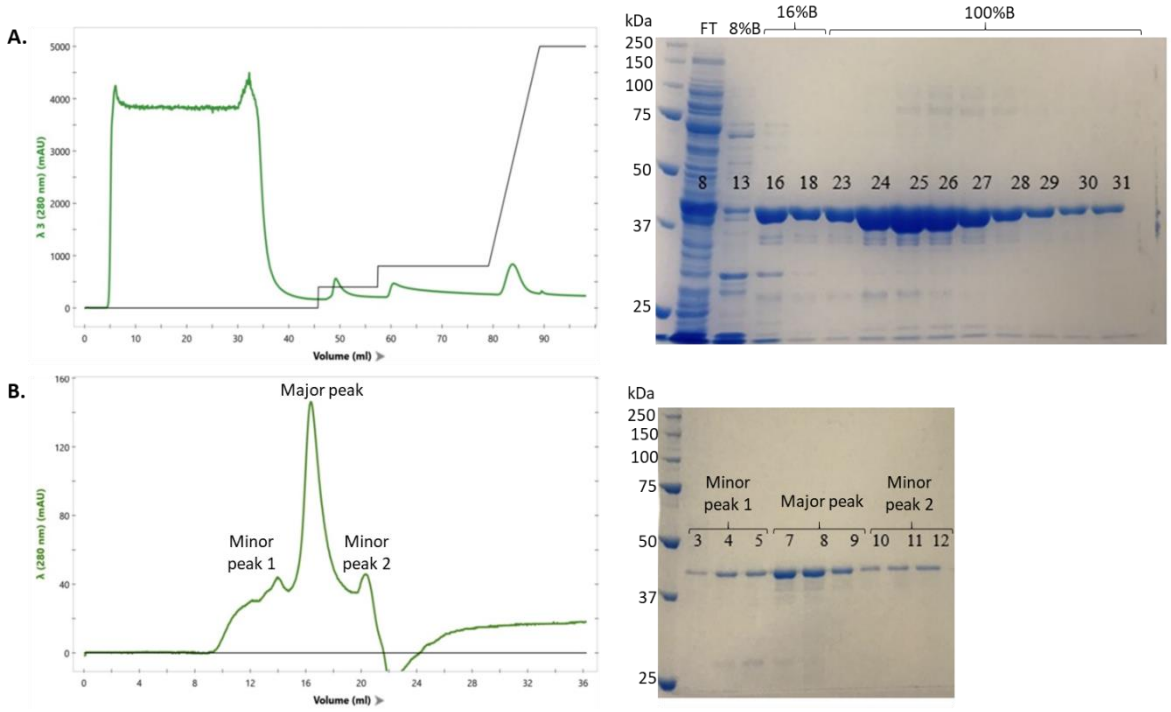


Figure S16: *IMAC (A) and SEC (B) purification of H256A+H289A mutant from a 1L culture. A) Wells from SDP-PAGE analysis: Flow through (FT), elution fractions from 8%, 16% and 100% imidazole concentration. B) SDS-PAGE analysis from the minor and major peaks.*

Chapter 5: Application of glycosyltransferases in synthetic membranes

Defrost cycle	Protein concentration (mg/ml)
0	2.058
1	2.026
2	1.850
3	1.564

Figure S17: *Assay on the stability of LgtC after several defrosting cycles. Buffer conditions: 20 mM Tris, 5mM TCEP, pH 8.5. The initial protein concentration was 2.058 mg/ml. Protein is defrosted on ice and centrifuged at 15000 x g for 10 minutes at 4°C. Protein concentration is then measured on a NanoDropTM ND-2000 spectrophotometer, using the parameters 61225 M^{-1} cm^{-1} and 34.85 kDa. Between each defrost cycle the protein sample was stored at -20°C for 2 days.*

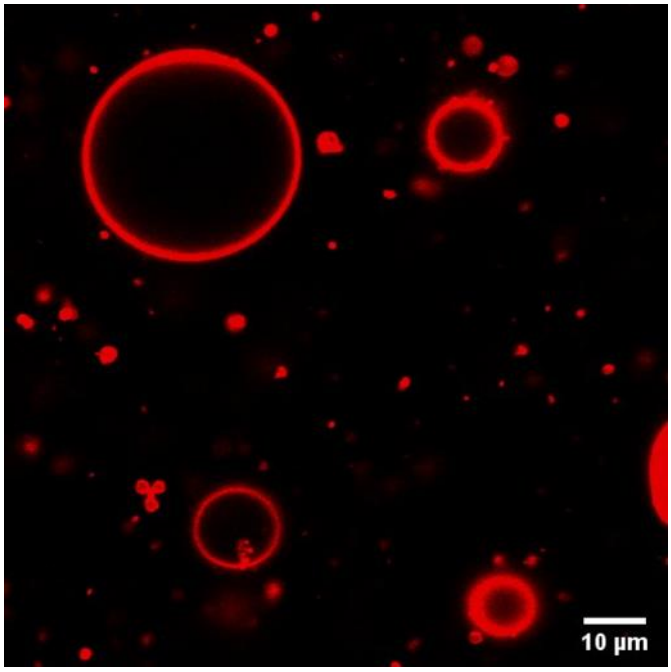


Figure S18: *Poor quality GUVs from 30 mol-% MGDG under confocal microscopy. Lipid composition: 49.7 mol-% DOPC, 10 mol-% Cholesterol, 30 mol-% MGDG, 10 mol-% PG, 0.3 mol-% Atto647 DOPE. GUVs formed according to the Material and Methods.*

SUMMARY

Glycosyltransferases (GTs) are enzymes able to glycosylate a wide range of molecules. For this reason, GTs are extensively used in therapeutic drug development. In this PhD project, we explored the application of GTs as tools for synthetic glycobiology, with a focus on the glycosylation of synthetic membranes. This project is split into two main parts. The first part studies the donor specificity and catalytic mechanism of a model GT which works on a lipid membrane, MGD1 from Arabidopsis thaliana. MGD1 is a galactosyltransferase that transfers a galactose from UDP-Gal to diacylglycerol (DAG), and needs anionic lipids such as phosphatidylglycerol (PG) or phosphatidic acid (PA) for activity. We used a structure and sequence-guided approach to engineer a MGD1 mutant altering its donor specificity to UDP-Glc. We performed mutational studies with the aim to consolidate the previously proposed PG-His catalytic dyad and probe the binding site for the allosteric activator PA. The second part explored the application of LgtC, a bacterial galactosyltransferase, to synthesize the Gb3 antigen on synthetic membrane systems. The capability of LgtC to generate Gb3 on giant unilamellar vesicles (GUVs) was demonstrated by confocal microscopy using fluorescently-labelled Shiga toxin B subunit (StxB), a lectin specific for Gb3. Quartz crystal microbalance with dissipation (QCM-D) allowed quantitative evaluation of Gb3 formed on supported lipid bilayer (SLB). The potential of LgtC and MGD1 to transfer a functionalised galactose (Gal-N3) to their respective acceptors was also evaluated. The knowledge gained from this research has value in the engineering of GTs as tools in synthetic glycobiology applications.

RESUME

Les glycosyltransférases (GT) sont des enzymes capables de glycosyler un large éventail de molécules. Pour cette raison, elles sont largement utilisées dans le développement de molécules thérapeutiques. Dans ce projet de thèse, nous avons exploré le potentiel des GTs comme outils pour la glycobiologie synthétique, en nous concentrant sur leur capacité à glycosyler des membranes synthétiques. Ce projet est divisé en deux parties. La première partie porte sur l'étude de la spécificité vis-à-vis du nucléotide sucre donneur et du mécanisme catalytique d'une GT modèle qui fonctionne sur une membrane lipidique, la MGD1 d'*Arabidopsis thaliana*. MGD1 est une galactosyltransférase qui transfère un galactose de l'UDP-Gal au diacylglycérol (DAG), et qui a besoin de lipides anioniques tels que le phosphatidylglycérol (PG) ou l'acide phosphatidique (PA) pour son activité. Nous avons utilisé une approche basée sur l'analyse de séquences et de structures de GT pour obtenir un mutant MGD1 pouvant utiliser l'UDP-glucose. Par des études de mutagenèse dirigée, nous avons cherché à consolider l'hypothèse d'une dyade catalytique PG-His chez MGD1 et à identifier le ou les sites de liaison de l'enzyme au PA. Dans la deuxième partie du projet, nous avons testé la capacité de la LgtC, une galactosyltransférase bactérienne, à synthétiser l'antigène Gb3 sur des systèmes membranaires synthétiques. La capacité de la LgtC à générer du Gb3 sur des vésicules unilamellaires géantes (GUVs) a été démontrée par microscopie confocale en utilisant une lectine spécifique du Gb3, la sous-unité B de la Shiga-toxine (StxB) marquée par fluorescence. L'utilisation de la microbalance à cristal de quartz avec mesure de dissipation (QCM-D) a permis une évaluation quantitative du Gb3 formé sur une bicouche lipidique supportée (SLB). Le potentiel de LgtC et MGD1 à transférer un galactose fonctionnalisé (Gal-N3) à leurs accepteurs respectifs a également été évalué. Les connaissances acquises dans le cadre de cette recherche ont des retombées pour des applications potentielles des GTs en glycobiologie synthétique.