University of Southampton

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Chemistry

An investigation into an unusual glycan branching enzyme from *Mycobacterium tuberculosis*

by

Jessica Sophie Gusthart

Thesis for the degree of Doctor of Philosophy

October 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Chemistry

Thesis for the degree of Doctor of Philosophy

An investigation into an unusual glycan branching enzyme from *Mycobacterium tuberculosis*

Jessica Sophie Gusthart

Mycobacterium tuberculosis (Mtb), the pathogen that causes tuberculosis (TB), is showing increasing resistance to current drug therapies. It is therefore essential that novel drug targets and drug therapies are discovered. Methylglucose lipopolysaccharides (MGLPS) are glycoconjugates produced by *Mycobacteria* identified as essential for viability. The genes involved in the biosynthesis of MGLPs have been shown to be essential for the survival of the bacterium. Initial stages of the biosynthetic pathway to MGLPs are thought to include the transfer of a glucose unit to glucosyl glycerate (GG), forming di glucosyl glycerate. The gene responsible for this reaction has been identified as *Rv3031* and is thought to encode a GH57 glucan branching enzyme (GBE). Known enzymes belonging to this family act upon large substrates, however, the proposed biosynthetic pathway to MGLPs implies that the MtGBE works on much smaller carbohydrate units.

The full chemical synthesis of several carbohydrate compounds was performed, including the synthesis of GG and a novel synthetic route to maltosyl glycerate (MalG). Compounds were characterised using NMR spectroscopy and mass spectrometry and GG and MalG were tested in enzymatic assays designed to identify the substrates of MtGBE and MtGlcT.

The co-expression of MtGBE with the chaperone system GroEL-GroES is described, as well as the purification of the protein by Ni-NTA affinity column. Various assay techniques were utilised to investigate the nature of the activity of MtGBE and to identify the glycan donor. The glycosyl transferase encoded for by *Rv3032* was expressed and isolated from inclusion bodies to investigate potential co-operation between this and MtGBE. Though no hydrolysis or branching activity was observed and no donors identified, this is the first time MtGBE protein has been expressed and studied. This work can rule out certain types of activity thus clearing the way to discovering the true nature of the MtGBE enzyme.

Abstract

Contents

List of Figures

List of Tables

Accompanying materials

A CD containing full spectroscopic data (Extended Appendix A) accompanies this thesis

DECLARATION OF AUTHORSHIP

I, Jessica Sophie Gusthart, declare that this thesis entitled 'An Investigation into an unusual glycan branching enzyme from *Mycobacterium tuberculosis*' and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed: ……

Date: ……

Declaration

Acknowledgements

First and foremost, my thanks go to my supervisor Dr Seung-Seo Lee for all your guidance, patience and support throughout my PhD. This was an amazing opportunity that has helped me grow not just as a researcher but also as a person. Thanks also to EPSRC for the funding.

I feel I could not have completed my work without my amazing colleagues and lab-mates.

Rachael, Alice and Josh were invaluable in my venture into microbiology and biochemistry. They have also helped me remain sane throughout the process with their amazing friendships.

I also thank Clement, Julien and JB all of whom not only became close friends but also mentors throughout my project and I count myself lucky to have shared a lab with them all.

I couldn't have asked for better group members than Pushpak and Gu. Two very special people I have greatly benefitted from knowing. Not just fountains of knowledge but providers of laughs, fun and a few drinks now and then! Thanks for becoming my chemistry brothers!

Thank you also to Profs Peter Roach and Bruno Linclau for helping me become a better scientist and person. Special thanks to Prof George Attard who became an invaluable mentor to me – whether intentional or not! Thank you for believing in me when I failed to do so.

For Mum and Dad, I am forever grateful. Your love and support has been more helpful than you can imagine. Your financial support has also been a huge help! I am so lucky to have such wonderful human beings as parents. I am also grateful for my two Brothers for always letting me know how proud they are of me and for always having my back!

I was also lucky enough to meet Andy, a very special person whose patience and understanding has helped me through the process of writing a thesis.

There are many others I would like to thank for their friendships which have made my time at Southampton so wonderful. Especially Lucy, Chris, Cam, Lisa and Rob. Friends I hope to have for life.

A special thank-you to Southern Communications Ltd for being so flexible and providing me with a job whilst I wrote this thesis. I'm not sure how many companies would have been so understanding and its been a pleasure each time you've let me back.

Finally, for providing the tea that fuelled me each day, thank you to Anne and Jenny. Two people who are vitally important to the Chemistry department!

List of Abbreviations

2x YT	Yeast Extract Tryptone
3-PGA	3-phosphoglyceric acid
5FGlcF	5-fluoro-α-D-glucopyranosyl fluoride
ACN	Acetonitrile
ACP	Acyl carrier protein
ADP	Adenosine diphosphate
AG	Arabinogalactan
AIBN	Azobisisobutyronitrile
Ala	Alanine
APS	Ammonium persulfate
Ara*f*	Arabinofuranose
Asp	Aspartic acid
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
CAZy	Carbohydrate Active enZyme
CD	Circular dichroism
Chlor	Chloramphenicol
COSY	COrrelated SpectroscopY
CSR	Conserved sequence region
DAST	Diethylaminosulphur trifluoride
DCM	Dichloromethane
DEPT	Distortionless nhancement by polarization transfer
DGG	Diglucosy lglycerate
DMAP	Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPA	Decaprenyl-monophosphoryl-D-arabinose
E. coli	*Escherichia coli*
EA	Ethyl acetate
ESI+/-	Electrospray ionisation +/-
ETB	Ethambutol
FPLC	Fast protein liquid chromatography
FQ	Fluoroquinolone
G1P	Glucose-1-phophate
G6P	Glucose-6-phosphate
Gal*f*	Galactofuranose
GalN	Galactosamine
GBE	Glycan branching enzyme
GDP	Guanosine diphosphate
GG	Glucosyl glycerate
GgS	Glucosyl glycerate synthase

GH	Glycosyl hydrolase
Glc	Glucose
GlcNAc	*N*-acetylglucosamine
Glc*p*	Glucopyranose
GlcT	Glycosyl Transferase
Glu	Glutamic acid
GPG	Glucosyl-3-phosphoglyerate
GpgP	Glucosyl-3-phosphoglyerate phosphatase
GpgS	Glucosyl-3-phosphoglycerate synthase
GT	Glycosyl Transferase
H. subflava	*Hoyosella subflava*
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
His6-Tag	Hexa-histidine tag
HIV	Human immunodeficiency virus
HMBC	Heteronuclear Multiple-Bond Correlation spectroscopy
HPLC	High performance liquid chromatography
HR-MS	High resolution mass spectrometry
HSQC	Heteronuclear single quantum coherence spectroscopy
IBs	Inclusion bodies
IPA	*iso*-propyl alcohol
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IZD	Isoniazid
Kan	Kanamycin
LAM	Lipoarabinomannan
L-Ara	L-Arabinose
LB	Luria broth
LDH	Lactate dehydrogenase
LM	Lipomannan
M. bovis	*Mycobacterium bovis*
M. burtonii	*Methanococcoides burtonii*
M. smeg	*Mycobacterium smegmatis*
M.tb	*Mycobacterium tuberculosis*
MA	Mycolic acids
mAG	Mycolyl arabinogalactan
MALDI	Matrix-assisted laser desorption/ionization
MalG	Maltosyl glycerate
Man*p*	Mannopyranose
MDR	Multi drug resistant
MES	2-(*N*-morpholino)ethanesulfonic acid
MGLP	Methylglucose lipopolysaccharide
MMP	3-*O*-methylmannose polysaccharides
MS	Mass spectrometry
MtGBE	*Mycobacterium tuberculosis* glycan branching enzyme
MTX	5-deoxy-5-methylthio-xylofuranose
MurNAc	*N*-acetylmuramic acid

MurNGlyc	*N*-glycolylmuramic
MWCO	Molecular weight cut off
NAD+	Nicotinamide adenine dinucleotide (oxidised)
NADH	Nicotinamide adenine dinucleotide (reduced)
NADPH	Nicotinamide adenine dinucleotide phosphate
NBS	*N*-bromosuccinimide
NDP	Nucleotide diphosphate
NGP	Neighbouring group participation
Ni-NTA	Nickel- nitrilotriacetic acid
NIS	*N*-iodosuccinimide
NMR	Nuclear magnetic resonance
O.D	Optical density
ORF	Open reading frame
P. marina	*Persephonella marina*
PBS	Phosphate buffered saline
PCR	Polymerise chain reaction
PEP	Phosphoenolpyruvate
PG	Peptidoglycan
PGL	Phenolic glycolipid
PGL-tb	Phenolic glycolipid from *Mycobacterium tubreculosis*
*p*HBAD	*p*-hydroxybenzoic acid derivatives
Pi	Inorganic phosphate
PIM	Phosphatidylinositol mannosides
PK	Phosphate kinase
*p*NP	*p*-nitrophenol
*p*NPGlc	*p*-nitrophenyl glucose
PYR	Pyrazinamide
R. qingshengii	*Rhodococcus qingshengii*
Rha*p*	Rhamnopyranose
RIF	Rifampicin
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RT	Room temperature
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SKKU	Sung Kyun Kwan University
S_N2	Nucleophilic substitution
SOC	Super optimal broth with catabolite repression
STR	Streptomycin
Strep. Sp.	*Streptomyces sp. AA4*
T. kodakarensis	*Thermococcus kodakaraensis* KOD1
T. thermophilius	*Thermus thermophilius*
TAE	Tris-base, acetic acid and EDTA
TB	Tuberculosis
TBAF	Tetra-*N*-butylammonium fluoride

TBDPS	*tert*-butyl(diphenyl)silane
TDM	Trehalose dimycolate
TDP	Thymidine diphosphate
TEMED	Tetramethylethylenediamine
TF	Trigger factor
TfOH	Trifluoromethanesulphonic acid
THF	Tetrahydrofuran
TLC	Thin layer chromatography
Tm	Melting temperature
TMM	Trehalose monomycolate
TMS-OTf	Trimethylsilyl trifluoromethanesulphonate
TOF	Time of flight
TPS	Trehalose-phosphate synthase
TraSH	Transposon site hybridization
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer ribonucleic acid
Trp	Tryptophan
Tyr	Tyrosine
UDP	Uridine diphosphate
UTP	Uridine triphosphate
UV	Ultra violet
WHO	World Health organisation
XDR	Extensively drug-resistant

1. Introduction

1.1 Tuberculosis

Tuberculosis (TB) is an ancient disease, depicted by the early Egyptians, described by Hippocrates (using the name phthisis) and is even mentioned in the bible (using the ancient Hebrew word schachepheth).[1] By the 19th Century, Europe was suffering a TB epidemic, during which time the disease was studied in depth by renowned physicians such as René Laennec and Robert Koch, but the disease was eventually seen as curable following the introduction of antibiotics starting with isoniazid in 1952.[1-3] Despite a period of time during which it was thought TB no longer posed a major threat[4], in recent decades the disease has returned, becoming one of the top 10 causes of death worldwide.[5-7] Indeed, TB is now such an issue that the World Health Organisation (WHO) have launched the 'End TB strategy' which aims to reduce TB deaths by 95% (in comparison to 2015) and TB incidence rates by 90% by 2035.[8]

Caused by the bacterial agent *Mycobacterium tuberculosis* (*M.tb*), TB mainly affects the lungs (pulmonary TB) but can also infect other areas in the body (extrapulmonary TB).[6] *M.tb* is a gram positive bacterium with a cell wall consisting of a two layers (see Figure 1. 1), the innermost of which (the cell wall core) is composed of peptidoglycan (PG) linked to arabinogalactan (AG) and this is then linked to mycolic acids.[9, 10]

The outer layer contains free lipids, amongst which are scattered the cell wall proteins, and it is this outer layer which is believed to play a role in signalling interactions, whilst the cell wall core is crucial in maintaining cell viability.[10] Once in the lungs, *M.tb* is phagocytosed into macrophages, however the bacterium has developed the ability to survive inside the macrophage, avoiding degradation.[11] Further immune response then leads to the formation of granulomas, in which the bacterium can survive in a dormant state for many years.[12]

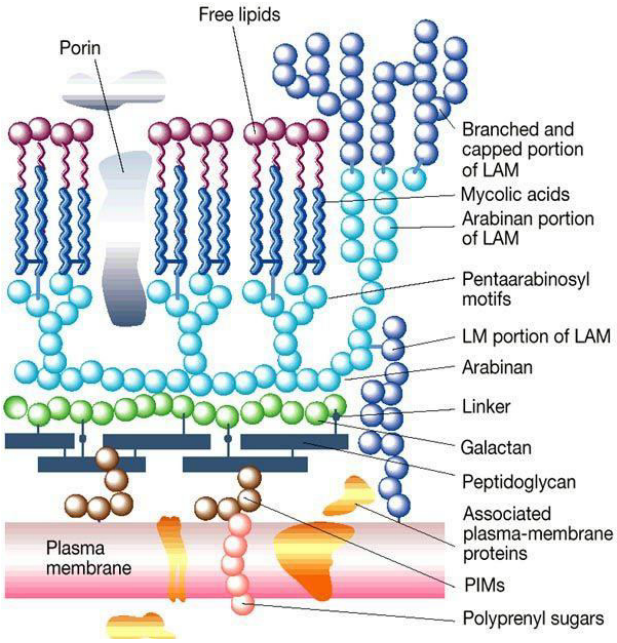


Figure 1. 1 Schematic of the Mycobacterial tuberculosis cell envelope, copied, with permissions, from reference 11. LAM = lipoarabinomannans, LM = lipomannans, PIM = phosphatidylinositol mannosides.[13]

The first line of treatment currently used to combat *M.tb* infections (summarised in Figure 1. 2) consists of a six-monthly course of four antibiotics: isoniazid (IZD), rifampicin (RIF), pyrazinamide (PYR) and either streptomycin (STR) or ethambutol (ETB).[5, 14] These drugs not only have to be taken for a long period of time but also can cause severe, undesirable side effects in patients.[14] Resistance of *M.tb* to two of the key drugs in first line treatment, isoniazid and rifampicin, is termed multidrug-resistant (MDR) TB and the term extensively drug-resistant (XDR) TB is used to describe MDR-TB which is also resistant to any fluoroquinolone (FQ) and any of the following three compounds: capreomycin, kanamycin or amikacin.[15] These resistant strains, along with the strong synergy between TB and the HIV virus, are a key contributing factor to the current need for new anti-TB drugs. [5-7]

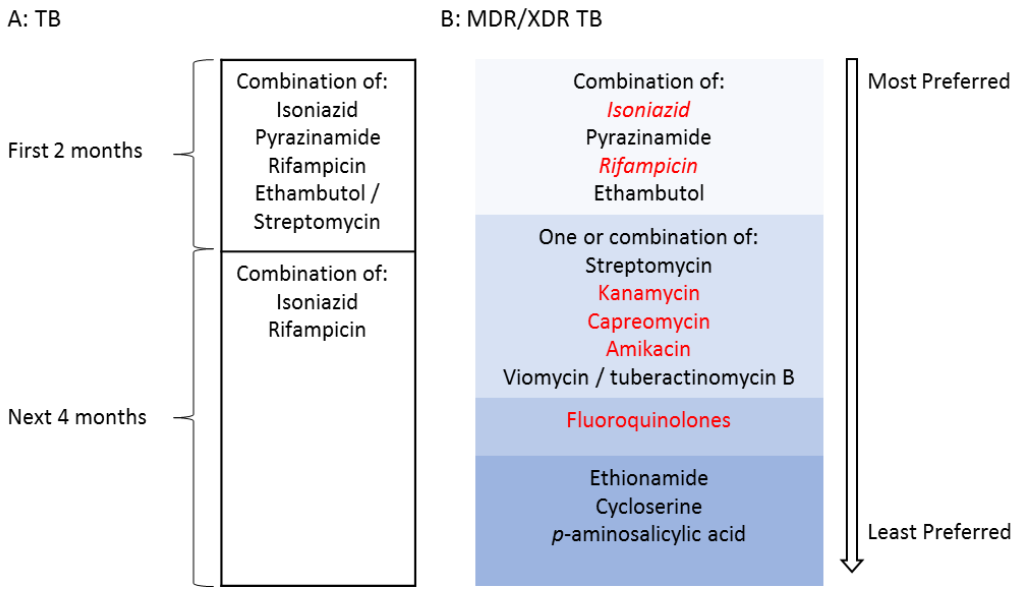


Figure 1. 2 A: Course of treatment for TB infections; B: Options for treating MDR/XDR TB cases in order of preference, italics = drugs MDR TB is resistant to, red = drugs XDR TB is resistant to. Figure created using information from Janin.[14]

The aforementioned components of the cell wall of *M.tb* consist of lipids[16], proteins and carbohydrates.[10] Carbohydrates play an important role not just in the cell wall, but in other biological functions of *M.tb* such as survival of the cell inside the host[17], metabolism[18] and virulence.[19] A selection of important carbohydrates found in *Mycobacteria* will be discussed in the following text.

1.1.1 Key anti-TB drugs

Current drug targets of anti TB therapies (depicted in Figure 1. 3) include biosynthetic pathways involved in cell wall synthesis, translation and transcription, amongst others.[14, 20-24]

IZD (**1.1**) and its structural relative ethionamide are both prodrugs, with isoniazid being activated by KatG to form adducts with NADH.[14, 25] The target of the activated compounds is believed to be InhA, an enoyl-ACP reductase involved in long chain fatty acid and mycolic acid biosynthesis.[25, 26] It has also been noted that KatG activation of isoniazid produces NO• radicals that can attack the bacterium.[27] It is believed that mutations in the *kat*G gene are the major cause of IZD resistance in MDR strains of TB.[28, 29]

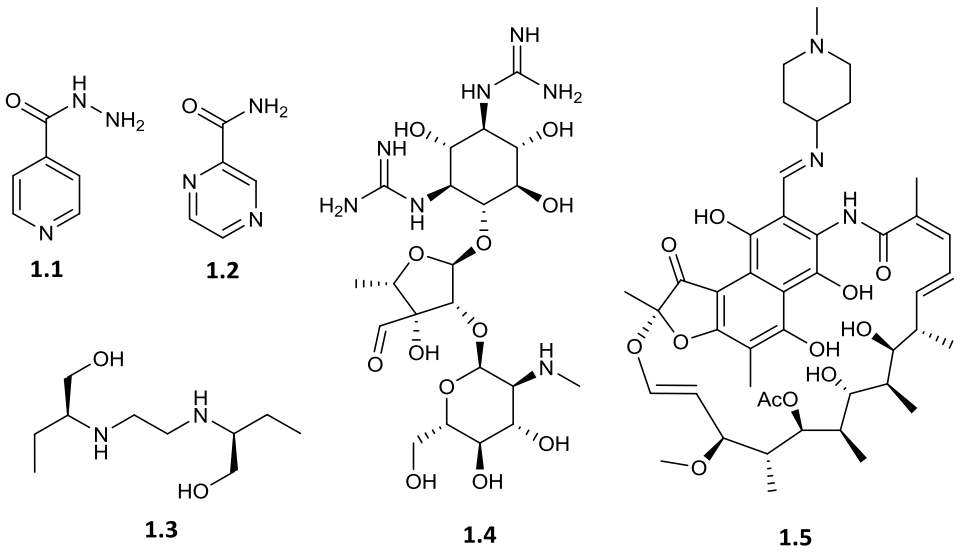


Figure 1. 3 Structures of key anti-TB drugs. Reproduced from Janin.[14]

PYR (**1.2**), similarly to isoniazid, is also a prodrug that is hydrolysed to an active form, pyrazinoic acid, which was thought to work by causing an increase in cellular acidity resulting in cell death.[14] Recent studies suggest that instead, there are two mechanisms of action: decrease in the essential cofactor acetyl CoA and inhibited biosynthesis of a virulence factor.[21] There aren't currently any molecular mechanisms of action of PYR and this is still a heavily studied area.[30] The enzyme that converts pyrazinamide to its active form within TB is known as PncA and is encoded for by the *pncA* gene.[31] It is the presence of mutations within this gene that cause most cases of PYR resistance[32] but mutations in other genes have also been observed in resistant strains.[33, 34]

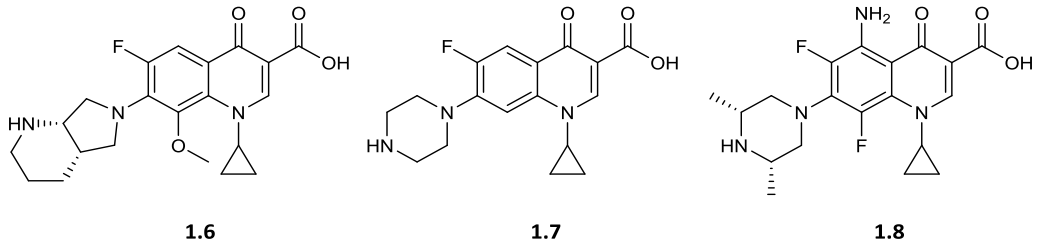
The targets of ETB (**1.3**) are the arabinosyl transferases involved in arabinan biosynthesis.[22, 35] Arabinan is an important component of arabinogalactan which in turn is a key component of the bacterial cell wall which, as previously discussed, is vital for cell viability.[22] It is within the gene cluster *embCAB,* containing the genes for the relevant arabinosyl transferases, in which mutations leading to ETB resistance are found.[36, 37]

STR (**1.4**) is classed as an aminoglycoside and belongs to the family of drugs with that name.[14] It targets translation, weakening the proof reading ability of the ribosome as well as lowering the accuracy with which tRNA binds, thus leading to an increase in

error during protein synthesis.[23] Kanamycin, amikacin, capreomycin and viomycin also target translation.[14, 38] Resistance to STR is mainly caused by mutations in the *rpsL* and *rrs* genes, which encode the S12 protein and 16S rRNA respectively.[39]

RIF (**1.5**), a large polyketide containing a napthoquinone core, targets the β sub-unit of bacterial RNA polymerase to which it binds, inhibiting RNA transcription by blocking RNA chain elongation at the 5' end.[20] The β sub-unit of RNA polymerase, is encoded for by the gene *rpoB*, and it is in this gene that 95% of mutations leading to RIF resistance are located.[40]

FQs, a large family of drugs that includes moxifloxacin (**1.6**), ciprofloxacin (**1.7**) and sparfloxacin (**1.8**) (see Figure 1. 4), target DNA gyrase, which is essential for effective replication, recombination and transcription of DNA.[24]



1.6 **1.7** **1.8**

Figure 1. 4 Chemical structure of some of the members of the fluoroquinolone family, recreated fromJanin.[14]

M.tb DNA gyrase consists of two sub units encoded by the genes *gyrA* and *gyrB*.[41] It is due to mutations in these genes that confer resistance to FQs in *M.tb* and it was found that these mutations can interact with each other to cause differing levels of resistance.[42]

Some of the above mentioned anti TB drugs either target carbohydrate biosynthetic pathways (for example, ETB) or contain carbohydrates in their structure (e.g. STR) and it is the biosynthesis of a carbohydrate containing molecule that is the focus of this study. The following section will therefore discuss carbohydrates, their structure, importance in nature and key chemical features.

1.2 Introduction to carbohydrates

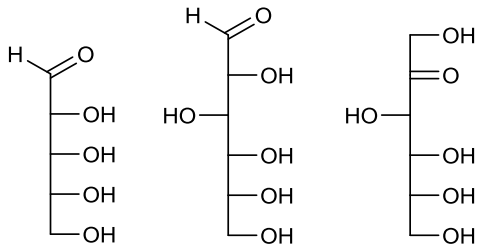
Ubiquitous throughout nature, carbohydrates are essential molecules in biological compounds and processes including; DNA and RNA[43], cell signalling[44] and recognition[45, 46], generation of energy in metabolism[47], structural components of cells[48, 49], reproduction[50], the immune system,[51] including the inflammatory response[52], and much more.[53] Carbohydrates are diverse in ring size, hydroxyl group modifications and anomeric configuration and, in addition, carbohydrate monomers can be linked in a variety of ways, all leading to an extremely high number of possible carbohydrate structures.[54]

Carbohydrates are important components in many diseases and infectious agents such as in diabetes[47], cancer[55-57], viral infections and bacterial infections[58, 59]. In bacteria, carbohydrates are involved not just in metabolism[60] and infection but are also essential cell wall components[61, 62] and are involved in sensing changes in the external environment to respond to stress.[63] In fact, carbohydrate binding proteins are often drug targets due to the sheer importance of carbohydrates in the survival of the bacterium.[64, 65]

Carbohydrates are already used for important medical applications such as therapeutic antibodies[66], anti-cancer and anti-bacterial drugs[67, 68] and vaccines[69-72] . However, due to their complex structures and only a more recent understanding of glycobiological processes, it is felt that the area of carbohydrates as drugs has not been given as much attention as other families of molecules.[73]

1.2.1 Basic structural properties of carbohydrates

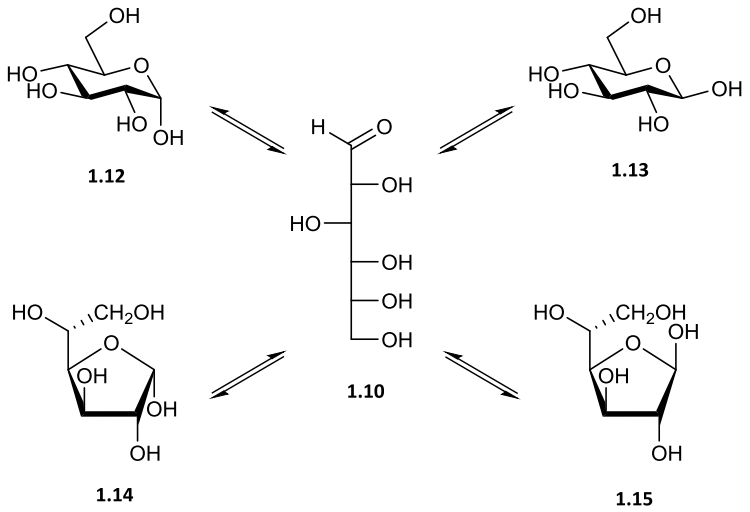
Carbohydrates are organic molecules containing carbon, oxygen and hydrogen, generally in the ratio of $C_n(H_2O)_n$.[74] They contain many hydroxyl groups and can be described by taking into account a few structural properties, for example; sugars containing 4 carbon atoms are termed tetroses, those with 5 carbon atoms in the chain are described as pentoses, 6 carbon atoms and it is a hexose (see Figure 1. 5).[75]



1.9 **1.10** **1.11**

Figure 1. 5 Examples of various classes of sugars; ribose (**1.9**) is a pentose, glucose (**1.10**) is a hexose and fructose (**1.11**) is a ketose (ribose and glucose are aldoses). Recreated from[75]

In addition, sugars in solution can exist in different forms, based either on tetrahydropyran (pyranose form), tetrahydrofuran (furanose form) or in an acyclic, straight chain form.[76] These forms are connected and are in chemical equilibrium with each other (see Figure 1. 6).[77] When in the solid state, however, the furanose or pyranose structures are permanent.[77] Depending on the functional group present in the straight chain structure, a ketone or an aldehyde, sugars can be further classified as ketoses or aldoses respectively (see Figure 1. 5).[75]



1.12 **1.13** **1.10** **1.14** **1.15**

Figure 1. 6 Furanose (**1.14** and **1.15**) and pyranose (**1.12** and **1.13**) forms of D-glucose, recreated from[76]

The stereochemistry of the hydroxyl groups is also key feature in sugars. When in one of the two ring configurations, a hemi-acetal is formed and it is the carbon atom of this hemi-acetal that is termed the anomeric position and labelled position 1 when numbering the ring.[75] The stereochemistry at the anomeric centre is termed as either

α or β (axial or equatorial respectively), and the stereochemistry of the chiral carbon centre furthest from the anomeric position (in the linear form) determines the absolute configuration of the sugar molecule, denoted as either D- , the most common configuration in natural sugars or L-, found to a lesser extent throughout nature (see Figure 1. 7) .[75]

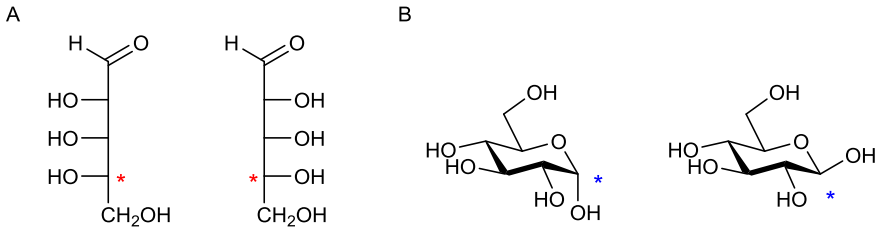


Figure 1. 7 A. the L- (**1.16**) and D- (**1.9**) forms of ribose with the centre at which absolute configuration is determined highlighted (*); B. the α (**1.12**) and β (**1.13**) forms of glucopyranose with the anomeric centre highlighted (*).

The stereochemistry of the other hydroxyl groups (those that are not the anomeric) determines the name of the sugar in question. For example, in the pyranose form of glucose, the hydroxyl group at position 4 is in an equatorial configuration, however if this same hydroxyl group is in an axial position, the sugar is now galactose. This is known as epimerisation and galactose (**1.17**) is an epimer of glucose (**1.12**) at position 4 (see Figure 1. 8).[75]

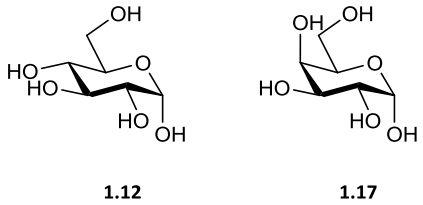


Figure 1. 8 Galactose is an epimer of glucose at position 4.

1.2.2 The anomeric effect

One important feature of carbohydrates is the anomeric effect. This is an observation that an electronegative anomeric substituent in a pyranose ring prefers to be in the axial position rather than the less sterically hindered equatorial position when in

equilibrium.[78-80] The cause of this observation is still studied today, though two widely accepted theories are discussed in the following text.[81]

One explanation for the anomeric effect is stabilisation by hyperconjugation depicted in Figure 1. 9. It is thought that in the axial position, the σ* orbital of the C-X bond of the anomeric substituent (X) is aligned with the lone pair of the endocyclic oxygen, allowing for electron donation and the lowering of the overall energy of the system.[77] In the beta position, the σ* orbital of the substituent would not be in alignment with the lone pair on the endocyclic oxygen, meaning no stabilisation could take place.[77]

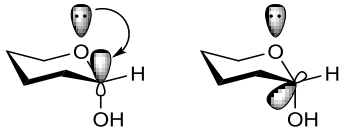


Figure 1. 9 Proposed role of hyperconjugation in the anomeric effect. Recreated from[82]

Another theory, illustrated in Figure 1. 10, is that the dipole moment of an axial substituent would almost oppose that of the neighbouring endocyclic oxygen, stabilising the configuration.[77] Were the anomeric substituent in the beta position, however, the dipole moments would be close to alignment and therefore would repel each other, destabilising the system.[77]

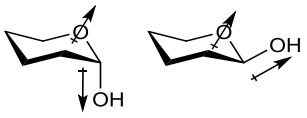


Figure 1. 10 The proposed role of dipole moments in the anomeric effect. Recreated from[83]

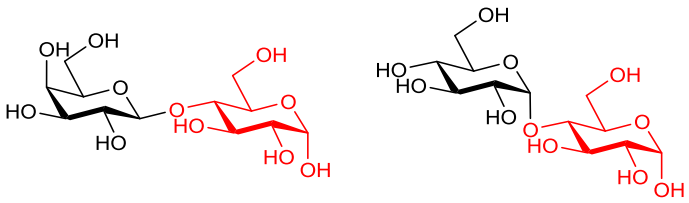
In nature, the configuration at the anomeric centre is a key feature that plays a part in enzyme recognition, with carbohydrate active enzymes being selective of a particular anomer. For example, α-glucosidase only works on α(1→4) bonds, like those found in maltose (see figure 1.11).[84] Therefore, when synthesising substrates for enzymes, it is important to obtain the correct anomer in a pure form. Ways of controlling the anomeric stereochemistry in synthesis include taking into consideration effects from

solvent, how a glycosyl donor is activated, steric hindrance, metal ion coordination and many other effects.[85] Some of these will be discussed in the proceeding chapters.

1.2.3 Disaccharides, oligosaccharides and polysaccharides

Sugars are able to link together *via* glycosidic bonds and when two simple sugars form a glycosidic bond, a disaccharide is formed.[86] Chains containing a few sugar monomers are known as oligosaccharides and really long chains are termed polysaccharides. All forms have many uses throughout nature, for example, oligosaccharides are often found in glycoproteins[87] and polysaccharides can be used to store glucose, as in the case of glycogen or starch.[88]

Disaccharides can be linked in a number of ways (see Figure 1. 11), depending on the configuration of the anomeric position (α or β), through which hydroxyl groups they are linked (for example 1→4 or 1→6) or through which atom they are linked (O-, N-, S- or C- linked etc.).[89] The end of the chain where the anomeric position is still free is termed the reducing end, whilst the other end of the chain is the non-reducing end. Enzymes that act upon disaccharides, oligosaccharides or polysaccharides can be highly selective for the type of glycosidic bond they act upon[76], including the position of the bond within the chain, for example, if it is in the middle of the chain or at the end.[89]



1.18 **1.19**

Figure 1. 11 Example of a β(1→4) linked (**1.18**) and an α(1→4) linked (**1.19**) disaccharide. Reducing end sugars highlighted in red.

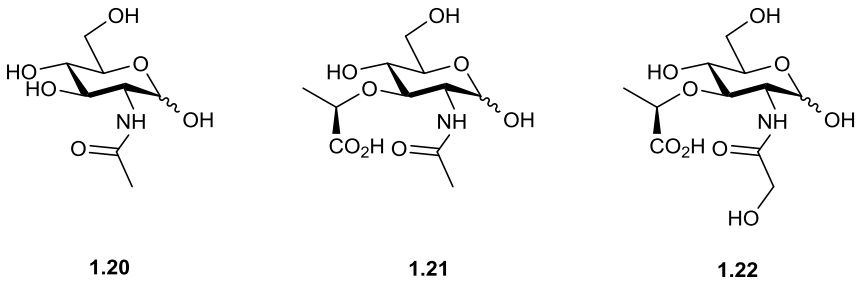
In bacteria, polysaccharides play an important role in survival, for example peptidoglycan, the main component of the cell wall.[62] Other polysaccharides, found in *Mycobacteria*, deemed essential for bacterial cell survival are 6-*O*-methylglucose lipopolysaccharides (MGLPs) and 3-*O*-methylmannose polysaccharides (MMPs).[90]

These polysaccharides were discovered in the 1960's but are still the subject of contemporary research which will be discussed in the following sections.[91, 92]

1.3 Carbohydrates in *Mycobacteria*

1.3.1 The cell wall

One of the major components of all bacterial cell walls, peptidoglycan (PG) consists of a carbohydrate backbone which is cross-linked by peptide chains.[93, 94] Generally made up of *N*-acetylglucosamine (GlcNAc) (**1.20,** Figure 1. 12) β(1→4) linked to *N*-acetylmuramic acid (MurNAc) (**1.21**)[94], the sugar backbone of PG is slightly different in *Mycobacteria*; rather than being *N*-acetylated, the muramic acid units are *N*-glycolylated resulting in *N*-glycolylmuramic acid (MurNGlyc) (**1.22**).[95]



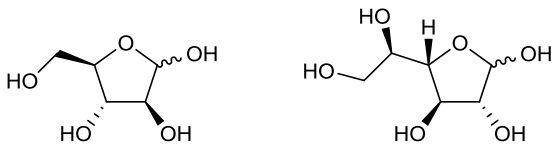
1.20 **1.21** **1.22**

Figure 1. 12 The structures of GlcNAc (**1.20**), MurNAc (**1.21**) and MurNGlyc (**1.22**), components of peptidoglycan. Adapted from Warth *et al.*[96]

Biosynthesis of PG begins with the addition of phosphoenolpyruvate (PEP) to the 3-OH position of UDP-GlcNAc a reaction catalysed by the MurA enzyme (encoded for by *Rv1315* in *M.tb*).[93, 97] The enoylpyruvyl group of this product is then reduced using NADPH by the enzyme MurB (*Rv0482*)[93] to form UDP-MurNAc[98] before glycolylation performed by NamH leads to UDP-MurNGlyc.[95] A number of peptides are then added to UDP-MurNAc or UDP-MurNGlyc *via* the action of MurC, MurD, MurE and MurF, all ATP dependant ligases, until a muramyl pentapeptide has been formed, known as Park's nucleotide.[99, 100] MurX (*Rv2156c*), also known as MurY, attaches Park's nucleotide to the membrane acceptor undecaprenyl phosphate to form a compound known as Lipid I.[101] A GlcNAc unit is then transferred to Lipid I by the action of MurG (*Rv2153c*), an *N*-acetylglucosaminyl transferase to form Lipid II.[101] Lipid II is then transported form the cytosol to the outer side of the cell membrane, however it is not yet clear which enzyme facilitates this in *Mycobacteria*.[102] Once outside the

membrane, Lipid II is used as the monomeric building block for the assembly of the PG polymer.[103]

In *M.tb*, around 10-12% of the MurNGlyc units within PG are linked *via* a special linker unit to arabinogalactan (AG).[93, 104] This linker consists of α-L-rhamnopyranosyl-(1→3)-α-D-*N*-acetylglucosamine-1-phosphate (α-L-Rha*p*-(1→3)-α-D-GlcNAc-(1→P) which connects the galactan portion of AG with the C-6 position of the relevant MurNGlyc component.[104] AG consists of a galactan domain, composed of β-D-galactofuranoside (β-D-Gal*f*) (**1.24**, Figure 1. 13) units linked alternatively *via* the 5 or 6 positions in a linear fashion, and an arabinan domain consisting of branched chains of arabinofuranose (Ara*f*) (**1.23**).[105, 106] The inner core of this arabinan domain also contains galactosamine (GalN) and succinyl ester substituents on some Ara*f* residues.[107-109]

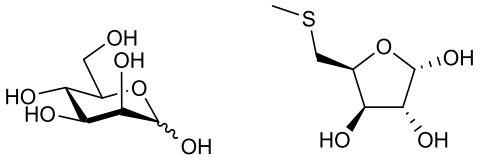


1.23 **1.24**

Figure 1. 13 Structures of Ara*f* (**1.23**) and Gal*f* (**1.24**), the main components of arabinogalactan.

As with PG, AG biosynthesis begins in the cytosolic side of the cell membrane and is completed on the outer side.[110] The linker unit is first synthesised, connected to decaprenyl phosphate, which is in turn used as the acceptor for the elongation of the Gal*f* chain.[110, 111] 30 Gal*f* units are linked together before transportation to the other side of the membrane.[112] The initiation and elongation of the galactan chain is catalysed by the galactofuranosyltransferases GlfT1 (*Rv3782*) andGlfT2 (*Rv3808c*) respectively, each of which utilise the donor UDP-Gal*f*.[113, 114] Once transported, the arabinan domain is built beginning with the transfer of Ara*f* units onto the galactan-linker section.[115] The Ara*f* donor for this is decaprenyl-monophosphoryl-D-arabinose (DPA) which is used a substrate by an undetermined number of Ara*f* transferases.[115] Branching occurs at α(1→3) branch points, the formation of which in *Mycobacterium smegmatis* (*M.smeg*) was found to be catalysed by the branching enzyme AftC (*Rv2673*).[116]

Arabinan is also a constituent of another cell wall component, lipoarabinomannan (LAM).[117] This consists of mannopyranose units joined *via* α(1→6) linkages forming a backbone which is interspersed with branches of single mannopyranoside (Man*p*) units (**1.25**, Figure 1. 14).[118] Further glycosylation of the mannan component occurs as a linear Ara*f* chain is attached.[119, 120] This is then branched with either a linear tetra-arabinoside or a branched hexa-arabinoside.[121-123] These are then finally capped with, in the case of *M.tb*, two α(1→2) linked Man*p* units forming ManLAM.[124] It has also been observed that *M.tb* ManLAM also contains a unique 5-deoxy-5-methylthio-xylofuranose (MTX) (**1.26**) moiety linked to the Man cap.[125, 126] LAM and its precursors are thought to play a role in the modulation of the host immune response. [127-130]



1.25 **1.26**

Figure 1. 14 Structures of Man*p*, the monomeric unit of mannan, and MTX, the unique capping glycan found in *M.tb*. Reproduced from Turnbull *et al*.[131]

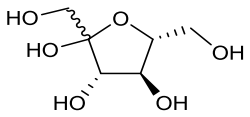
The bacterial cell wall of *Mycobacteria* is surrounded by an outer capsule, the main component of which is a glycogen-like α-glucan.[132, 133] This α-glucan is comprised of an α(1→4) linked α-D-glucopyranose (Glc*p*) core which contains branches of α(1→4) linked Glc*p* at regular α(1→6) branch points.[134] The capsule is believed to play a role in the survival of *M.tb* within the macrophage, as well as in the modulation of the host immune response.[135-137] The classical pathway to α-glucan is the GlgC-GlgA pathway.[134, 138] This pathway begins with the formation of a glucose nucleotide (NDP-Glc) from glucose-1-phosphate (G1P), catalysed by GlgC, a nucleotide diphosphoglucose pyrophosphorylase.[138] This then undergoes polymerisation by GlcA then subsequent branching which is performed by GlgB.[138] An alternative pathway to glycogen and α-glucan biosynthesis, the GlgE-GlgB pathway, has been more recently described.[139] This pathway involves trealose synthase and maltose kinase which generates maltose-1-phosphate from trehalose.[134] Maltose-1-phosphate then serves

as the substrate used in polymerisation of the glucan chain which is elongated by GlgE.[134, 140] As with the classical pathway, it is believed that branch points are then introduced by GlgB.[134]

1.3.2 Metabolism

It has been shown that *M.smeg* can utilise many sugars as carbon sources including both hexoses and pentoses.[141, 142] In alignment with this is the identification of many carbohydrate transporters within the *M.smeg* cell membrane.[143] In contrast, only five carbohydrate transporters have been identified in *M.tb* compared to the twenty-eight identified in *M.smeg*, thus suggesting a lack of diversity in carbohydrate nutrients within the macrophage where *M.tb* resides during infection.[143]

Studies have identified that *M.tb* will consume either galactose, glucose, arabinose or fructose (**1.27**, Figure 1. 15) when these are present in growth media *in vitro*.[18] Though it has been reported that *M.tb* mostly utilises fatty acids as a carbon source,[143, 144] the disruption of glycolysis by knocking out pyruvate kinase prevents growth of *M.tb* when grown on fermentable carbon sources.[145] This suggests that glycolysis is an important metabolic pathway in *M.tb*.[145] In addition, the phosphorylation of glucose has been shown as important for *M.tb* growth during mice infections suggesting *M.tb* has access to glucose during infection.[146]



1.27

Figure 1. 15 Structure of fructose (**1.27**).

Interestingly, it has also been observed that *M.tb* can co-catabolise different carbon sources (glucose and glycerol) at the same time by putting each source down a separate metabolic pathway, a novel type of behaviour found in bacteria at the time.[147]

As with most bacteria, the storage of glucose in *M.tb* is facilitated by the synthesis of glycogen, a branched polymer of α(1→4) linked Glc*p* units with branches introduced at

α(1→6) linkages.[134] the biosynthesis of glycogen is discussed in the previous section (1.3.1).

1.3.3 Virulence and persistence

Mycolic acids (MAs) form a highly protective layer around the outer membrane of *M.tb*, thought to be a major factor in the resistance of the bacterium to toxic compounds such as antibiotics.[143, 148] In the biosynthesis of MA, an important role is played by trehalose, a disaccharide consisting of α(1→1) linked glucose units.[149] Trehalose acts as a carrier for MAs as it is attached to lipids to form trehalose monomycolate (TMM) which is then transported from the cytoplasm to the periplasm.[150, 151] TMM is then used in the synthesis of mycolyl-AG and trehalose dimycolate (TDM) (**1.28**, Figure 1. 16), each components of the cell wall, during which trehalose is released.[152] Both mAG and TDM have been linked to the virulence and antibiotic resistance of *M.tb*.[153-155] Indeed, TDM has been identified as the most abundant lipid released by virulent *M.tb*.[152]

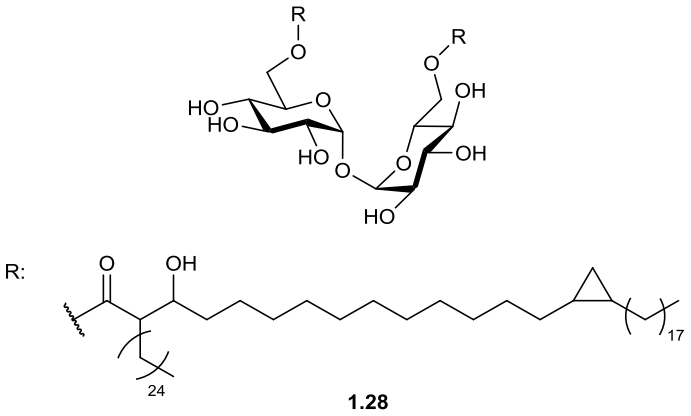


Figure 1. 16 Structure of TDM (**1.28**). Reproduced from Spargo *et al*.[156] and Glickman *et al*.[157]

In addition to its role in virulence, trehalose also takes part in other key biological functions such as acting as a carbon source and providing protection against environmental stresses such as osmotic stress or freezing.[152, 158]

Also found on the cell surface, phenolic glycolipids (PGLs) and *p*-hydroxybenzoic acid derivatives (*p*HBADs) (**1.30** and **1.31**) are thought to play vital roles in the virulence and infectivity of *M.tb* (see Figure 1. 17).[19, 159] PGLs consist of a lipid core ω-terminated by a phenolic moiety.[160, 161] The phenolic moiety is then glycosylated; in the case of *M.tb*

(PGL-tb) (**1.29**), this is with a trisaccharide composed of 2,3,4-tri-*O*-methyl-α-L-fucopyranosyl-(1→3)-α-L-rhamnopyranosyl-(1→3)-2-*O*-methyl-α-L-rhamnopyranoside.[160, 162] *p*HBADs are much smaller compounds which consist of a glycosyl unit attached to a *p*-hydroxybenzoic acid moiety.[160] The glycosyl units of *p*HBADs are either the same as that of PGL-tb or structurally related.[160]

PGL-tb is only found in a few *M.tb* strains and its exact role in these strains is unclear.[160] However, studies suggest that it enhances infectivity of *M.tb* and acts as a vital virulence factor.[159, 163, 164] In addition, evidence also suggested that PGL-tb may play a role in the hyper-virulence of clinical isolates belonging to the W. Beijing family.[161, 164, 165]

*p*HBADs are thought to have a significant impact on the host immune response by inhibiting pro-inflammatory responses in infected macrophages.[160]

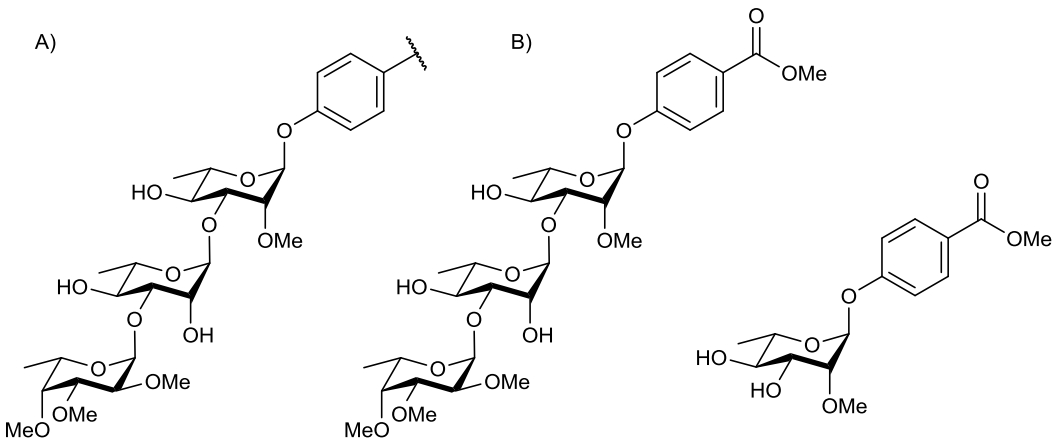
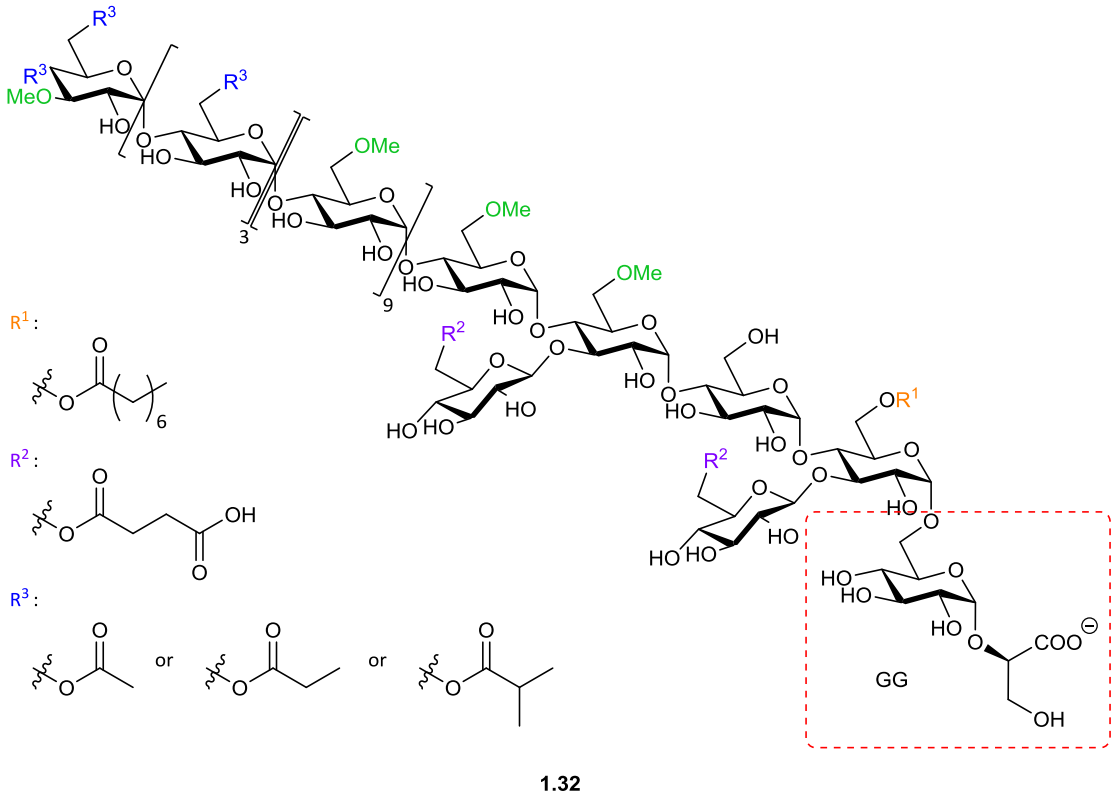


Figure 1. 17 A) the glycosylated region of PGL-tb (**1.29**) and B) Examples of *p*HBADs (**1.30** and **1.31**). Figure reproduced from Barnes *et al*.[19]

1.3.4 6-*O*-methylglucose lipopolysaccharides (MGLPs) as drug targets in *M.tb*

Thought to regulate fatty acid synthesis in *Mycobacteria*[166], MGLPs (**1.32**) are acylated polysaccharide chains consisting of 6-*O*-methylglucose α(1→4) linked units with an α(1→6) linked glucose unit at the reducing end (see Figure 1. 18).[167] At varying points in the glucose chain, 6-*O*-substituted glucose monomers are attached via β(1→3) links and the reducing end glucose contains a glycerate moiety at the anomeric position in an α configuration.[167] This reducing end sugar is termed glucosyl glycerate (GG).

Whilst early structures suggested that the carboxylic acid of GG was octanoylated, more recent studies have concluded that the octanoate group is actually attached to the second sugar in from the reducing end.[168]



1.32

Figure 1. 18 Structure of MGLPs (**1.32**) with reducing end glucosyl glycerate highlighted. Reproduced from[167, 168]

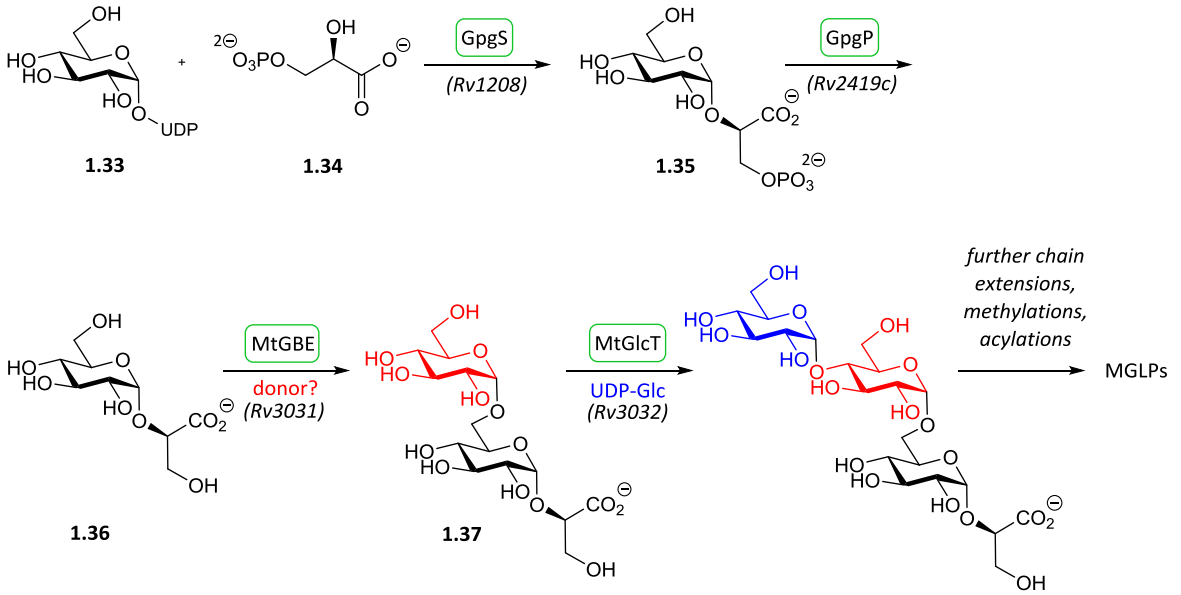
Genetic studies by Sassetti *et al*. utilised the technique of transposon site hybridization (TraSH) to produce a comprehensive list of genes essential to the growth of *Mycobacteria*.[169] Within their list of genes identified as 'required for optimal growth' resides a number of genes responsible for the biosynthesis of MGLPs, suggesting that MGLPs may be essential compounds for Mycobacterial survival.[169] This makes the biosynthesis of MGLPs a potential novel drug target for the development of new anti TB drugs and this project will focus on one of the early steps in this process.

1.4 Enzymes involved in MGLP synthesis

A model for the biosynthesis of MGLPs was first proposed in the late 80's[170] but it wasn't until 2007 that the genetic basis for the biosynthetic pathway was reported (see Scheme 1. 1).[90] By utilising the Carbohydrate-Active enZYmes (CAZy) database

(www.cazy.org), a gene cluster was identified that the authors believed was responsible for the synthesis of MGLPs. They proposed that the genes encoding the enzymes involved in MGLP biosynthesis were as follows: *Rv3032* which encodes a glucosyl transferase; *Rv3031* which encodes a GH57 family branching enzyme; *Rv3030* thought to encode an *S*-adenosylmethionine-dependant methyltransferase; *Rv3034c*, thought to encode an acetyltransferase and *Rv3037c*, believed to encode another *S*-adenosylmethionine-dependant methyltransferase.[90]

The same authors were also able to confirm that *Rv3032* encoded for an α(1→4) glucosyltransferase that was not only the main enzyme responsible for elongating MGLPs but was also involved in glycogen synthesis.[90] It was also suggested that the *S*-adenosylmethionine-dependant methyltransferase encoded for by *Rv3030* was responsible for the main, 6-*O* methylation of MGLP.[90, 171]



Scheme 1. 1 Current proposed biosynthetic pathway of MGLP synthesis. Figure reproduced from[167]

A year later, Empadinhas *et al*. purified and characterised a glucosyl-3-phosphoglycerate synthase (GpgS) encoded for by *Rv1208*.[172] The crystal structure of GpgS was also published in the same year[173, 174], followed by mechanistic studies later on.[175, 176] This enzyme is responsible for the formation of glucosyl-3-phosphoglyerate (GPG) (**1.35**) with optimal substrates *in vitro* found to be UDP-glucose (**1.33**) and 3-phosphoglyceric acid (3-PGA) (**1.34**).[172] It was suggested that GPG was a pre-cursor to

di-glucosyl glycerate (DGG) (**1.37**) which itself is a precursor to MGLP synthesis.[172] In addition, *Rv1208* was among those genes found essential for growth in *M.tb*, leading the authors to postulate that MGLP synthesis could not be initiated without the presence of GPG.[169, 172]The suggestion that GPG was vital for MGLP synthesis was given further credit by studies that found a direct link between GpgS and the synthesis of MGLPs.[177]

The next enzyme involved in MGLP biosynthesis to be characterised was the glucosyl-3-phosphoglyerate phosphatase (GpgP) encoded for by *Rv2419c* which converts GPG to GG (**1.36**) and so is thought to be involved in the early stages of MGLP synthesis.[178] Classified as belonging to the histidine phosphatase superfamily, structural and mechanistic studies led to the elucidation of a 2 step mechanism of action for GG synthesis.[179]

Although the characterisation of enzymes involved in MGLP biosynthesis is not complete, a picture of the overall process can start to be mapped.

1.4.1 *Rv3031* encodes an apparent glucan branching enzyme

The gene *Rv3031* was not only identified as being essential for growth of *M.tb*[169], but was also found to reside in the gene cluster associated with MGLP biosynthesis.[90] It is believed that *Rv3031* encodes an enzyme that catalyses the formation of di-glucosyl glycerate (DGG) starting from GG and an unknown donor.[90, 167, 171]

Sequence analysis and alignment shows strong sequence homology with branching enzymes that are also members of the CAZy family GH57 (those that have been characterised include those isolated from *Thermococcus kodakaraensis* KOD1[180], *Thermus thermophilus*[181] and, more recently reported, *Pyrococcus horikoshii*[182]). This would suggest that *Rv3031* could also encode a branching enzyme belonging to the GH57 family. However, whilst these sequence-similar proteins all act upon large, lengthy polysaccharide chains,[180-182] it would appear that, given which step it is thought to catalyse in the MGLP pathway, the protein encoded for by *Rv3031* actually acts upon much smaller substrates, which is unusual for a branching enzyme.

Characterisation of this enzyme could not only lead to further understanding of the biosynthetic pathway of MGLPs in *Mycobacteria*, but might also reveal novel branching enzyme behaviour.

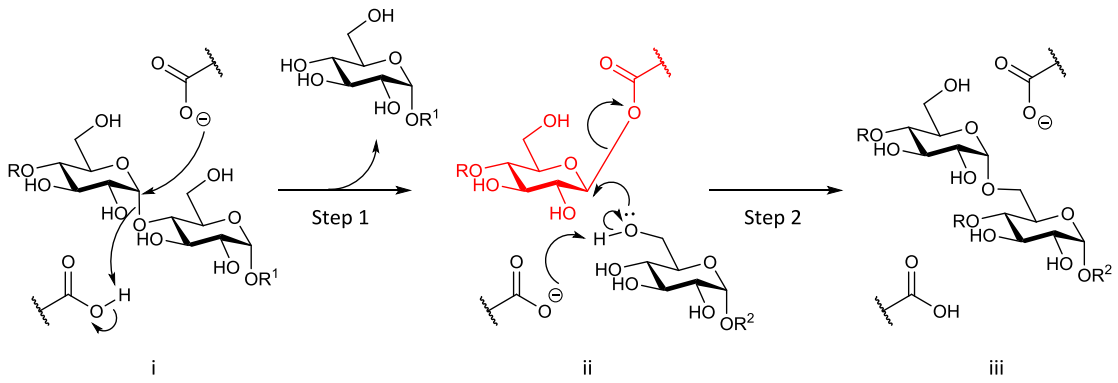
1.5 Project Aim

The primary aim of this project is to characterise the enzyme encoded for by *Rv3031*, suspected of being a glucan branching enzyme (GBE) and hence referred to as MtGBE. Characterisation would include identifying substrates, activity and eventually optimal conditions for assays leading to kinetic characterisation.

A secondary objective is to identify any inhibitors of MtGBE from which novel anti TB therapeutics could be developed.

1.5.1 Mechanistic hypotheses

Branching enzymes undergo a retaining 2-step transferase mechanism, shown in Scheme 1. 2, in which the α(1→4) glucan substrate is cleaved, forming a covalently bound enzyme-substrate intermediate where the reducing end portion of the cleaved glucan is bound to a carboxylic acid moiety of an active site residue.[183] In the GH57 family of enzymes, this residue is a glutamic acid (Glu).[184] In the following step, the enzyme bound chain is transferred to the hydroxyl group in the 6 position of either the same glucan chain (intrachain transfer) or another glucan chain (interchain transfer), creating a new α(1→6) branch point.[185] Should a water molecule, rather than the 6-hydroxy group, cleave the covalent enzyme-substrate bond then the result is a cleaved α(1→4) glucan chain in a hydrolysis reaction, hence branching enzymes are classified into glycosyl hydrolase groups.[184]



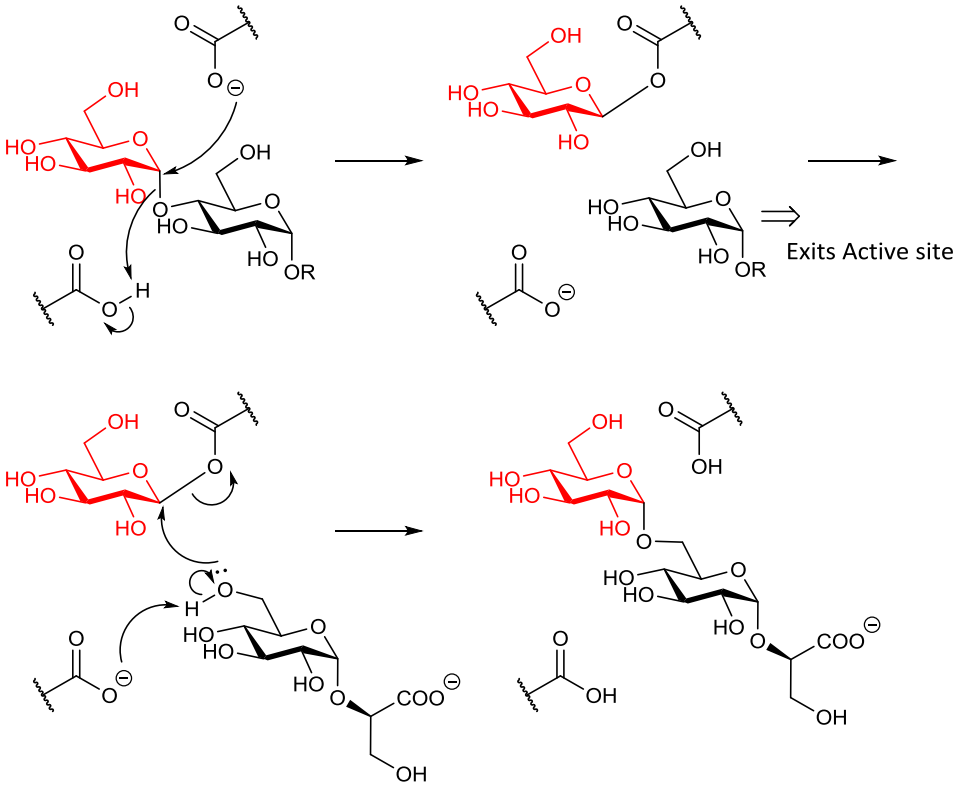
Scheme 1. 2 Retaining mechanism of glucan branching enzyme. Step 1: α(1→4) link is cleaved (i) and a covalently bound enzyme-substrate intermediate is formed (ii, highlighted in red); Step 2: 6-hydroxy group of a glucan chain attacks intermediate to form a new α(1→6)link (iii). R,R^1,R^2 = glucan chain. Figure reproduced from[183]

As MtGBE is a GH57 branching enzyme, it is thought that it would also undergo a retaining transferase mechanism and that one substrate is GG, given that the biosynthetic route to MGLPs suggests this molecule is made first before the chain is elongated.[167] However, as most branching enzymes have large glucan chains as their substrate, the fact that MtGBE appears to utilise small substrates is unusual.[185]

Based on the hypothesis that MtGBE follows the retaining transferase mechanism but also works on at least one small substrate (GG), two potential mechanisms have been proposed, each with a different donor, as the source of the additional glucose unit in DGG has not yet been confirmed.

1.5.1.1 Proposed mechanism A

In the first of the postulated mechanisms, herein referred to as mechanism A, the donor substrate is proposed to be an α(1→4) glucan chain (see Scheme 1. 3). It is suggested that perhaps the glucan chain will bind and form the enzyme-substrate intermediate with the non-reducing end sugar unit, with the 6-hydroxy group of GG then attacking to form the α(1→6) link of DGG.



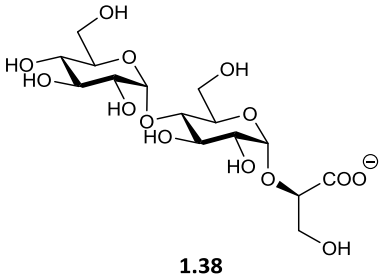
Scheme 1. 3 Proposed mechanism A where the donor is an α(1→4) linked glucan chain.

It is thought that the acceptor could either enter the active site after the intermediate formation, or it could become part of a ternary complex. If the latter occurs, it would suggest that the formation of the covalently bound intermediate would induce a conformational change of the active site in order to put the acceptor in a good position for carrying out the transglycosylation reaction as described.

This proposed mechanism however, suggests that MtGBE behaves as an exo-glycosidase in that it may hydrolyse the glycosidic bond at the end of the glycan chain. However, known GBEs generally show endo-glycosidase activity; hydrolysing glycosidic bonds in the middle of a chain to produce two shorter chains.[184]

1.5.1.2 Proposed mechanism B

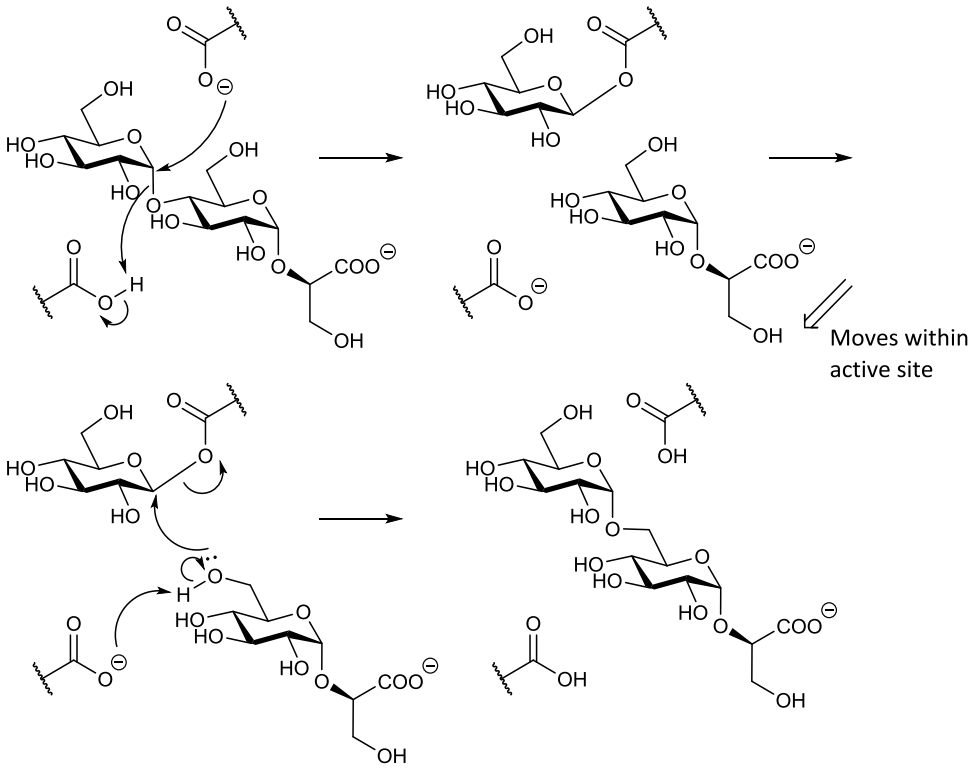
The second hypothetical mechanism, mechanism B, suggests the donor carbohydrate is a maltose unit with a glycerate moiety at the anomeric position, here termed 'maltosyl glycerate' (MalG) (**1.38**, Figure 1. 19). This suggested compound is similar to DGG but with an α(1→4) glycosidic bond, instead of the α(1→6) bond that is present in DGG.



1.38

Figure 1. 19 The proposed substrate for MtGBE in mechanism B, maltosyl glycerate (**1.38**).

It is proposed here that MtGBE forms a covalently bonded enzyme-substrate intermediate with the reducing end glucose (Scheme 1. 4). Following this, once the α(1→4) glycosidic bond is hydrolysed, the newly liberated GG could then move within the active site so as to be in the best position to now form an α(1→6) bond. The 6-hydroxy group of GG can then perform nucleophilic attack at the anomeric position of the bound glucose unit, forming DGG.



Scheme 1. 4 Proposed mechanism B, where maltosyl glycerate is the substrate

This mechanism, if it were to occur, would require the enzymatic synthesis of maltosyl glycerate before it could act as a substrate for MtGBE. A likely candidate could be the glycosyl transferase (GlcT) enzyme encoded for by *Rv3032*. This transferase enzyme has been reported to show transferase activity using UDP-glucose (UDP-Glc) and short chain α(1→4) glucans.[171] It could be that this GlcT utilises UDP-Glc and GG as substrates to form maltosyl glycerate.

1.5.2 Investigating the proposed mechanisms

The first step in determining the mechanism of MtGBE will be to determine the currently unknown donor substrate. As GG is the acceptor substrate, this compound would also need to be acquired.

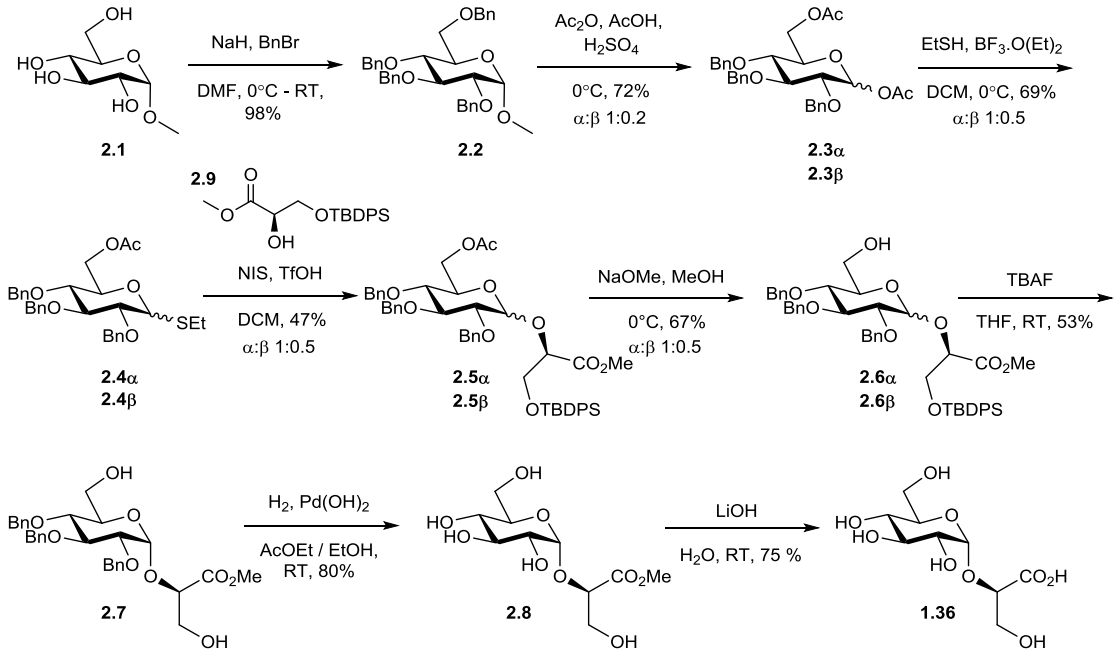
To investigate substrate breadth, purified recombinant protein will be tested with a number of compounds and detection of any potential product carried out by a number of techniques, such as TLC, mass spectrometry and various assays. In order to investigate mechanism A, various α(1→4) glucans which are commercially available will be tested. However, the proposed substrate for mechanism B is not commercially available nor is there a reported synthesis of this compound. Therefore, the synthesis of maltosyl glycerate will need to be undertaken in order to test this with MtGBE.

For testing of compounds against recombinant MtGBE, the recombinant protein will be over-expressed and purified. Furthermore, investigations into a suitable activity assay will be undertaken in order to determine activity and any effect various compounds may have on this. If the type of enzymatic activity can be determined, this will assist in determining the potential mechanism of MtGBE.

2. Chemical synthesis of potential substrates and inhibitors of MtGBE

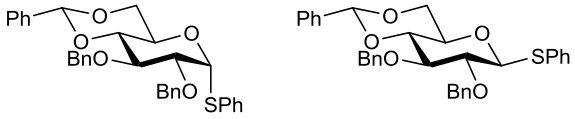
2.1 Glucosyl Glycerate

The synthesis of GG was based upon that of Lourenço *et al.* who reported the full synthesis in 2009.[186] The final synthetic route to GG undertaken in this project is depicted in Scheme 2. 1.



Scheme 2. 1 The synthetic pathway to glucosyl glycerate based on that reported by Lourenço *et al.*[186]

In their search for a glycosylation reaction with high α selectivity, the authors opted to follow the procedure developed by Crich and co.[186-188] Crich *et al.* had already developed a method for selectively forming β-mannopyranosides without having to employ neighbouring group strategy[189-192] and later proposed to use the same strategy to form β-glucosides, however they found this was not possible.[187] Instead it was found that *S*-phenyl 2,3-di-*O*-benzyl-4,6-*O*-benzylidene-1-deoxy-1-thia-D-glucopyranosides (**2.10** and **2.11**, Figure 2. 1) activated with triflic acid in DCM led to the selective formation of α glucosides when reacted in the presence of 2,6-di-tert-butyl-4-methylpyridine at -78 °C.[187]



2.10 **2.11**

Figure 2. 1 α (**2.10**) and β (**2.11**) anomers of S-phenyl 2,3-di-O-benzyl-4,6-O-benzylidene-1-deoxy-1-thia-D-α-glucopyranoside, both found to react with alcohols to yield predominantly α oriented products when in the presence of 2,6-di-tert-butyl-4-methylpyridine by Crich et al.[187]

When Lourenço *et al.* attempted this method in the formation of GG, they obtained the glycosylation product with a 72% yield with no β anomer detected.[186] Despite high yields and stereoselectivity, the lability of the benzylidene protecting group proved problematic for future use of the compound in subsequent coupling reactions.[186] Therefore, the next strategy employed was that in which the sugar with all hydroxyl groups protected except for the 6-OH could be formed (**2.6**). In this case, the authors prepared ethyl 6-*O*-acetyl-2,3,4-tri-*O*-benzyl-1-thio-α/β-D-glucopyranoside (**2.4**) *via* direct treatment of the per-benzylated methyl glucoside (**2.2**) with sulphuric and acetic acid, simplifying a more convoluted procedure previously reported in the literature.[193]

With **2.4** in hand, the Crich method of activation was attempted, however it was found that activation with *N*-iodosuccinimide (NIS) and triflic acid (TfOH) provided more desirable results with higher anomeric selectivity.[186] The use of the acetyl protecting group at the 6 position allowed for the 6-hydroxyl to be deprotected, which could then be used to form a 1→6 linked disaccharide.[186] In this work, it was thought that **2.4** could be used for future derivatives that may be investigated and thus it was decided this route would be followed. Furthermore, it is postulated that the acetyl group in the 6 position of **2.4** may aid in the formation of the α anomer by electronically shielding the top face of the glycosyl donor.[194]

2.1.1 Thioglucoside donor

Methyl-α-D-glucopyranose (**2.1**) is a readily available resource so was used as the starting point for this synthesis, though the anomeric configuration would not be retained further on in the synthesis and thus a mixture of anomers could have been used. Benzylation using the commonly used conditions of sodium hydride and benzyl bromide in dimethylformamide (DMF) readily yielded the tetra-*O*-benzylated methyl

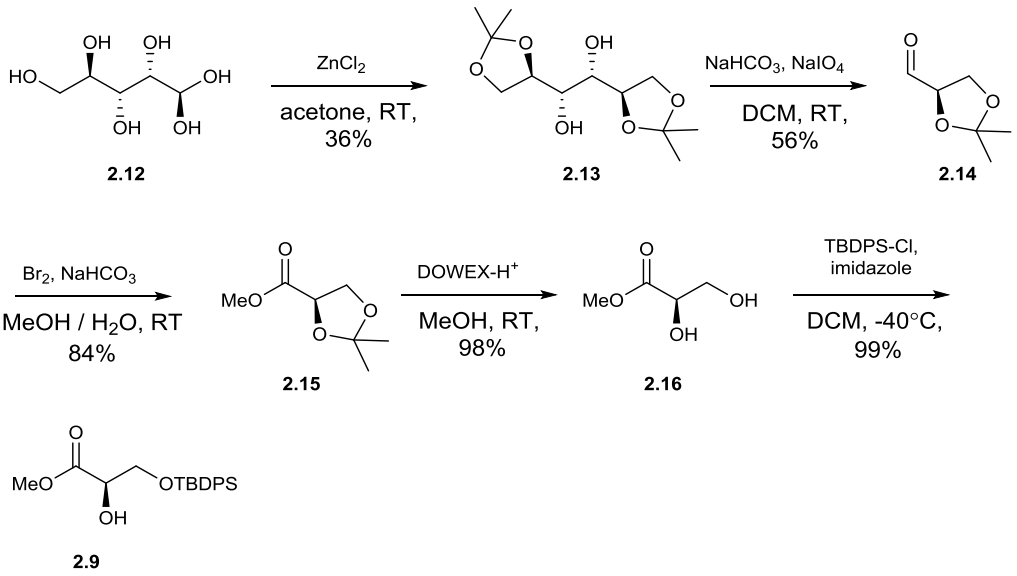
glucoside (**2.2**). During work up, diethyl ether was used during extraction and thorough washing of this with water allowed for removal of most of the DMF. Flash chromatography easily allowed for removal of residual DMF and mineral oil that had stabilised the sodium hydride.

Regioselective acetolysis of **2.2** using a mixture of acetic acid and acetic anhydride in the presence of a catalytic amount of sulphuric acid was used to produce the 1,6-di-*O*-acetyl-2,3,4-tri-*O*-benzyl glucoside (**2.3**) in a mixture of anomers (**2.3α** and **2.3β**). The reaction was carried out by Laura Jowett, an undergraduate student who worked on this project, who obtained **2.3** with an overall yield of 72% after purification by column chromatography.

In the third step, the formation of **2.4** was achieved by activating **2.3** with boron trifluoride in the presence of ethanethiol. Pure product took the form of a white waxy solid once purified by column chromatography, and **2.4** was obtained in 69% yield. With **2.4** in hand, the desired glycosylation donor was prepared and attention turned to the preparation of the acceptor.

2.1.2 Glyceric acid acceptor

The glycerate moiety of GG is derived from glyceric acid, though coupling of unprotected glyceric acid could lead to undesired products due to it having two free hydroxyl groups. To selectively couple the donor (**2.4**) with the 2-hydroxyl group of glyceric acid, the primary alcohol was protected selectively as a *tert*-butyldiphenylsilyl (TBDPS) ether. The bulkiness of TBDPS compared to other silyl protecting groups allows for selective protection of the primary alcohol, due to steric hindrance of the secondary hydroxyl group. The acid group of the glyceric acid is also protected in the form of a methyl ester (**2.9**).



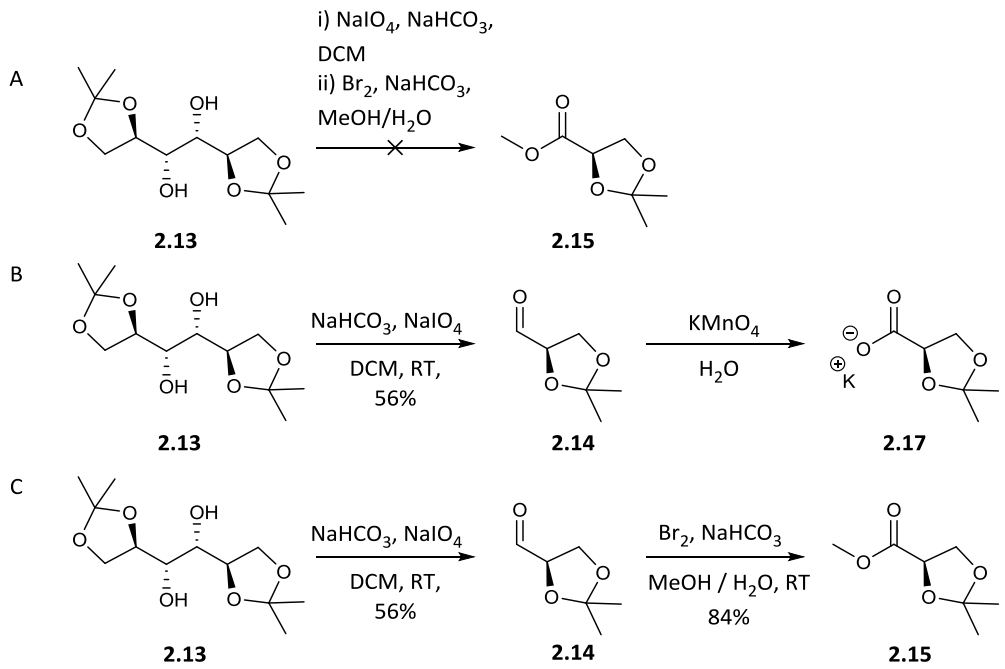
Scheme 2. 2 Synthetic route to glycerate acceptor **2.9**.

To form the protected glycerate **2.9** (Scheme 2. 2), a synthetic route starting from D-mannitol (**2.12**) was employed. Initially, 1,2:5,6-di-*O*-isopropylidine-D-mannitol (**2.13**) was prepared by stirring mannitol in acetone in the presence of zinc chloride. This reaction was performed on large-scale in which 30 g of mannitol was used to generate a large amount of **2.13**. The literature procedure referenced for this reaction described the use of recrystallization from ethyl acetate as a purification method.[195] Whilst this method was indeed successful, with pure crystals being formed, ethyl acetate was also the solvent used in the extractive work up in the preparation of the compound, meaning that the desired product dissolved well in this solvent and so using the minimum amount of boiling ethyl acetate in the recrystallization was critical. This may have been one reason the yield obtained in this work was not as high as that reported.[195] Another reason may be the lack of anhydrous acetone, however access to a still to prepare this was not available and considering the large scale of the reaction, the 36% yield of **2.13** still gave enough material to carry through.

Oxidative cleavage of **2.13** yielded two molar equivalents of the protected glyceraldehyde (**2.14**). This was achieved using sodium periodate in DCM and saturated sodium bicarbonate solution. Due to periodate not being soluble in DCM, the reaction was stirred vigorously to allow for thorough mixing of the aqueous and organic layers. In this case, the procedure reported by Sugisaki *et al.* was followed in

which the product was distilled under reduced pressure.[196] At the time this work was carried out, the crude product was combined with that of a colleague working with the same compound, for ease of access to the distillation equipment. Therefore, it was not possible to determine the overall yield after purification, however the crude yield for **2.14** was 56%, lower than that reported by Sugisaki *et al*.[196]

As **2.14** is an unstable compound[197], the next reaction in the synthetic pathway was carried out as soon as possible post distillation. Indeed, signals were observed in the ^{1}H NMR, particularly around 4.20-3.90 ppm and 1.38-1.35 ppm, which may have arisen from polymerisation, even after distillation. It may have been for this reason that Ladame *et al.* carried out oxidation of **2.14** to a methyl ester (**2.15**) immediately after the oxidative cleavage, without having isolated the aldehyde in-between steps.[197] When attempted in this work, the immediate oxidation of **2.14** to **2.15** was unsuccessful, resulting in a small amount of brown residue after work up. It was therefore decided that another route to **2.15** be taken, by first oxidising to an acid then subsequently carrying out esterification. This was attempted with potassium permanganate using a procedure reported by Qu *et al*. to get the acid product[198], a reaction that was successful with a crude yield of 76%, however the crude product required further purification. Another procedure, as per Tanaka *et al.*, was trialled in an attempt to obtain the potassium salt (**2.17**) which would result in a crystalline product.[199] It was thought a solid would be easier to handle and purify (*via* recrystallization), however in this case the solid was very fine and proved difficult to filter off. Alongside forming the acid or the salt, a third reaction was carried out in which purified **2.14** was oxidised using the soft oxidative conditions used by Ladame *et al*. in the aforementioned work[197]. This was to see if isolating and purifying **2.14** before further oxidation was more of a success than the one pot method reported in the literature. A summary of the methods trialled to obtain **2.15** and **2.17** is shown in Scheme 2. 3.



Scheme 2. 3 A: One-pot reaction combining oxidative cleavage and oxidation as reported by Ladame et al,[197] unsuccessful in our case; B: Route to methyl ester *via* the carboxylic acid as per Qu et al. [198]; C: Most successful route to methyl ester achieved by separating the cleavage and oxidation steps.

It would appear that separating the oxidative cleavage and further oxidation steps proved more effective than attempting both reactions in immediate succession. These soft oxidative conditions were achieved using bromine in water and concomitant esterification was achieved thanks to the presence of methanol. Following the quenching of bromine and the extraction of the product with DCM, **2.15** was obtained without need for further purification in 84% yield.

The removal of the acetonide protecting group of **2.15** was performed using acid hydrolysis, and the use of an acidic resin (Dowex®) made this a simple reaction to carry out, the only work up required being the removal of the resin *via* filtration. Again, no purification steps were required and the diol (**2.16**) could now be used in the final step of this synthesis.

Following the reported procedure from Lourenço and co,[186] TBDPS chloride was added to **2.16** in the presence of imidazole to produce **2.19** with an almost quantitative yield (99%). The thick syrup that resulted was not purified any further as ^{1}H NMR

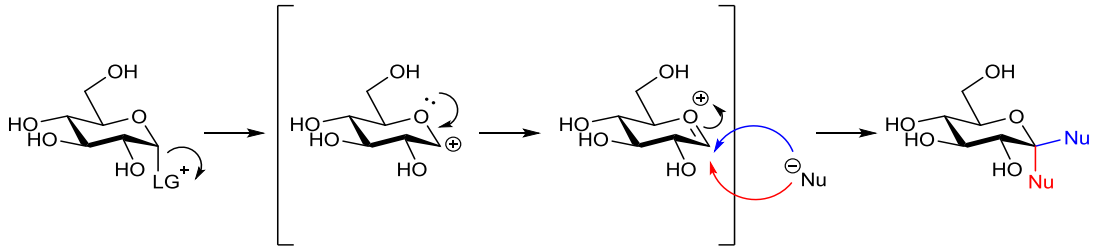
spectroscopy suggested the product was pure enough to be taken through to glycosylation.

2.1.3 Glycosylation and final steps to glucosyl glycerate

With both the thioglycoside donor (**2.4**) and the glycerate acceptor (**2.9**) in hand, glycosylation was the next step towards GG. Due to the need to control the stereochemical outcome of the coupling reaction, this is a challenging step in the synthesis.

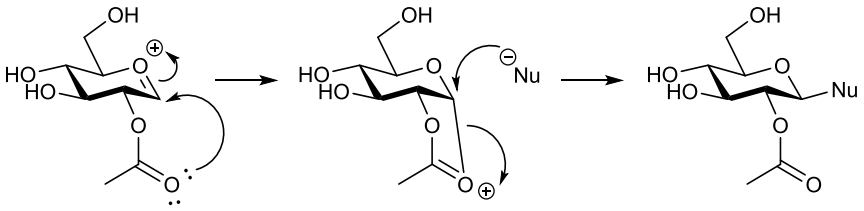
Natural glucosyl glycerate contains an α-glucopyranoside linkage and this configuration must be replicated in the chemical synthesis. Controlling the stereochemical outcome of glycosylation reactions has been studied in detail and there are a number of ways to attempt to selectively obtain the α (or 1,2-*cis*) anomer.[85, 200, 201]

During the glycosylation reaction, the activated leaving group of the sugar donor departs and thus forms a glycosyl cation (Scheme 2. 4). The latter evolves through resonance to the more stable oxocarbenium ion that is then attacked by the nucleophilic acceptor compound, either from the bottom face or the top face, forming either an α or β anomer respectively.



Scheme 2. 4 General reaction mechanism for chemical glycosylation showing the formation of both alpha and beta anomers.

Controlling the reaction at this point can lead to stereochemical control of the product. The presence of an acyl containing functional group at position 2 on the sugar ring can aid stereoselectivity *via* a process known as neighbouring group participation (NGP), depicted in Scheme 2. 5.[202, 203] In NGP, the lone pair from the carbonyl oxygen of the acyl group attacks the anomeric position of the oxocarbenium ion forming an acetoxonium ion, which blocks the bottom face of the ring, forcing the nucleophile to attack from the top face and form the β product.



Scheme 2. 5 Neighbouring group participation of a C2 acyl substituent.

In the case of 1,2-cis glycosylation NGP from a group at position 2 is highly undesirable and so, in the case of this synthesis, the donor is benzylated at position 2, a group which does not contain acyl functionality and so cannot partake in NGP. As well as position 2, it is thought that NGP may also occur with groups at position 6 on the sugar ring.[204-206] For example, in their synthesis of GG, Lourenço *et al.* found that having a 6-OAc group on the thioglycoside donor gave greater selectivity of the α anomer.[186, 207] In a similar manner to NGP at position 2, the 6-*O* acetyl group can electronically shield the top face of the sugar ring, driving the attacking nucleophile to attack from below.[194] As this work follows their synthesis closely, the 6-OAc group was also utilised in this work.

Another variable that can be manipulated in order to control the stereochemical outcome of glycosylation is the solvent used in the reaction. It is generally believed that polar solvents promote the formation of the β anomer and non-polar solvents promote formation of the α anomer.[85] In addition, it has also been observed that ether-containing solvents promote formation of the α product[205, 208], supposedly through the formation of an intermediate in the equatorial position.[85, 194] Conversely, nitrile containing solvents have been observed to favour formation of the β anomer *via* formation of a nitrilium cation that adopts an axial conformation.[85] In this synthesis, the reaction solvent was kept as DCM, a non-polar solvent that led to a yield of 63% and an alpha selectivity of α:β 2:1. Diethyl ether was trialled but gave a low yield of 9%. Lourenço *et al*. also experimented with solvent effects on glycosylation with similar compounds to GG, however whilst they found diethyl ether greatly promoted 1,2-*cis* glycosylation with smaller, more reactive acceptors, they did not report improvement in α selectivity in their glycosylation reaction involved in the synthesis of GG.[194]

Finally, a third consideration is the temperature of the reaction. Due to the anomeric effect described in chapter 1 (section 1.2.2) the alpha anomer of unprotected sugars is favoured thermodynamically and so generally forms at warmer temperatures.[85] Generally, during glycosylation reactions, lower temperatures favour formation of the β-anomer [209], however there is also evidence of the opposite situation occurring.[210] In this case, the reaction was carried out at 0 °C as per the literature.[186]

Initially, separation of anomers of **2.5** was not achieved. A gradient of ethyl acetate in hexane starting from 30 % was initially tried but with no success when isolating anomers. TLC in various solvent systems then showed promise with 90:9:1 DCM/ toluene/ EtOAc though again, no separation of the anomers was achieved. However, later on in the project, the synthesis was revisited in order to produce more material and it was then found that a system of 10% ethyl acetate in hexane was found to lead to almost pure α anomer when a slow gradient was used up to 15% ethyl acetate.

The first deprotection reaction carried out was the removal of the 6-*O*-Acetyl group using sodium methoxide in methanol, known as Zemplén conditions. With a reaction temperature of 0 °C, **2.6** was obtained with yield of 67%. This yield was lower than the literature procedure which was followed.[186] Lourenço *et al.* used sodium metal in methanol rather than a stock solution of sodium methoxide[186] and therefore *in situ* generation of sodium methoxide may prove more effective in the reaction in future.

The second deprotection step was the removal of the TBDPS group using tetrabutylammonium fluoride (TBAF) to give **2.7**. As before, the high yields reported were not repeated in this work (the highest achieved yield was 53%). It was initially found that separation of the anomers after the glycosylation was not possible and so a mixture of anomers was taken on to this step in the early stages of the project. Interestingly, following removal of the TBDPS group, the α anomer was the only anomer observed both in the crude material and after purification.

Following removal of the TBDPS protecting group, the benzyl groups were then removed to give **2.8**. The literature reported this was possible using hydrogen and a palladium catalyst with a mix of ethyl acetate and ethanol as the reaction solvent all under 35 psi.[186] Due to the lack of access to equipment required to carry the reaction

out under pressure, this was first attempted at atmospheric pressure. Though some benzyl groups were removed (as interpreted from TLC) the reaction did not reach completion even after 3-4 days with addition of fresh catalyst and hydrogen. To try and improve this result, different catalysts were used, changing the initial catalyst of palladium on carbon to palladium black or palladium hydroxide on carbon. Neither changing the catalyst or the reaction solvent improved the reaction efficiency, all conditions attempted are summarized in table 2.1. Even once access to a pressurised reaction vessel was obtained, the literature results of complete reaction after 4 hours could not be replicated. In a further attempt, triethylsilane was used to generate hydrogen gas *in situ.* However, this did not produce the fully de-benzylated product.

Table 2. 1 Conditions trialled in attempts to remove benzyl ether protecting groups from GG.

Catalyst	**Solvent**	**Reaction time**	**Pressure**	**Result**
Pd/C (10 %)	EtOAc/EtOH (2:1)	3 hrs	2 atm	Mix of products
Pd/C (10 %)	EtOAc/EtOH (2:1)	48 hrs	1 atm	Mix of products
Pd/C (10 %)	EtOAc/EtOH (2:1) (+ AcOH)	4 days	1 atm	Mix of products
$Pd(OH)_2/C$ (20 %)	EtOAc/MeOH (2:1)	4 days	1 atm	Mix of products
Pd black	EtOAc/MeOH (2:1)	24 hrs	1 atm	Mix of products
Pd black	EtOAc/MeOH (2:1)	24 hrs	5 atm	Mix of products
Pd/C (10 %) (+TES)	EtOH	9 days	1 atm	Mix of products
$Pd(OH)_2/C$ (20%)	EtOAc/EtOH (2:1)	18 hrs	1 atm	Desired product

However, when attempted again at standard pressure but with more vigorous degassing and increased volumes of hydrogen bubbled through the system, the reaction was taken to completion after 18 hours with palladium hydroxide as the catalyst and a mix of ethyl acetate and ethanol as the reaction solvent. Methods used

previously had involved using a hydrogen balloon and a needle to allow the gas to bubble through the reaction mixture before another hydrogen balloon was used to maintain a saturated atmosphere. The new method involved exposing the system to vacuum until the solvent began to boil and then releasing hydrogen gas into the system. This was repeated multiple times before a balloon used to maintain a saturated atmosphere as before and proved the most successful method. Once the reaction was complete, the palladium hydroxide could be filtered off over Celite® which was washed with methanol. NMR suggested no further purification was required due to lack of significant peaks in the aromatic region. The yield for this first successful reaction was 80% of **2.8**, and repetition of the reaction led to a yield of 95% of **2.8** when it was left for 48 hours.

The final step towards **1.36** was the removal of the methyl ester to free the carboxylic acid group on the glycerate moiety of the compound. This was achieved in a quantitative yield *via* ester hydrolysis with a 1.0 M solution of lithium hydroxide. NMR and mass spectrometry matched that reported by Lourenço and co., though it was also observed in some cases that the NMR would show slightly different coupling patterns, thought to be due to conformational effects (Figure 2. 2).

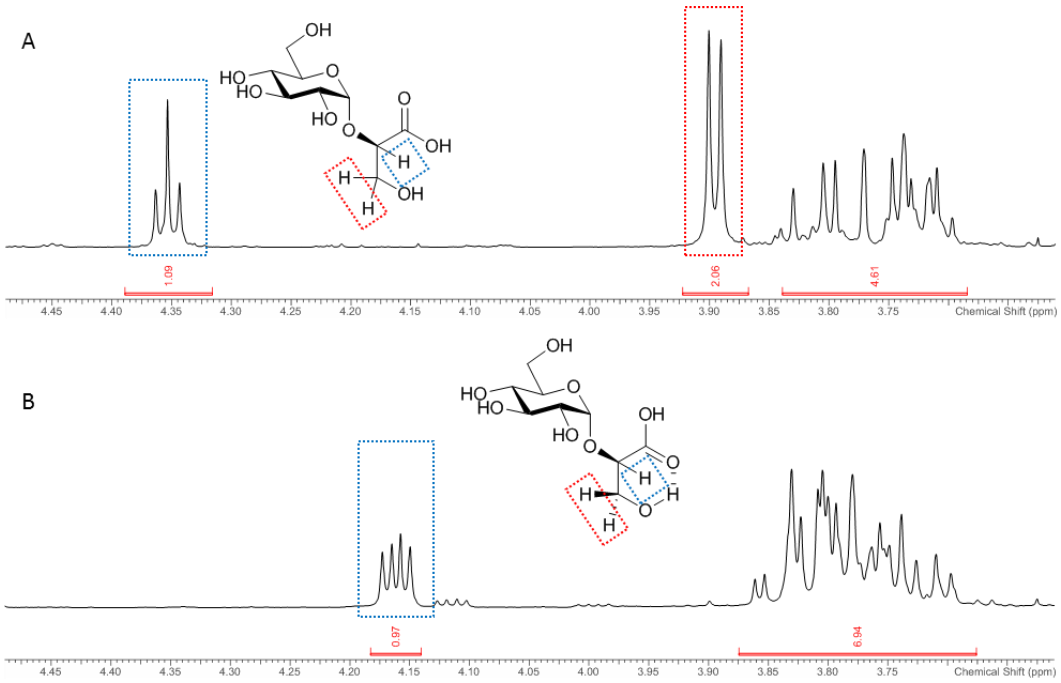


Figure 2. 2 Section of ^{1}H NMR of GG with relevant peaks highlighted; A: the spectrum obtained that matches that reported by Lourenço et al.; B: the spectrum thought to arise from intermolecular hydrogen bonding, locking the conformation of the compound; blue square = H7, red square, H8.

For example, in the ^{1}H NMR spectrum reported by Lourenço[186] and also in initial spectra in this work (Figure 2. 2 A), an apparent triplet is observed at 4.35 which can be assigned to the C-H environment in the glycerate region (in this report assigned as H7). This triplet couples with the signal at 3.9 ppm which is seen here as a doublet (though reported as a multiplet in the literature[186]) and integrates to 2 protons, leading to its assignment as the CH_2 environment in the glycerate part of the compound (here designated H8). However, in the ^{1}H NMR spectrum for the same compound taken later on in this project (Figure 2. 2 B), the signal for H7 now shows upfield at 4.16 ppm and is now a clear doublet of doublets. Similarly, the doublet of H8 has also shifted upfield and is now within the multiplet that appears from 3.83-3.70 ppm, becoming less distinguishable.

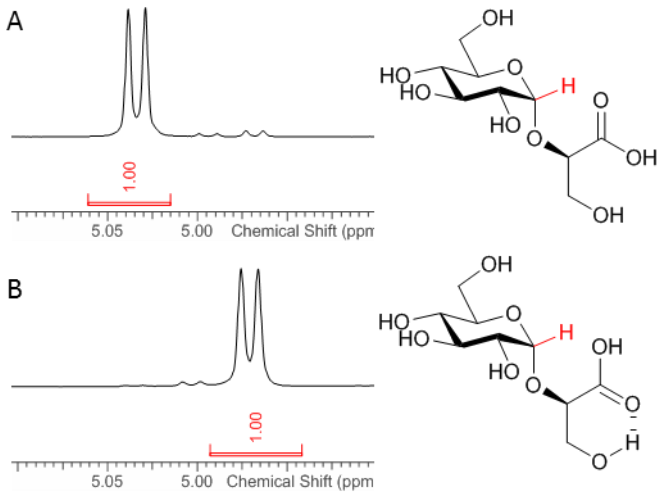


Figure 2. 3 Anomeric signals in GG ^{1}H NMR spectrum; A: the spectrum obtained that matches that reported by Lourenço et al.; B: the spectrum thought to arise from intramolecular hydrogen bonding, locking the conformation of the compound. Both spectra obtained with samples dissolved in D_2O.

Other chemical shifts remain similar, except for the anomeric signal that shifts slightly from 5.03 ppm to 4.97 ppm (Figure 2. 3). It was thought that these changes in the relevant peaks may come around from the two H8 protons being fixed in their environments. This may occur if the carbonyl group of the acid were to rotate and hydrogen bond with the hydroxyl group, locking the conformation and hindering rotation in that part of the molecule (Figure 2. 3 B).

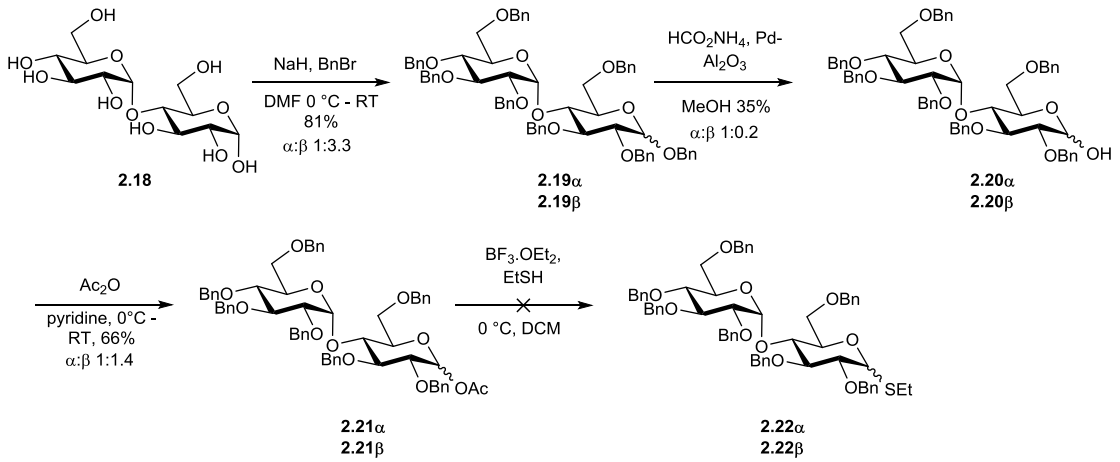
2.2 Maltosyl glycerate

As described in chapter 1 (section 1.5.1.2), substrate investigations would include the testing of the compound maltosyl glycerate (MalG, **1.38**). A literature search suggested that this compound had not been reported before and so a novel synthesis had to be designed.

It was initially thought that a route similar to that of GG could be taken (Scheme 2. 6), starting with maltose (**2.18**). However, though the idea of using a thioglycoside activated by NIS and TfOH was carried over into this synthesis, the route to make the donor had to be modified. The GG synthesis started from the commercially available methyl glucoside which could then be benzylated and the anomeric and 6-OH positions selectively acetylated. This could not be done in the case of MalG however, as methyl

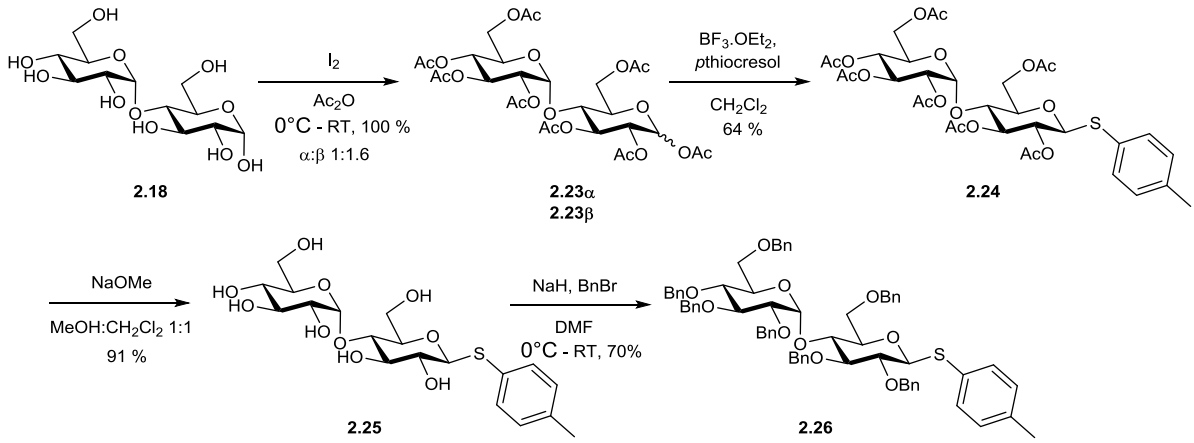
maltoside is not a readily available resource. This could have been prepared from maltose, but it was deemed that the synthesis would be more efficient if this step was not included.

When deciding on protecting group strategy, the effects of protecting groups on the reactivity of the glycosylation donor were taken into consideration. Works have been carried out investigating the influence of protecting groups on the reactivity of glycosyl donors with the aim of being able to tune the properties of the donor in use.[211-213] It has been found that esters have a 'disarming' effect on the sugar ring, making reactions at the anomeric centre slower, whilst ethers have an arming effect, leading to faster reaction times in glycosylation.[76] Therefore, the use of benzyl ether groups to protect the thiomaltoside donor would be desirable. However, to generate the thiomaltoside, the anomeric position must be free of a benzyl group as this would not make a good leaving group when coupled with a thiol. As in the GG synthesis, it was thought that should the anomeric protecting group be converted to an acetyl group, the donor could be easily formed. Literature reports do suggest it is possible to selectively remove the anomeric benzyl group from fully benzyl protected maltose using *in situ* generation of hydrogen.[214] Bieg *et al.* used ammonium formate and a palladium catalyst to produce the 2,3,6,2',3',4',6'-*O*-benzyl maltose (**2.20**) from the fully protected sugar (**2.19**).[214] Therefore the initial synthesis plan was based on benzylation, followed by selective deprotection of the anomeric position, then acetylation and coupling with a thiol to produce a thiomaltoside donor for the glycosylation with the glycerate acceptor (Scheme 2. 6).



Scheme 2. 6 Initial synthetic plan to thiomaltoside donor.

However, the selective removal of the anomeric benzyl group to give **2.20** was inefficient when attempted in this work. Therefore, a different approach to the synthesis of the thiomaltoside was designed (Scheme 2. 7). This involved per-*O*-acetylation of maltose to give **2.23**, which could then be coupled with a thiol, followed by removal and replacement of the acetyl groups with benzyl protecting groups, forming the armed donor.



Scheme 2. 7 Revised synthetic plan of thiomaltoside donor.

As maltose (**2.18**) is a disaccharide of two glucose units, it was thought the same glycosylation conditions for GG could also be applied in the case of the MalG synthesis. It was also thought that separation of anomers could be carried out at this point. Following on from the glycosylation reaction, it would then take 3 deprotection steps

to achieve MalG; removal of the TBDPS group, removal of the benzyl ethers and finally hydrolysis of the methyl ester.

Once maltosyl glycerate had been obtained, it was then planned that this would be tested for activity with purified MtGBE protein.

2.2.1 Initial attempt to prepare thiomaltoside donor

The synthesis of MalG was first attempted *via* benzylation of **2.18**. This was carried out using standard benzylating conditions, using benzyl bromide in the presence of sodium hydride with DMF as the reaction solvent. The reaction was deemed complete *via* TLC after 3 hours which led to **2.19** with a yield of 81% following purification *via* column chromatography.

The selective removal of the anomeric benzyl group was attempted multiple times with ammonium formate and 10% palladium on alumina in methanol. **2.20** was obtained with mild success as yields ranged from 10-35 % and scale-up to 2.69 mmol (2.86 g) of starting material only yielded 12 % of desired material. The procedure followed in the literature reported a 65% yield of **2.20** after 10 hours.[214] However, when this was attempted in this work on the same reaction scale, after 8 hours a yield of just 12% was achieved. In fact the highest yield obtained (35%) was achieved after 5 hours on a 0.11 mmol scale, however this was still too low to enable an efficient synthesis. After each attempt the starting material (**2.19**) could be recycled *via* column chromatography using a gradient of 0-20% diethyl ether in toluene and reacted again to obtain more product. Thus, enough **2.20** was obtained to be taken through to the subsequent step.

Once the anomeric benzyl group had been removed, the next step was the acetylation of the anomeric hydroxyl group to give **2.21** which was carried out with acetic anhydride in pyridine. After overnight reaction, TLC suggested complete conversion of the starting material though following column chromatography, starting material was isolated as a minor fraction. The highest obtained yield of **2.21** was 66%.

Attempts at then preparing **2.22** were unsuccessful. The reaction was attempted at 0 °C using boron trifluoride diethyl etherate as the activating agent with DCM as the reaction solvent. Reactions were first carried out on small scale and though TLC

suggested consumption of the starting material, no product was seen in mass spectrometry analysis of the crude mixture. Due to the difficulty in forming **2.22** and the low yields in the regioselective benzyl removal, this synthetic route to MalG was redesigned. It was also decided that a different thiol be used to prepare the thiomaltoside donor to improve the ease of handling the malodourous smells given off by these compounds. Therefore, *p*-thiocresol was chosen to replace ethanethiol as the solid is naturally less volatile and the smell less potent.

2.2.2 Successful synthesis of maltosyl glycerate

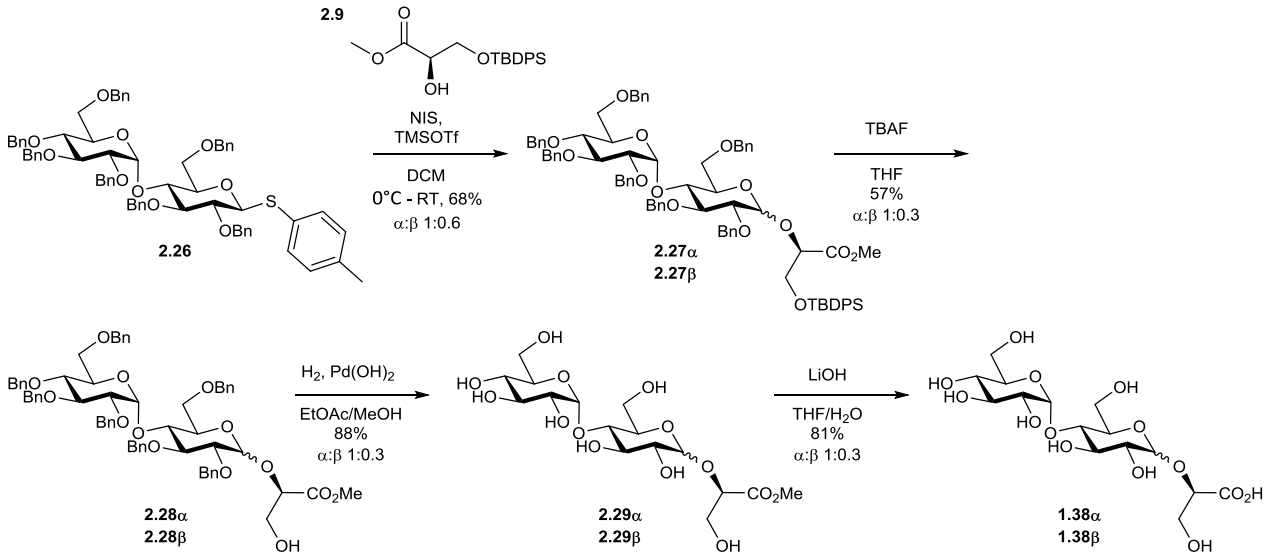
2.2.2.1 Thiomaltoside donor

Donor **2.26** was obtained in four steps, the first of which was peracetylation of maltose to give **2.23**. Kartha *et al.* had reported full acetylation of maltose, among other substrates, using iodine as a catalyst after 10 minutes.[215] This was deemed a faster acting and greener alternative to the standard method of acetylation involving pyridine, acetic anhydride and 4-dimethylaminopyridine (DMAP) and therefore this procedure was followed. The reaction was carried out on large scale (20.0 g, 55.5 mmol of maltose) with acetic anhydride as the reaction solvent. When a catalytic amount (0.07 molar equivalents) of iodine was added the reaction became hot and was therefore then cooled with an ice bath. After initial reaction of approximately 5 minutes, the ice bath was removed and the reaction carried out at room temperature. Though the literature procedure had said they found complete acetylation after ten minutes, this reaction was left for 1.5 hours due to the vastly increased scale. Following work up, **2.23** was obtained to an acceptable level of purity meaning no further purification was carried out and the material taken through to the next step.

With **2.23** in hand, it could then be coupled with *p*-thiocresol to form **2.24**. Boron trifluoride was used to activate **2.23** and the reaction was carried out in DCM at room temperature for 18 hours. Guan *et al.* reported this reaction was complete after just 2 hours[216], however it was found in this case that starting material was still present after 2 hours and so the reaction was left overnight. **2.24** was successfully obtained in a 64% yield after purification *via* column chromatography, similar to the reported yield of 73%.

In order to create an armed donor, the global protecting groups were changed from acetyl groups to benzyl ethers. The acetyl groups were removed using Zemplén conditions which led to **2.25** in a 91% yield after column chromatography. Following deprotection, global benzylation of the donor was achieved using sodium hydride and benzyl bromide which, after column chromatography, led to **2.26** with a yield of 70%.

2.2.2.2 Glycosylation and deprotection to give MalG

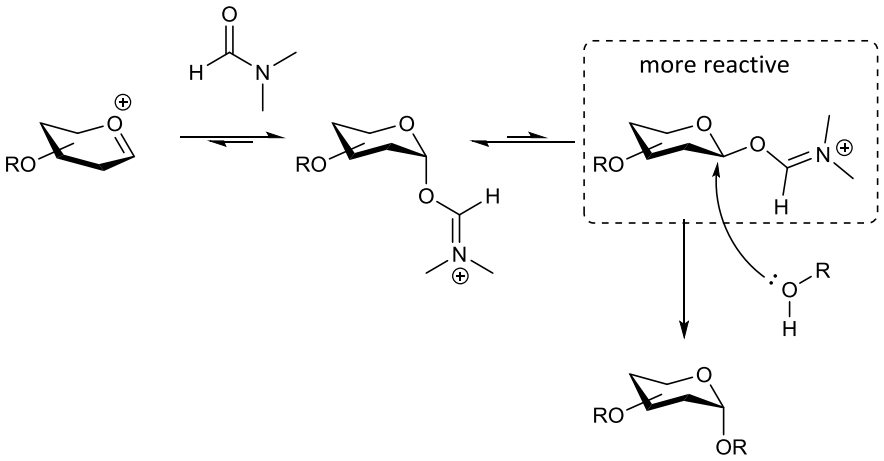


Scheme 2. 8 Remaining synthetic steps towards MalG.

With **2.26** in hand, the glycosylation reaction with **2.9** could now be performed (Scheme 2. 8). The conditions chosen for glycosylation were those of the glycosylation in the GG synthesis as they had previously successfully yielded a majority of α anomer.[186] Therefore, **2.26** was activated with NIS and trimethylsilyl trifluoromethanesulfonate (TMS-OTf) in the presence of **2.9** with the reaction solvent being DCM. TMS-OTf in this case replaces TfOH as it is easier to store and to handle. The reaction was successful, with the most successful yield of **2.27** being 68%, however this was the overall yield for a mixture of both α and β anomers, the ratio of which was found to be 1:0.6 α:β *via* ^{1}H NMR. Separation of anomers was then attempted *via* column chromatography using a system of ethyl acetate in hexane, however complete separation was not achieved. Therefore, further purification using a longer silica column and a slower solvent gradient was used in an attempt to obtain the desired α anomer (**2.27α**). Whilst this was successful to a certain degree, mixed fractions were

always still obtained. It may have been possible to achieve complete separation using automated chromatography, such as with a Biotage® system, or *via* HPLC, however due to time considerations, it was decided to proceed with the synthesis using both the mix of anomers (**2.27**) (which were now in a ratio of 1:0.3 α:β) and the pure α material that had been obtained (**2.27α**).

Alongside the attempts to separate the anomers of the glycosylation product, a literature report was found in which DMF was used to promote the formation of the α product of glycosylation reactions.[217] Lu *et al.* found that pre-activation of a perbenzyl thiogalactoside formed an intermediate glycosyl imidate, the β anomer of which they proposed to be more reactive upon addition of the acceptor, leading to the formation of the α glycosidic linkage in the final desired product (Scheme 2. 9). Although the authors found that increased molar equivalents of DMF in the pre-activation step increased the ratio of α:β in the products, the effect was minimal, however with more sterically constrained donors, the ratio could be greatly improved upon.[217]



Scheme 2. 9 Suggested mechanism for DMF mediated stereoselective glycosylation. An oxocarbenium ion forms both α and β intermediates with DMF, the latter of which reacts faster with the incoming acceptor to lead to the formation of the α product. Figure recreated from Mong et al.[217]

This method was attempted in this work with the aim of increasing the yield of the α glycosylation product. Therefore, following the literature procedure, 1.5 molar equivalents of **2.26** were pre-activated with DMF (6 molar equivalents) before addition of **2.9**. The reaction was found to have produced **2.27** as detected *via* mass

spectrometry, however following purification *via* column chromatography, the product could not easily be separated from other by-products of the reaction. In addition, the mass of the impure product fraction would only have led to a 14 % yield of **2.27** and for this reason, coupled with the appearance of various other side-products, this procedure was not pursued further.

The first deprotection step undertaken after glycosylation was the removal of the TBDPS group to give **2.28**, achieved using 1.0 M TBAF in THF. The yield of this reaction following purification was low, with **2.28** produced in a 57% yield. This may have been due to reagent quality, had any moisture been present in the TBAF solution or in the solvent (bought commercially as anhydrous) then the TBAF may have decomposed. However, TLC had shown the starting material to have been fully consumed after 3 hours, suggesting another possible reason for the lower than desired yield.

Following TBDPS removal, hydrogenolysis was performed in order to remove all 7 benzyl ethers to give **2.29**. This was initially performed using palladium on an alumina support. It was thought that alumina may have a larger surface area, which may increase adsorption of the substrate. In addition, this was also the catalyst used in the literature report using hydrogenolysis to regioselectively remove the anomeric benzyl group as previously mentioned.[214] However, the reaction was not complete even when left for 7 days, with varying degrees of deprotected sugar being present. Therefore different catalysts were explored, including palladium on carbon, palladium hydroxide on carbon and a mix of both, the latter of which had been reported to be a more efficient catalyst than either component used separately.[218] Small scale reactions were successful when using palladium hydroxide alone, and scale up was successful when a large amount of catalyst was employed (1.5 molar equivalents). This method of using large equivalents of catalyst in a mixed solvent system of ethyl acetate and methanol was thus used to convert **2.28** to **2.29** with the highest yield achieved being 88%, with no further purification deemed necessary following work up.

Additionally, an alternative deprotection strategy found in the literature was tested in which sodium bromate and sodium dithionite alongside a biphasic solvent system are employed to provide oxidative conditions for benzyl ether removal.[219] This procedure was tested on small scale (0.03 mmol) but although TLC suggested reaction

progression, NMR spectroscopy showed the presence of benzyl groups after column chromatography of the crude material. Therefore this procedure was not repeated.

The final step to obtaining MalG involved hydrolysis of the methyl ester in **2.29**. This was achieved using a 1.0 M solution of lithium hydroxide in a mixture of water and THF. This was successfully achieved with an 81% yield to produce **1.38** as determined by 1D and 2D NMR and high resolution mass spectroscopy (HRMS) (see appendix A, A.5).

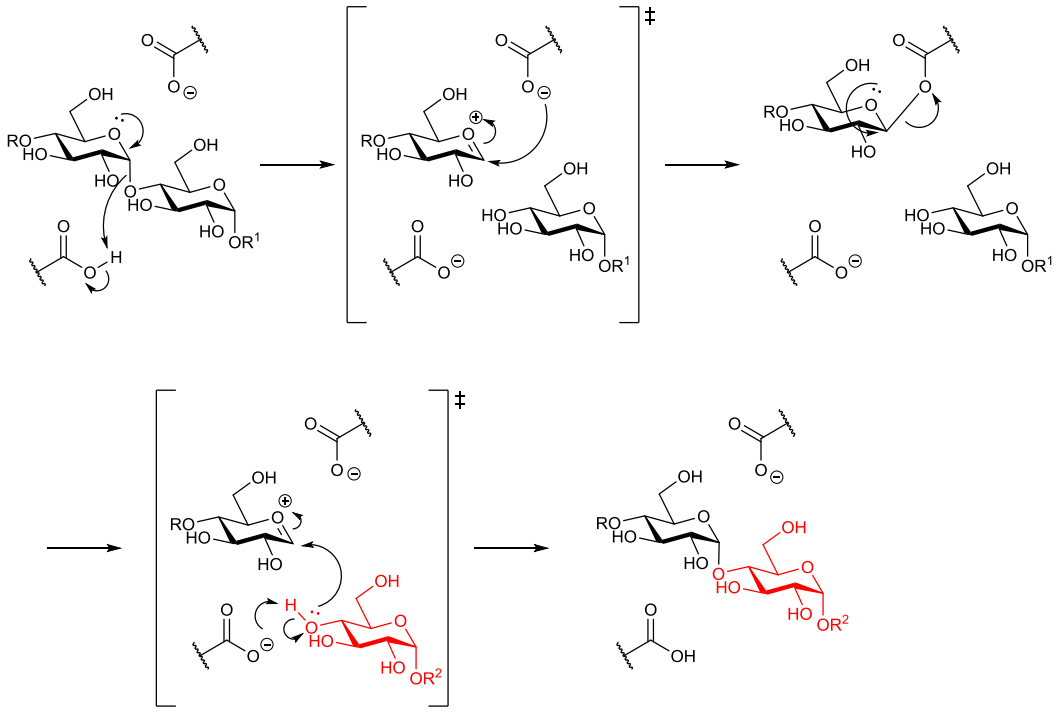
In this way, both the pure α anomer of maltosyl glycerate and a mixture of anomers containing predominantly α maltosyl glycerate were obtained. These were then taken through to check for any activity with MtGBE.

2.3 Fluorinated inhibitors

For decades now, glycosyl fluorides have been known to act as substrates for glycosidases,[220] usually with fast reaction rates.[221] They also have an advantage over other popular glycosyl derivatives in that there are few cases, if any, of a glycosidase not working on the corresponding glycosyl fluoride substrate.[221] This advantage is highlighted in the case of the mammalian glycosidase pancreatic α-amylase wherein the appropriate α-maltooligosaccharyl fluorides are useful substrates[222, 223] but the corresponding aryl glycosides are deemed poor substrates for this protein.[221]

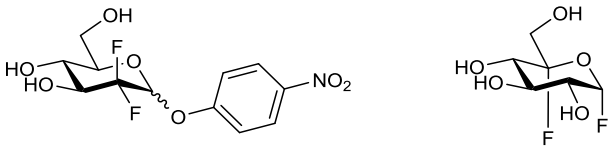
Glycosyl fluorides have also proven useful tools in investigating and identifying the key catalytic residues in enzyme active sites, due to their not requiring acid catalysts in hydrolysis.[224] Should a mutated residue still show activity on the glycosyl fluoride but not the native *O*-glycoside, the residue is likely involved in acid catalysed hydrolysis.[224]

Fluorinated carbohydrates have also been widely used in studying the covalent enzyme-substrate intermediate for both α and β retaining glycosidases.[183, 225, 226] The established mechanism for glycosidases suggests that this intermediate occurs *via* a transition state of considerable oxocarbenium ion character (Scheme 2. 10).[183, 221]



Scheme 2. 10 Generic mechanism for a retaining glycosidase showing transition states. Figure reproduced from Uitdehaag et al.[183]

One of the original tactics for trapping covalent intermediates of glycosidases was to replace the hydroxyl group at C-2 with a fluorine atom.[221] The strong electronegativity of fluorine destabilises the transition state so much so that glycosylation is really slow. Due to this effect, 2-deoxy-2-fluoro carbohydrates have been used to inhibit enzymes mechanistically.[221] However, it was found that these 2-deoxy-2-fluoro carbohydrates act as better inhibitors for β glycosidases compared to α glycosidases, thought to be partly due to the fact that the α configured intermediate for the β glycosidase is more stable than the β configured intermediate of the α glycosidases.[227] It was to tackle this selectivity that Withers *et al*. developed 2,2-difluoro sugars, such as **2.30** (Figure 2. 4), which were tested and found to be good inhibitors of both classes of glycosidases.[227]



2.30 **2.31**

Figure 2. 4 Examples of fluorinated carbohydrates designed to trap enzyme-substrate intermediates.[221]

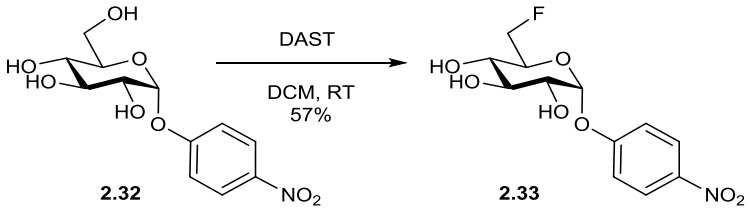
Another class of fluorinated inhibitors found to be useful for trapping intermediates are 5-fluoro-glycosyl fluorides of which 5-fluoro-α-D-glucopyranosyl fluoride (5FGlcF) (**2.31**) is an example.[221, 228] It is thought that the highly electronegative fluorine atom at C-5, being positioned next to the ring oxygen where most charge is built up, again destabilises the transition state, making glycosylation extremely slow.[229] It has been reported that most charge is found on the ring oxygen compared to the anomeric carbon during the transition state in α glycosidases, and that this may contribute to the fact that 5-fluoro-glycosyl fluorides trap intermediates for both α and β.[183, 230] It is believed that the use of **2.31** will have a similar effect on the MtGBE enzyme, allowing the observation of an intermediate.

In addition, it was also planned that *p*-nitrophenyl 6-deoxy-6-fluoro-α-D-glucopyranoside (**2.33**) be synthesised and tested as a potential inhibitor of MtGBE. Though the synthesis was completed, the compound was not able to be tested with the protein. The synthesis followed that of Card and Reddy who reported that the α anomer was not stable enough for characterisation.[231] In this case however, the α anomer was synthesised and characterised with both NMR spectroscopy and Mass spectrometry.

2.3.1 *p*-nitrophenyl 6-deoxy-6-fluoro-α-D-glucopyranoside

Diethylaminosulphur trifluoride (DAST) has long been used in the synthesis of fluorinated carbohydrates, due to its use in one-step mono- and difluorination reactions of unprotected carbohydrate substrates.[231-233] Therefore, DAST fluorination of *p*-nitrophenyl α-D-glucopyranoside (**2.32**), which requires no protection or deprotection steps and can provide the desired product in one step, is an attractive route to **2.33** (Scheme 2. 11). The procedure for this reaction was reported by Card

and Reddy, however they did not fully report on the synthesis of the α anomer, claiming it was unstable.[231] Despite this, the reaction was performed in order to obtain the material. This work was carried out by an undergraduate student Laura Jowett who contributed to this project as part of a Master's degree programme.



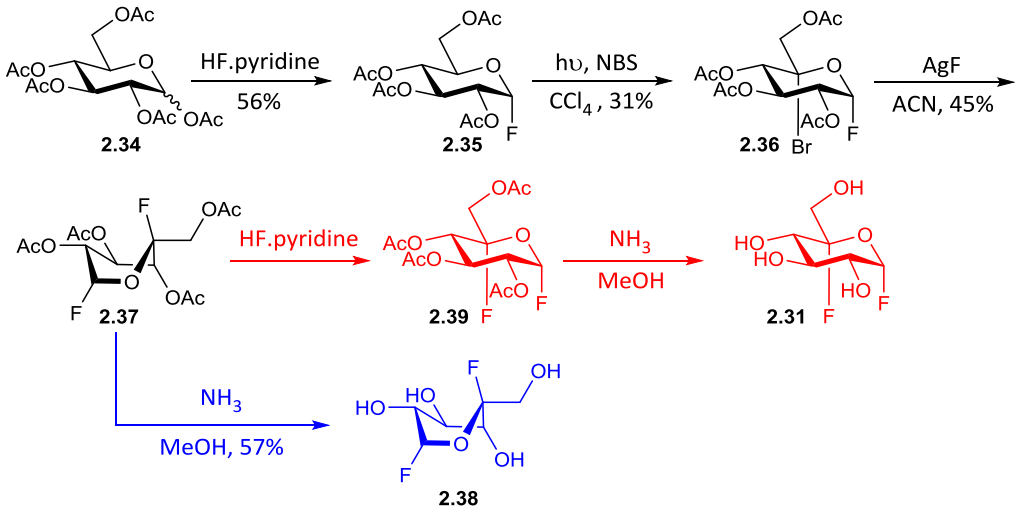
Scheme 2. 11 One-step synthesis of **2.33**, performed by Laura Jowett, following the procedure of Card and Reddy.[231]

The reaction was carried out in DCM with an excess of DAST (5.7 molar equivalents) to produce **2.33** in a 57% yield following purification *via* column chromatography. The DAST selectively fluorinated the 6-hydroxy position, because the S_N2 of the DAST intermediate by fluoride ions is faster on the primary position than on the secondary ones on the sugar ring. The compound was stable enough to allow for characterisation by both ^{1}H and ^{19}F NMR spectroscopy and mass spectrometry. A doublet in the ^{1}H NMR at 5.71 shows the presence of the α anomer due to the small coupling constant of 3.6 Hz, as would be expected due to the reaction having no effect on the anomeric position of the substrate. Furthermore, the position of the fluoride at the 6-position is confirmed due to the large coupling constants found in the multiplet arising from the CH_2 protons at position 6. Further evidence is seen in the ^{13}C DEPT-135 spectrum, where the CH_2 signal of C6 (δ = 97.9 ppm) is split due to coupling with the fluorine at this position.

2.3.2 5-fluoro-β-L-idopyranosyl fluoride

Initially, it was planned that 5-fluoro-α-D-glucopyranosyl fluoride (**2.31**) would be synthesised and tested as a potential inhibitor of MtGBE should a substrate be identified and a suitable assay developed. However, it was then decided that **2.31**, as a potential inhibitor, would be of good use in crystallographic studies of the protein and was therefore to be sent to collaborators to attempt co-crystallisation with MtGBE. However, due to low yields in the synthesis (see Scheme 2. 12), and therefore low

amounts of materials, it was the 5-fluoro-β-L-idopyranosyl fluoride (**2.38**) that was synthesised, as this could be achieved in the same synthesis but with 2 fewer steps than required for **2.31**. It was thought that the idosyl derivative, an epimer of glucose at position C-5, may also act as an inhibitor useful for co-crystallisation studies.



Scheme 2. 12 Synthetic route to both 5-fluoro-α-D-glucopyranosyl fluoride (**2.31**) and 5-fluoro-β-L-idopyranosyl fluoride (**2.38**) based on Thanna et al.[234]

The synthesis of both **2.31** and **2.38** has been reported by Thanna *et al.* as well as McCarter.[234, 235] Starting from the commercially available 1,2,3,4,6-penta-*O*-acetyl-D-glucopyranose (**2.34**), **2.38** was obtained after four steps, the first of which was to form the 2,3,4,6-tetra-*O*-acetyl-D-glucopyranosyl fluoride (**2.35**). This was achieved using HF.pyridine at room temperature overnight to exclusively produce **2.35**, exclusively as the α anomer in a 56% yield after purification by column chromatography. In this reaction, the anomeric effect (described in more detail in Chapter 1, section 1.2.2) is thought to be the main factor as to why only the α anomer is formed. This was confirmed by the ^{1}H NMR of **2.35** in which the anomeric signal at 5.76 ppm is a doublet of doublets with a large coupling constant (52.9 Hz) due to $^{2}J_{H1\text{-}F}$ and a smaller coupling constant (2.8 Hz) due to coupling between the equatorial anomeric proton and H2 ($^{3}J_{H1eq\text{-}H2ax}$).

The second step in the synthesis used light irradiation to facilitate photobromination of the C5 position of **2.35** to form **2.36**, a process described early on by Ferrier *et al.*[236, 237]

The authors proposed that the reaction proceeded by way of a tertiary radical formed at position C-5 which is then brominated from the axial direction.[237] Light was produced *via* a 400 W tungsten bulb which was used to heat a solution containing **2.35** and *N*-bromosuccinimide (NBS) in carbon tetrachloride to reflux for 9 hours. An initial attempt of this reaction was left overnight, however the following day, the solvent had entirely evaporated despite the use of a reflux condenser, leaving a brown residue on the flask which had to be discarded. Therefore, for all future repeats, the reaction was not left overnight again. The literature procedure that was followed reports a 47% yield[234] and the highest yield of **2.36** obtained in this work was 31%. This is in agreement with the fact that radical reactions are not generally high yielding due to the possibilities of many other side reactions occurring. In addition, starting material was still recovered, so the reaction did not reach completion. However, Thanna *et al.* described using 2,2'-Azobis(2-methylpropionitrile)[234] (AIBN), a radical initiator, which was not used in this case. Use of AIBN in future work may lead to reduced reaction time. Analysis of the ^{1}H NMR for **2.36** (Appendix A, A8) shows the lack of a C5 proton signal compared to the starting material, which, along with mass spectrometry, confirms installation of a bromine atom at this position in the ring.

After formation of **2.36**, halogen exchange with fluorine was then possible. This reaction had been reported by Withers and co. who used silver fluoride in acetonitrile as the fluorinating reagent.[228] This method was also utilised by Thanna *et al.* who reported a 68% yield.[234] However, in this case the highest yield of **2.37** achieved was 45%. The reaction had to be carried out in the dark due to the sensitivity of silver fluoride so it is possible that some of the reagent had decomposed before reacting. The reaction was carried out overnight and column chromatography allowed for purification of the desired product. The reaction causes inversion of the stereochemistry at C5, thus producing the epimer of the glucosyl product (idosyl). This is confirmed *via* comparison with the ^{1}H NMR of both the ido- and gluco- forms found in the literature, the former of which matched almost exactly with the data obtained in this work[235] (Appendix A, A9). The presence of a second fluorine atom in the sugar ring was confirmed by ^{19}F NMR of **2.37** which now showed two signals at -106.3 and -138.6 ppm in accordance with the literature data.[235] The coupling constants in the ^{1}H NMR

for **2.37** also suggest that the sugar ring is in boat conformation rather than a chair. The $^{3}J_{H2\text{-}H3}$ coupling constant, 8.3 Hz, suggests an approximate *anti*-periplanar relationship between H2 and H3. In addition there is a smaller coupling constant between H3 and H4 ($^{3}J_{H3\text{-}H4}$ = 1.7 Hz) suggesting a more gauche-like relationship between the two. Similar coupling constants were reported by Nelson when describing the structure of 2,3,4-tri-*O*-acetyl-D-xylano-1,5-lactone, which was suggested to adopt a $^{2}B_{5}$ boat-like conformation.[238] For this reason, **2.37** and **2.38** are depicted as having this $^{2}B_{5}$ conformation.

The final step to **2.38** was deacetylation for which a methanolic solution of ammonia was used. A solution of ammonia in methanol was used as a substitute for bubbling ammonium gas through the cooled solution, as reported in the literature.[234, 235] Basic deprotection of acetyls is a commonly used method[239] and so the yield of just 57% for **2.38** achieved following purification *via* column chromatography is considered low. Despite this, 1H NMR confirmed the presence of the desired product due to the absence of any methyl peaks.

Once **2.38** had been obtained, it was sent to collaborators for attempts at co-crystallisation.

2.4 Conclusions

This chapter has described the successful synthesis of GG (**1.36**), MalG (**1.38**), *p*-nitrophenyl 6-deoxy-6-fluoro-α-D-glucopyranoside (**2.33**) and 5-fluoro-β-L-idopyranosyl fluoride (**2.38**). In the case of GG, the majority of the synthesis followed that of Lourenço and co. with slight changes made to their procedures, mainly in the removal of the benzyl ether protecting groups.[186] This step proved troublesome, though improvements to the experimental technique led to eventual success. Overall, the yields reported in this work were not as high as those reported in the literature, though despite this, GG was obtained with enough material for testing in enzymatic assays. In addition, the synthesis of the glycerate acceptor used in the synthesis of GG was successfully synthesised from D-maltose. Adaptation of a procedure by Ladame and co. led to methyl (*R*)-(+)-2,2-dimethyl-1,3-dioxolane-4-carboxylate (**2.15**) directly from the aldehyde which was then converted to the desired acceptor (**2.9**).

With no literature precedent for MalG, a synthesis was planned which followed a similar path to that of GG in which a thioglycoside donor was formed before coupling with the glycerate acceptor. Though the synthesis that was initially planned had to be revised, the final synthetic route led successfully to MalG, though not as anomerically pure as desired. Whilst some α anomer was obtained, further purification techniques, such as HPLC, would be required to separate the α and β anomers completely. Samples of both pure α anomer and a mixture of anomers were taken through for testing with the enzyme.

In order to produce potential inhibitors of the enzyme MtGBE, two fluorinated compounds were prepared. The first, **2.33**, was prepared by project student Laura Jowett *via* a simple one-step DAST reaction. Following purification, the yield of this reaction was 57%, a good yield for this type of procedure, especially considering data on the α anomer had not been previously reported.

The second fluorinated compound obtained was **2.38**, which was synthesised following the works of both McCarter[235] and Thanna *et al*.[234] The synthesis contained low-yielding steps; the radical photobromination of position C5 and the subsequent substitution of the installed bromine atom with fluorine. Overall the low-yielding synthesis led to stopping at the idosyl sugar rather than continuing the synthesis to obtain the gluco form of the product. However, samples were sent to collaborators for co-crystallisation with the enzyme, though results of this have not yet been reported.

3. Expression and purification of *Mycobacterium tuberculosis* glucan branching enzyme (MtGBE)

3.1 Initial attempts at expression of recombinant MtGBE

3.1.1 Vector design and plasmid transformation

The plasmid containing the gene *Rv3031* was ordered from GenScript and consisted of a pET-28a(+) vector into which the codon optimised sequence for *Rv3031* was cloned using the restriction sites NdeI and XhoI. As seen in the plasmid map in Figure 3. 1, the gene sequence included a hexa histidine-tag (His_6-tag) as well as a thrombin cutting site between the His_6 -tag and the target protein. The pET-28a(+) system utilises the IPTG promotor system and also contains a kanamycin (Kan) resistance marker.

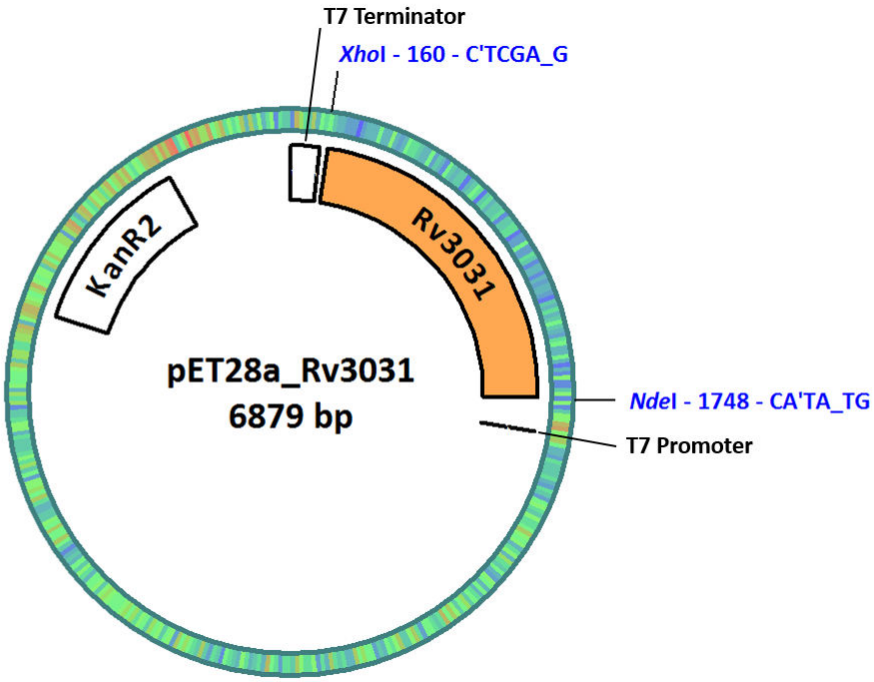


Figure 3. 1 Plasmid map of pET28a(+)::Rv3031 generated using pDRAW32 software.

The vector pET28a(+)::*Rv3031* was transformed into chemically competent *E. coli* BL21(DE3) cells (for preparation of cells, see method 8.2.1.3, chapter 8), *via* the heat shock technique (method 8.2.1.4, chapter 8) which proved successful upon the first attempt. Incubation of the transformants on agar plates containing Kan (50 µg/mL) led to the successful growth of many colonies.

In order to increase plasmid stocks, pET28a(+)::*Rv3031* was amplified and purified *via* a mini-prep technique using the Invitrogen™ ChargeSwitch® -Pro plasmid mini prep kit (see procedure 8.2.1.1, chapter 8). The purified plasmid DNA was then analysed by agarose gel electrophoresis (see procedure 8.2.1.2, chapter 8), the results of which are shown in Figure 3. 2. The samples analysed included the in-tact plasmid, as well as the products of digesting the plasmid with either NdeI, XhoI or both (for digestion procedure, see procedure 8.2.2.1, chapter 8). The fragments produced in the digest with both restriction enzymes appears as roughly 2000 bp long, which is in conjunction with the actual size of the gene sequence that was ordered (1715 bp). For the full sequence, see appendix B. To confirm that this was indeed the desired plasmid and that the competent cells had not introduced errors during amplification, purified plasmid was sent to Eurofins Genomics for DNA sequencing. The results confirmed the plasmid was the same sequence as the original that was ordered (results shown in appendix B).

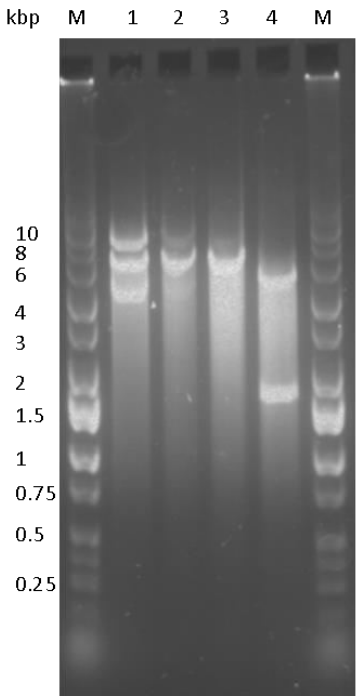


Figure 3. 2 1% agarose gel of NdeI and XhoI restriction digestion products. Lane 1: no restriction enzymes present; Lane 2: NdeI only; Lane 3: XhoI only; Lane 4: NdeI and XhoI.

3.1.2 Expression of MtGBE

Following transformation, attempts were then made to express the recombinant MtGBE protein. Initially, expression studies were performed on a 500 mL scale. To

determine the optimal conditions for protein expression, both the temperature of expression and the concentration of ITPG used in induction were varied.

During the overexpression of recombinant proteins in bacterial cells, it is possible for the tertiary structure to fold incorrectly. Misfolded proteins are recognised by the cell and are gathered into insoluble packages termed inclusion bodies (IBs).[240] When proteins are expressed in IBs they are difficult to isolate and refold. Therefore, it is desirable to prevent misfolding of the proteins during expression and one such approach to achieving this is to slow down expression. This can be achieved by lowering the temperature. After induction, differing temperatures of 18 °C and 37 °C overnight were used to compare levels of protein expression. Another method employed to slow expression was to lower the concentration of IPTG added upon induction. Concentrations of 0.1 mM, 0.5 mM and 1.0 mM IPTG were tested.

Following protein expression, the cells were harvested and lysed (see chapter 8, sections 8.2.1.7 and 8.2.1.8). The cells debris was then separated *via* centrifugation (12 krpm, 30 minutes, 4 °C) leaving behind all soluble proteins in the lysate.

In order to analyse the effect of varying the temperature and concentration of IPTG used in expression, the lysate, the insoluble fraction and the whole cell pellet before lysis from each experiment were analysed using SDS-PAGE and compared (Figure 3. 3). No increase in the amount of soluble protein was observed upon the addition of a lower concentration of IPTG. This is indicated by the weak band observed at around 55 kDa in each of the wells corresponding to the cell lysate in each experiment. The results show that varying the temperature of expression also has no effect on the amount of soluble protein expressed.

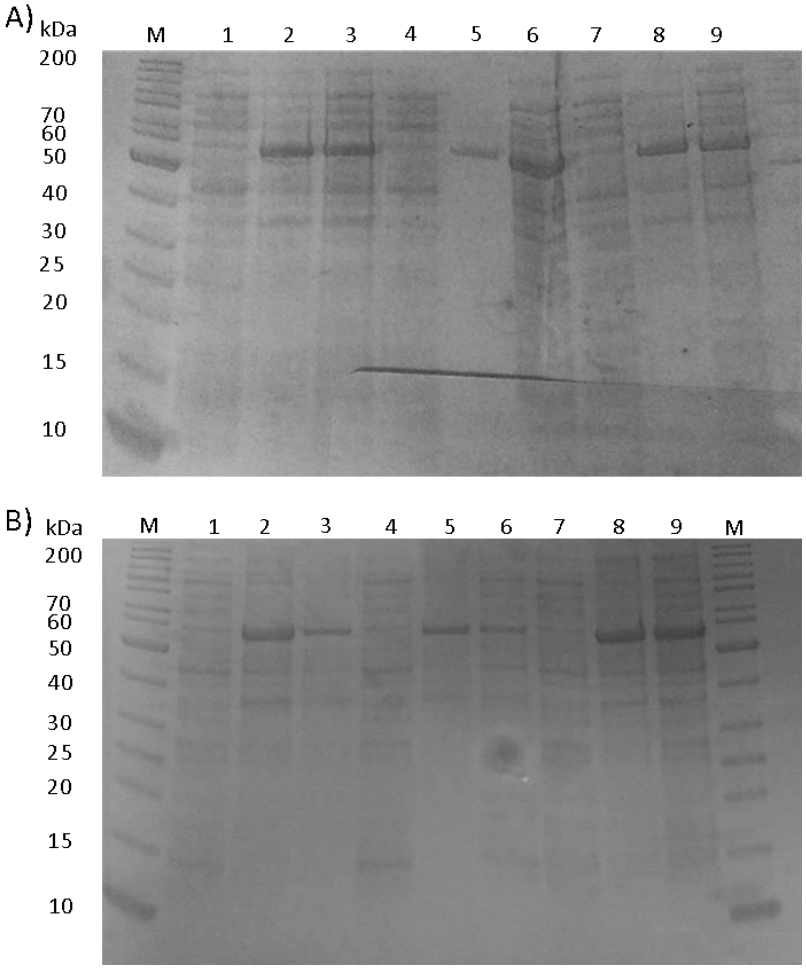


Figure 3. 3 SDS PAGE results of initial expression studies of MtGBE. A: 18 °C expression; B; 37 °C expression; Lanes 1-3: 0.1 mM IPTG, lysate, insoluble, whole cell; Lanes 4-6: 0.5 mM IPTG, lysate, insoluble, whole cell; Lanes 7-9: 1.0 mM IPTG, lysate, insoluble, whole cell; M: marker. MtGBE = 57 kDa.

From the results of the SDS-PAGE in Figure 3. 3, there appears to be no significant increase in the amount of soluble protein in any of the experiments. The corresponding band from the experiment using 0.1 mM IPTG is slightly stronger than that of the other experiments, but the majority of the protein remains in the insoluble fraction.

In an attempt to obtain a small amount of protein, a 1.0 L scale expression was carried out at 18 °C overnight using 0.1 mM IPTG for induction. An attempt to purify the soluble protein from this experiment was then made using a nickel affinity column (see procedure 8.2.1.10, chapter 8) (see Figure 3. 4). In this first purification attempt, the

fractions collected at the time of high imidazole concentration were analysed by SDS-PAGE (Figure 3. 5).

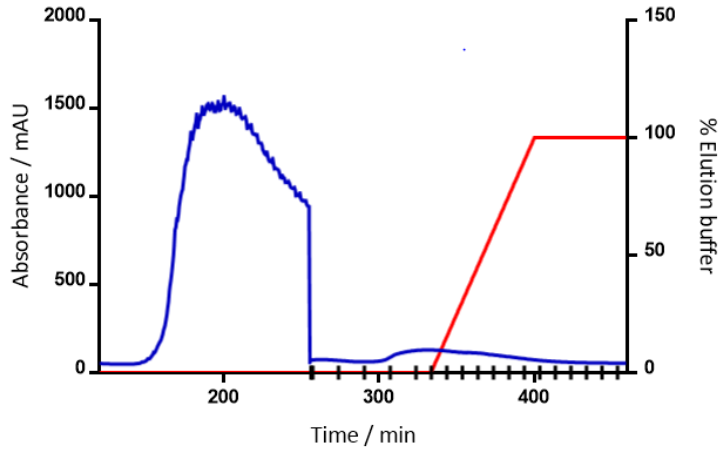


Figure 3. 4 Ni-NTA purification of MtGBE protein: UV-vis trace (blue); % elution buffer (red); fractions (black).

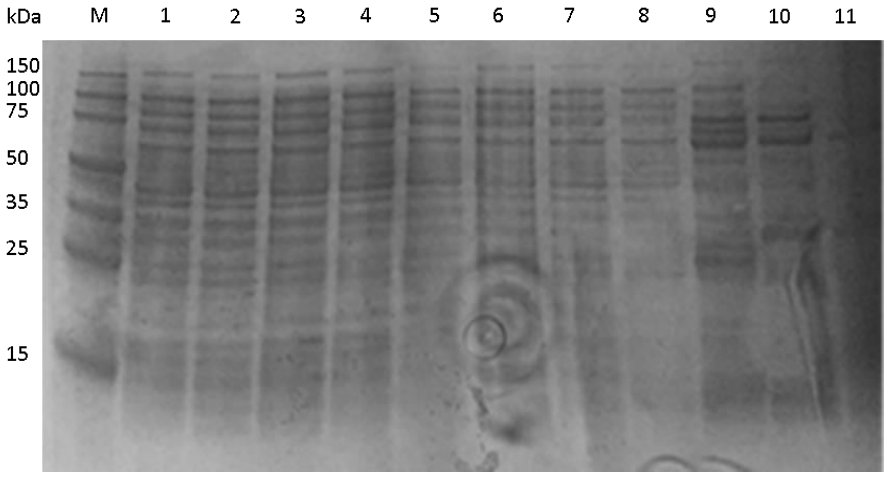


Figure 3. 5 SDS-PAGE of eluted fractions from the first attempt at purifying MtGBE protein. Lanes 1-11: fractions 1-11; M: marker. MtGBE = 57 kDa.

The results of the SDS-PAGE shown in figure 3.5 suggest that the purification was unsuccessful. Although the desired band appears in the fractions, it does so with a lot of other bands, meaning the fractions contain a mix of proteins still. As there was a low

level of target protein, a larger number of *E. coli* proteins could bind to the column, thus leading to impure fractions.

3.2 Attempts to solubilise MtGBE

As described in section 3.1.2, initial attempts to express MtGBE in soluble form were unsuccessful. Therefore, solubilisation techniques were utilised in an attempt to solve this issue. The first technique employed was the recovery of misfolded protein from IBs. However this method has disadvantages and so co-expression with chaperone proteins was also attempted.

3.2.1 Recovery of proteins from inclusion bodies

During the over-expression of recombinant proteins there is a chance that the target protein can become mis-folded during synthesis and will either be degraded or form insoluble aggregates with other incorrectly folded proteins.[241, 242] These aggregates come together in one site within the cell to form IBs (illustrated in Figure 3. 6), the composition of which has been shown to be heterologous, containing proteins in varying states of folding or degradation.[241, 242]

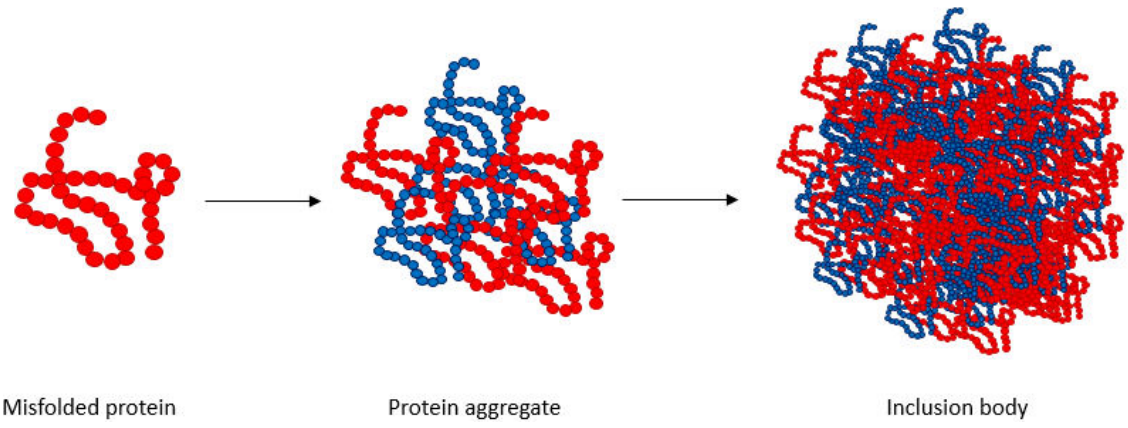


Figure 3. 6 Cartoon depicting one way in which IBs are formed. Figure reconstructed from Kopito.[242]

As inclusion bodies can contain large amounts of the desired protein, there have been many reports on the extraction and re-folding of proteins from IBs.[243-249] Commonly used agents used in the solubilisation of IB proteins are shown in Figure 3. 7 and include urea (**3.1**) and guanidine hydrochloride (**3.2**), known as chaotropic agents, as well as surfactants such as sodium dodecyl sulphate (SDS) (**3.3**) or sodium lauroyl

sarcosinate (sarkosyl) (**3.4**). These reagents are typically used in high concentrations to solubilise the protein and are then subsequently removed in order to affect re-folding of the protein.[243] However, high concentrations of chaotropic agents have been shown to increase the formation of aggregates during the refolding process and so milder solubilisation techniques are preferred.[244]

Figure 3. 7 Compounds used in the solubilisation of proteins from IBs; A: chaotropic agents urea and guanidinium hydrochloride; B: surfactants SDS and sarkosyl.

In this project the recovery of MtGBE from IBs was attempted using sarkosyl. A low concentration (0.2 %) of sarkosyl was used, following a procedure reported by Klutts et al., providing mild conditions to allow for recovery of correctly folded protein.[250]

3.2.1.1 Solubilisation of MtGBE from inclusion bodies using sarkosyl

Firstly, MtGBE protein was expressed as per procedure 8.2.2.3 (chapter 8), using 1.0 mM IPTG to induce expression which was then performed at 37 °C overnight. 2.0% sarkosyl solution (final concentration 0.2%) was added at various stages of protein extraction, illustrated in Figure 3. 8, in order to determine at which point the addition of sarkosyl would give the highest yield of soluble protein. The first stage at which sarkosyl was added was once the cells had been harvested and were re-suspended in lysis buffer prior to sonication and subsequent centrifugation (pre-lysis). The sarkosyl was added as part of the lysis buffer itself. The second stage was after the intact cells had been sonicated and centrifuged to yield the insoluble cell debris. This insoluble fraction was re-suspended in the presence of 0.2% sarkosyl (post-lysis) *via* stirring at 4 °C overnight. A control experiment in which no sarkosyl was added was run in parallel to the others.

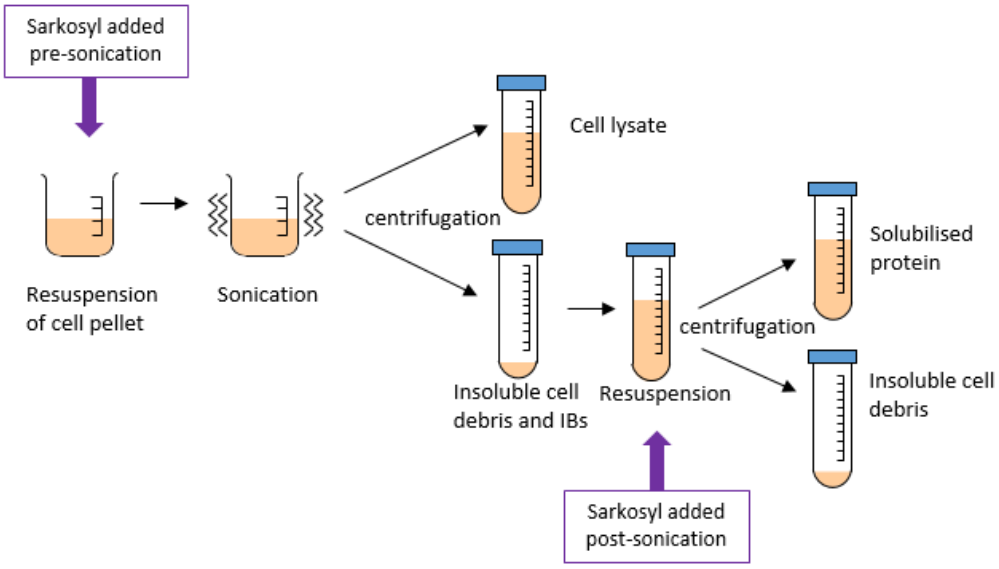


Figure 3. 8 Representation of the different points at which sarkosyl was added in an attempt to solubilise MtGBE from IBs.

In the pre-lysis sarkosyl experiment, sonication and centrifugation led to two fractions, the cell lysate and the insoluble fraction. In the post-lysis sarkosyl experiment, the resultant resuspension was centrifuged again to yield fractions termed the soluble (supernatant) and insoluble (pellet) fractions. These fractions, as well as the whole cell fractions from each experiment, were analysed by SDS-PAGE alongside the whole cell, cell lysate and insoluble fraction from the control experiment in order to assess the effects, if any, of sarkosyl addition.

Figure 3. 9 shows that very low levels of protein were isolated in the soluble fraction. The experiment in which sarkosyl was added pre-lysis also shows low levels of soluble protein. However, post-lysis addition of sarkosyl shows a strong band of desired protein in the soluble fraction. The results therefore suggest that the best point at which to add the sarkosyl is once the insoluble cell debris and IBs have already been separated from the crude cell extract, which can be re-suspended in the presence of sarkosyl in order to solubilise the protein.

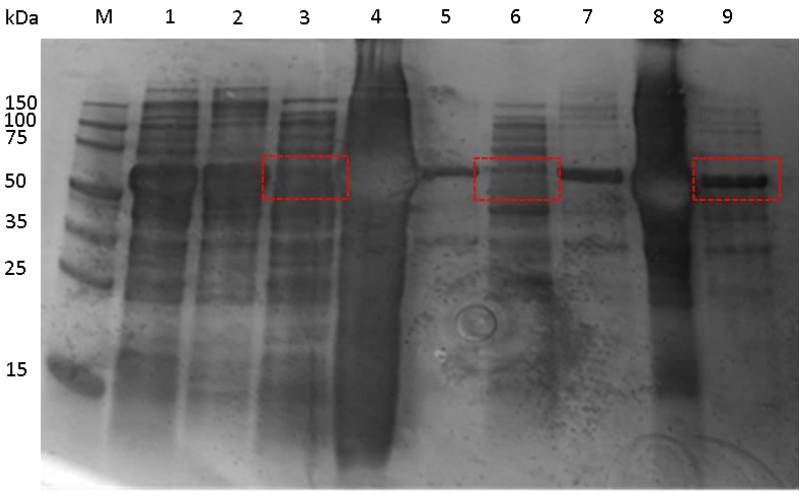


Figure 3. 9 SDS-PAGE of the effect of sarkosyl and the stage at which it is added. Red boxes highlight soluble MtGBE. Lanes 1-3: no sarkosyl added, whole cell, insoluble, soluble; Lanes 4-6: pre-sonication addition of sarkosyl, whole cell, insoluble, soluble; Lanes 7-9: post-sonication addition of sarkosyl, whole cell, insoluble, soluble; M: marker. MtGBE = 57 kDa.

As the solubilisation of the insoluble MtGBE from inclusion bodies using sarkosyl appeared promising in initial experiments, solubilisation was combined with Ni-NTA purification. In this experiment, the insoluble cell debris fraction obtained from protein expression of MtGBE was resuspended in 20 mL deionised water before being centrifuged (12,000 rpm, 30 minutes, 4 °C). This was to remove any contaminant proteins that were already soluble. Next, the pellet was resuspended in solubilisation buffer (chapter 8, section 8.2.6.12). The suspension was then divided into two and both portions were gently mixed on a rocker at 4 °C. One portion was mixed in this way for 1 hour, the other for 4 hours. Once mixing was complete, the protein solutions were dialysed against Tris buffer (20 mM Tris HCl pH7.4, 0.5 M NaCl) in order to remove the sarkosyl. To remove any particulates, the samples were centrifuged before SDS-PAGE analysis was run to determine if purification by Ni-NTA affinity column was feasible.

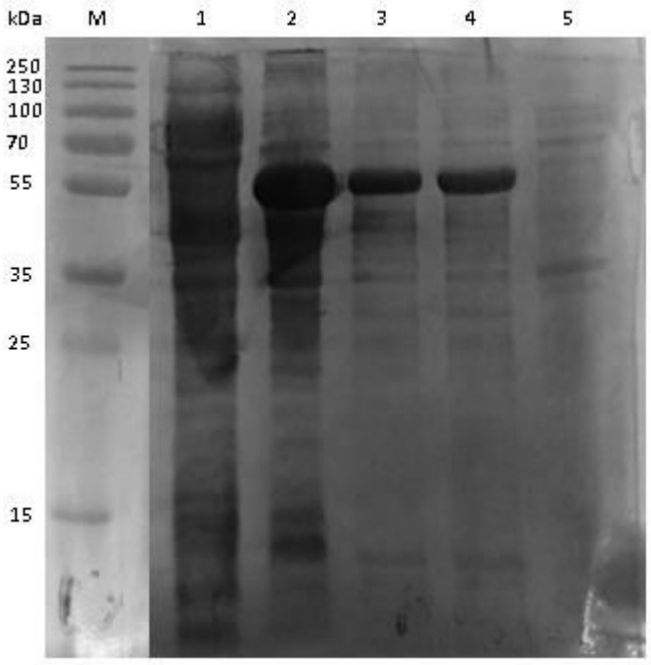


Figure 3. 10 SDS-PAGE of solubilisation of MtGBE from IBs. Lane 1: cell lysate; Lane 2: Insoluble cell debris after lysis; Lane 3: solubilisation on 0.2% sarkosyl, 1 hour mixing; Lane 4: solubilisation on 0.2% sarkosyl, 4 hours mixing; Lane 5: supernatant after washing of IBs with water; M: marker. MtGBE = 57 kDa.

The SDS-PAGE results in Figure 3. 10 show thick bands of soluble protein of a mass between 55-72 kDa, indicative of MtGBE protein, for both the 1 and 4 hour solubilisation experiments. As impurities were still present, the sample was further purified by Ni-NTA affinity column. The protein solutions were loaded onto the Ni-NTA resin and eluted using a gradient of 0-0.5 M imidazole. In each case there was a small peak shortly after the initiation of the elution gradient. These fractions were analysed by SDS-PAGE, however, the analysis showed no protein was present in any of the eluted fractions. This may have been due to incorrect folding of the protein upon renaturation (dialysis to remove the sarkosyl). If the protein were not folded correctly, there is a possibility the His_6-tag may have ended up 'hidden' within the protein structure and not able to bind to the nickel resin. Therefore, this method of protein solubilisation was deemed unsuccessful and other methods of obtaining soluble protein were explored. Due to the lack of requirement of a solubilisation and refolding process, co-expression with chaperone proteins was chosen as an alternative technique.

3.2.2 Molecular chaperone proteins

In the highly saturated environment of the cell, chaperone proteins can assist in the correct folding of large multi-domain and complex proteins.[251] Two chaperone systems that have been widely studied are the DnaK-DnaJ-GrpE and GroEL-GroES systems, the latter of which is used in this project.[252] GroEL belongs to the family of group I chaperonins and works in conjunction with the co-chaperonin GroES to aid in the correct folding of proteins post translation.[253] GroEL acts upon proteins between 20-60 kDa in size and has an open barrel-like structure in which misfolded proteins can enter.[251] The misfolded proteins generally present hydrophobic areas on the surface which bind to hydrophobic residues in the apical region of GroEL.[251] Following substrate binding, GroES binds, acting as a lid and forming a closed cage structure, the interior of which is hydrophilic and provides an ideal environment in which the substrate protein can refold correctly (see figure 3. 11).[253] The binding of GroES is reversible and regulated by ATP.[253] Once GroES leaves, opening the lid of the cage, the substrate protein is released; should the substrate still not be folded correctly, it can bind again for further attempts at refolding in the chaperonin system.[251]

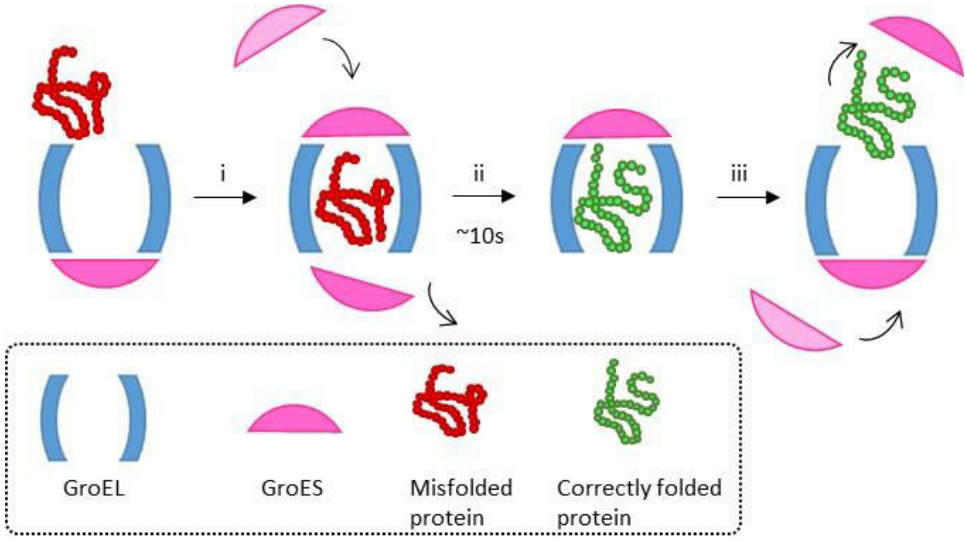


Figure 3. 11 simplified representation of how the GroEL-GroES system aids in protein refolding. i) misfolded proteins enter the barrel-like GroEL via hydrophobic interactions; ii) GroES binds forming a 'lid', the protein is held within a hydrophilic environment for roughly 10 seconds during which refolding can occur; iii) correctly folded protein can now leave as the GroES 'lid' is removed. Adapted from Hartl et al.[251]

In 1998, Nishihara *et al.* reported the construction of plasmid systems that allowed for independent expression of the GroEL-GroES system and today, plasmids of a similar function are commercially available. One such example, the pGro7 plasmid, was obtained from Takara Bio Inc. and contains the genes encoding GroEL and GroES under the control of an *araB* promotor which allows for the induction of protein expression in the presence of L-arabinose (L-ara). The pGro7 plasmid is 5400 base pairs in length and contains a chloramphenicol (Chlor) resistance marker. This was co-transformed with pET28a(+)::*Rv3031* into *E. coli* BL21(DE3) chemically competent cells for co-expression of the chaperone system with MtGBE.

3.2.2.1 Co-transformation of pGro7 and pET28a(+)::*Rv3031* plasmids

The pGro7 plasmid was initially amplified and purified. It was then transformed into chemically competent *E. coli* BL21(DE3) cells. In this case, the resistance marker was Chlor and so transformants were grown on agar plates containing 30 μg/mL Chlor. Though few, transformation resulted in the growth of colonies, one of which was used to grow an overnight culture for use in a plasmid mini prep. Analysis using a NanoDrop® ND-1000 spectrophotometer suggested a plasmid concentration of 189 ng/μL.

The pGro7 plasmid was co-transformed into BL21 (DE3) cells with the pET28a(+)::*Rv3031* plasmid described previously. Two methods of co-transformation were trialled; the first involved transforming both plasmids into the competent cells at the same time, whilst the second involved transforming one plasmid, converting the successful transformants into competent cells and then transforming the second plasmid into the new competent cells (see procedure 8.2.2.5, chapter 8). The second method led to a higher number of colonies grown on the agar plate, suggesting it was more successful than the first method of co-transformation.

In order to confirm the success of the co-transformation of the chaperone plasmid with the vector containing the *Rv3031* gene, plasmid amplifications and purifications were carried out on BL21 cells transformed with either pET28a(+)::*Rv3031* or co-transformed with both plasmids. The purified plasmid DNA from each experiment was then analysed *via* agarose gel electrophoresis.

Initially, *E. coli* BL21(DE3) cells were transformed with pET28a(+)::*Rv3031*. A co-transformation was also carried out in which both plasmids were transformed in one experiment. Colonies from each transformation were used to grow an overnight culture (conditions detailed in table 3.1) which was used in a plasmid mini prep.

Table 3. 1 Conditions used in the experiments to confirm success of co-transformation

Plasmids transformed	**Antibiotics in overnight media (concentration)**
pET28a(+)::*Rv3031*	Kanamycin (30 µg/mL)
pET28a(+)::*Rv3031* + pGro7	Kanamycin (30 µg/mL) and chloramphenicol (20 µg/mL)

Once the plasmids were isolated, they were then analysed using agarose gel electrophoresis, the results of which are shown in figure 3. 12.

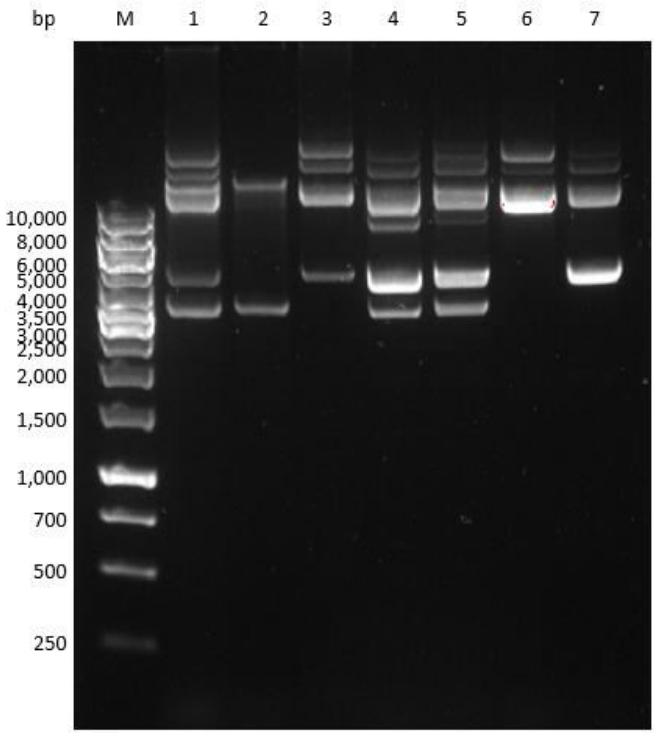


Figure 3. 12 Agarose gel of pET28a(+)::Rv3031 transformation and pET28a(+)::Rv3031 co-transformation with pGro7. Lane 1: pET28a(+)::Rv3031 and pGro7 mix of stock plasmids; Lane 2: pGro7 stock plasmid; Lane 3: pET28a(+)::Rv3031 stock plasmid; Lane 4 + 5: co-transformation; Lane 6 + 7: pET28a(+)::Rv3031 transformation; M: marker.

Lanes 1-3 are samples of the plasmid stocks used to transform the cells. Lane 1 contains a 50:50 mix of each plasmid whilst Lanes 2 and 3 contain samples of the pGro7 stock and the pET28a(+)::*Rv3031* stock respectively. Lanes 4 and 5 represent plasmid DNA purified from colonies that were co-transformed and each contains the same bands as the mixed standard suggesting both plasmids are present in each colony. Lanes 6 and 7 contain purified plasmid DNA from colonies that were only transformed with pET28a(+)*::Rv3031*. Each lane contains bands that are also present in the reference sample, though Lane 6 seems to contain more supercoiled DNA, as indicated by the strong band in the 10,000 bp region. In contrast, the sample in Lane 7 appears to contain more single stranded circular DNA, as indicated by the strong band that appears in the 5,000 bp region. Overall, this analysis strongly suggests that the co-transformation of the target gene and the chaperone plasmid was successful.

3.2.2.2 Small scale co-expression of MtGBE with GroEL-GroES chaperone proteins

Although it produced fewer colonies, the first co-transformation method was faster than the second method of transforming one plasmid after the other. Therefore colonies grown from the first transformation method were used in co-expression experiments.

A 1.0 L culture of cells was used to co-express MtGBE with the chaperones with 0.5 mM IPTG used in induction. Protein was expressed at 18 °C, 180 rpm overnight. Cells were then harvested and lysed. Analysis by SDS-PAGE, shown in figure 3. 13, showed a thick band at around 60 kDa suggesting soluble MtGBE was present.

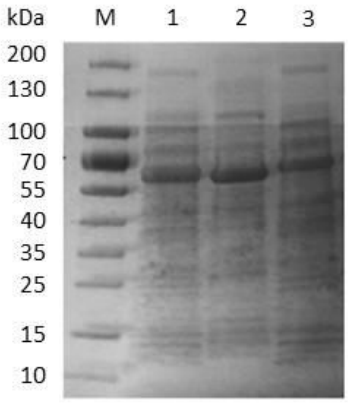


Figure 3. 13 SDS-PAGE of first attempt at the co-expression of MtGBE and the GroEL-GroES chaperone system. Lane1: whole cell; lane 2: insoluble; Lane 3: lysate; M: marker. MtGBE = 57 kDa.

As there was soluble protein present, an attempt was made to purify it in order to obtain MtGBE for testing. The soluble cell lysate was loaded onto a Ni-NTA affinity column and eluted over a gradient of 0-0.5 M imidazole over 60 mL (figure 3. 14). Although the UV trace from the purification did not show a large peak for protein elution, a small peak was observed that may have been the target protein. To confirm, the fractions collected were analysed *via* Bradford assay (see procedure 8.2.1.11, chapter 8) in order to determine the protein concentration of each fraction. The results of the Bradford assay are shown in figure 3. 15.

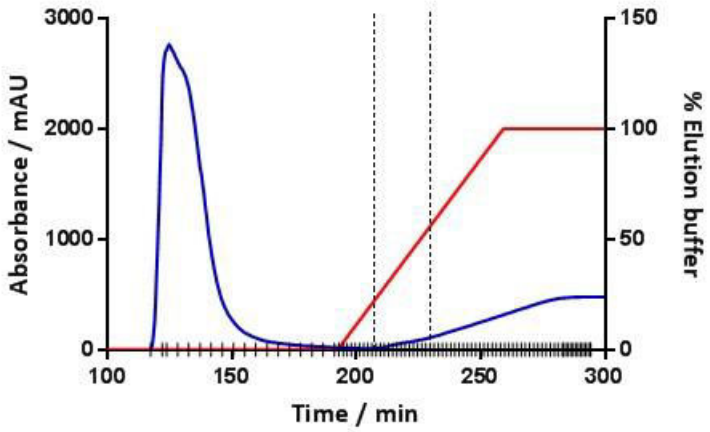


Figure 3. 14 Ni-NTA purification of MtGBE from first attempted co-expression: Absorbance (blue); % elution buffer (red); fractions (black). Dashed lines show region in which fractions were collected and pooled.

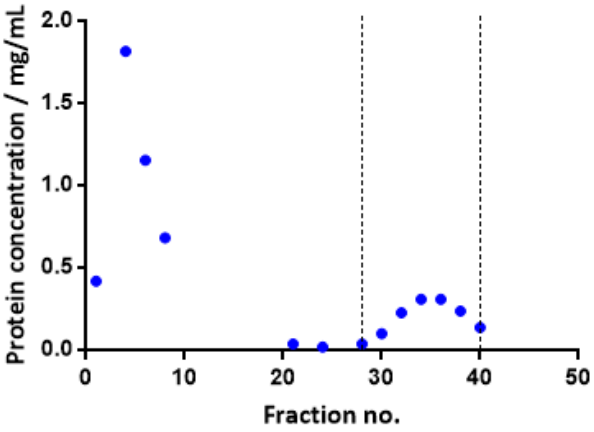


Figure 3. 15 Protein concentration of fractions analysed *via* Bradford analysis, dotted lines indicate fractions pooled and dialysed.

The Bradford analysis suggested that protein was present in fractions 28-40, and a cross section of these fractions was analysed *via* SDS-PAGE, as shown in figure 3. 16.

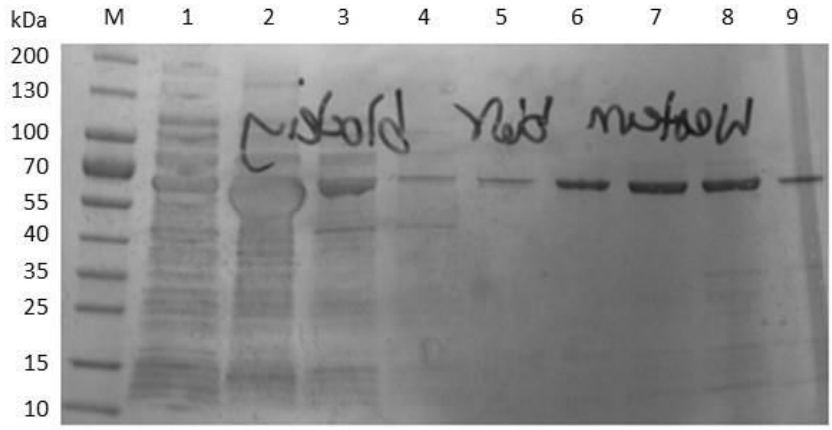


Figure 3. 16 SDS-PAGE analysis of fraction from purification of first co-expression of MtGBE. Lane 1: lysate; Lane 2: insoluble; Lane 3: fraction 4; Lane 4: fraction 6; Lane 5: fraction 24; Lane 6: fraction 30; Lane 7: fraction 34; Lane 8: fraction 38; Lane 9: fraction 40; M: marker. MtGBE = 57 kDa.

As the SDS-PAGE result showed pure protein of the desired mass, fractions 28-40 were pooled and dialysed to remove the imidazole and high salt content of the elution buffer. The protein solution was then concentrated *via* spin concentration using Amicon® Ultra – 15 (Merk) centrifugal units (10 kDa MWCO) at 4,000 rpm, 4 °C for runs of 10 minutes until the desired volume of 10 mL was met. Subsequent Bradford analysis suggested a final concentration of 0.5 mg/mL. The overall yield of this expression and purification was 5 mg protein per litre culture (5 mg/L). This is a low yield though unsurprising given the low solubility of the expressed protein.

Collaborators from the group of Prof. Kyeong Kyu Kim at Sung Kyun Kwan University (SKKU) were working on the crystallisation of MtGBE. Communications from them suggested a new method of co-expression in which the L-Ara was added to the overnight culture rather than in the large scale growth. Therefore, this method was trialled in which 600 µg/mL L-Ara was added to the overnight culture. 2.5 mL was used to inoculate 1.0 L LB media and the protein was expressed at 18 °C overnight. Once the cells had been harvested and lysed, the cell lysate was purified by Ni-NTA affinity column. No peak was observed in the UV trace and so this method was deemed unsuccessful. However, this did suggest that adding L-Ara to the overnight culture alone was not enough to express soluble MtGBE.

3.2.2.2.1 Investigations into the effects of L-Ara concentration on the amount of soluble MtGBE protein expressed

Given that there was a large amount of insoluble protein of the size expected for MtGBE in the SDS-PAGE analysis of the co-expression, it was thought that increasing the volume of chaperone proteins may also increase the amount of solubilised protein during expression. As the expression of the chaperones was moderated by the addition of L-Ara to the culture, an investigation was carried out to determine if increases in the concentration added also increased the amount of soluble target protein.

To investigate the effect of increasing the concentration of L-Ara in the growth medium, three experiments were carried out where three different concentrations of L-Ara were present in the growth during co-expression: 600 µg/mL, 700 µg/mL and 800 µg/mL. Each growth gave approximately the same mass of pellet (600 = 0.47 g, 700 = 0.52 g and 800 = 0.47 g) suggesting that cell growth was not affected by the increase in presence of L-Ara. Once lysed *via* sonication, the whole cell extract, cell lysate and insoluble cell debris of each experiment was analysed *via* SDS-PAGE (figure 3. 17).

Although each lane on the gel is saturated, it can be seen that the addition of 700 and 800 µg/mL of L-Ara yields a larger band in the soluble cell lysate between 56 and 72 kDa. This may be due to the increase in presence of GroEL, which is itself 57 kDa in size.[254] However, the insoluble fractions of both the 700 and 800 µg/mL experiments show a smaller band at 56-72 kDa, suggesting that more insoluble MtGBE is being converted to soluble protein. Therefore, as a result of this experiment, future co-expression of MtGBE contained 800 µg/mL L-Ara in the large scale culture.

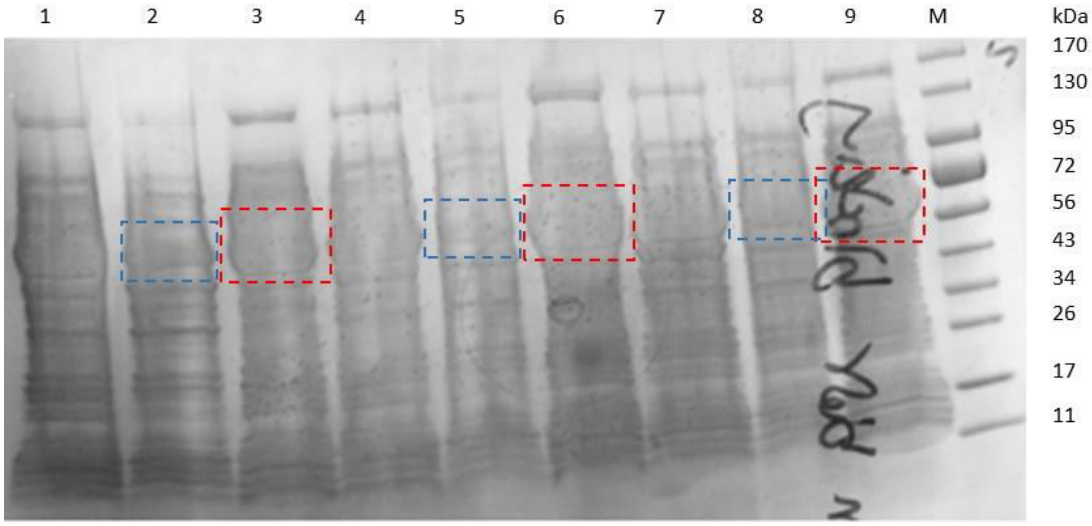


Figure 3. 17 SDS-PAGE of the effects of increasing L-Ara concentrations in co-expression. Lanes 1-3: 600 µg/mL L-Ara, whole cell, insoluble, lysate; Lanes 4-6: 700 µg/mL L-Ara, whole cell, insoluble, lysate; Lanes 7-9: 800 µg/mL L-Ara, whole cell, insoluble, lysate; M: marker. Insoluble (blue box) and soluble (red box) fractions for each experiment highlighted for clarity. MtGBE = 57 kDa.

A final small scale co-expression (1.0 L) was carried out under these conditions to check the levels of soluble protein expressed and determine a purification protocol. The cell lysate was loaded onto a Ni-NTA column and the protein was eluted over a gradient from 0-0.5 M imidazole over 150 mL. A Bradford analysis was carried out on each fraction collected to give an indication of protein concentration (figure 3. 18).

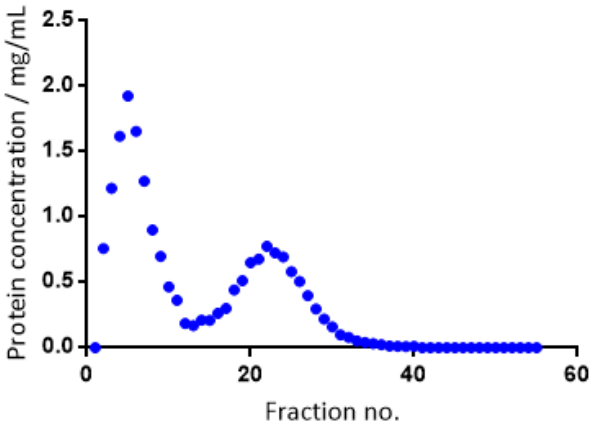


Figure 3. 18 Protein concentration of each fraction from purification of MtGBE 1.0 L co-expression with 800 µg/mL L-Ara as determined by Bradford analysis.

SDS-PAGE was then carried out to analyse fractions of interest (figure 3. 19).

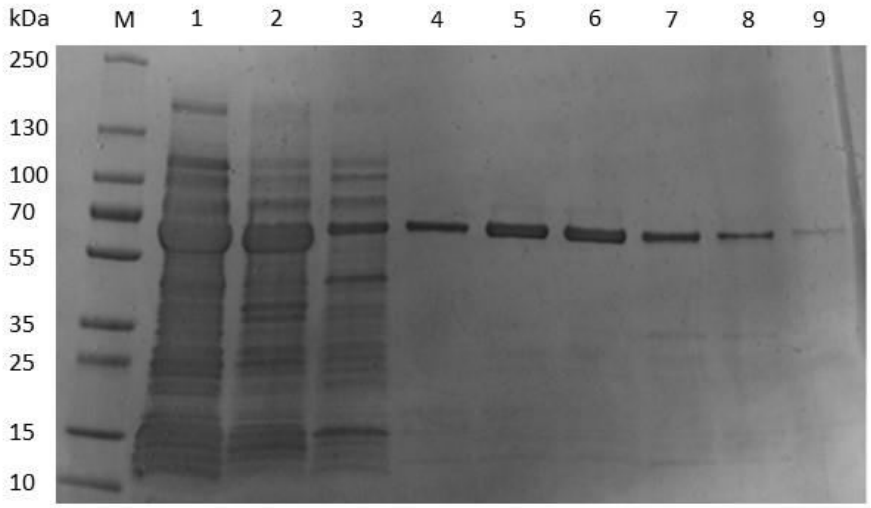


Figure 3. 19 SDS-PAGE analysis of purification of MtGBE. Lane 1:lysate; Lane 2: insoluble; Lane 3: fraction 5; Lane 4: fraction 15; Lane 5: fraction 22; Lane 6: fraction 24; Lane 7: fraction 30; Lane 8: fraction 34; Lane 9: fraction 44; M: marker. MtGBE = 57 kDa.

As seen in the SDS-PAGE, fractions 15-44 contain pure protein with a mass between 55 and 70 kDa, suggestive of MtGBE. Therefore, fractions 15-30 (those showing the highest concentrations of protein) were pooled and dialysed to remove imidazole and high salt concentrations present in the elution buffer. During dialysis, the buffer was also exchanged from Tris-HCl to HEPES buffer (50 mM, pH7) as this was the buffer to be used in future assays. The protein solution was then concentrated *via* spin filtration

using Amicon® Ultra – 15 (Merk) until the desired volume of 1.0 mL was reached. Bradford analysis suggested a final protein concentration of 14 mg/mL. This gave an overall protein yield of 14 mg/L, an improvement on the previously obtained 5 mg/L. These results indicate that increased L-Ara concentration does in fact lead to a higher protein yield.

3.2.2.3 Large scale co-expression and purification of MtGBE

Once the conditions for co-expression had been optimised, a large scale growth of 5.0 L was carried out. L-arabinose was present in the expression culture at a concentration of 800 µg/mL and 0.5 mM IPTG was used in induction. Protein was expressed at 18 °C, 170 rpm overnight. Cells were then harvested and lysed *via* sonication. The soluble cell lysate was then obtained and loaded onto a Ni-NTA affinity column. The His_6-tagged target protein was eluted over a gradient from 0-0.5 M imidazole. Bradford analysis of each fraction collected during elution (figure 3. 20) suggested a large quantity of eluted protein in fractions 11-16 which were subsequently analysed *via* SDS-PAGE (figure 3. 21).

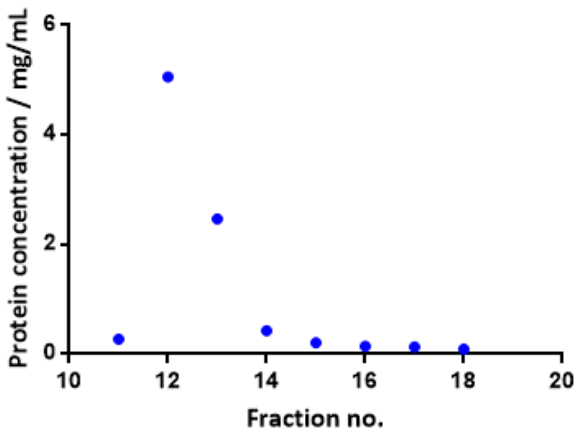


Figure 3. 20 protein concentration of fraction from 5.0 L co-expression of MtGBE as determined by Bradford analysis.

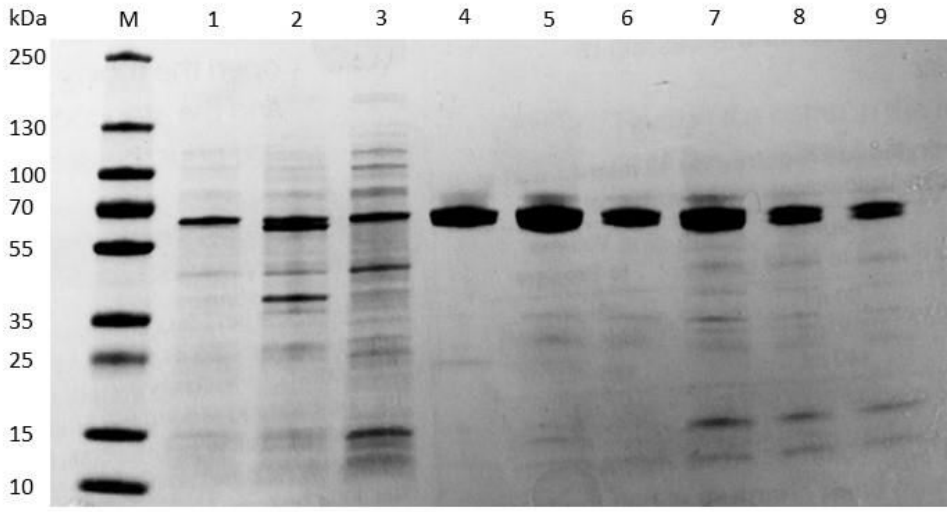


Figure 3. 21 SDS-PAGE of 5.0 L co-expression of MtGBE. Lane 1: S; Lane 2: In; Lane 3: fraction 2; Lanes 4-9: fractions 11;-16; M: marker. MtGBE = 57 kDa.

The Bradford analysis and SDS-PAGE results in figures 3.20 and 3.21 led to fractions 11-16 being pooled and dialysed against HEPES buffer (50 mM, pH 7.4) for further concentration and use in assays. The protein solution was concentrated to a final concentration of 4.0 mg/mL and stored at -80 °C for future use.

3.2.2.4 Experiment to confirm GroEL is not purified alongside MtGBE

SDS-PAGE analysis is unable to differentiate between proteins of a similar size. As MtGBE and GroEL are both 57 kDa in mass, experiments were designed to determine if the presence of GroEL was affecting analysis of SDS-PAGE gels.[254] To investigate this, an experiment in which three protein expressions were performed; one in which

MtGBE was expressed on its own, one where the chaperone proteins were expressed on their own and one where MtGBE was co-expressed with the chaperone system. A negative control was run in parallel in which neither the gene for MtGBE nor that for the chaperone system was present, i.e. 'blank' *E. coli* BL21(DE3) cells were grown under same the expression conditions. For conditions for each expression, see chapter 8, section 8.2.2.9. The soluble cell lysate and insoluble cell debris from each expression were analysed *via* SDS-PAGE, the results of which are shown in figure 3. 22.

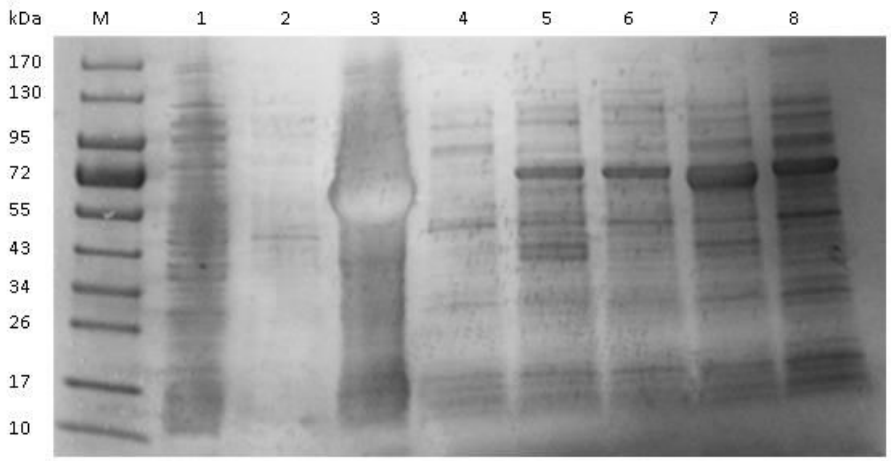


Figure 3. 22 SDS-PAGE analysis of the effect chaperones have on the solubility of MtGBE. Lanes 1 + 2: BL21 blank, insoluble, lysate; Lanes 3 + 4: MtGBE only, insoluble, lysate; Lanes 5 + 6: chaperones only, insoluble, lysate; Lanes 7 + 8: co-expression, insoluble, lysate; M: marker. MtGBE = 57 kDa.

From the thick band that appears in the 55-72 kDa region in the soluble fraction of the co-expression, it can be seen that soluble protein is being expressed that is not normally present when MtGBE is expressed on its own. This could be soluble MtGBE appearing but an identical band also appears in the soluble fraction of the experiment in which the chaperone proteins were expressed with no MtGBE present. In order to investigate further, more in depth experiments were carried out in which the same set of protein expressions were performed on a larger scale in order to then purify the cell lysates obtained *via* Ni-NTA affinity column. Whilst in theory, the chaperone proteins, which do not contain a His-tag, would not be eluted from the column alongside MtGBE, it was still considered prudent to confirm that this is indeed the case in order to ensure the protein being purified was in fact MtGBE.

For the large scale experiments, transformations, overnight cultures and expression conditions were the same as in the small scale experiments. This time the larger culture was 2.0 L in volume. Cell lysates were loaded onto a Ni-NTA column and protein was eluted *via* a gradient of 0-0.5 M imidazole over 150 mL.

The UV absorbance traces from each purification are shown in figure 3. 23. For the purification of MtGBE expressed on its own, a small peak is observed shortly after the gradient begins. For the co-expression experiment, this peak appears at the same time, though much larger in size. For the expression of the chaperone proteins, a peak is also observed shortly after elution begins, this appears to be approximately the same size as that from the MtGBE only expression and so cannot been seen as insignificant in this context. The peak is split into two overlapping signals and so SDS-PAGE analysis was done on all fractions across the whole signal.

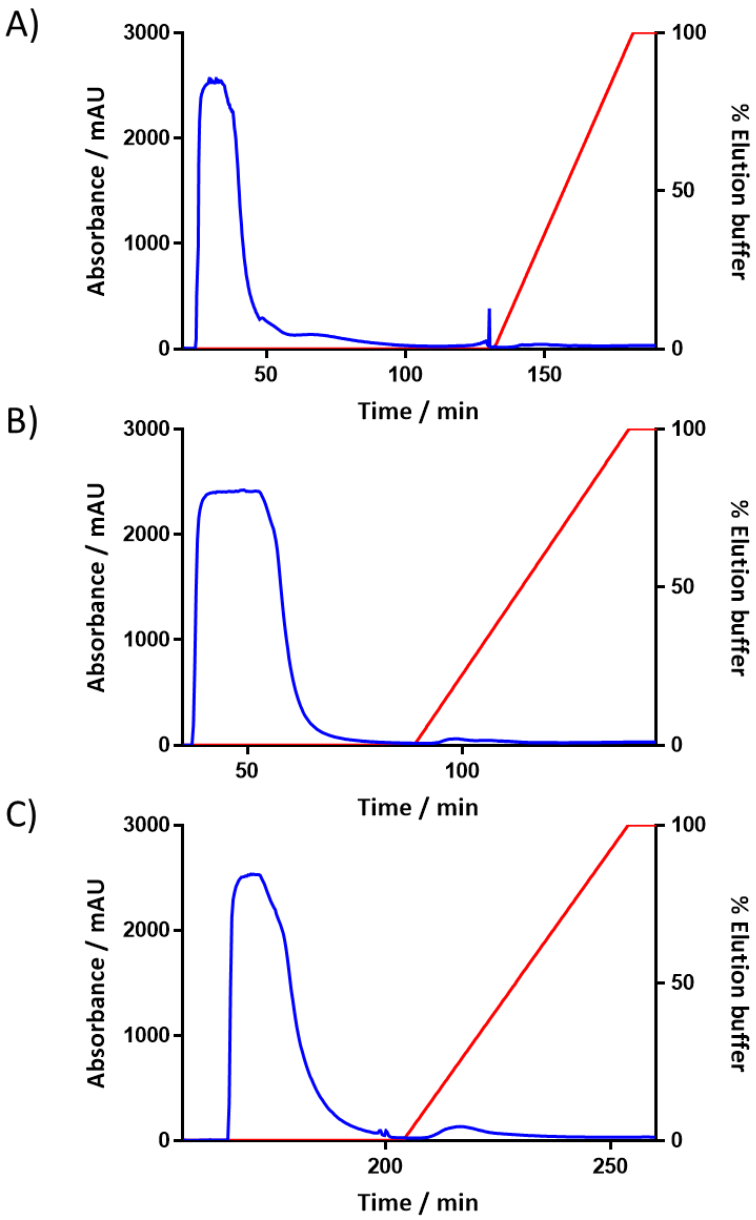


Figure 3. 23 Experiment to confirm chaperone proteins are not co-eluted with MtGBE. Purification of: A: MtGBE only; B: chaperone proteins only; C: co-expression. Absorbance trace in blue; % elution buffer in red.

SDS-PAGE analysis was run on all fractions containing the eluate that led to the relevant signals seen on the purification traces. These are shown in figure 3. 24.

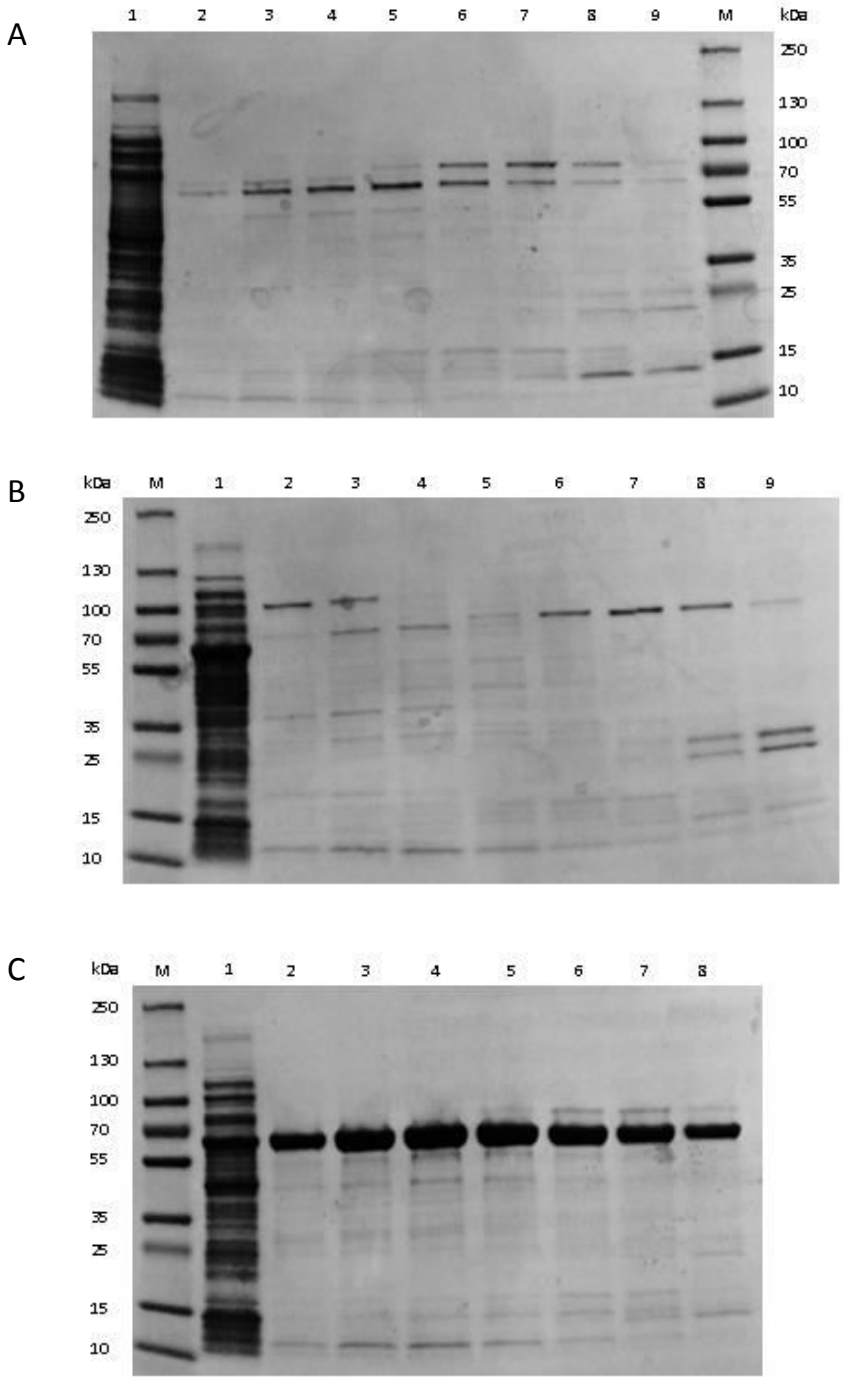


Figure 3. 24 Experiment to confirm chaperone proteins are not co-eluted with MtGBE. SDS-PAGE analysis of purifications. A: MtGBE only, Lane 1: wash through; Lanes 2-9: eluted fractions; B: chaperone proteins only, Lane 1: wash through; Lanes 2-9: eluted fractions; C: co-expression, Lane 1: wash through; Lanes 2-9: eluted fractions; M: marker. MtGBE (= GroEL) = 57 kDa.

From the SDS-PAGE, it is clear that a larger amount of soluble MtGBE was eluted in the co-expression. This suggests that even if chaperones were co-eluting with MtGBE, they

are not the majority of the protein seen on the gel. Moreover, the faint bands that can be seen on the SDS-PAGE of the chaperones only expression do not appear to be of the relevant size, i.e. they are not between the 55-72 kDa bands on the marker, but seem to be equal to 72 kDa or heavier. In addition, in the MtGBE only experiment, and to a lesser extent in the co-expression, these same, heavier bands appear in later fractions, suggesting that rather than a product of the chaperone proteins themselves, they could be *E. coli* proteins.

Though the SDS-PAGE from the purifications of each experiment suggested that the chaperone proteins were not being co-eluted with MtGBE, the fractions containing protein of interest were pooled and dialysed in order to be analysed by spectroscopic methods as well as in assays, so that this could be fully confirmed. These results will be discussed in the following sections, as well as in the following chapter.

3.3 Spectroscopic analysis of MtGBE

3.3.1 Mass spectrometry

In an attempt to characterise MtGBE, mass spectrometry was carried out in order to confirm the mass of the protein that had been expressed and purified. According to the gene sequence, the protein, including the His_6-tag, should have a mass of 59.979 kDa (calculated using ExPAsy peptide mass tool: https://web.expasy.org/peptide_mass/). To see if this was true for the purified MtGBE obtained in the co-expression experiments, protein was dialysed into ammonium formate buffer (20 mM) and diluted to 50 µg/mL. This was then run on a MaXis (Bruker Daltonics, Bremen, Germany) mass spectrometer which was equipped with a time of flight (TOF) analyser (figure 3. 25).

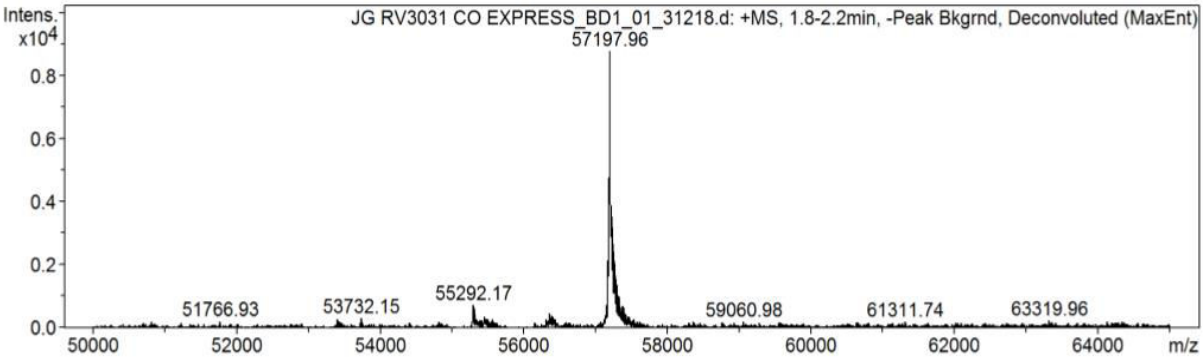


Figure 3. 25 Deconvoluted mass spectrum (ESI+) of MtGBE co-expressed with chaperone proteins.

The mass spectrum shows a main peak at 57,198 Da, which is not in agreement with the mass expected of MtGBE. The expected mass of MtGBE, including the His_6-tag is 59,979 Da, meaning a mass of around 3000 Da is unaccounted for. One reason for this may be that the protein was not correctly translated. Sequencing of the plasmid used in experiments confirmed the correct DNA sequence but errors may have occurred during protein synthesis. However, analysis of the DNA sequence and other possible open reading frames (ORFs) which could lead to off-target expression, does not produce a protein of the mass observed in MS. It may also be that the protein is degraded by proteases despite the presence of protease inhibitors. If this were the case, due to the His_6-tag appearing to remain in place, the fragment cleaved would have to be removed from the C-terminus. Mass calculations of peptide sequences removed from this end would suggest that the protein would be cleaved at site 521 (this would leave behind a protein of mass 57,126 Da). A search using the ExPASy peptide cutter tool (https://web.expasy.org/peptide_cutter/) has identified only one peptide cleavage site at position 521, which would be for proteinase K. However, there are also many other cutting sites for this protease and it is unlikely only one site out of 278 options would be cleaved.

As discussed previously, experiments were run in which MtGBE and chaperone proteins were expressed separately alongside the co-expression of both MtGBE and the chaperone proteins. The cell lysates from each experiment were purified by Ni-NTA column and fractions containing protein, as determined by SDS-PAGE, were pooled. In order to analyse the proteins obtained from each purification, the relevant fractions

were dialysed against 20 mM ammonium formate buffer and run on a MaXis (Bruker Daltonics, Bremen, Germany) mass spectrometer which was equipped with a time of flight (TOF) analyser. The mass spectra for MtGBE expression in the absence of chaperones and for chaperone expression only are shown in figure 3. 26 below.

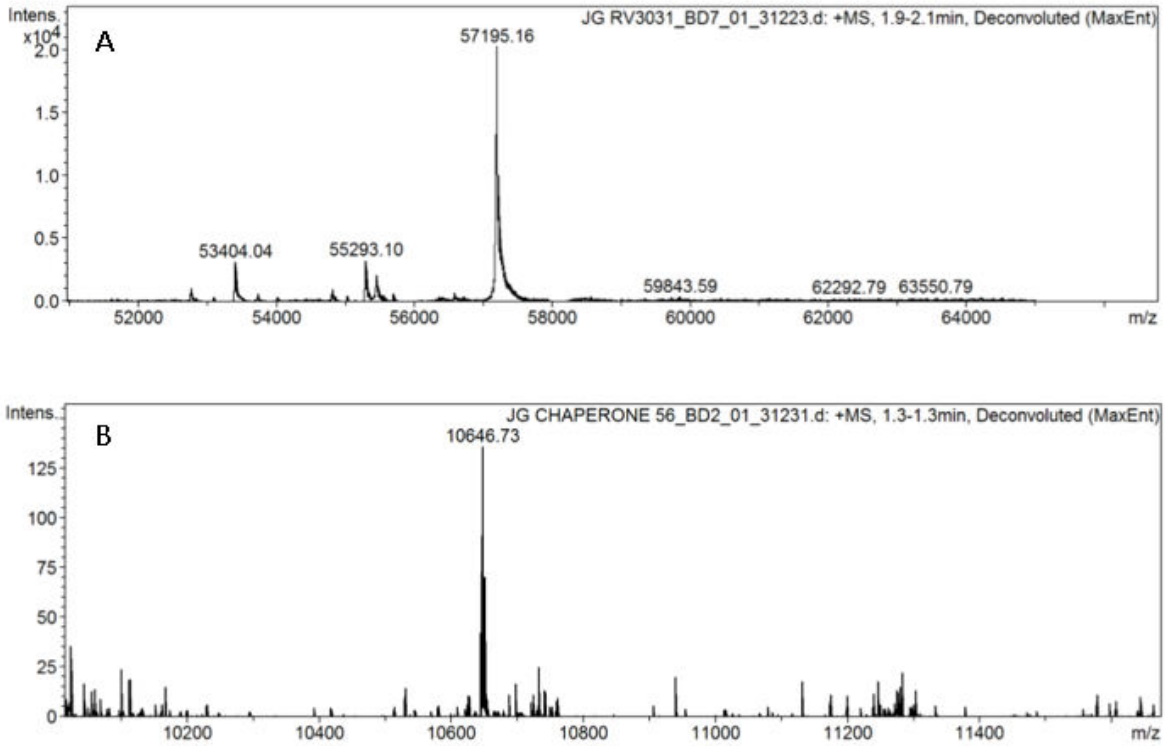


Figure 3. 26 Deconvoluted mass spectrum (ESI+) of A: MtGBE expressed alone and B: chaperone proteins expressed alone.

These results indicate that the co-expressed MtGBE is the same mass as the protein obtained when MtGBE is expressed on its own without any chaperones present. In addition, the protein isolated from the chaperone protein is much less than the mass obtained from the co-expression, strongly suggesting that chaperone proteins are not co-eluted with MtGBE. Thus, it was concluded that the protein obtained in the co-expression did not contain chaperone proteins.

3.3.2 Circular Dichroism

Circular dichroism (CD) is the spectroscopic technique in which circularly polarised light is shone at a sample. In the case of proteins, the sample is an aqueous solution which must be clear of any components that have a strong absorption in the 180-260 nm (far-UV) range.[255] The relationship between the absorption of a sample and the

variables involved in the measurement is characteristic of the molecule and is represented as the molar extinction coefficient, ε. For molecules that absorb at a certain wavelength, ε will have a different value depending on the rotation direction of the polarised light (left or right). In CD, the difference between the value of ε for left circularly polarised light and right circularly polarised light is measured and this is what is observed in the spectrum.[255] The absorption comes mainly from the amide backbone of the protein, as amide bonds have a strong electronic absorption at 190 nm.[255] The secondary structure of the protein will have an effect on the electronic environment of the backbone and therefore its absorbance, thus making CD spectroscopy sensitive to the overall secondary structure of the protein.[256] Examples of typical CD spectra for proteins with pure secondary structures consisting of α-helices, β-sheets and random coils are shown in figure 3. 27.

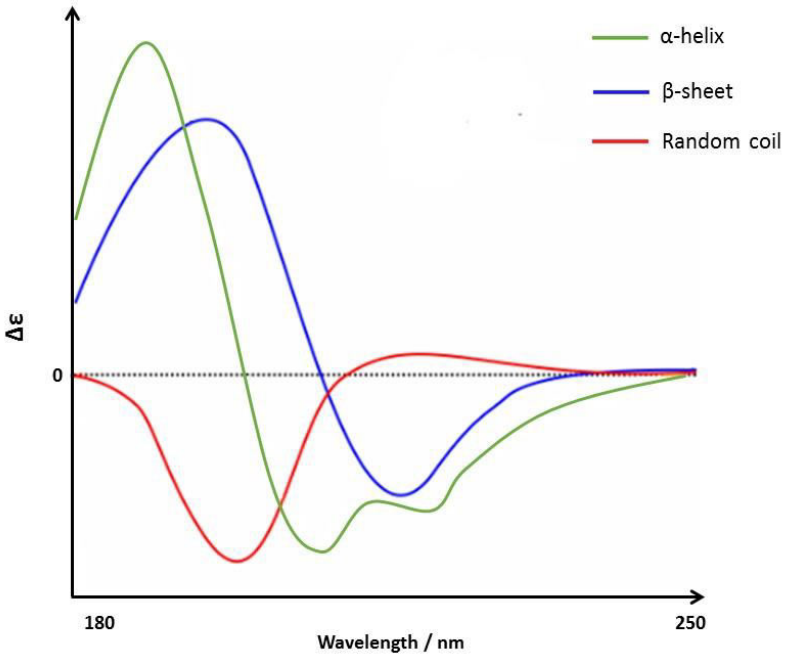


Figure 3. 27 Example CD spectra for pure proteins with certain structural characteristics. Reconstructed from Johnson[255] and proteinchemist.com.[257]

The characteristic shapes of the CD spectra as a result of certain structural features has led to the widespread use of CD in the characterisation of protein structure.[256] Usually, part of the analysis involves comparison to a set of data from proteins of a known structure.[256] It has also been shown useful to include denatured protein samples in these reference sets.[258] In this work, however, CD was used as a quick reference to

compare the proteins purified from the various expression experiments and so further analysis was not carried out.

Following purification, samples were dialysed into 20 mM potassium phosphate buffer in order to remove any traces of components that may absorb in the far-UV region. Samples were then either concentrated or diluted to 0.5 mg/mL except for the sample from the MtGBE only expression, which, due to lack of material, was only concentrated to 0.1 mg/mL. Samples were loaded into a 1 mm sample cell and run on a Jasco J720 CD spectrophotometer from 190-260 nm with a band width of 2 and an accumulation of 10. The resulting spectra are shown in figure 3. 28.

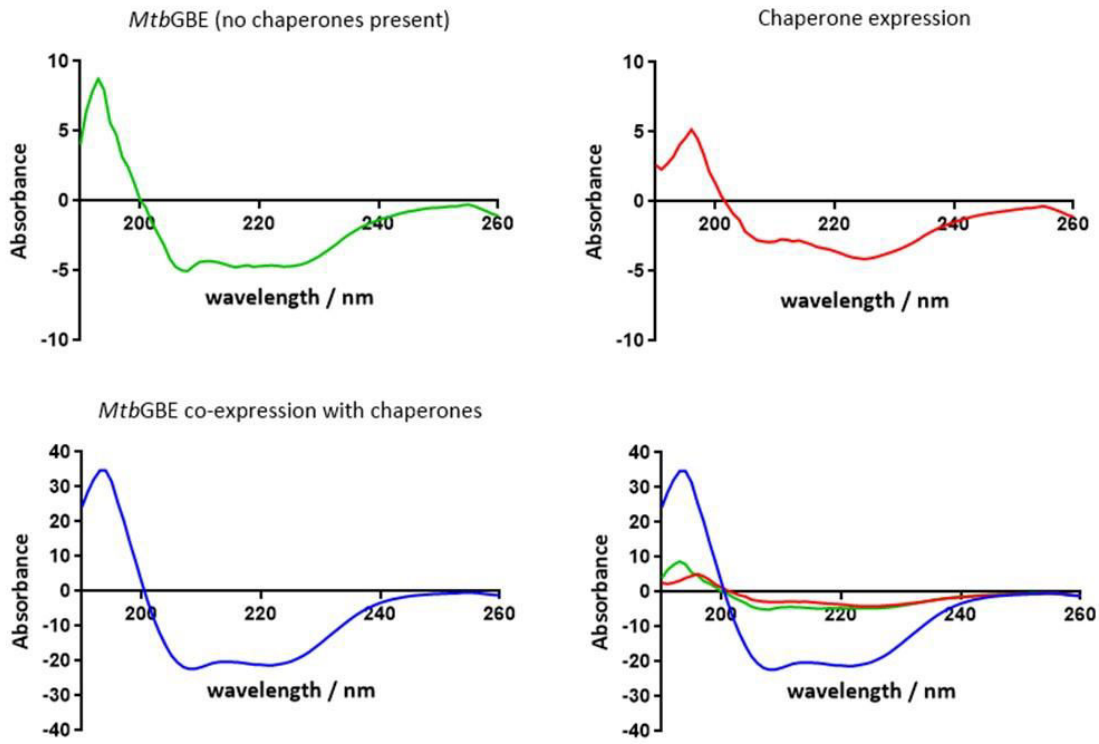


Figure 3. 28 Spectra obtained from circular dichroism experiments on protein obtained from MtGBE expression (green), chaperone expression (red) and MtGBE co-expression with chaperones (blue). Data were smoothed using 2nd order smoothing (4 neighbours) on GraphPad Prism 7.03.

The CD spectrum of the co-expressed protein gives a graph with a distinct shape which suggests high α-helix content.[255] In comparison, the MtGBE which was expressed on its own has a much weaker signal, potentially due to a lower protein concentration in the

sample. The shape resembles that of the co-expressed protein but it cannot be said to be the same with certainty. However the absorbance maxima is the same as that of the co-expressed protein at 193 nm. Similarly the signal from the chaperone protein expression also has a weak signal, again, likely due to low protein concentration. The chaperone expression product also has a similar shape graph to the other two suggesting it may also have high α-helix content. The signal maxima however, is not the same as the other two graphs, being of a slightly higher wavelength (196 nm). Whether or not the MtGBE protein is correctly folded is also not possible to deduce from these experiments without a sample of denatured protein to compare these results to. However, when it was attempted to denature a sample, precipitation kept occurring which meant the sample could not be run on the spectrometer. However, if such a sample could be obtained, this could give an indication in future as to the folding state of the MtGBE protein obtained from the co-expression experiment.

3.4 Conclusions

Though transformation of the plasmid vector containing the *Rv3031* gene was successful, as shown by agarose gel and sequencing of the subsequently purified plasmid, expression of the MtGBE protein led to the formation of inclusion bodies within the bacteria. Two methods were attempted to obtain correctly folded soluble protein: solubilisation using sarkosyl and co-expression with chaperone proteins.

Overall, the most successful method proved to be the co-expression. Though sarkosyl was able to solubilise apparent MtGBE from IBs, attempts at purification of the resulting solution proved unsuccessful, possibly due to incorrect refolding during renaturation.

Co-transformation was achieved successfully *via* two separate methods, one in which both plasmids were transformed at the same time, and one in which cells transformed with one plasmid were made chemically competent and then transformed with the other. Though less efficient in terms of transformants produced, the first method was used in future experiments due to its taking less time to achieve. Agarose gel electrophoresis and sequencing of the purified plasmid from the transformants confirmed both plasmids were successfully taken up by the competent cells.

Co-expression with the chaperone proteins required the presence of L-arabinose in the growth medium, the concentration of which was optimised in small scale experiments. SDS-PAGE analysis showed that the more L-Ara was in the culture, the less target protein appeared in the insoluble cell fraction. After initial small-scale experiments to find the best conditions for co-expression, protein was produced using a large scale (5.0 L) experiment. This was carried out by expressing the proteins overnight at 18 °C using 0.5 mM IPTG and 800 µg/mL L-Ara to induce MtGBE and chaperone protein expression respectively. Protein was then purified *via* Ni-NTA FPLC. Protein preparations were deemed pure enough for use in the assay experiments that were planned.

Finally, it was shown that the band at approx. 60 kDa in the SDS-PAGE from protein purification was not due to the chaperone proteins being co-eluted during Ni-NTA FPLC. Parallel expressions were carried out in which MtGBE was expressed on its own, the chaperone proteins were expressed on their own and where both the target protein and chaperones were co-expressed. Purification of each of the cell lysates and subsequent analysis of the eluates *via* SDS-PAGE, suggested that the protein obtained in the co-expression was indeed the target protein and not the larger of the two chaperones (GroEL = 57 kDa). LC-MS of each protein fraction also supported this conclusion. CD spectra were less conclusive though they do show that the MtGBE co-expressed protein may have a secondary structure of high α-helical character.

4. Investigations to identify the unknown donor substrate of MtGBE

With MtGBE successfully expressed and purified (see chapter 3), investigations into its possible activity and natural substrates could now be undertaken. As discussed in chapter 1 (section 1.4), the reaction thought to be catalysed by MtGBE is the addition of a glucose monomer at the 6 position of glucosyl glycerate (GG).[90, 167, 171] If true, GG would be one of the natural substrates for this enzyme. However, the donor substrate for this reaction remains unknown and two main hypotheses were proposed (see chapter 1, section 1.5) in which the donor was either an α-glucan, or maltosyl glycerate (MalG). It was thought that despite sequence similarities between MtGBE and branching enzymes belonging to the GH57 family, the fact that at least one substrate is a small monosaccharide may hint at novel branching behaviour in which small molecules act as substrates for a branching enzyme. Therefore, experiments were first aimed at probing such a possibility. However, the likelihood of MtGBE exhibiting typical branching behaviour could not be ruled out completely and was therefore also investigated. Once the synthesis of GG and MalG were complete (see chapter 2), assays to probe the substrates of MtGBE began.

4.1 Sequence similarities between MtGBE and GH57 branching enzymes

A Basic Local Alignment Search Tool (BLAST) (http://www.uniprot.org/) was used to find proteins with a similar sequence to MtGBE in order to determine the potential activity of the enzyme. The results showed strong sequence homology with α-1,4 branching enzymes and hydrolase enzymes, particularly from the CAZy database family, GH57. A sequence alignment was performed using MultAlin[259] and the results illustrated in JalView. Figure 4.1 shows the alignment of MtGBE (A) with: a 1,4-α-glucan branching enzyme from *Rhodococcus qingshengii* showing 69% sequence homology (B); a family 57 glycosyl hydrolase from *Hoyosella subflava* showing 64% sequence homology (C); a family 57 glycosyl hydrolase from *Streptomyces sp. AA4* showing 56% sequence homology (D); a 1,4-α-glucan branching enzyme from *Thermus thermophilius* showing 33% sequence homology (E) and a 1,4-α-glucan branching enzyme from *Thermococcus kodakarensis* showing 30% sequence homology (F).

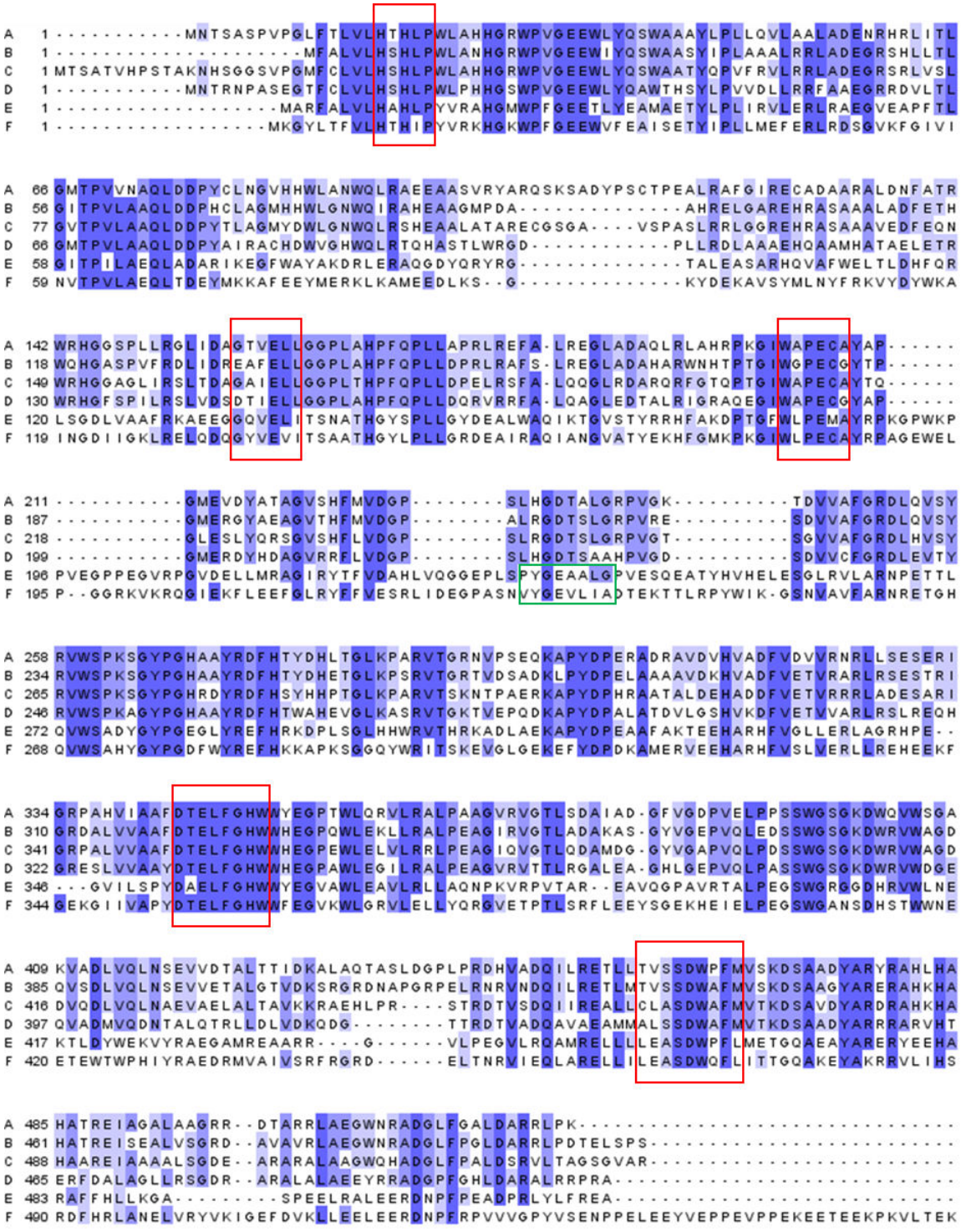


Figure 4. 1 Sequence alignment of MtGBE (A) with a GBE from *R. qingshengii* (B); a GH57 glycosyl hydrolase from *H. subflava* (C); a GH57 glycosyl hydrolase from *Strep. sp. AA4* (D); a GBE from *T. thermophilius* (E) and a GBE from *T. kodakarensis* (F). Alignments performed with MultiAlin[259] and presented using JalView. Red boxes indicate conserved sequence regions shared in GH57 enzymes.[260] Green box indicates loop important to branching enzymes, identified in known structures.[260]

4.1.1 The GH57 family

Part of the carbohydrate active enzyme database (CAZy, www.cazy.org) and created in 1996, the family GH57 (glycosyl hydrolase 57) was formed to accommodate α-amylase enzymes whose structures were not similar with any other known amylases and so did not sort into existing GH families.[261] The family contains proteins that can be grouped into 5 families according to activity: alpha amylase, amylopullulanase, branching enzyme, 4-alpha-glucanotransferase and alpha galactosidase.[260] A comprehensive study by Zona *et al.* identified 5 conserved sequence regions (CSR)[262] of GH57 proteins, one of which was updated in 2012 in a study by Blesak *et al.*[260] In the same study, the catalytic residues were identified as Glu and Asp (Glu291 and Asp394 in the amylopullulanase from *Thermococcus hydrothermalis*) which act as the nucleophile and proton donor respectively.[262, 263] However, not all members of the family have been found to contain these residues, though all contain a $(\beta/\alpha)_7$-barrel as their catalytic domain.[262, 263] In addition, Palomo *et al.* have demonstrated that GH57 enzymes follow a retaining mechanism.[181]

Common to all GH57 enzymes, an α-helical domain, termed domain C, that follows the $(\beta/\alpha)_7$-barrel structure is deemed critical for activity.[181, 260, 264, 265] This arrangement is found in branching enzymes along with a few additional features. For example, 2 α-helical inserts within the $(\beta/\alpha)_7$-barrel are found in branching enzymes, 1 of which is believed to be unique to the branching enzymes within GH57 and not found in any other GH57 enzymes that display one of the other 4 classes of activity.[260] Additionally, a loop has been identified as being important for branching activity.[181] These features are highlighted in figure 4.1 based on the work reported by Blesak *et al.*[260]

According to the aforementioned BLAST search, MtGBE shares the highest sequence similarity (69%) with GH57 branching enzymes. The structure and biochemical characterisation of three of these enzymes have been reported. These are the branching enzymes from *P. horikoshii*[182], *T. kodakarensis*[180] and *T. thermophilius*[181] and will be discussed in the following section.

4.1.2 Structurally and biochemically characterised GH57 branching enzymes

The first GH57 branching enzyme to be described was that from *T. kodakarensis*.[180] Noting the enzyme contained the five CSRs of GH57 proteins, the authors also identified the catalytic residues Glu183 and Asp354.[180] Additionally, these residues were surrounded by histidine and tryptophan residues, with Trp407 and Trp416 thought to prevent water or the acceptor sugar from approaching from the Glu183 direction.[180] A later study by Santos *et al.* focussed on more detailed structural analysis, which , amongst other interesting features, identified the $(\beta/\alpha)_7$-barrel and α-helical domain important to catalytic activity in GH57 proteins.[266] The residue Tyr233 was also identified in the catalytic loop region and this residue is also conserved in the GH57 family.[266]

In subsequent years, Palomo *et al.* reported the crystal structure and biochemical analysis of branching enzyme from *T. thermophilius*.[181] The report described the $(\beta/\alpha)_7$-barrel and α-helical domain, identifying the catalytic residues as Glu184 and Asp353.[181] Additionally, Trp274 and Trp404 which flank the active site were found to be essential to activity, indeed, modelling studies carried out by the group suggested they are involved in substrate recognition.[181] Again, the conserved residue Tyr236 was found in the flexible loop which was reported essential for enzymatic activity as the enzyme was found to be inactive once Tyr236 was mutated to Ala.[181] Product analysis showed that the GH57 branching enzymes utilise a retaining mechanism and modelling studies suggested that initial cleavage of the donor substrate had to be completed before the acceptor substrate could bind and form the product.[181]

Most recently, the crystal structure and mechanistic studies of a GH57 branching enzyme from *P. horikoshii* was reported.[182] The enzyme was shown to contain the $(\beta/\alpha)_7$-barrel and two α-helical domains common amongst GH57 branching enzymes.[182] In addition, the authors identified the catalytic resides, Glu185 and Asp355, as well as many tryptophan, histidine and phenylalanine residues near the active site residues thought to be involved in substrate recognition.[182] The flexible loop containing the conserved residue Tyr236 and thought to be involved in catalytic mechanism was also noted by the authors to have potential involvement in chain length distribution of products.[182]

4.1.3 Applications in this project

MtGBE shares the conserved CSR regions found in family GH57 as well as apparently sharing some key residues identified in the other branching enzymes (figure 4.1). These were identified using sequence alignment and include the catalytic residues Glu205 and Asp344 and the important residues flanking the active site Trp260 and Trp396. However, the conserved residue in the flexible loop (Tyr 236 in *P. horikoshii* and *T. thermophilius* and Tyr233 in *T. kodakarensis*) appears not to be present in MtGBE and instead a His is found at this apparent position. Further residues common to MtGBE and the other GH57 branching enzymes are discussed and summarised in section 4.4.1, table 4.3.

These shared features may indicate that MtGBE has GH57 branching activity. However, the small acceptor substrate in the reaction that MtGBE catalyses (GG) is at odds with the usual large substrates of the GH57 branching enzymes already characterised. Therefore, it was thought that MtGBE may follow a novel branching mechanism. This possibility was explored before the potential of branching activity was investigated as per the reports of the proteins discussed in the preceding section.

4.2 Assays employed to identify the activity and potential substrates of MtGBE

Discussed in chapter 1, the reaction believed to be catalysed by MtGBE involves the transfer of one glucose unit to a small monosaccharide, GG.[90, 167, 171] The similarity in sequence between MtGBE and GH57 branching enzymes discussed in section 4.1 also suggests MtGBE may act as a branching enzyme. However, as branching enzymes do not work on small substrates but on large polysaccharides, it may be a possibility that a novel form of branching activity could be displayed by MtGBE. Due to this conflict between the proposed activity (branching enzyme) and the apparent substrates (small molecules) of MtGBE, hypotheses were formed suggesting an alternative mechanism to those of the known branching enzymes. These are discussed in detail in chapter 1, section 1.5. Both suggest initial hydrolysis behaviour in which a glucose unit is cleaved from a donor to then be transferred to GG. Therefore, to test the viability of these hypotheses, hydrolysis activity would was investigated.

Initially, experiments were carried out using reconstituted enzyme that was sent from collaborators (Prof. Kyeong Kyu Kim Group at Sung Kyun Kwan University (SKKU)) and then with freshly expressed enzyme from the experiments discussed in chapter 3.

The absorbance assays carried out in this project with *p*-nitrophenyl-α-D-glucopyranoside (pNPGlc) did not suggest any hydrolytic activity was associated with MtGBE. Therefore, different possibilities were investigated, mainly focussing on branching activity which will be discussed in the following text.

When no branching activity was observed different glucose donors were investigated and one such donor investigated was glucose-1-phosphate (G1P). An assay to test for release of phosphate is the malachite green assay and so this was also attempted with the enzyme and G1P.

Alongside the spectroscopic assays, substrate activity was tested for using thin layer chromatography (TLC) analysis of the starting materials and reaction components. In this way, the proposed substrate MalG was tested. Whilst not conclusive, TLC can indicate product formation or starting material consumption.

The following sections detail the assays performed and the results of each alongside a brief introduction to each assay technique.

4.2.1 Absorbance assays with *p*-nitrophenyl-α-D-glucopyranoside

4.2.1.1 MtGBE absorbance assay with *p*-nitrophenyl-α-D-glucopyranoside

Initial enquiries into potential substrates and activity of MtGBE focussed on identifying any hydrolysis activity. This was attempted by monitoring the reaction between *p*-nitrophenyl-α-D-glucose (pNPGlc) and MtGBE at 400 nm in order to detect any release of *p*-nitrophenol.

At the outset, MtGBE prepared by collaborators (Prof. Kim Group, SKKU) was used in the assays. The enzyme came in lyophilised form which was reconstituted with sterile water. These first assays were also used as a test to see if the presence of metal ions was required for activity as well as which concentration of enzyme would be best to use. In this case, KCl was used as a source of K^+ and two enzyme concentrations were used. The conditions tested are summarised in table 4.1.

Table 4.1 Experimental conditions used in initial assay attempts.

Experiment no.	[MtGBE] / mg/mL	[KCl] / mM	[pNPGlc] / mM
1	0.1	0.0	0.1
2	1.0	0.0	0.1
3	0.1	100.0	0.1
4	1.0	100.0	0.1

Each experiment was run in 50 mM HEPES buffer at pH 7 at a total volume of 100 µL. Reactions were run for 10 minutes at 37 °C and absorbance readings taken at 400 nm from the addition of the enzyme. Results are shown in figure 4.2.

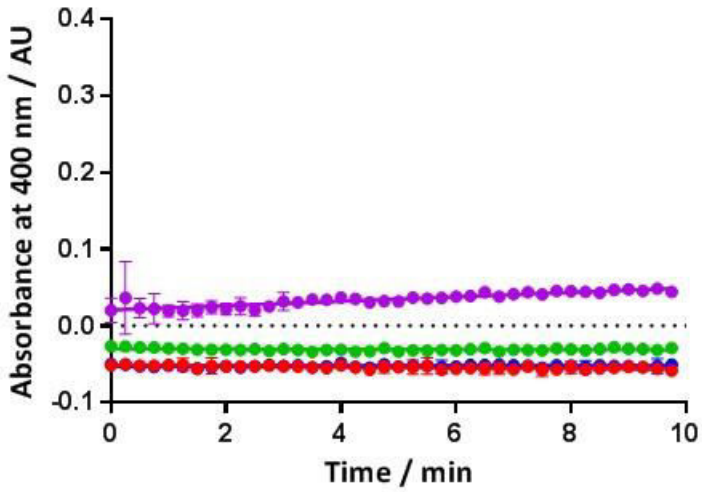


Figure 4. 2 Results of first assay attempt with MtGBE and pNPGlc. Reactions run in 50 mM HEPES, pH 7 at 37 °C. Various conditions trialled: 0.1 mg/mL MtGBE, no KCl (red); 1 mg/mL MtGBE, no KCl (blue); 0.1 mg/mL MtGBE, 100 mM KCl (green); 1.0 mg/mL MtGBE, 100 mM KCl (purple). Assays performed in duplicate.

All but one experiment show no increase over time. Experiment 4 (see table 4.1) did show a slight increase (gradient of line of best fit = 0.00288 ± 0.0002039) compared to the other three experiments carried out (gradient of lines of best fit = 0). Though this increase is small, these results may be indicative of KCl needing to be present in the reaction mixture and that, due to the reaction rate being very slow, larger

concentrations of enzyme or longer reaction times were required. Thus, further optimisation was attempted.

In an attempt to obtain more conclusive results, the substrate concentration was increased to 1.0 mM. Additionally, a wider range of enzyme concentrations was investigated. The concentrations tested were: 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 mg/mL. Assays were run as before, with a negative control run in which no MtGBE was present. Results are shown in Figure 4.3.

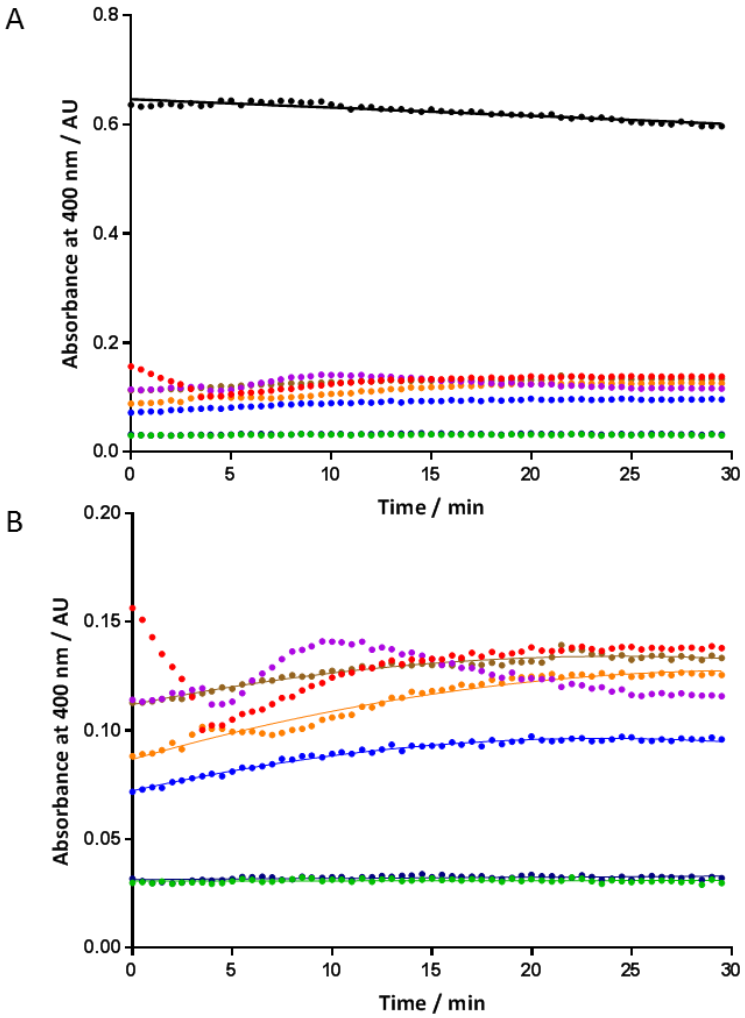


Figure 4. 3 Absorbance assays run with MtGBE and 1.0 mM pNPGlc. Enzyme concentrations tested: 0.1 mg/mL (blue), 0.2 mg/mL (red), 0.3 mg/mL (green), 0.4 mg/mL (purple), 0.5 mg/mL (orange) 0.6 mg/mL (black) and 0.7 mg/mL (brown); negative control with no enzyme present (dark blue); A: all data sets, B: data set for 0.6 mg/mL removed for clarity.

As seen in figure 4.3, results from varying the enzyme concentration are inconsistent. The concentrations that appeared to show a clear trend were 0.1 mg/mL, 0.5 mg/mL and 0.7 mg/mL. Of these, 0.5 mg/mL was chosen as the enzyme concentration to be used in following assays. However, the change in absorbance is still very low, suggesting that either the reaction rate is low or the signal emitted from the product is low and more optimisation is needed.

The reaction catalysed by MtGBE involves transfer of a glucose unit to GG. If the reaction mechanism of MtGBE involves the hydrolysis of a donor and then transfer of the resulting glucose monomer to GG, the presence of GG may facilitate this transfer and therefore increase activity.

To investigate this, an absorbance assay was carried out with 0.5 mg/mL MtGBE, 1.0 mM pNPGlc and either 0 or 1.0 mM of GG. All other components were present as described for the previous experiments. Reactions were carried out at 37 °C with readings taken at 400 nm, every 30s for 60 cycles. The results are illustrated in figure 4.4.

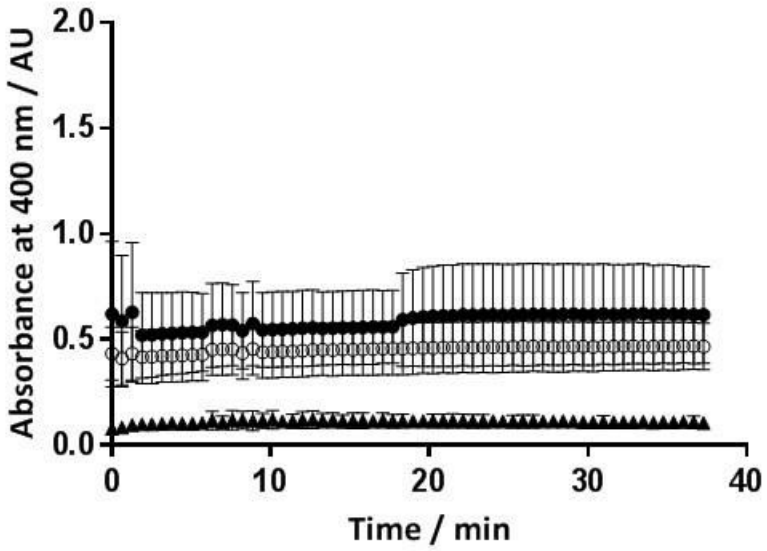


Figure 4. 4 Absorbance assay of MtGBE and pNPGlc either in the presence (●) or absence (○) of GG. Control experiment run without enzyme present (▲). Assays performed in duplicate.

From the graph shown in figure 4.4, it was not possible to conclude whether the addition of GG had an effect on the reaction rate. The error for each experiment overlap and therefore render the results inconclusive. What can be seen is that the starting absorbance between the control experiment without enzyme present and the other two experiments is different (approximately 0.4 AU). One possibility for this could be that the reaction has gone too fast and the beginning has been missed by the time measurements are taken. This seems quite unlikely however, as complete reactions would have a large amount of pNP present in the reaction mixture which

would turn the solution yellow, a colour change visible to the naked eye. However, after the experiments had been run, no yellow colour was observed, suggesting the reaction had no gone to completion. A further explanation for the difference in starting absorbance may be that the enzyme is absorbing at 400 nm. To explore this theory, further experiments would need to be carried out to test this possibility (see conclusions).

Following these experiments, a fresh sample of protein was sent by collaborators (Prof. Kim Group at SKKU) in lyophilised form, as before. This fresh batch of enzyme was tested to see if any activity could be observed and if previous failed attempts were due to the degradation of the earlier protein sample. In addition, these assays were carried out in a UV-vis spectrometer (Agilent Technologies) using a quartz cuvette, rather than the plate reader and 96-well plate setup used previously in an attempt to see if more accurate results could be obtained.

The assays with the fresh sample of MtGBE protein were carried out as before, with 1.0 mM pNPGlc, 100 mM KCl in 50 mM HEPES pH 7. As a positive control, the same reaction with α-glucosidase rather than MtGBE was performed. The carbohydrate active enzyme α-glucosidase naturally occurs in the mammalian gut and breaks down glycogen and other glucose based oligosaccharides.[84] It hydrolyses both α(1→4) and α(1→6) bonds and will therefore cleave the α(1→4) bond of pNPGlc used in this assay.[267] Additionally, a reaction using the original sample of protein was also run in order to compare the two batches.

As can be seen in the results (figure 4.5), the positive control displays a clear trend, showing what one would expect as the glucosidase breaks the glycosidic bond and releases pNP. None of the reactions with MtGBE appear to show any activity. The assay was also repeated in the presence of GG, however this produced the same result – no significant increase in absorbance (results can be seen in appendix C, C.1).

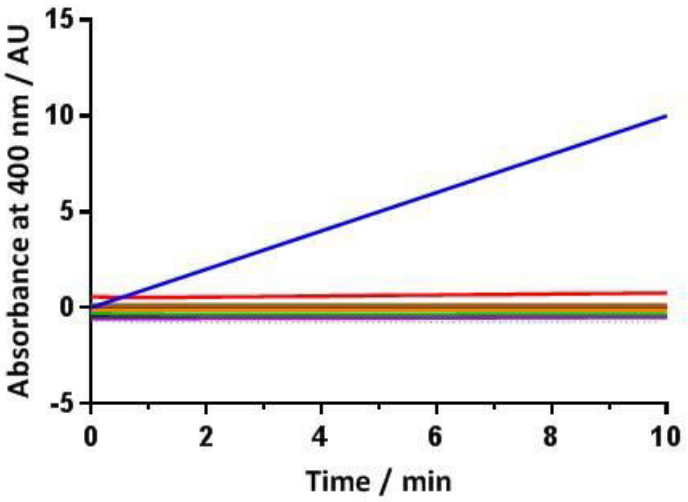


Figure 4. 5 Absorbance assays of MtGBE and pNPGlc performed with fresh enzyme samples sent from collaborators. Reactions run at 37 °C for 10 minutes with 1.0 mM pNPGlc. Enzyme concentrations varied: 0.5 (green), 0.8 (purple), 1.0 (orange), 1.2 (black), 2.0 (brown) mg/mL. Control reactions: no enzyme present (deep red); α-glucosidase (blue); sample with previous batch of enzyme (0.5 mg/mL, red).

Therefore, further assays were postponed until MtGBE protein could be generated freshly on site in this project. It was also concluded that the UV-vis spectrometer was no more accurate than monitoring the reaction with the plate reader and so further assays were carried out on the latter instrument.

Once soluble protein was obtained from the co-expression with chaperone proteins (see chapter 3), assays with fresh enzyme stocks could be carried out. To further optimise the assay conditions and see if pH played a vital role in the activity of the enzyme, reactions were carried out in HEPES buffer covering a range of pH values including: 6.8, 7.0, 7.5 and 8.0. To increase any weak activity, the concentration of both enzyme and substrate was increased to 1.0 mg/mL and 10 mM respectively. Apart from the changes mentioned, the assay was carried out as described previously.

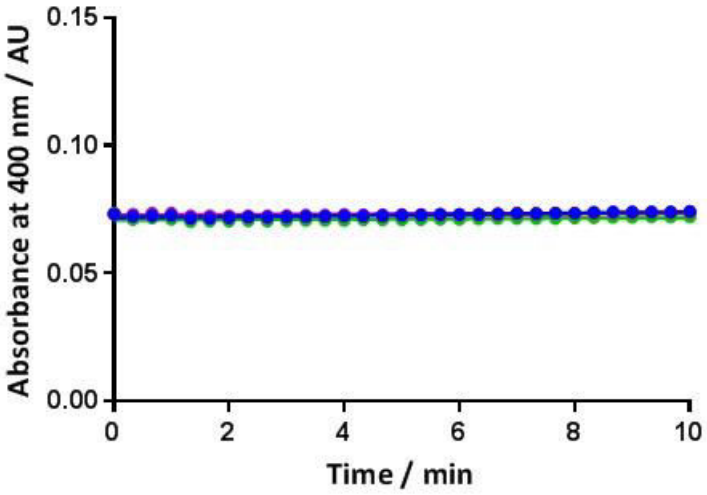


Figure 4. 6 Absorbance assay with freshly prepared MtGBE (1.0 mg/mL) and pNPGlc at various pH levels: pH 6.8 (blue); pH 7.0 (red); pH 7.5 (green) and pH 8.0 (purple).

The results in figure 4.6 suggest that none of the pH ranges tested had any effect on the activity of the enzyme. There was very little to no increase in absorbance at all (gradient of lines; pH 6.8 = 0.0002297 ± 0.0000246; pH 7.0 = 0.0001733 ± 0.0000243, pH 7.5 = 0.0000912 ± 0.0000390 and pH 8.0 = 0.0000714 ± 0.0000189).

As HEPES buffer has a useful pH range between 6.8 and 8.2, lower pH levels were not able to be tested in the initial pH optimisation assay. To check lower pH ranges, MES buffer was used (useful pH range of 5.5-6.7) to test for any activity at pH levels 5.5, 6.0 and 6.5. The assays were then run as before with 1.0 mg/mL enzyme, though substrate concentration was 2.0 mM. All other conditions were the same. Blank experiments were run with no enzyme present.

Upon first examination of the results in figure 4.7, the reaction carried out at pH 5.5 displays apparent activity as there is a sharp increase in absorbance which then levels out after 20 minutes. If this is indeed activity, it would also appear that the beginning of the reaction has been missed and may therefore need to be slowed down for initial activity to be observed in future experiments.

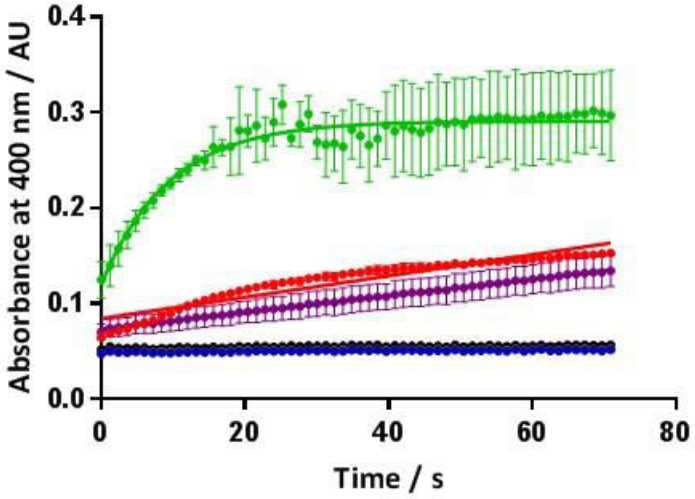


Figure 4. 7 Absorbance assay with freshly prepared MtGBE (1.0 mg/mL) and pNPGlc at lower pH levels: pH 5.5 (green); pH 6.0 (red) and pH 6.5 (purple). Control experiments run with no enzyme present: pH 5.5 (blue); pH 6.0 (brown) and pH 6.5 (black).

At the other two pH levels, there does also appear to be an increase in absorbance, which occurs more steeply at the lower pH level (pH 6.0). Although the data in the first 20 minutes of the assay show a trend, it cannot be ignored that at pH 5.5 the data is less conclusive from 20 minutes onwards, with errors becoming large. This is probably explained by the formation of a white solid that appeared in the wells for these reactions after the assay was complete. It is assumed that this is precipitated enzyme as this did not occur in the control experiment at the same pH level. If this is the case and the protein is precipitating, what may appear to be activity in the initial part of the assay may be invalid. To test this, the experiment was repeated exactly as before, though with a lower enzyme concentration of 0.5 mg/mL. Should the apparent activity been a valid result in the first assay at pH 5.5, the beginning of the reaction should be seen now it has been slowed down with less enzyme present. The control experiments for this assay were changed from having no enzyme present to having no substrate present. Thus, if the trend observed was indeed due to unstable protein, the control should show the same trend. The results of this assay are shown in figure 4.8.

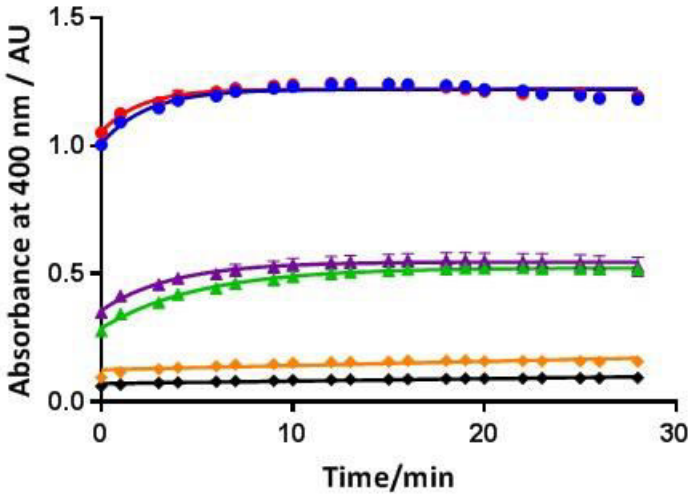


Figure 4. 8 Attempt to repeat 'positive' assay result. Absorbance assay with MtGBE and pNPGlc (1.0 mM) at low pH values in MES buffer (50 mM). Enzyme concentration is 0.5 mg/mL. pH levels tested: pH 5.5 (red); pH 6.0 (purple) and pH 6.5 (black). Control experiments run with no pNPGlc present: pH 5.5 (blue); pH 6.0 (green) and pH 6.5 (orange). Experiments run in triplicate.

The trends observed in figure 4.8, suggest that the 'positive' results seen in the first assays carried out at low pH (figure 4.7) were not positive, but due to the protein, possibly being unstable in these conditions. In all reactions in which all components are present, the trend is almost exactly the same as the equivalent control reaction in which no substrate is included. Therefore, the increase in absorbance observed is not due to the hydrolysis of pNPGlc. It is also interesting to see that again, there are differences between the starting absorbance readings for each experiment. This may suggest possible absorbance of the protein at 400 nm.

The assays at various pH levels had so far been carried out with no GG present at all. To see if GG was required for activity, assays were carried out with both pNPGlc and GG present at varying pH levels. As before, for pH 7 and 8, HEPES buffer was used, and for pH 6, MES buffer was used. The concentrations of pNPGlc and GG were each 1.0 mM and MtGBE was present at a concentration of 0.5 mg/mL. Results are depicted in figure 4.9.

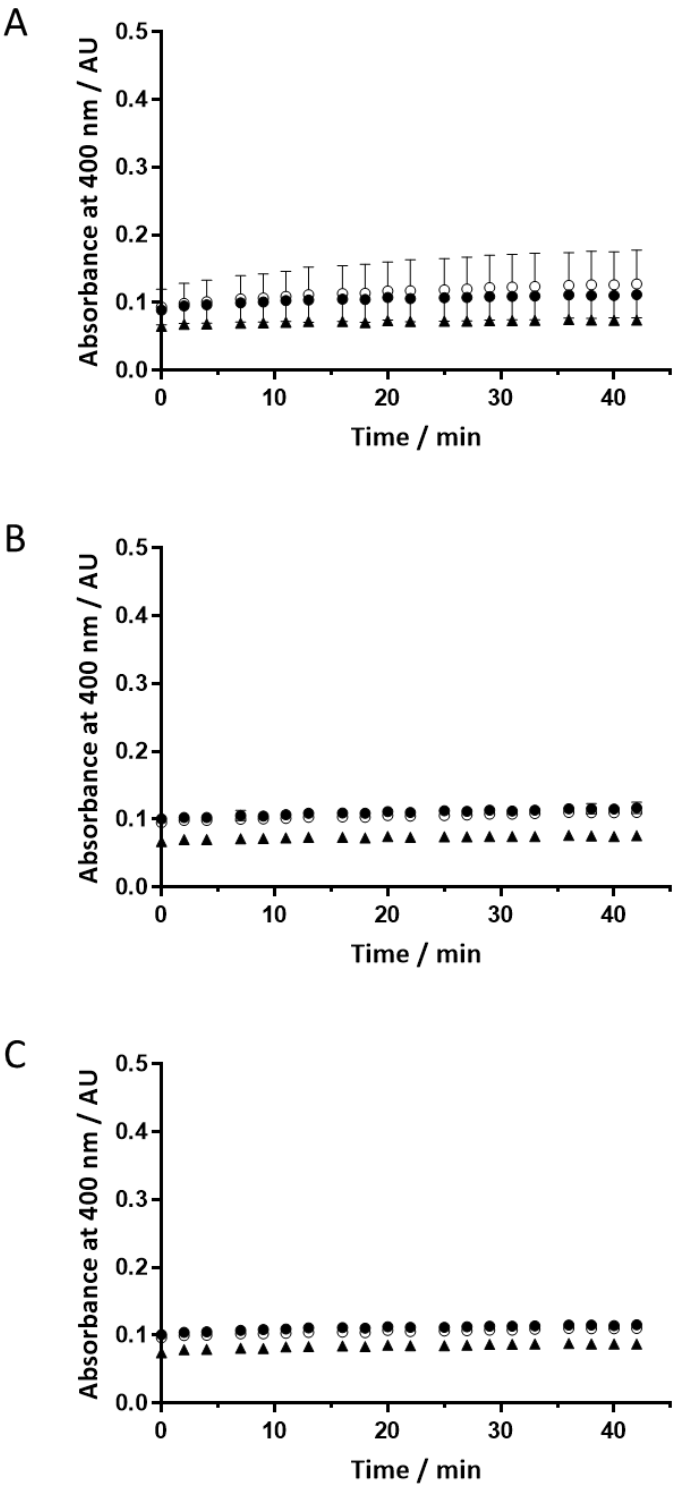


Figure 4. 9 Absorbance assays with MtGBE (0.5 mg/mL) and pNPGlc (1.0 mM) in the presence (●) or absence (○) of GG (1.0 mM) at various pH levels: A: pH 6; B: pH 7; C: pH 8. Control experiments run with no enzyme present (▲).

As seen in figure 4.9, the presence of GG does not lead to any significant increase in absorbance, suggesting that no activity is occurring.

All the absorbance assays attempted so far have shown either no or very little increase in absorbance over time or that increases in absorbance may be due to the presence of the protein itself and not due to its activity (for example, figure 4.7). Therefore, the investigations into hydrolysis activity ended and attention was turned towards other possible activity that MtGBE may possess. MtGBE has been classified as a branching enzyme in family GH57 and for this reason, the potential for MtGBE to show branching behaviour was explored.

4.2.1.2 Absorbance assays to test effect of co-expression on MtGBE

In the previous chapter (chapter 3) it is described how MtGBE had to be co-expressed with chaperone proteins so as to obtain soluble protein. Due to one of the chaperones having a similar mass (57 kDa) to MtGBE, experiments were run in which MtGBE was expressed on its own, the chaperones were expressed on their own and in which both chaperones and target protein were co-expressed. The resulting lysates from each experiment were purified via Ni-NTA column and samples analysed by both MS and CD. All these experiments and their results are discussed in chapter 3, sections 3.2.2.4 and 3.3.

In addition to the aforementioned purification, SDS-PAGE and spectroscopic data, the purified protein from each experiment was also used in absorbance assays to explore any effect the co-expression may have had on the results of the previous absorbance assays. In each assay, the protein concentration used was 0.5 mg/mL for each sample and 5.0 mM pNPGlc was used. The assays were run in 50 mM HEPES buffer at pH 7 in the presence of 100 mM KCl. The absorbance at 400 nm was measured every 30 minutes over a 15 hour period. Assays were run overnight to account for any activity that may be slow and therefore missed at shorter time periods. The results of each assay are shown in figure 4.10.

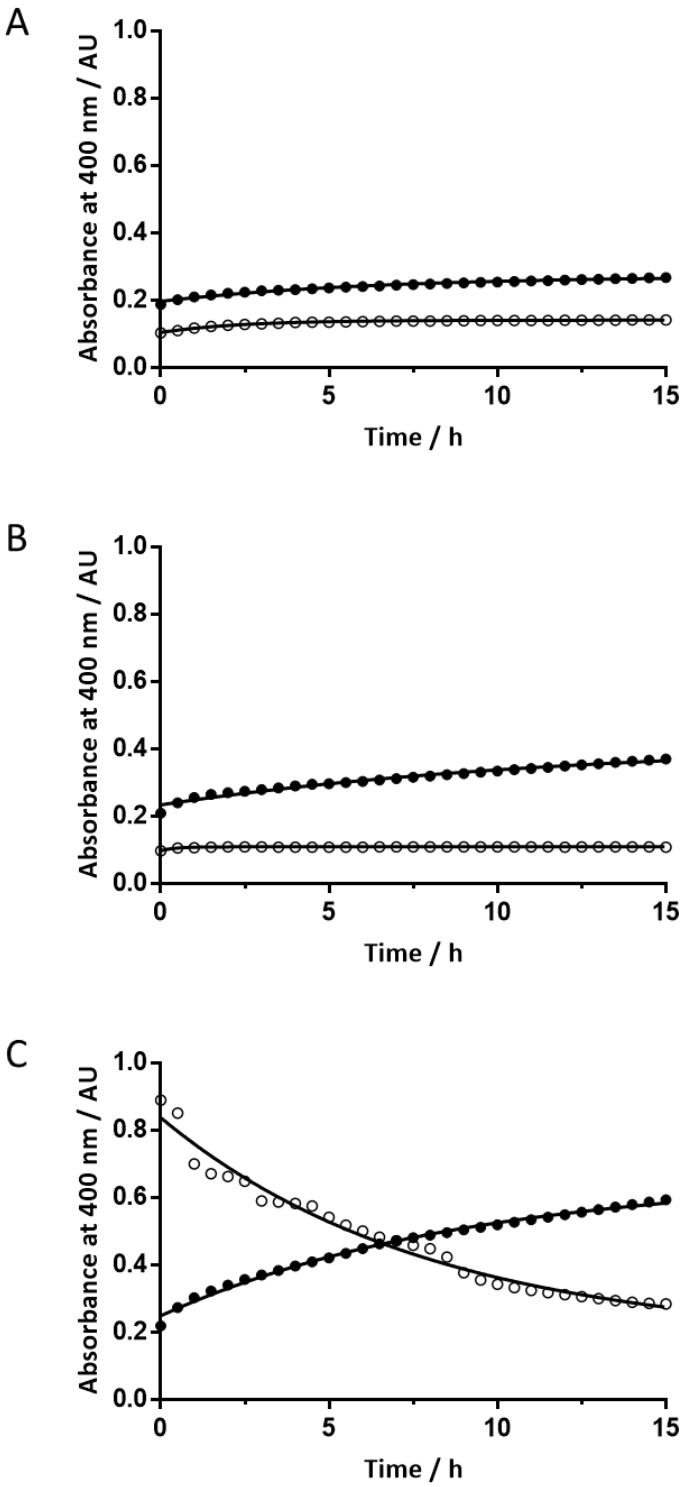


Figure 4. 10 Absorbance assays with 5.0 mM pNPGlc (●) and A: co-expressed MtGBE; B: MtGBE expressed alone and C: protein obtained after chaperone only expression and purification. Control experiments run in which no pNPGlc was present (○). Experiments run in duplicate, except for controls.

In figure 4.10, both the co-expressed and non co-expressed MtGBE give similar trends, showing a slight increase in absorbance over time. However, given the results of the absorbance assays discussed so far in this section, it seems unlikely this is due to any hydrolysis activity. Indeed, for the co-expressed protein, the control experiment in which no substrate is present has a very similar result to that in which substrate is present. For MtGBE expressed independently, this difference is less obvious, though the small increase in absorbance still suggests no hydrolysis activity from the target protein. The only graph that shows a significant increase in absorbance is that of the chaperone proteins. However, the control in which no substrate is present gives an anomalous trend, suggesting the protein present is unstable. Conversely, this behaviour hasn't been seen in any of the experiments in which co-expressed enzyme is present, suggesting that this 'activity' is either anomalous or that the proteins leading to this signal are not present in the samples that have been used for previous assays.

Overall, it was thought that, from looking at these experiments, the chaperone proteins used in co-expression do not seem to be affecting the results seen in the absorbance assays as there is no significant difference between the co-expression and the expression of MtGBE on its own. Furthermore, the similarity in trend between the MtGBE experiments and that containing no MtGBE protein (the chaperone only experiment) supports the idea that the increase in absorbance seen in earlier experiments is not necessarily due to any hydrolysis activity of MtGBE.

4.2.2 Assays designed to detect branching activity: Iodine staining assay

It was concluded that no hydrolysis activity was detected for MtGBE in the experiments discussed in section 4.2.1. Therefore, investigations into another possible behaviour of MtGBE were carried out. As discussed in the preceding text (section 4.1), MtGBE shows sequence homology with branching enzymes from the GH57 family, thus branching assays were performed.

Branching enzymes act upon α(1→4) linked glucan chains, cleaving an α(1→4) link to produce a short chain which is then attached to either elsewhere on the original chain, or onto another chain, *via* an α(1→6) glycosidic bond in a transglycosylation reaction (illustrated in figure 4.11).[185] The result is a highly branched α(1→4) glucan chain, with α(1→6) branch points.[185]

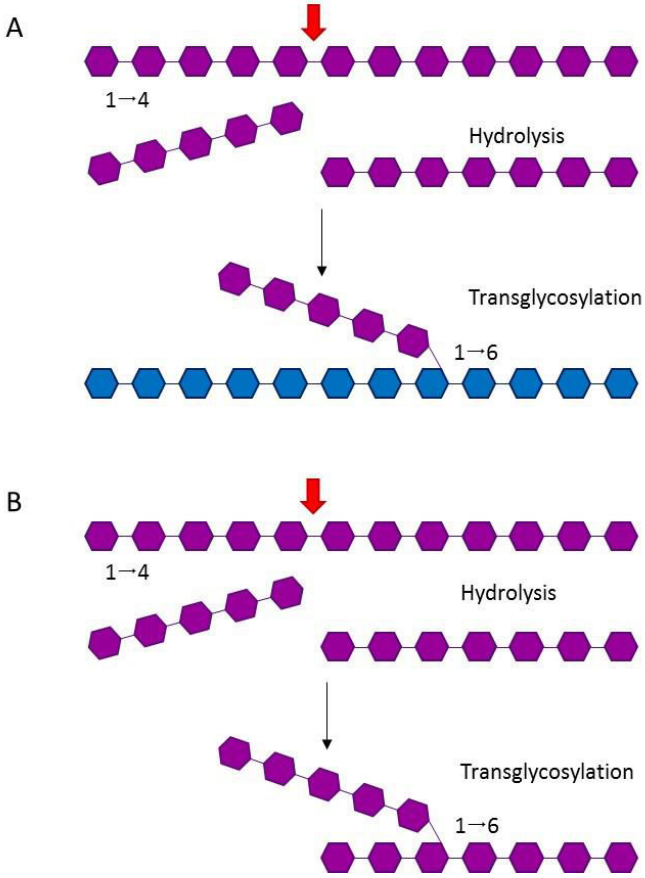


Figure 4. 11 Cartoon depicting general mechanism of branching enzymes. A: inter-chain branching; B: intra-chain branching. Figure reconstructed from Tetlow.[185]

Of the branching enzymes found in family GH57, three have been characterised with both crystallographic and biophysical data. These enzymes are branching enzymes from *P. horikoshii*[182], *T. kodakarensis*[180] and *T. thermophilius*[181], the latter two of which show around 30% sequence homology with MtGBE. Each of these enzymes were investigated using the iodine staining assay, a technique widely reported in the literature and used to characterise branching enzymes that are active upon starch or glycogen.[268-271]

The assay utilises the colour change that occurs when a solution of iodine and iodide comes into contact with α-glucans. The colour is a result of polyiodide complexes sitting within the helical structure of the glucan; in the case of amylose, the species I_5^- forms a dark blue complex, and in the case of amylopectin, I_3^- forms a pink/purple colour complex.[272] The intensity of this colour changes depending on the length of the

α glucan chain with which the iodide is complexed and so the total activity (transglycosylation and hydrolysis) of the branching enzyme can be monitored.[269]

If MtGBE were to utilise the branching activity displayed by other GH57 branching enzymes, the iodine staining assay could be used to identify polysaccharide substrates such as amylose or amylopectin. These are the most common polysaccharides used in the iodine staining assay[180, 181, 268] and were therefore used as potential substrates for MtGBE.

4.2.2.1 MtGBE iodine staining assays

Following the methods reported by Takata *et al.*,[268] Murakami *et al.* [180] and Palomo *et al.*,[181] MtGBE was first tested with amylose and amylopectin as potential substrates in the iodine staining assay. The assay components were prepared with the same ratio and concentrations described by Murakami *et al.*[180] with alterations as to the type of buffer used. The amount of amylopectin used was based on the procedure detailed by Palomo *et al.* [181] A substrate stock was prepared containing amylose (1.2 % w/v in DMSO) or amylopectin (2.5 % w/v in DMSO) in HEPES (0.5 M, pH 7.4) buffer. Enzyme solution was prepared by diluting the stock to 0.2 mg/mL with 50 mM HEPES pH 7.4. This was further diluted to yield solutions of the following concentrations: 20, 18, 16, 14, 12, 10, 8, 6, 4 and 2 µg/mL. For the reaction, the enzyme solution, substrate preparation and GG were mixed and incubated for 10 minutes at 37 °C. To stop the reaction after 10 minutes, 0.4 M HCl was added followed by iodine reagent (see chapter 8, section 8.2.6.13). The absorbance at either 530 nm (amylopectin) or 660 nm (amylose) was measured for each concentration of enzyme used.

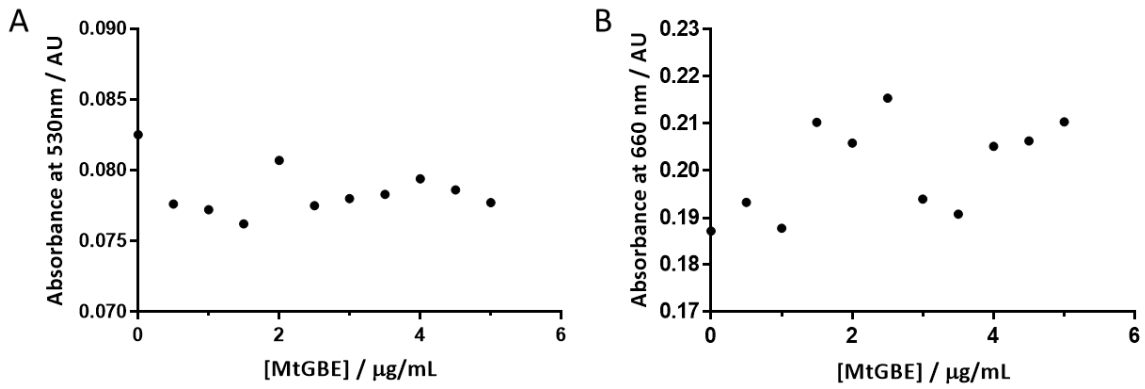


Figure 4. 12 Initial iodine staining assay attempts with MtGBE and A: Amylopectin or B: Amylose.

The results of these experiments are shown in figure 4.12 and in each, no clear trend in the data is observed. It is expected that the absorbance would decrease as the glucan chains are cut and branched thus becoming shorter. For both amylopectin and amylose however, there seems to be no overall decrease in absorbance nor does there appear to be an increase.

To investigate whether the lack of trend in the data was due to the low concentration of enzyme used, the assay was repeated using much larger concentrations of MtGBE (final concentrations used: 1.0, 0.5, 0.25, 0.125 and 0.0625 mg/mL). The results are shown in figure 4.13.

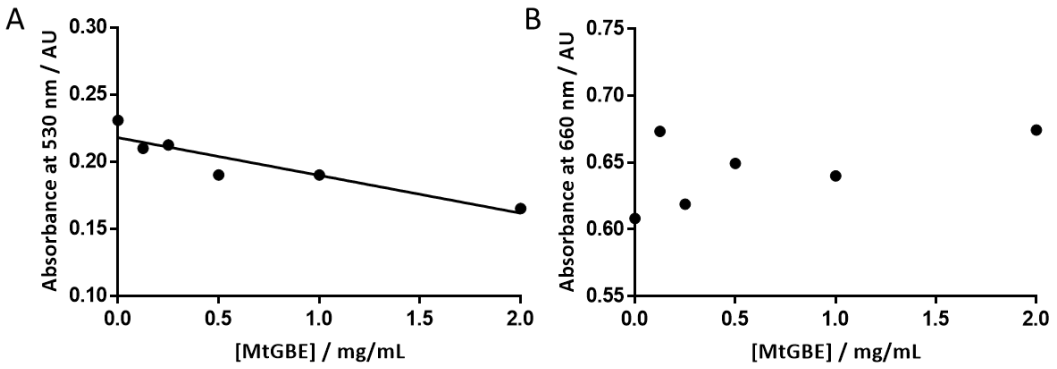


Figure 4. 13 Iodine staining assay with increased enzyme concentrations of MtGBE. Substrate was either A: Amylopectin or B: Amylose.

A trend is now observed in the data. For amylopectin, the absorbance decreases as the enzyme concentration is increased, suggesting possible branching activity. A linear regression was performed using GraphPad Prism 7, as shown in figure 4.13, A. However, as before, there is still no clear trend in the data when amylose is used as the substrate. This may be due to lack of activity for this substrate, but to investigate the possibility that the concentration of amylose may be too high, an experiment was run with half the amount of amylose. In addition, experiments were run in triplicate, unlike previous attempts. The results of this, compared to results of an experiment using the same amount of amylose as before (0.06%) are shown in figure 4.14.

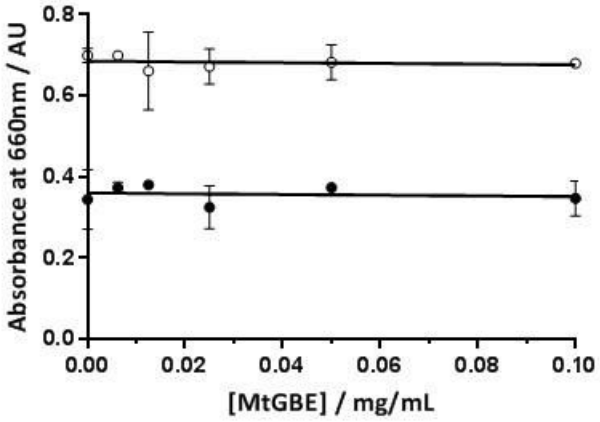


Figure 4. 14 Iodine staining assay with MtGBE and either 0.03 % Amylose (●) or 0.06 % amylose (○). Experiments performed in triplicate.

The results again show no overall decrease in absorbance with increasing enzyme concentration for either concentration of amylose. From this it was concluded that amylose was not acting as a substrate for MtGBE and was therefore not further investigated.

Amylopectin, however, was taken forward as a potential substrate for optimisation of the staining assay. In the literature, branching activity from this assay is taken as the amount of enzyme that is able to decrease the absorbance by 1% in a minute.[268] Practically, the data are collected after 10 minutes, and if it is assumed that the reaction is in the linear stage, it could be estimated that the absorbance would therefore decrease by 10% per unit enzyme activity. Using the data for amylopectin in figure 4.13, A, according to the linear regression, the initial absorbance (y intercept, x = 0 mg/mL enzyme) is 0.22. A decrease in 10% would lead to an absorbance of 0.198. Interpolation using the linear regression leads to an enzyme concentration of 0.353 mg/mL. This would lead to an activity of 2.83 U/mg. However, the conditions for this assay are not optimised, and so investigations into temperature and pH optimisation were carried out. Temperatures investigated included 30, 37, 40 and 45 °C. It was thought that 37 °C was likely the optimal temperature due to this being body temperature and so temperatures either side of 37 °C were explored. The results are shown in figure 4.15.

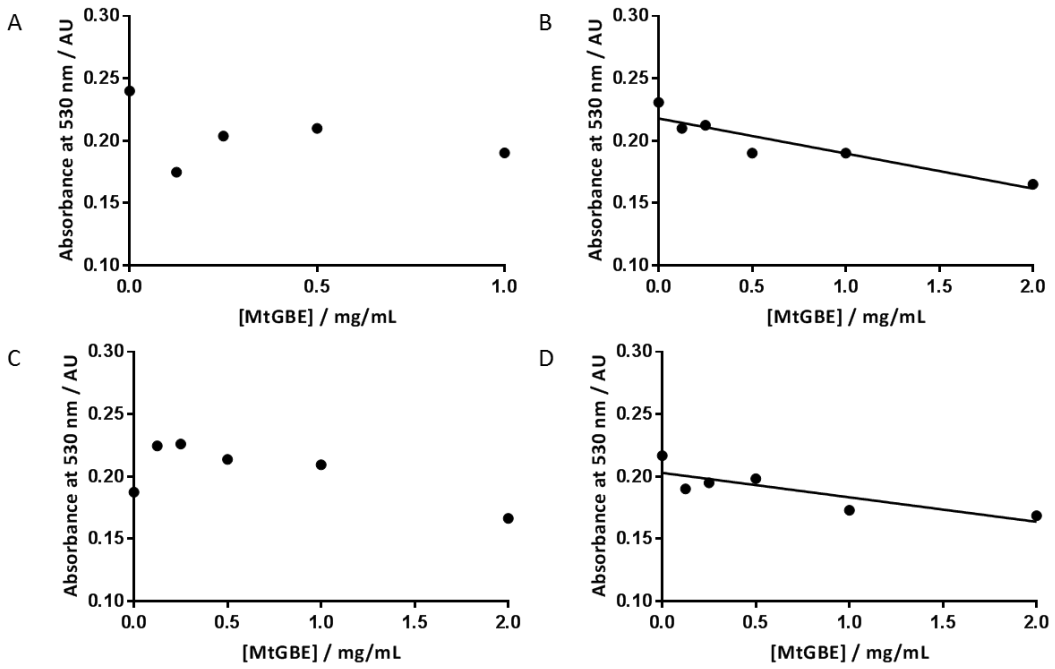


Figure 4. 15 Iodine Staining assay with MtGBE and amylopectin at different temperatures. A: 30 °C; B: 37 °C; C: 40 °C; D: 45 °C.

For each data set, linear regressions were performed and the activity of the protein under each set of conditions was measured as described above. The activity was then plotted against the temperature. Results are shown in figure 4.16.

The results suggest that 37 °C is the optimal temperature as at this temperature, MtGBE showed highest units activity per mg protein.

Another condition to be optimised was the pH of the reaction. To test this, the assay was carried out at varying pH levels in either MES (pH 5.5) or phosphate buffer (pH 7, pH 8). The results are depicted in figure 4.17.

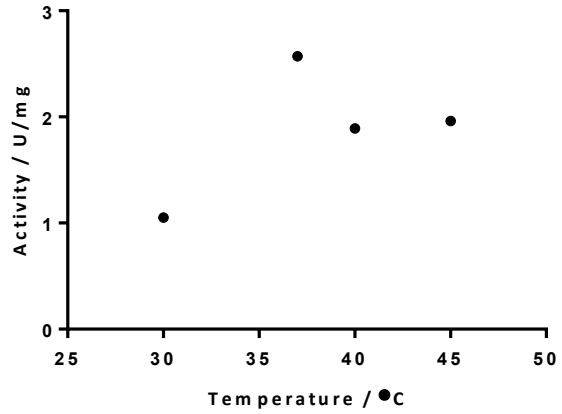


Figure 4. 16 Activity of MtGBE on amylopectin at various temperatures. Linear gradient = 0.05534 ± 0.0585.

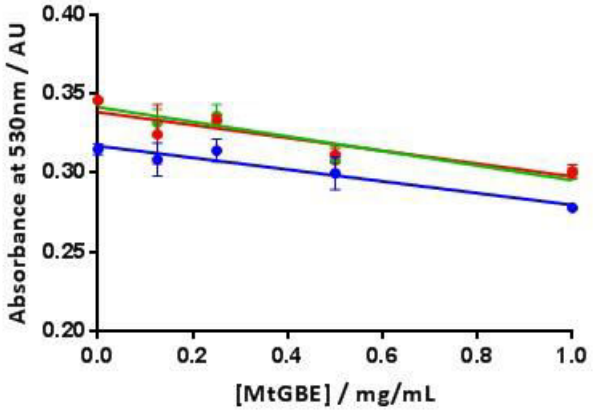


Figure 4. 17 Iodine staining assay with MtGBE and amylopectin at different pH levels: pH 5.5 (blue), pH 7.0 (red) and pH 8.0 (green). Experiments run in triplicate with the exclusion of two anomalous data points in pH 8 experiment.

As described previously, the activity at each pH level was determined from the linear regressions and then plotted as a function of pH (figure 4.18).

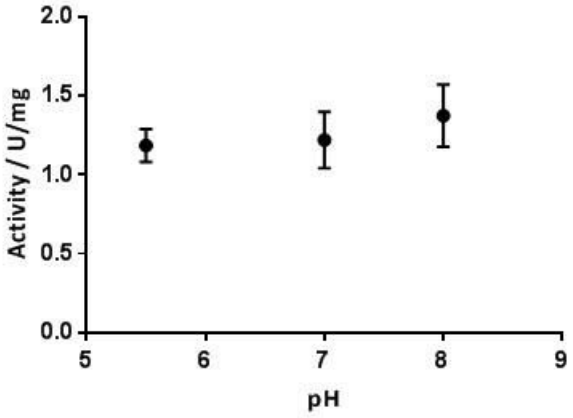


Figure 4. 18 Activity of MtGBE at various pH values.

The results in figure 4.18 suggest that changing the pH does not have a significant effect on the activity of the enzyme. Although the activity appears slightly higher at pH 8 than it is at pH 6 or 7, the value is within the error range of pH 7, making it unlikely that this indicates optimal pH. Therefore, all future experiments were carried out as before, at pH 7.4 in HEPES buffer.

To investigate whether kinetic data could be obtained *via* the iodine staining assay, an experiment was carried out in which the branching reaction was initiated and incubated for 30 minutes, with aliquots at certain time points being taken. The aliquots were treated with 0.4 M HCl to stop the reaction, as described previously, and then mixed as before with the iodine reagent. Absorbance was then measured for each aliquot and plotted as a function of time. To try and optimise the enzyme concentration needed for this type of reaction, the assay was carried out with various concentrations of protein. The results in figure 4.19 show the results of this assay both in the presence and absence of 1.0 mM GG. The lines of best fit were generated using GraphPad Prism 7 and the 1st order exponential decay function.

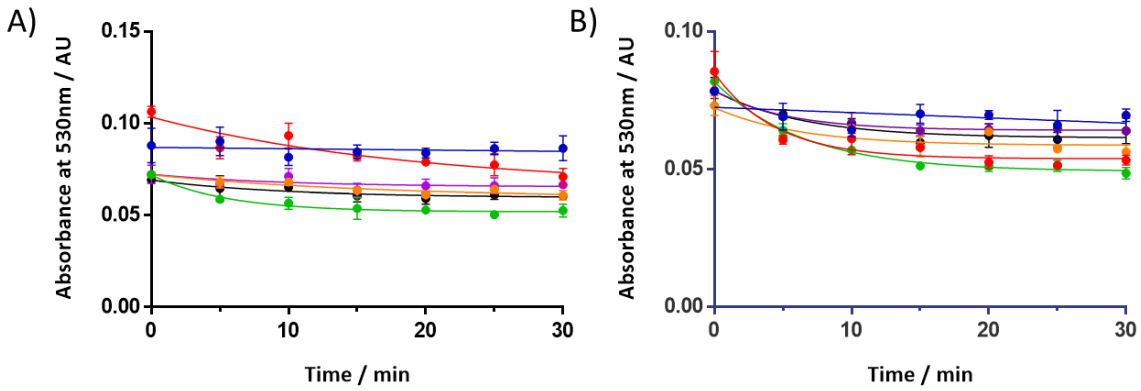


Figure 4. 19 Iodine staining assay kinetics experiment with varying concentrations of MtGBE and amylopectin, with (A) and without (B) 1 mM GG. Enzyme concentration used: 0.0 mg/mL (blue); 2.0 mg/mL (red); 1.0 mg/mL (green); 0.5 mg/mL (orange); 0.25 mg/mL (purple); 0.125 mg/mL (black).

To try and visualise any effect the change in enzyme concentration may have on the initial rate of reaction, the initial gradient for each line of best fit was added and plotted against enzyme concentration. For the data sets which fit a 1 phase exponential decay curve, the initial gradient was calculated as -kY0, where k is the rate and Y0 is the Y intercept, each of which are calculated by the GraphPad software. As the absorbance decrease is being measured, the gradient values are negative.Therefore, the gradient values were multiplied by -1 to give positive values of the same magnitude (figure 4.20).

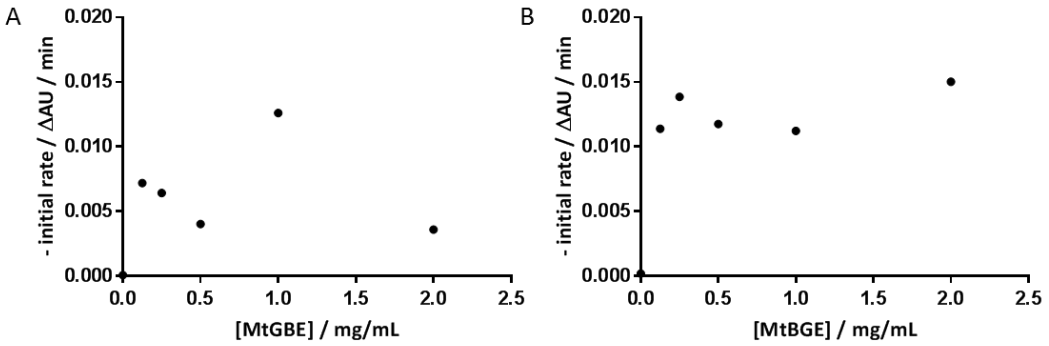


Figure 4. 20 Initial rates of MtGBE branching reactions for different concentrations of enzyme A: with 1.0 mM GG; B: no GG present. Initial rates calculated using line of best fit for each data set (1 phase exponential decay for all except 0 mg/mL for which linear regression was performed).

In the case of the initial rate of reaction in which GG is present in the reaction mixture (figure 4.20, A) it would appear that no clear trend has emerged. In the case in which GG is not present (figure 4.20 B), an overall increase is observed, though this increase is small. Overall, neither graph gives a clear, conclusive result and this method was not deemed suitable for attempting to obtain Michaelis-Menten type kinetic data.

To gain more accurate data, it was decided that the kinetics would be run with a change in substrate concentration whilst keeping the enzyme concentration constant. An enzyme concentration of 0.25 mg/mL was chosen as this gave a clear data trend in the previous reactions (i.e. not appearing to react too fast or too slow). The data are shown in figure 4.21, which includes initial rate plotted as a function of time.

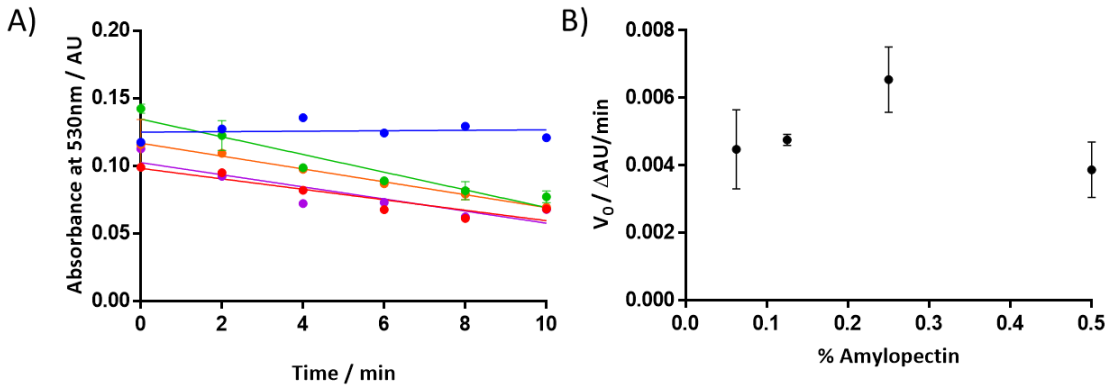


Figure 4. 21 Iodine staining assay attempt at kinetic measurements of MtGBE and amylopectin. A: Amylopectin concentrations used: 0.5% (red), 0.25% (green), 0.125% (orange) and 0.0625% (purple). Control experiment run with no MtGBE present (blue). B: Initial rate of reaction as a function of % amylopectin.

Whilst all the experiments with enzyme present led to a decrease in absorbance over time when compared to the control (no enzyme present), there does not seem to be a clear trend in initial rate of reaction and increasing substrate concentration. The initial rate increases slightly from 0.0625% to 0.125% and then steeply at 0.25%. However, the rate then appears to decrease greatly when the substrate concentration is 1.0%. This does not fit a Michaelis Menten kinetic model[273] and therefore this method was not deemed as appropriate for gathering kinetic data.

In addition to the kinetic experiments run for amylopectin, glycogen and maltodextrin were investigated as other potential substrates using both the iodine staining assay and the kinetic iodine staining method described above. For the iodine staining assay, the colour that formed when the iodine stain was added to glycogen or maltodextrin was a red/brown colour and the absorbance of these complexes was measured at 395 nm. Moulay describes how measuring the glycogen iodine complex at this wavelength has been used to quantify glycogen in the past.[272] However in the same article, the issues with this method for glycogen determination are described; these include the fact that absorbance can change not just with glycogen concentration but also with temperature and iodine concentration.[272] These variables are however kept constant in this case and so this method was used as before with amylose and amylopectin.

Absorbance scans indicated the maltodextrin complex absorbed at the same wavelength as the glycogen complex and therefore the absorbance was also measured at 395 nm (for absorbance scans, see appendix C, C.2). In the assays for each substrate, the final glucan concentration was 0.125% and experiments were run in 50 mM HEPES pH 7 at 37 °C. Each assay was run both with or without 1.0 mM GG. The results are shown in figure 4.22.

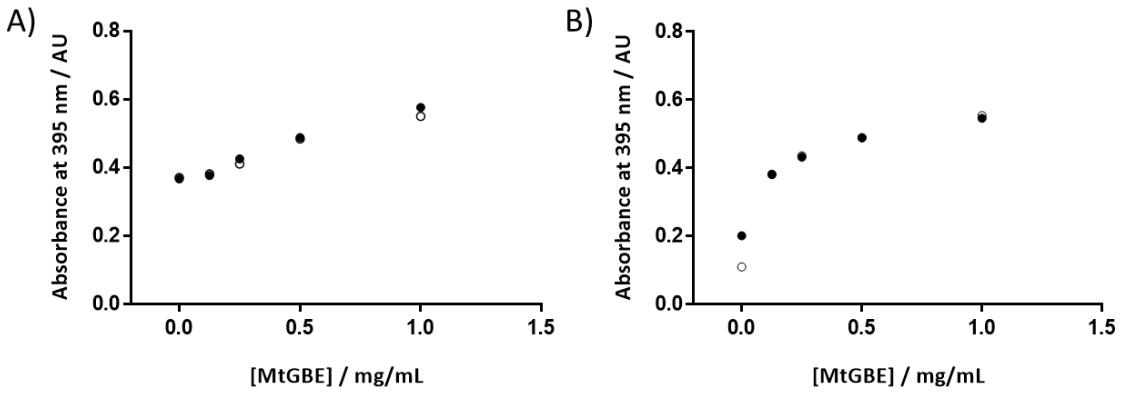


Figure 4. 22 Iodine staining assays with MtGBE and A: Maltodextrin or B: Glycogen. With (○) and without (●) the addition of GG (1.0 mM).

Interestingly, rather than a decrease in absorbance with increasing enzyme concentration, the absorbance increases. Given that the reported branching assays with amylose and amylopectin rely on a decrease in absorbance due to the different lengths of glucan chains being produced in the branching reaction, it is unclear what the relationship is here. It may be possible that the complex formed becomes more intense as the glycan chains become more branched, or it could be that it is not branching activity that is leading to the observed trends. One conclusion that can be drawn from these results is that the presence of GG has no effect on the absorbance values obtained, though this may be due to no reaction occurring.

To investigate if the increase in absorbance could still be seen with constant enzyme concentration, the same reactions were run but this time aliquots were taken at different time points, the reaction stopped, and the solution mixed with the iodine staining solution. The absorbance of each time point was then plotted. This was

repeated for reactions with different substrate concentrations ranging from 0.5%-0.0625% and the results can be seen in figure 4.23.

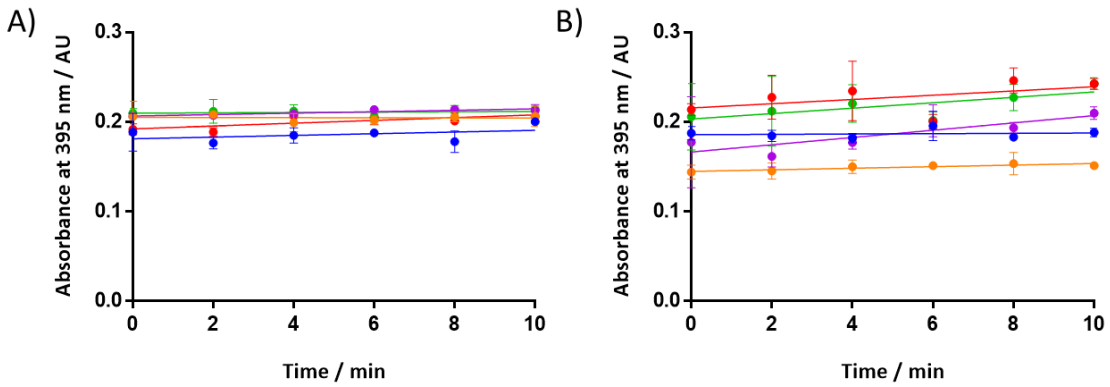


Figure 4. 23 Iodine staining assay kinetics experiments with MtGBE and A: Maltodextrin or B: Glycogen. Concentrations of substrate used: 0.5% (red), 0.25% (green), 0.125% (purple), 0.0625% (orange). Control experiments run in which no enzyme was used (blue).

The data show that, for each substrate, there is no significant increase in absorbance. For each data set a linear regression was performed, and the gradients are summarised in table 4.2.

Table 4. 2 Summary of gradients from figure 4.23: iodine staining assay kinetics with maltodextrin and glycogen

	Gradient	
% substrate	**Maltodextrin**	**Glycogen**
0.5	0.001546 ± 0.0007143	0.002372 ± 0.002037
0.25	0.000199 ± 0.0004295	0.003027 ± 0.001366
0.125	0.0008052 ± 0.0001999	0.00406 ± 0.001291
0.0625	-0.0000776 ± 0.0004429	0.000879 ± 0.0002236
0	0.0009452 ± 0.001047	0.0001871 ± 0.0006497

As the gradient of each reaction in which substrate is present is not significantly bigger or is less than that of the control experiment, it was concluded that neither glycogen nor maltodextrin acted as a substrate for MtGBE.

4.2.3 Assays designed to detect branching activity: Branching assay with isoamylase

When reporting on the activity of the *Thermus thermophilius* GH57 branching enzyme, Palomo *et al.* used a debranching assay alongside the iodine staining assay in order to determine the number of α(1→6) branch points introduced during the branching reaction.[181] In order to determine if there was any correlation between the observations in the iodine staining assays and legitimate branching activity (i.e. if branch points are actually being introduced), this assay was performed for MtGBE with amylopectin and amylose.

Isoamylase is a debranching enzyme that catalyses the hydrolysis of α(1→6) branch points in α-glucans (figure 4.24).[274] Palomo *et al.* describe the use of isoamylase in a branching assay designed to detect the number of new branch points introduced by the branching enzyme in question.[269] The authors carry out the branching reaction, stopping it at different time points. The number of reducing end sugars are then measured before isoamylase is added to the reaction products. Once debranching has been allowed to occur, the number of reducing end sugars is then measured again, indicating how many branch points were introduced.[269] For example, the more branch points that are introduced, the more α(1→6) linkages are introduced. Once de-branched using isoamylase, the α(1→6) linkages are now cleaved, leaving behind short chains with more reducing end sugars available for detection.

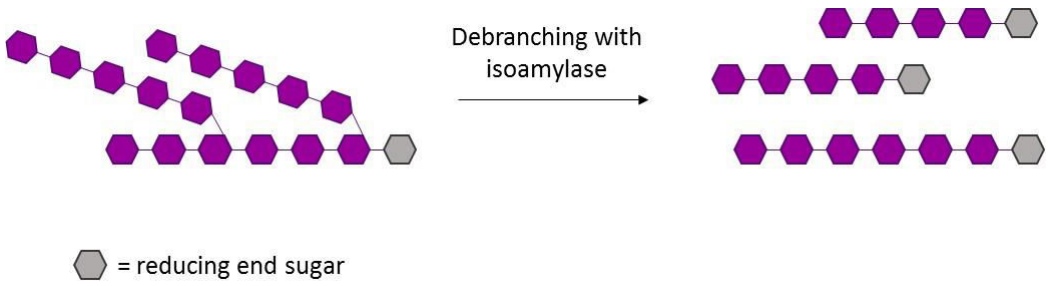


Figure 4. 24 Cartoon depicting how reducing end sugar content increases after isoamylase debranching

The number of reducing ends present was measured using bicinchoninic acid (BCA), two molecules of which can complex Cu^{+} ions to give a purple colour solution, the intensity of which is dependent upon the concentration of Cu^{+} present.[275, 276] When a solution of Cu^{2+} is introduced to a solution containing reducing end sugars, the sugars

can reduce the copper ions to Cu^{+}, and thus the number of reducing end sugars present can be estimated.[275] To estimate the amount of reducing sugars present, a standard curve is performed using known concentrations of reducing sugar. In this work, the sugar used to generate a standard curve was maltotriose. Solutions of maltotriose were prepared in the following concentrations: 500, 400, 300, 200, 100, 50, 25, 10 µM. The standard curve was measured as per section 8.2.3.6 in chapter 8 and the results are shown in Appendix C.3. By quantifying the difference between the number of reducing end sugars before and after debranching, the number of branch points that were introduced during the branching reaction can be assessed.[269]

4.2.3.1 MtGBE branching assays

Following the procedure reported by Palomo *et al.*[269], the branching assays were carried out as per chapter 8, section 8.2.3.7.

For amylopectin, the substrate concentration used was 0.5% and for amylose it was 0.25 %. These concentrations were chosen after initial attempts with less substrate gave inconclusive results. In the same way, the enzyme concentration was chosen to be 0.25 mg/mL. The results are shown in figure 4.25.

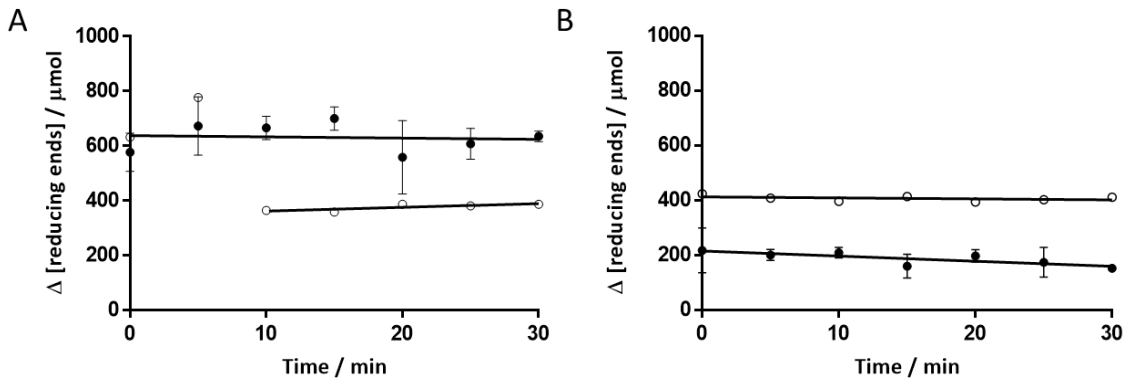


Figure 4. 25 BCA branching assay. MtGBE with A: Amylopectin and B: Amylose. Control experiment carried out in which no enzyme was present (○). Linear regression of amylopectin control excludes first two values which were deemed anomalous.

If the reaction were a success, the number of branch points introduced over time would increase as branching activity increases. This would lead to a greater number of reducing end sugars being present once isoamylase has removed all 1,6-α-linked side

chains. However, the data show that neither amylopectin nor amylose show an increase in the number of branch points introduced and, therefore, that no branching activity is occurring. Each set of data was analysed using linear regression and although the data for amylose appear to show a decrease in the number of branch points introduced, the gradient is low (gradient = -1.864 ± 0.6285) given the scale of the graph. Each experiment shows the same trend as the control, which in this case was the substrate sugar without MtGBE present. Therefore, no further assays to detect branching activity were pursued.

4.2.4 Investigations into other potential substrates: glucose-1-phosphate

As no hydrolysis or branching activity had been detected, another potential donor was investigated. The sugar phosphate G1P is commonly used throughout nature as a building block and donor for many reactions[277-279], including the biosynthesis of glycogen[280] and starch.[281] As MtGBE is believed to contribute to the formation of a polysaccharide chain[167, 171], it is possible that G1P could act as a donor for the formation of di-glucosyl glycerate. If this were the case, it would be highly likely that the mechanism would include the formation of inorganic phosphate (P_i) as the glucose of G1P is transferred to GG.

The release of P_i from reactions can be measured using a solution containing malachite green dye and molybdate anions, an assay technique widely reported in the literature.[282-284] The assay works by measuring the intensity of the colour produced upon the formation of a phosphomolybdate complex with the malachite green dye.[283, 284] In the absence of P_i, the malachite green solution remains yellow, whilst in the presence of P_i, the solution turns blue/green.[283] The concentration of P_i can be determined by creating a standard curve in which the absorbance maxima of samples containing known concentrations of P_i are measured. In this case, the standard curve was made with solutions of potassium phosphate (monobasic) prepared in the following concentrations: 10, 20, 40, 50, 60, 80, 100 μM. Standard curves are shown in appendix C and were made in both water and HEPES to investigate any effect the buffer may have on the readings.

4.2.4.1 MtGBE malachite green assays

To test whether G1P could be a substrate for MtGBE, a malachite green assay was performed. First, the malachite green dye was prepared from malachite green carbinol base, acidified water, and ammonium molybdate tetrahydrate (see chapter 8, section 8.2.6.15). The dye was separated into aliquots and before each use, Triton-X 100 was added.

For the reaction, 1.0 mM GG and 1.0 mM G1P were combined in 50 mM HEPES (pH 7) to which a solution of MtGBE was added to give a final concentration of 1.0 mg/mL. The reaction was carried out at 37 °C over 60 minutes, during which time aliquots were taken and added to the malachite green dye. The dyed sample solution was left to incubate for 5 minutes before absorbance spectra were taken 3 times, each 2 minutes apart. Absorbance data was collected between 600-750 nm. From the absorbance spectra obtained, it was determined that the absorbance maxima for each sample was 630 nm and so the absorbance value for each experiment was plotted as a function of time. The results can be seen in figure 4.26.

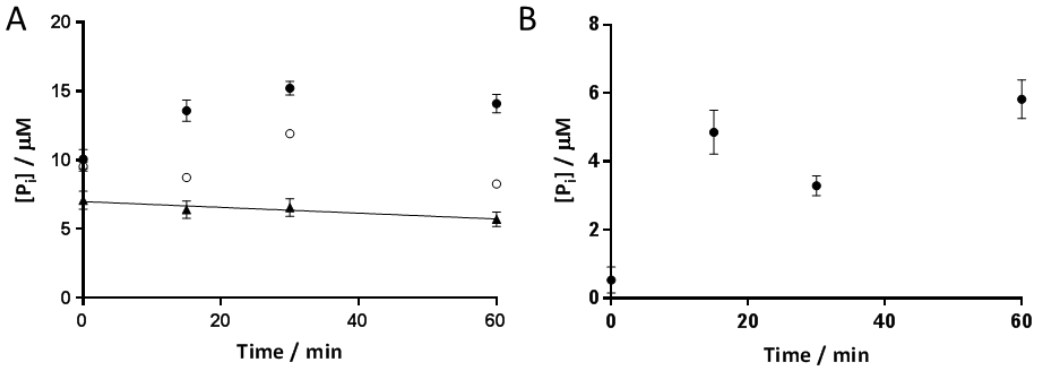


Figure 4. 26 A: Malachite green assay of MtGBE and G1P: reaction (●); control with no enzyme present (▲); control with no substrate present (○). B: Malachite green assay of MtGBE and G1P, difference between reaction data and control containing no substrate.

The data show that the concentration of P_i present increases gradually over time. However, the control experiment run with MtGBE and no substrates present also follows this trend, though with lower concentrations. The enzyme on its own could be reacting with the dye and producing absorbance readings. Therefore, in an attempt to

separate the signal from the enzyme and any signal relating to release of P_i in the reaction, the data set for the enzyme only control was subtracted from the data set for the reaction (figure 4.27, B). The results of this remove the trend previously observed. Although there is an increase in phosphate after the initial reading at 0 minutes, the increase does not continue; P_i concentration decreases before increasing again. Therefore, this particular reaction was deemed unsuccessful, though further investigations were not able to be carried out due to the time constraints of the project.

4.3 Substrate screening using thin layer chromatography

4.3.1 Oligosaccharides

To test the hypothesis that MtGBE may be active on a chain of α(1→4) linked glucose units (see chapter 1, section 1.5), reactions were carried out with oligosaccharides of such a nature. Those tested were maltose (consisting of 2 glucose units), maltotriose (3 glucose units), maltotetraose (4 glucose units) and maltoheptaose (7 glucose units). The reactions were carried out as per chapter 8, section 8.2.3.10. The results are shown in figure 4.27.

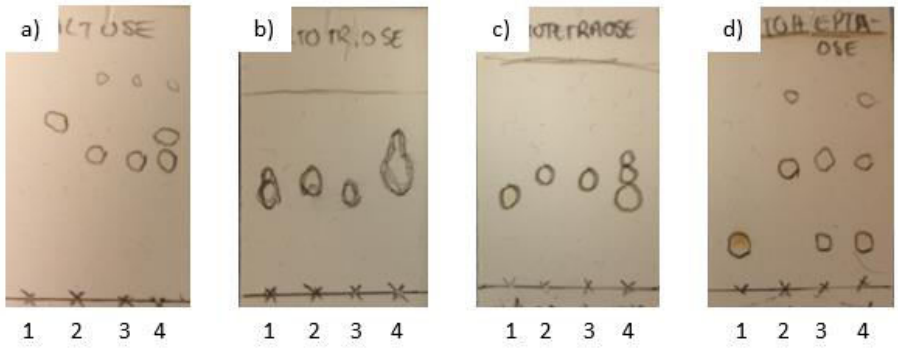


Figure 4. 27 TLC analysis of reaction with MtGBE, GG and; a) maltose, b) maltotriose, c) maltotetraose, d) maltoheptaose. Lane 1: glucan donor; Lane 2: GG; Lane 3: Crude reaction mixture; Lane 4: Mixed spot. Eluent used: ACN/EA/IPA/H_2O 85:20:50:50.

For reactions in which the donor used was either maltose, maltotriose or maltotetraose, the TLC results show that the donor is not present. This may be because the donor has been consumed in the reaction and it is possible that the main spot in the reaction mixture is due to the formation of a new product which is very

similar in polarity to GG. The main spot seen in these reaction mixtures is very similar in Rf to that of GG and it is plausible that DGG would be slightly lower in polarity. Conversely, it may be that the donor is not consumed but is not visualised due to its being present in low amounts in the reaction. In this case, the main spot observed in these reaction mixtures would be due to the GG that is still present. The mixed spot for these reactions does not contain a new product spot, supporting the latter hypothesis. More detailed analysis such as HPLC could be used in further studies to investigate the potential of these oligosaccharides as donors. For the reaction with maltoheptaose, it is clear that no new product spot has been formed as the donor glucan is still visible in the reaction mixture, thus conclusively ruling this out as a potential donor.

4.3.2 Melibiose and pullulan

Other enzymatic activities that have been described in the GH57 CAZy family include α-galactosidase and amylopullulanase activity[263] and one way to test for such activity is to incubate the enzyme with either melibiose (**4.1**, figure 4.28) (a disaccharide consisting of galactose α(1→6) linked to glucose) or pullulan (**4.2**, figure 4.28) (a polysaccharide consisting of maltotriose units linked *via* α(1→6) glycosidic bonds).[180]

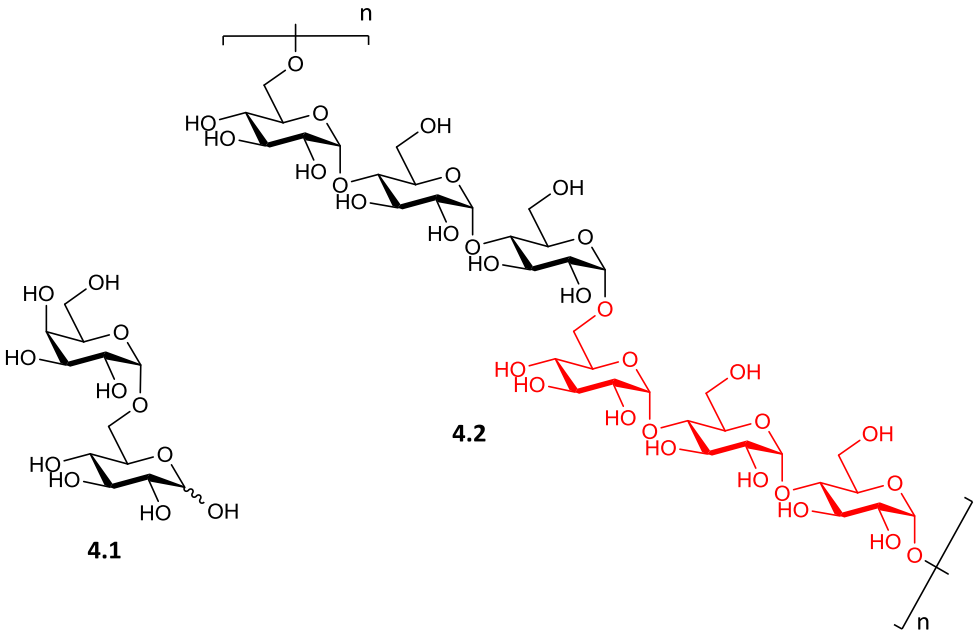


Figure 4. 28 Chemical structures of melibiose and pullulan (with one maltotriose monomer highlighted in red).

To test to see if MtGBE shows either of the aforementioned behaviours, the enzyme was incubated with each substrate for 6 hours at 37 °C. The reactions were carried out in HEPES buffer (50 mM, pH 7) in the presence of 1.0 mM GG either with or without 100 mM KCl. The enzyme concentration used was 0.1 mg/mL and potential donors were present at a concentration of 1 mg/mL. The TLC was run in an eluent consisting of iPA/acetone/H_2O in a ratio of 2:2:1. The stain used to dip the plate was 20% H_2SO_4 in ethanol. The results are shown in figure 4.29.

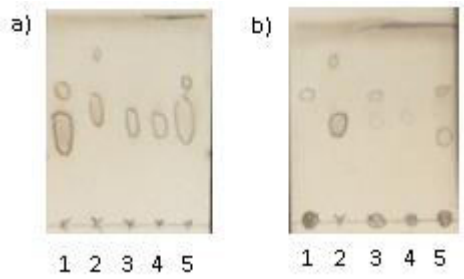


Figure 4. 29 TLC analysis of reaction with MtGBE, GG and; a) melibiose, b) pullulan. Lane 1: glucan donor; Lane 2: GG; Lane 3: Reaction mixture without KCl present; Lane 4: Reaction mixture with KCl present; Lane 5: Mixed spot. Eluent used: IPA/acetone/H_2O 2:2:1.

In the analysis of the reaction with melibiose, only one spot is visible in the reaction mixture lane. There is no difference between the reaction mixtures with or without KCl present. The spot observed occurs at a similar Rf value as both the substrates (melibiose or GG). This could mean that the components were not detected by the methods used or that a product spot has appeared and both reagents consumed. To confirm the results, more precise methods could be used, for example, HPLC or mass spectrometry. The pullulan reaction analysis is clearer. Though faint, the reaction mixture lanes (in which KCl was and in which KCl was not present) show a spot corresponding to GG as well as a strong spot on the baseline which may correspond to pullulan. As no new product spot appeared, it was concluded that MtGBE does not exhibit amylopullulanase activity.

4.3.3 Maltosyl glycerate

The compound maltosyl glycerate (MalG) was synthesised towards the end of the project and therefore comprehensive assays with MtGBE were not able to be performed. Therefore, the reaction was analysed by TLC to check for any signs of activity. The reaction was carried out as per chapter 8, section 8.2.3.10. The results are shown in figure 4.30.

Figure 4. 30 TLC analysis of reaction between MtGBE and MalG. Lane 1: MalG; Lane 2: control reaction with no enzyme present; Lane 3: mixed spot; Lane 4: reaction mixture. Eluent used: ACN/EA/IPA/H_2O 85:20:50:50.

The TLC results are inconclusive. The reaction mixture contains a spot corresponding to the starting material and nothing else, though the spot is not well defined and so is not clear. Repeated attempts at TLC varying the amount of product spotted gave no

clearer results. The spot has an elongated shape which may suggest a product formed of a similar Rf, but more likely this is due to the nature of the reaction mixture conditions. For example, the control reaction in which no enzyme is present also has an elongated shape, suggesting the same could have occurred in the enzymatic reaction sample. Additionally, the mixed spot does not show any sign of an additional product spot, though this is not conclusive due to the streaking. Overall it was concluded that the results were not clear and therefore inconclusive. More accurate techniques would be needed in future to determine if MalG is a substrate of MtGBE.

4.4 Generation of MtGBE mutants using site directed mutagenesis

One technique used to identify key residues involved in the catalytic mechanism of an enzyme is to replace amino acids thought to play an important role with an 'inert' residue such as glycine or alanine.[285-287] To identify key residues in MtGBE, site directed mutagenesis (SDM) was employed to replace amino acid residues identified as being important in sequence alignment studies.

4.4.1 Mutant selection and primer design

Two GH57 branching enzymes, isolated from *T. thermophilius* (*Tt*GBE) and *T. kodakarensis* (*Tk*GBE), show sequence homology with MtGBE and have been structurally characterised.[180, 181, 266] Alignment of these sequences with MtGBE (see figure 4.1) highlighted conserved areas and sites that aligned with key amino acid residues in the structurally characterised proteins. A summary of potentially key amino acid residues identified *via* sequence alignment is given in Table 4.3:

Table 4.3 Summary of potential key residues in MtGBE based on sequence alignment.

Residue in MtGBE	Corresponding residue in *Tt*GBE	Corresponding residue in *Tk*GBE	Function of residue according to known structures
W29	W21	W22	Forms part of hydrophobic channel, possibly part of substrate recognition, conserved in GH57 family.[266]
E205	E184	E183	Catalytic site residue, conserved in GH57 family[181, 266]
H232	Y236	Y233	*Tt*GBE: Part of the flexible loop essential for branching activity.[181] *Tk*GBE: involved in substrate binding.[266]
W260	W274	W270	Lines the catalytic site, involved in aromatic stacking interactions.[181]
F275	F289	F285	Forms part of hydrophobic channel, possibly part of substrate recognition.[266]
D344	D353	D354	Catalytic acid/base residue, conserved in GH57 family.[181]
W351	W360	W361	Lines active site cleft.[181]
W352	W361	W362	*Tt*GBE: Forms part of lid which flanks active site cleft, (highly conserved in GH57).[181] *Tk*GBE: buried in structure as part of hydrophobic core.[266]
W396	W404	W407	Lines the active site cleft, involved in hydrogen bonding interactions, conserved in GH57 family.[181]
F465	F463	F470	Lines the active site cleft, conserved in GH57 family[181, 266]

To investigate the residues identified in Table 4.2, mutant proteins were designed in which these residues were replaced with Ala or Tyr. Primers were designed to be 21 base pairs long, with 9 base pairs either side of the codon of interest, and were codon optimised for *E. coli*. Melting temperatures (Tm) and GC % of each set of primers were calculated using the ThermoFisher Tm calculator (for URL see appendix C, C.6). Tm was kept around or below 70 °C except for that of W29A which had a Tm of 75 °C. GC content was generally around 60-70 % though W29A contained 80.95 % GC. Primers were purchased from IDT®.

4.4.2 Site Directed Mutagenesis (SDM)

In collaboration with this project, the group of Prof. Kyeong Kyu Kim at SKKU aided in the development of the protein co-expression as well as pursuing crystallographic studies on MtGBE. In a visit to the Kim group at SKKU, initial SDM attempts were made to generate the mutant plasmids.

The PCR was carried out as per chapter 8, section 8.2.3.11. PCR products were transformed into chemically competent cells which were then used in plasmid amplification. Samples were then sent for sequencing, the results of which showed that 8 reactions were successful first time (W29A, E205A, H232A, W260A, D344A, W351A, W352A, W396A).

Selected reactions that were not initially successful were then attempted back at The University of Southampton. Analysis using the IDT® OligoAnalyzer 3.1 (https://www.idtdna.com/calc/analyzer) suggested that the H232Y mutant primers were able to dimerise with each other. Two attempts were then made to resolve this issue. First, the concentration of the primers were reduced. A lower concentration of primers reduces the probability of dimerization and increases the chance of binding to the template. Therefore, mutant H232Y was repeated using different concentrations of primers in addition to the conditions used previously. The primers were diluted by a factor of 2 or 4. Following DpnI digestion, the remaining DNA in each experiment was analysed *via* agarose gel electrophoresis. The gel revealed no bands, revealing that the experiments had failed once again.

It has been reported that the addition of DMSO can prevent base pairing and therefore may prevent dimerization between primers and improve PCR results.[288] Therefore, the experiments were repeated once more with the addition of 0.25% DMSO. In this experiment, the primers for the H232Y were not diluted. As DMSO effects base pairing, the annealing temperature of the primers may also be affected. To optimise the reaction conditions, reactions were carried out with varying annealing temperatures. The F465A mutant (primers may also dimerise, analysed with OligoAnalyzer 3.1) was also attempted in this way. The agarose gel analysis is shown in figure 4.31.

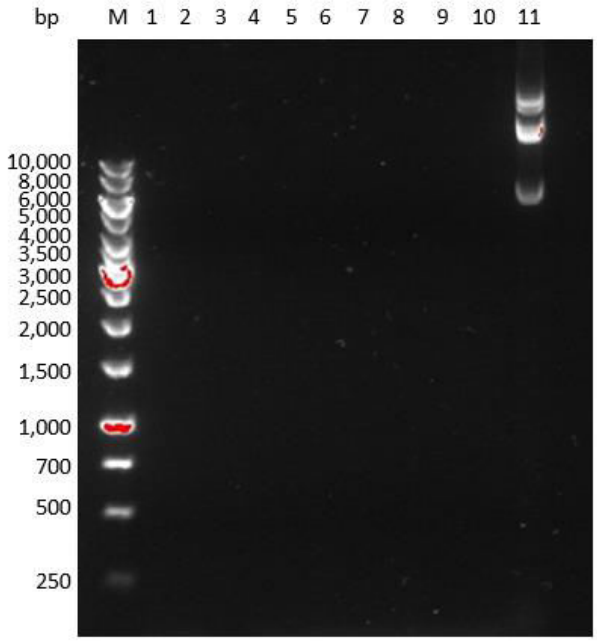


Figure 4. 31 Agarose gel of PCR products after digestion with DpnI. Lanes 1-5 attempt at F465A mutant; Lanes 6-10 attempt at H232Y mutant; Lane 1 and 6: annealing temp = 40 °C; Lane 2 and 7: annealing temp = 45.4 °C; Lane 3 and 8: annealing temp = 50.7 °C; Lane 4 and 9 : annealing temp = 56 °C; Lane 5 and 10: annealing temp = 60.6 °C; Lane 11: template plasmid.

As can be seen in figure 4.31, the reactions were unsuccessful. Due to the ongoing search for substrates for MtGBE, mutant proteins would not be able to be tested and so these SDM reactions were not pursued further.

4.5 Conclusions

4.5.1 Absorbance assays

The first assays carried out with MtGBE were designed to test for potential hydrolysis activity that may support the hypothesised mechanisms discussed in chapter 1, section

1.5. These assays involved the use of pNPGlc as a substrate which, if hydrolysis activity were present, would be degraded into glucose and pNP, increasing the absorbance reading when measured at 400 nm. Initially a small increase in absorbance was regarded as promising, as no natural substrates were present, low activity was expected. However, attempts to optimise the conditions by changing pH, temperature and component concentrations were unsuccessful, suggesting that initial conclusions were incorrect. Additionally, attempts to gather kinetic data by changing the substrate concentration were unsuccessful. These results suggest that the increase in absorbance signal over time that was occasionally observed may be due to another source, perhaps an impurity in the enzyme solution. This could be addressed by further purification of the protein following expression. The protein was purified by Ni-NTA affinity column but faint bands seen in SDS-PAGE analysis suggest further purification would be needed for 100% purity. This could be carried out in future studies using size exclusion chromatography, for example.

Additionally, the same assays with protein expressed in various ways were undertaken to investigate any effect the co-expression with chaperone proteins may have had on MtGBE. As the chaperone GroEL has a similar mass as MtGBE, it was possible that this may be co-eluted with MtGBE during purification of co-expressed protein and therefore any effect it could have in the absorbance assays was investigated. Protein fractions obtained from purification of MtGBE that had been expressed on its own and from co-expression were compared. In addition, protein obtained from a control expression in which the chaperone proteins were expressed were also tested. The results from these assays support the hypothesis that background impurities may be responsible for signals observed in other assays. For example, the protein fraction from the chaperone protein expression led to the largest increase in absorbance over time, suggesting that protein that was isolated after purification when no MtGBE is present, can give rise to signals that appear to show activity.

Overall, it is concluded that MtGBE does not show hydrolysis activity on small substrates such as pNPGlc, either in the presence or absence of GG. However, as this would have been novel behaviour for a GH57 branching enzyme, this is not surprising.

4.5.2 Branching assays

As no hydrolysis activity was observed, branching activity was next investigated. This was due to the classification of MtGBE as a GH57 branching enzyme. The first assay used to detect branching activity was the iodine staining assay. Initial results from these assays were promising. Though amylose shows no trend and thus appears not to be a substrate of MtGBE, amylopectin gave a decrease in absorbance with increasing enzyme, suggesting branching activity was present. Attempts at optimisation however, did not lead to significant changes in activity, and although the temperature optimisation was successful, change in pH had no effect on initial reaction rate which is in contrast to other GBEs which show clear trends in activity vs pH.[180, 181] In addition, experiments were run in which the change in absorbance over time was measured by taking time point aliquots of continual reactions. Whilst the data followed a trend in which absorbance decreases over time, this method did not allow the collection of any kinetic data when substrate or enzyme concentration was changed. Again, this suggests that either there is no suitable way of gaining Michaelis-Menten kinetic parameters with branching enzymes, or the initial branching activity observed with Amylopectin could be an anomaly due to impurities in the protein solution.

In addition to the iodine staining assays, another assay aimed at detecting the number of branch points introduced during the branching reaction was performed. Isoamylase was used in a debranching reaction and the number of reducing ends present before and after isoamylase digestion were measured. The results of this assay were not positive. Over the course of the reaction, the number of branch points introduced did not increase, suggesting no branching activity present. However, a positive control for these experiments was not carried out, which means that the activity of the isoamylase used could not be confirmed. A positive control experiment that could be used in future would involve the use of a known branching enzyme that acts on amylopectin or amylose.

In the iodine staining assays, both maltodextrin and glycogen were tested as potential substrates. Though both showed an increase in absorbance with increasing enzyme concentration, this is unlikely to have arisen from any activity as changes to the concentration of the substrate did not have an effect on reaction rate.

Overall, the iodine staining assay suggests that weak activity may be occurring on amylopectin but not on amylose, glycogen or maltodextrin. The isoamylase branching assays did not support this result from the iodine staining assays but a repeat of these experiments with a positive control would confirm this. In addition, other techniques such as HPLC analysis could further confirm the presence or absence of branching activity.

4.5.3 Malachite green assays and TLC substrate investigations

Malachite green assays were performed to assess the potential of G1P as a donor in the formation of DGG. As G1P is involved in the biosynthesis of glycogen and starch[280, 281], there is a possibility that it may be involved in the biosynthesis of MGLPs. However, these assays showed no increase in the amount of P_i released in the reaction, suggesting that G1P is not a substrate of MtGBE. Similarly to the isoamylase branching assays, no positive control was used due to no suitable enzyme being easily obtainable at the time. A positive control would confirm the negative results. However, the lack of any activity at all, strongly suggests G1P is not a substrate of MtGBE.

The TLC experiments confirmed that neither maltoheptaose, pullulan nor MalG are substrates for MtGBE. Due to no clear product spots forming and no consumption of starting material, none of these components are believed to be active with this enzyme. To confirm this, more quantitative analysis could be performed in future, particularly with MalG. If MalG were a substrate, it is proposed the mechanism involves isomerisation to GG and it is possible that separation by polarity is not achieved on the TLC plates used. Further techniques, such as reverse phase TLC, HPLC or even NMR could be employed to follow any potential reaction progress.

Other TLC results were inconclusive; those in which maltose, maltotriose, maltotetraose and melibiose were used as substrates. This could be due to a lack of separation of components and therefore HPLC could be employed to confirm these results. In the case of melibiose, other substrates could be employed to test for galactosidase activity, for example, α-galactosyl fluoride. This substrate would be much easier to visualise on TLC and thus monitoring of substrate consumption could be easier.

4.5.4 Site directed mutagenesis generation of mutant proteins

The generation of mutants had varying degrees of success. Of the initial number of mutants planned, 8 mutant plasmids were successfully obtained. Transformation of these plasmids into competent cells led to the generation of glycerol stocks which can be stored for future use should a suitable assay for MtGBE be found. For the unsuccessful mutants experimentation into the PCR conditions used did not yield any success. However, time constraints and the changing needs of the project meant SDM to generate mutant plasmids was not continued.

5. Expression, purification and initial characterisation of the glycosyl transferase encoded for by *Rv3032*

5.1 *Rv3032* and the role of MtGlcT in MGLP biosynthesis

The genes thought to be involved in the biosynthesis of MGLPs in *Mtb* are discussed in chapter 1, section 1.4. Included in this discussion is the gene *Rv3032* which encodes what has been identified as a glycosyl transferase (GT) belonging to the CAZy family GT4. [90, 289] Studies carried out by Stadthagen *et al.* suggest that this GT (MtGlcT) is responsible for the elongation of the MGLP chain, by acting upon DGG (**1.37**) to form **5.1** (shown in figure 5.1), as well as upon other α-glucans such as glycogen.[90] Furthermore, it has been shown that MtGlcT is involved in an alternative pathway to the GlgE biosynthetic route to α-glucan[139, 290], although this pathway is not essential to the growth of the bacterium.[138]

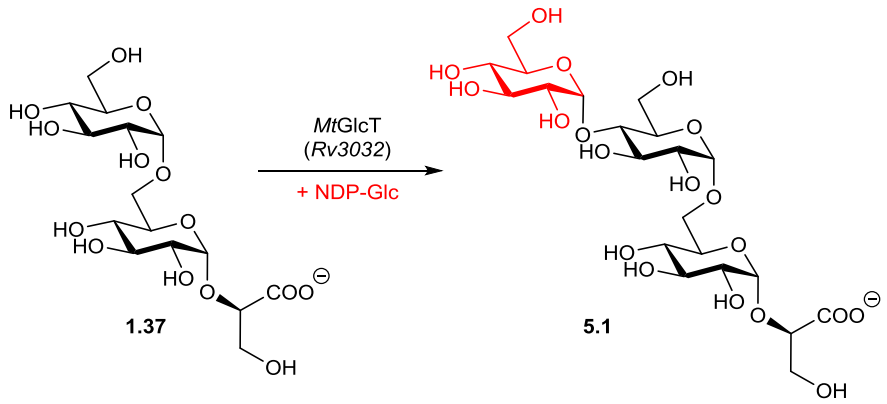


Figure 5. 1 One of the reactions in the MGLP biosynthetic pathway thought to be catalysed by MtGlcT. Recreated from Mendes et al.[167]

It was reported that, due to the fact that MGLP resides in the cytosol, the sugar donor used in elongation reactions is likely to be a sugar nucleotide (NDP-Glc). Preliminary experiments involving partially purified MtGlcT showed that the elongation of short chain α-glucans was possible with UDP-Glc as a donor.[90, 171, 290]

Due to the apparent involvement of MtGlcT in elongating α-glucan chains[90], it was proposed that this GT may be involved in the initial stages of MGLP synthesis as well as

in the chain elongation. For example, if MalG (**1.38**) was the donor for MtGBE (see chapter 1, section 1.5.1.2), it would need to be synthesised beforehand. This could be carried out by a GT enzyme that can transfer glucose to GG (**1.36**) and form an α(1→4) glycosidic bond as depicted in figure 5.2. Due to the nature of the involvement of MtGlcT in the synthesis of MGLPs[90], it was proposed that MtGlcT could be a likely candidate for the synthesis of MalG if this was the substrate utilised by MtGBE.

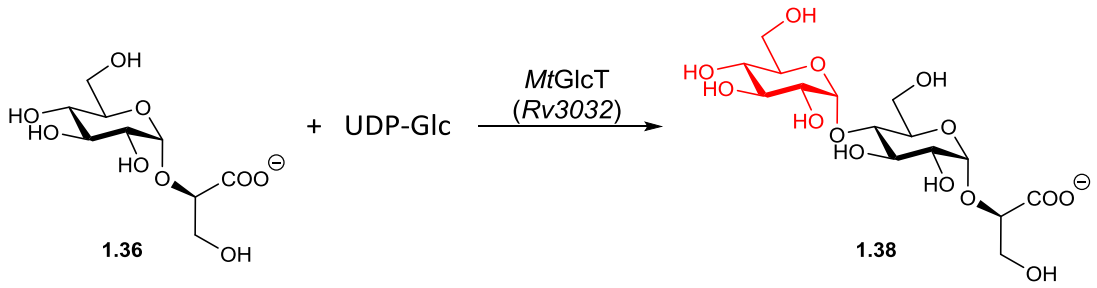


Figure 5. 2 Proposed biosynthesis of maltosyl glycerate, one hypothetical substrate for MtGBE.

To investigate this hypothesis, recombinant MtGlcT was required to test for GT activity as well as to investigate if GG could be used as a substrate with UDP-Glc to make MalG. Therefore, the expression and purification of MtGlcT was completed and initial coupled assays designed to test for GT activity.

5.2 Expression and purification of MtGlcT

5.2.1 Plasmid transformation and amplification

The *Rv3032* gene was synthesised and cloned into a pET-28a(+) vector with a kanamycin resistance marker by Genescript. The protein contained a His6-tag at the N-terminus for Ni-NTA purification. For plasmid amplification and purification, the pET28a(+)::*Rv3032* plasmid was transformed into chemically competent *E. coli* JM109 cells using the heat shock method (See chapter 8, procedure 8.2.1.4). The transformation was successful and one colony of transformants was used to amplify the plasmid DNA in a mini-prep as described in Chapter 8, section 8.2.1.1. The purified plasmid DNA was then used in future transformations for the protein expression experiments.

5.2.2 Initial attempts at MtGlcT expression

To determine the best conditions for protein expression, a series of small expression studies were carried out. MtGlcT was expressed in *E. coli* BL21 (DE3) cells. These experiments varied the amount of IPTG used to induce expression, the temperature at which the cultures were incubated during expression as well as the length of expression. The conditions trialled are summarised in Table 5.1.

Table 5.1 Conditions used in expression studies of MtGlcT

Experiment	[IPTG] used to induce expression / mM	Temperature of expression /°C	Time of expression / h
1	1.0	18	18
2	0.5	18	18
3	0.1	18	18
4	1.0	37	18
5	1.0	37	4

Initially, IPTG concentration was changed to investigate the effect of varying the amount used to induce protein expression on the amount of soluble protein produced. The ITPG concentrations tested were 1.0, 0.5 and 0.1 mM. Multiple small scale (100 mL) protein expressions were completed as per the procedure described in chapter 8, section 8.2.4.1. The soluble lysate fraction and the insoluble fraction from each experiment were analysed using SDS-PAGE.

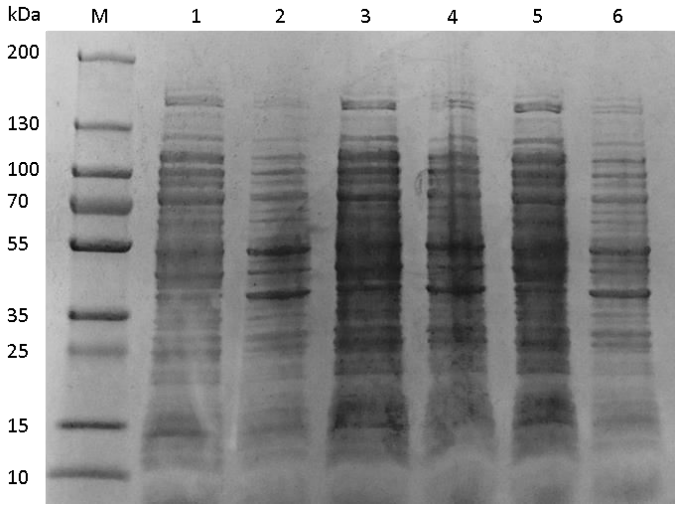


Figure 5. 3 SDS-PAGE analysis of expression studies on MtGlcT with changing IPTG concentration. Lanes 1-2: 1.0 mM IPTG, lysate, insoluble; Lanes 3-4: 0.5 mM IPTG, lysate, insoluble; Lanes 5-6: 0.1 mM IPTG, lysate, insoluble; M: marker. MtGlcT = approx. 45 kDa.

SDS-PAGE analysis, seen in figure 5.3, shows no major bands at the correct molecular weight for MtGlcT (45 kDa). There are thick bands at the 55 kDa mark as well as around the 40 kDa range. If the latter of these is the target protein (MtGlcT has a mass of approximately 45.6 kDa), then it would appear that there are extremely low levels of soluble protein for each IPTG concentration tested.

Levels of expression at 18 °C overnight were insufficient, so the next set of expression experiments investigated the effect of expression temperature and time. Therefore, MtGlcT was expressed on small scale (100 mL) at 37 °C, over time periods of either 4 hours or 18 hours. The results of this expression were then compared with that of the 18 °C overnight expression experiment *via* SDS-PAGE analysis, the results of which are shown in figure 5.4.

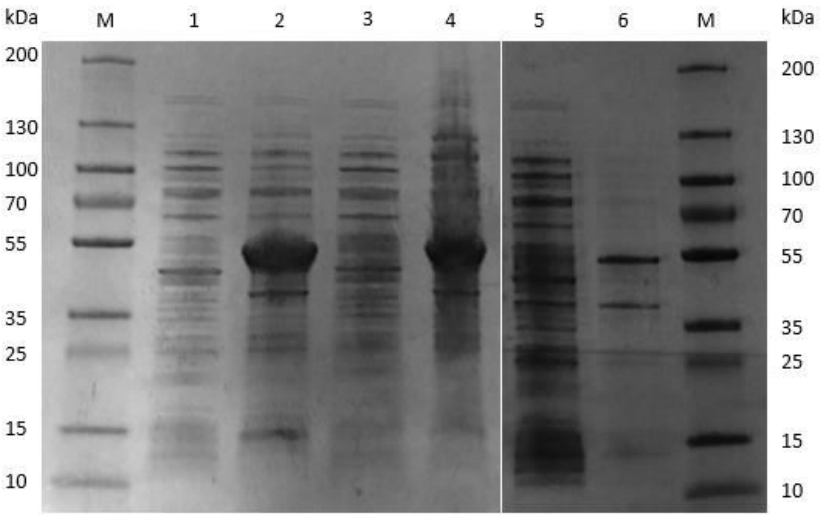


Figure 5. 4 SDS-PAGE results of further expression studies of MtGlcT. Lanes 1-2: 37 °C expression, 4 hours, lysate, insoluble; Lanes 3-4: 37 °C expression, 18 hours, lysate, insoluble; Lanes 5-6: 18 °C expression, 18 hours, lysate, insoluble; M: marker. MtGlcT = approx. 45 kDa.

When incubated at 37 °C, a large band could be seen at the 45 – 50 kDa range, suggesting the presence of over-expressed protein. However, the protein remains in the insoluble fraction, suggesting that the protein has been expressed but is contained within inclusion bodies.

Due to the successful use of the chaperone proteins GroEL and GroES in the solubilisation of MtGBE (see chapter 3, section 3.2.2), the same technique was applied in this case. Therefore, MtGlcT was co-expressed with the GroEL-GroES chaperone system.

5.2.3 Co-transformation and co-expression of MtGlcT with the GroEL-GroES chaperone system

The co-transformation of pET28a(+)::*Rv3032* and pGro7 was achieved using co-expression method 2 described in chapter 8, section 8.2.2.5. This method involved initial transformation of the pGro7 plasmid into cells which were then grown and once more made chemically competent. A subsequent transformation of pET28a(+)::*Rv3032* led to cells containing both plasmids.

The temperature and length of expression were varied to optimise conditions. Protein was expressed at either 18 °C or 37 °C over a period of either 4 or 18 hours. For each

experiment, the soluble lysate and the insoluble cell debris were analysed using SDS-PAGE (figure 5.5).

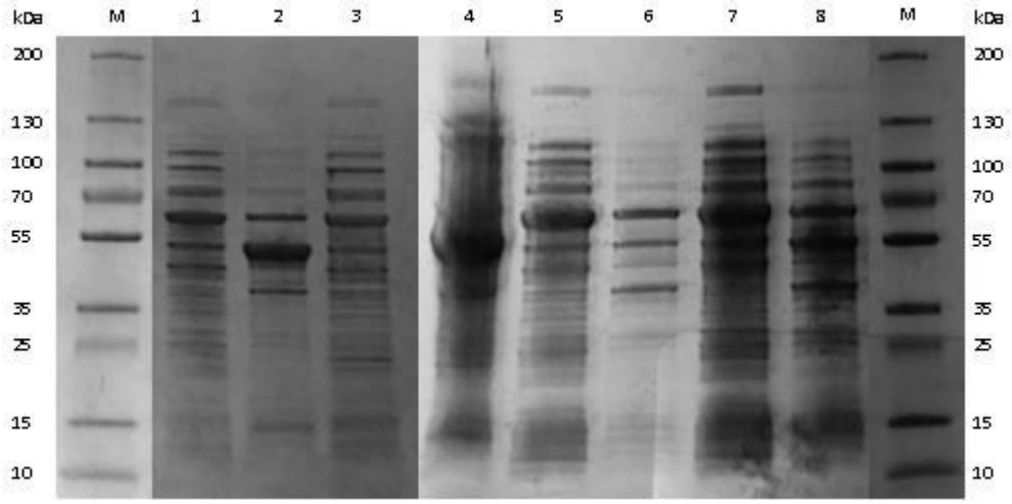


Figure 5. 5 SDS-PAGE results of co-expression studies of MtGlcT with GroEL-GroES system. Lanes 1-2: 37 °C expression, 4 hours, lysate, insoluble; Lanes 3-4: 37 °C expression, 18 hours, lysate, insoluble; Lanes 5-6: 18 °C expression, 4 hours, lysate, insoluble; Lanes 7-8: 18 °C expression, 18 hours, lysate, insoluble; M: marker. MtGlcT = approx. 45 kDa.

SDS-PAGE analysis shows a thick band with a mass of approximately 60 kDa which corresponds to the mass of GroEL (57 kDa), suggesting the chaperones were successfully expressed. The experiments at 18 °C show very low expression levels of target protein in both the lysate (lanes 5 and 7) and the insoluble fractions (lanes 6 and 8), suggesting this temperature is not compatible with the expression of MtGlcT. The experiments at 37 °C both show overexpression of MtGlcT, though the protein still resides mainly in insoluble inclusion bodies (lanes 2 and 4).

In a further attempt to achieve soluble MtGlcT, another chaperone system was investigated. This was also procured from Takara Bio Inc. and sent from collaborators (Prof. Kim group, SKKU). The pTf16 plasmid contains the *tig* gene which encodes the trigger factor (TF) chaperone protein. This protein was co-expressed with MtGlcT to see if this was more successful than the GroEL-GroES system.

5.2.4 Co-transformation and co-expression of MtGlcT with the trigger factor chaperone protein

5.2.4.1 The trigger factor chaperone protein

The TF protein is a 48 kDa chaperone protein that is found bound to the ribosome; the only chaperone protein found in bacteria to do so.[291] TF binds to the ribosome and thus interacts with the nascent protein chain as it is being transcribed, helping to fold the chain correctly as it forms.[292-294]

Although the exact mechanism has not been fully characterised, the crystal structure of TF has been solved and, from studies of the protein structure with peptide chains bound, hypotheses have been proposed regarding its mechanism of action.[295, 296] For example, Ferbitz *et al.* found that TF forms a folding space in which the peptide chain is shielded, potentially from proteases and aggregation.[295] Additionally, Baram *et al*. suggest that TF provides an alternative hydrophobic environment which prevents protein aggregation.[296]

More recently, molecular dynamics studies have potentially elucidated more detail in the potential mechanism, suggesting the TF protein 'pinches' the nascent protein chain and encapsulates it in a manner which is novel in comparison to other known chaperone proteins such as GroEL.[297] Additionally, Liu *et al.* used optical tweezers to observe protein folding and found that TF protects the emerging protein chain from unfavourable interactions with other, unfolded, parts of the same chain.[298]

As well as binding emerging peptide chains, it has also been proposed that TF can bind full length proteins in order to stabilise them and rescue them from misfolding.[299]

The mechanism and location of TF in the cell is different to that of the GroEL-GroES chaperone system and therefore provides an alternative solubilisation option in the goal of expressing soluble MtGlcT protein.

5.2.4.2 Co-expression with TF and purification of MtGlcT

Studies by collaborators from the Kim group in SKKU, South Korea, involved small scale co-expression studies on MtGlcT with various chaperone systems. Of these, the most successful co-expression was that with the TF chaperone when expression was carried out overnight at 18 °C, as seen in the SDS-PAGE in figure 5.6. The purification of co-

expressed MtGlcT was also investigated and these conditions that were repeated in this work in order to reproduce their results.

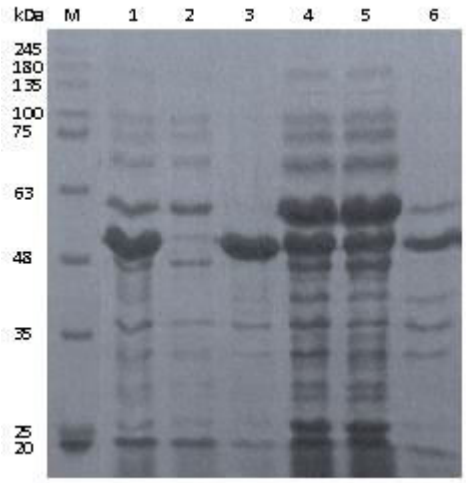


Figure 5. 6 SDS-PAGE of co-expression study of MtGlcT and TF chaperone protein. Lanes 1-3: 37 °C, whole cell, lysate, insoluble; Lanes 4-6: 18 °C, whole cell, lysate, insoluble; M: marker. Experiment and SDS-PAGE carried out by Kim group members at SKKU, South Korea. MtGlcT = approx. 45 kDa.

The plasmid pTf16 contains the same resistance marker (Chlor) and promotor system (*araB*) as the pGro7 plasmid. The plasmid was transformed into *E. coli* BL21 (DE3) chemically competent cells then amplified and isolated using a plasmid mini-prep. The purified plasmid was then used in the co-transformation with pET28a(+)::*Rv3032*.

The co-transformation was carried out by transforming both plasmids into *E. coli* BL21 (DE3) chemically competent cells at the same time (method 1, chapter 8, section 8.2.2.5). Colonies were successfully grown, suggesting the co-transformation was successful. The transformants were used in a 1.0 L growth and protein was expressed at 18 °C overnight. Soluble cell lysate and insoluble cell debris were analysed using SDS-PAGE analysis with the results shown in figure 5.7. A control experiment in which the pET28a(+)::*Rv3032* plasmid was not present, but the pTf16 plasmid was present, was carried out under the same conditions on a smaller scale (100 mL). The lysate and insoluble fractions were also analysed and the results are shown in figure 5.7

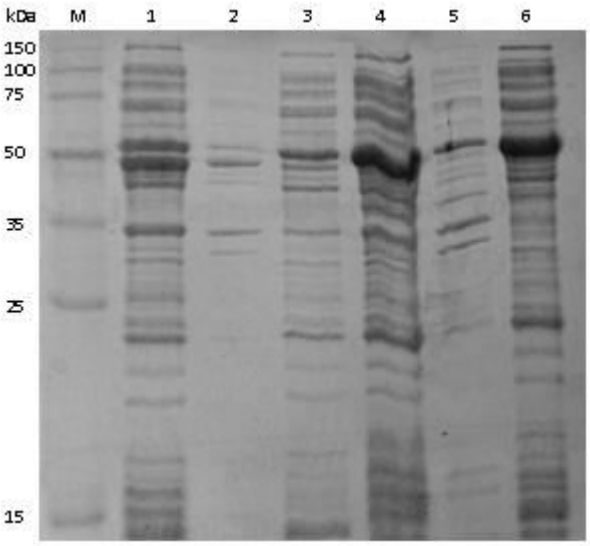


Figure 5. 7 SDS-PAGE of co-expression study of MtGlcT and TF chaperone (Lanes 1-3) as well as expression of TF protein only (Lanes 4-6). Lane 1 + 4: whole cell; Lane 2 + 5: insoluble; Lane 4 + 6: lysate; M: marker. MtGlcT = approx. 45 kDa.

SDS-PAGE analysis suggested that the expression of MtGlcT had occurred with a modest level of overexpression, due to the thick band in the whole cell sample just below the 50 kDa mark. However, the majority of expressed protein was still located in the insoluble fraction. The control experiment shows that the thick band above the 50 kDa mark is due to the chaperone proteins, as this appears both in the control without MtGlcT and in the co-expression.

Despite the low levels of soluble target protein in the lysate, purification was still attempted to test if this would be a viable method of obtaining MtGlcT. For purification, the lysate was loaded onto a Ni-NTA column following procedure 8.2.1.10 (chapter 8). Ni bound protein was then eluted over a gradient of 0-100% elution buffer. No absorbance peak was detected in any of the fractions collected during elution and later Bradford analysis of the peaks confirmed that no protein was present. Therefore, the co-expression with the TF chaperone protein was also deemed unsuccessful and was not further pursued.

5.2.5 Large scale expression of MtGlcT in an attempt to obtain soluble protein

To obtain soluble MtGlcT protein, of which there were low levels, a large scale protein expression (8.0 L) was carried out. MtGlcT was expressed in *E. coli* BL21 DE3 cells at 18

°C overnight following induction with 0.5 mM IPTG. MtGlcT was purified by Ni-NTA affinity chromatography with the chromatograph and SDS-PAGE analysis of fractions shown in figure 5.8.

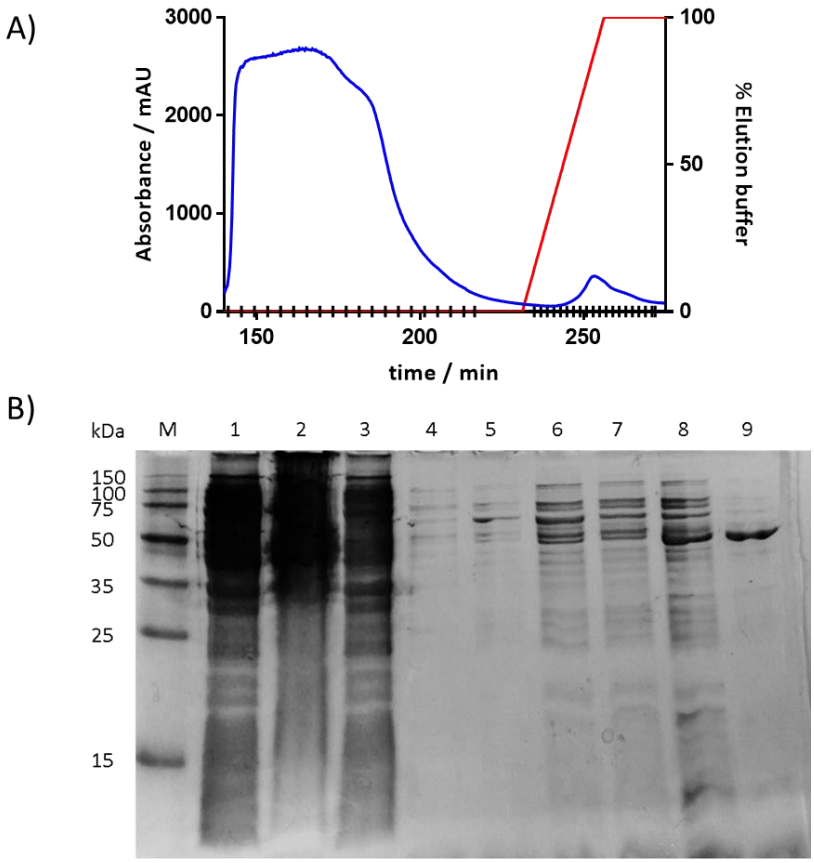


Figure 5. 8 A) Absorbance trace (blue), % elution buffer (red) and fractions (black) from Ni-NTA purification of MtGlcT from 8 L expression. B) SDS-PAGE analysis of fractions: Lane 1: lysate, Lane 2: insoluble; Lane 3: fraction 5 (flow through); Lane 4: Fraction 23; Lane 5: Fraction 26; Lane 6: Fraction 28; Lane 7: Fraction 30; Lane 8 Fraction 32; Lane 9: Fraction 34; M: Marker. MtGlcT = approx. 45 kDa.

The later fractions collected appeared to contain MtGlcT protein (bands just below 50 kDa). Following dialysis into 50 mM HEPES pH 7.4, a large amount of precipitate formed, suggesting the protein had been unstable and had crashed out of solution. Bradford assay determined a protein concentration in the supernatant of 0.1 mg/mL and therefore, the supernatant was concentrated further. However, further protein precipitation meant that the sample could not be concentrated. As the final volume was small and of low concentration, this method was not deemed viable due to the

lack of stability of the protein. The expression could have been attempted again using further precautions to attempt to prevent protein precipitation, however, another method involving solubilisation from IBs was being developed by Dr. Gu Yoo of the same lab which showed promise in obtaining recombinant MtGlcT. This potentially more successful route to MtGlcT was therefore pursued.

5.2.6 Isolation and purification of MtGlcT from inclusion bodies

Work carried out by Dr. Gu Yoo involved solubilisation of MtGlcT from inclusion bodies. This was achieved by re-suspending the insoluble cell debris into a solution of 0.2% sarkosyl at 4 °C for 18 hours. The sarkosyl was removed *via* dialysis into phosphate buffered saline (PBS). The solubilisation was analysed with SDS-PAGE, the results of which are shown in figure 5.9. This solution of re-solubilised protein was then used in this project. The solution obtained was cloudy, suggesting protein precipitation.

To confirm the presence of MtGlcT, whether it had precipitated and if any had remained in solution, the solid was removed via centrifugation (8000 rpm, 4 °C, 30 minutes) and kept separately to the supernatant. Once resuspended, the solid was analysed and compared to the supernatant using SDS-PAGE. The results can be seen in figure 5.10.

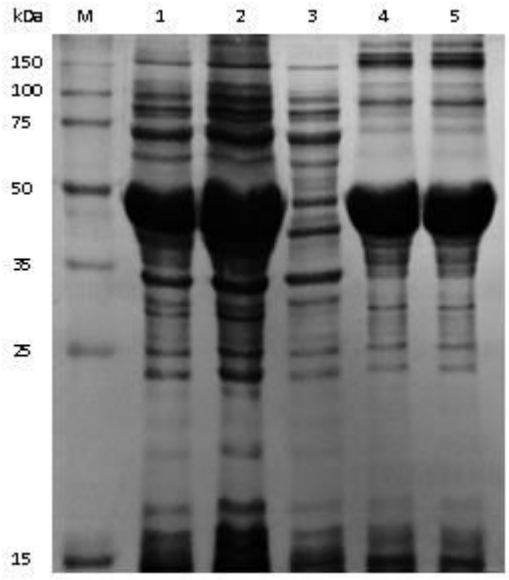


Figure 5. 9 SDS-PAGE of solubilisation of MtGlcT from IBs. Lane 1: whole Cell; Lane 2: After sonication; Lane 3: lysate; Lane 4: Solubilisation of insoluble pellet with 0.2% sarkosyl; Lane 5: Supernatant after dialysis into PBS buffer; M: marker. Solubilisation and SDS-PAGE carried out by Dr. Gu Yoo. MtGlcT = approx. 45 kDa.

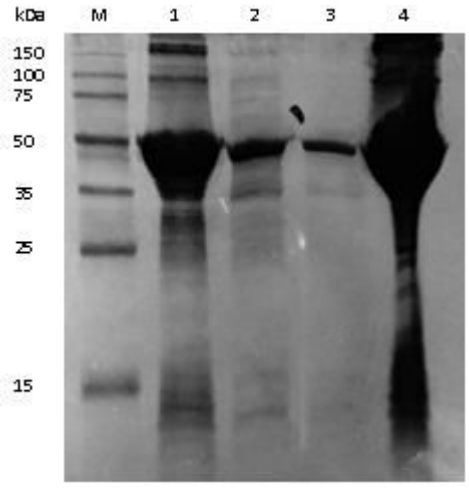


Figure 5. 10 SDS-PAGE of MtGlcT solutions. Lane 1: Solution of MtGlcT after dialysis (see Figure 5.9, Lane 5); Lane 2: Supernatant after removal of precipitate; Lane 3: supernatant after removal of precipitate, 2x dilution; Lane 4: Re-suspension of precipitate; M: marker. MtGlcT = approx. 45 kDa.

The SDS-PAGE shows that a lot of MtGlcT may have precipitated, however Bradford assay determined that 2.0 mg/mL protein remained in solution. Therefore, the

supernatant was concentrated to give a final concentration of 40 mg/mL. Partially purified MtGlcT was then used for initial activity assays.

5.3 Glycosyl transferase coupled assay with MtGlcT

5.3.1 Pyruvate kinase – lactate dehydrogenase coupled assay

Many GT enzymes exist in nature.[300] Common donors include sugar nucleotides (e.g. UDP-Glucose) or lipid linked sugars (e.g. dolicholphosphate mannose).[300] In the case of MtGlcT, it was reported that this GT showed activity when UDP-Glc was used as a donor in the elongation of α-glucan chains.[171]

Detection of GT activity on sugar nucleotide donors can be continuously monitored by coupling the transferase reaction with the activities of pyruvate kinase (PK) and lactate dehydrogenase (LDH).[301, 302] The biological role of the PK enzyme is to catalyse the formation of ATP and pyruvate from ADP and phosphoenolpyruvate (PEP) in the final step of glycolysis.[303] This enzyme can also utilise UDP in the same reaction to produce UTP.[301] LDH catalyses the reversible conversion of pyruvate to lactate.[304] This reaction also involves the oxidation of NADH to NAD^+ and is the final component of the PK/LDH coupled GT assay. The consumption of NADH can be followed by monitoring the decrease in absorbance at 340 nm.[301] A summary of the coupled reaction can be seen in Figure 5.11.

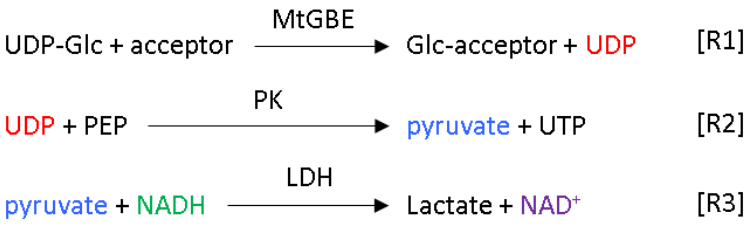


Figure 5.11 Reactions involved in the PK/LD coupled GT assay. Reconstructed from Palcic and Sujino.[301]

As MtGlcT confers a glucose unit to the acceptor molecule, 1 molecule of UDP is released, which is then utilised as a substrate by PK to produce pyruvate. This pyruvate is then converted to lactate by LDH which also oxidises one molecule of NADH, thus reducing the absorbance at 340 nm. Therefore, the formation of products from the GT

reaction [R1] is proportional to the amount of NADH that is oxidised in the LDH reaction [R3]. Thus, the GT activity of MtGlcT was investigated using this coupled assay technique.

5.3.2 MtGlcT assay

Preliminary studies mentioned by Jackson *et al*. suggest MtGlcT acts upon the substrates UDP-Glc and short chain α-glucans.[171] To confirm this, the PK-LDH coupled assay was employed in which MtGlcT activity was monitored in the presence of maltose and maltotriose as the putative acceptor molecules. To investigate if a single sugar could act as an acceptor, glucose and GG were also tested. The reaction mixture contained UDP-Glc and the acceptor being tested as well as PEP, NADH, PK and LDH. MtGlcT was used at 1 mg/mL and the reaction was performed at 37 °C. To measure the decrease in NADH, the absorbance was measured at 340 nm. Assay composition and protocol is described in chapter 8, section 8.2.4.6. The results are shown in figure 5.12.

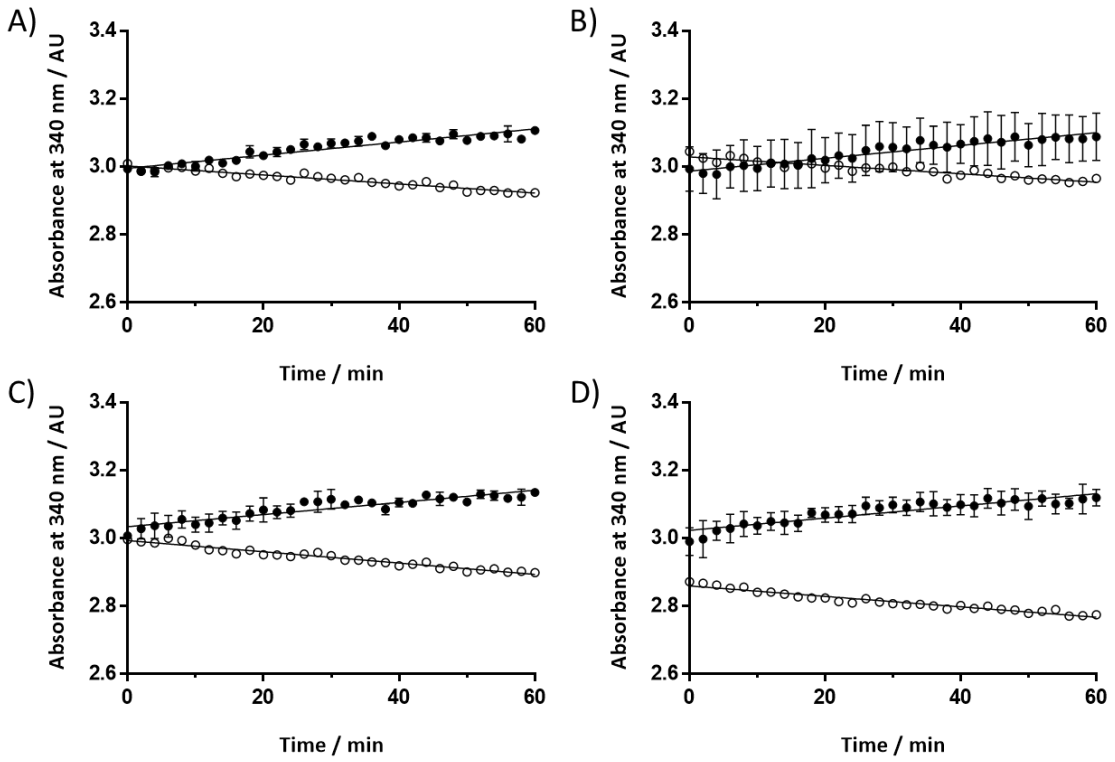


Figure 5. 12 PH/LDH coupled GT assay with MtGlcT and: A) Maltose; B) Maltotriose; C) Glucose; D) GG. Reaction with glucan acceptor (●) and control reaction with no MtGlcT present (○). Reactions performed in triplicate. Linear regression performed with GraphPad Prism 7.0.

Table 5. 2 Summary of gradients from figure 5.12: PH/LDH coupled GT assay with MtGlcT and various α-glucans.

	Gradient	
Acceptor substrate	**Reaction**	**Negative control**
Maltose (A)	0.001935 ± 0.000123	-0.001328 ± 0.000070
Maltotriose (B)	0.001889 ± 0.000112	-0.001240 ± 0.000085
Glucose (C)	0.001791 ± 0.000137	-0.001662 ± 0.000073
GG (D)	0.001799 ± 0.000152	-0.001535 ± 0.000073

The results show that no GT activity is exhibited by MtGlcT with any of the substrates tested. Whilst the control experiments show a slight decrease in absorbance at 340 nm (the gradients of the control reactions are all between -0.0012 and -0.0017, see table 5.2), the reactions in which MtGlcT is present show a slight increase in absorbance (gradients of each reaction are between 0.0017 and 0.002). The decrease observed in the control may be due to chemical degradation of the UDP-Glc at 37 °C over time, reflected by the very low level of decrease in absorbance observed. It is unclear what caused an increase in absorbance at 340 nm in the enzymatic reactions, however the gradients for each experiment (see table 5.2) are all similar, suggesting the increase may not be due to enzymatic activity. Interestingly, the gradients for the trend line in each experiment are almost of the same magnitude as the control but are positive rather than negative.

In addition to the single sugar and short chain glucans, longer chain glucans were also tested. Maltoheptaose, amylose, amylopectin and glycogen were all tested in the PK/LDH coupled assay with MtGlcT and UDP-Glc. All components were present in the same concentration as before, though in the case of polysaccharide chains, the amounts used are as follows: amylose 0.03%; amylopectin 0.25 %; glycogen 0.0625%. An additional control experiment was also run, in which MtGlcT was present but no acceptor sugar was added to the reaction mixture. The results are shown in figure 5.13.

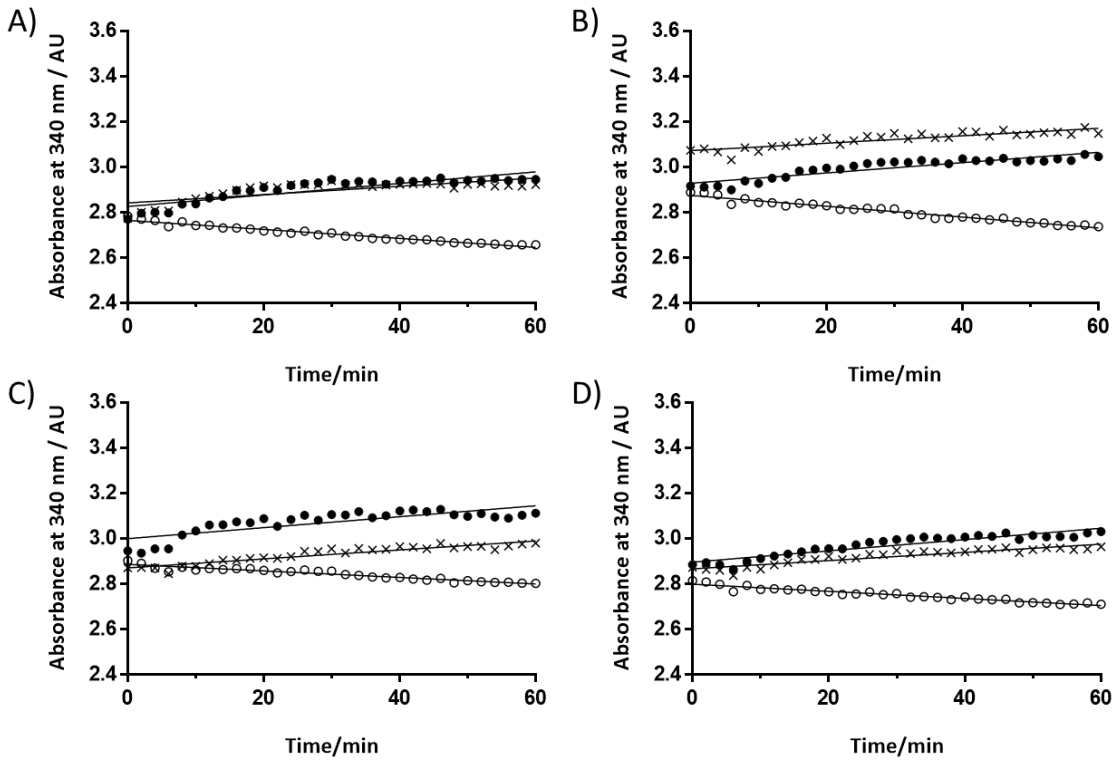


Figure 5. 13 PH/LDH coupled GT assay with MtGlcT and: A) Maltoheptaose; B) Amylose; C) Amylopectin; D) Glycogen. Reaction with glucan acceptor (●), control reaction with no MtGlcT present (○) and control reaction without acceptor present (×). Linear regression performed with GraphPad Prism 7.0.

Table 5. 3 Summary of gradients from figure 5.13: PH/LDH coupled GT assay with MtGlcT and various polysaccharides.

	Gradient		
Acceptor substrate	**Reaction**	**Negative control – no MtGlcT**	**Negative control – no acceptor**
Maltoheptaose (A)	0.002548 ± 0.0002699	-0.001984 ± 0.000074	0.001831 ± 0.0002956
Amylose (B)	0.002274 ± 0.0001983	-0.002393 ± 0.000093	0.001646 ± 0.0001645
Amylopectin (C)	0.002415 ± 0.000342	-0.001442 ± 0.000083	0.001979 ± 0.0001391
Glycogen (D)	0.002493 ± 0.0001885	-0.001571 ± 0.000079	0.001813 ± 0.0001426

The results show that MtGlcT does not exhibit GT activity when longer glucan chains are used as acceptors. The reaction without MtGlcT present shows a slight decrease in absorbance over time. Similarly to the experiments depicted in figure 5.12, gradients

of trend lines range between -0.0014-0.0024 (see table 5.3). The control reactions in which MtGlcT is present show no decrease but instead a slight increase. This suggests that the increase in absorbance could be due to the enzyme itself. In addition, the gradients of the control experiments in which MtGlcT was present (between 0.0017 and 0.0020) are close to those of the reaction containing all components (between 0.0023 and 0.0026) suggesting no significant difference between the experiments. Due to these results it was concluded that maltoheptaose, amylose, amylopectin and glycogen are not be substrates of MtGlcT.

5.3.3 MtGlcT assay in the presence of metal ions

Known GT enzymes that also belong to the GT4 family have been assayed in the presence of metal ions and MtGlcT may also require the presence of metal ions to display activity.[305-307] Therefore, reactions were repeated with either $MgCl_2$ (0.2 M) or KCl (0.2 M) present in the reaction mixture. Maltotriose and maltoheptaose were tested as well as maltotetraose which was used to test a chain length midway between the other substrates. All other assay components remained as before. The results are shown in figure 5.14.

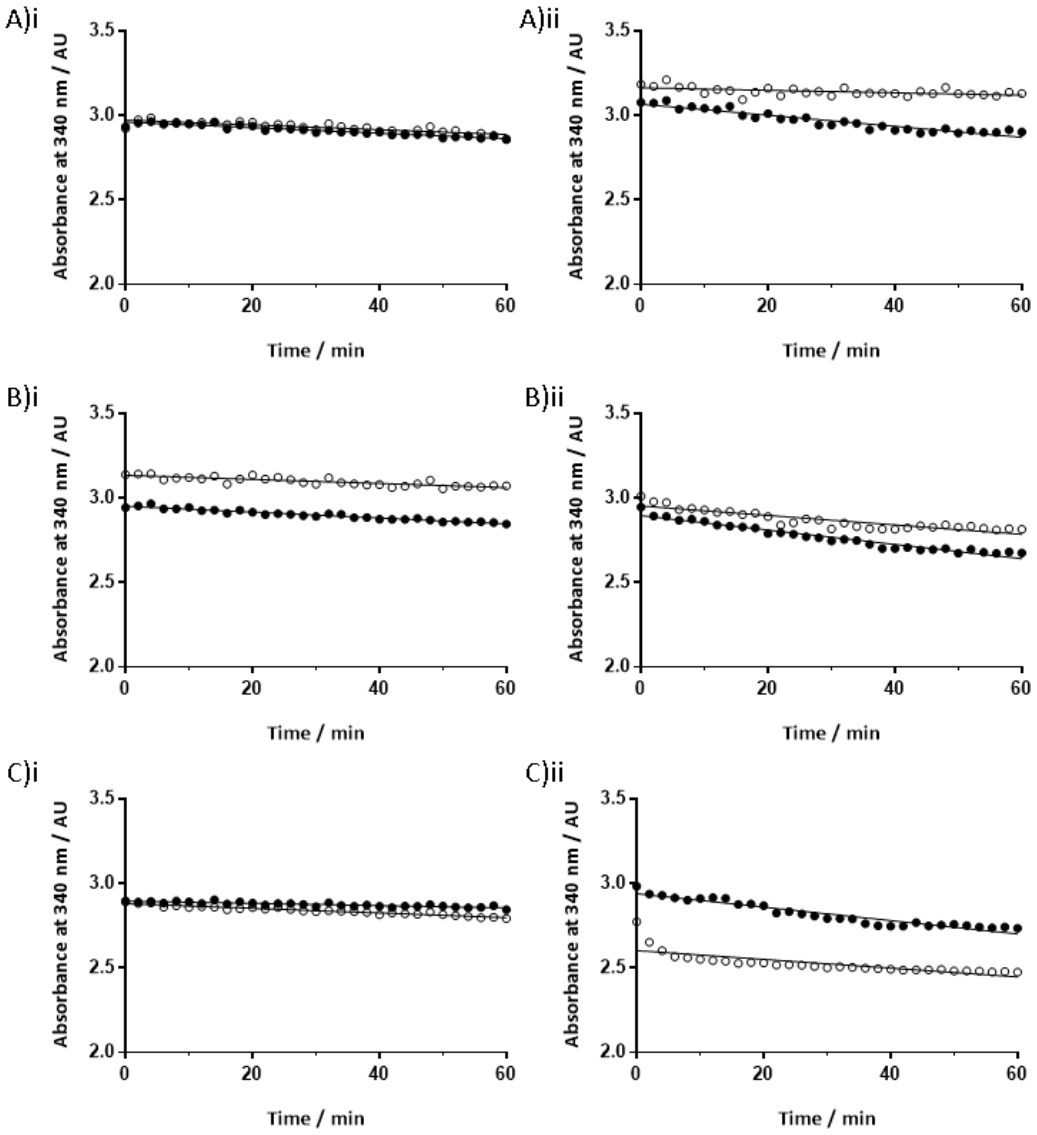


Figure 5. 14 PH/LDH coupled GT assay with MtGlcT and: A) Maltotriose; B) Maltotetraose; C) Maltoheptaose. Reactions carried out either in the presence of KCl (i) or $MgCl_2$ (ii). Reaction with glucan acceptor (●) and control reaction with no acceptor present (○). Linear regression performed with GraphPad Prism 7.0.

Table 5. 4 Summary of gradients from figure 5.14: PH/LDH coupled GT assay in the presence of either K^+ or Mg^{2+}.

	Gradient	
Acceptor substrate	**Reaction**	**Negative control**
Maltotriose with K^+ (A, i)	-0.001633 ± 0.000102	-0.001413 ± 0.000162
Maltotriose with Mg^{2+} (A, ii)	-0.003273 ± 0.000203	-0.000731 ± 0.000212
Maltotetraose with K^+ (B, i)	-0.001742 ± 0.000078	-0.001208 ± 0.000147
Maltotriose with Mg^{2+} (B, ii)	-0.004247 ± 0.000186	-0.002796 ± 0.000261
Maltoheptaose with K^+ (C, i)	-0.000669 ± 0.000073	-0.001377 ± 0.000069
Maltoheptaose with Mg^{2+} (C, ii)	-0.003981 ± 0.000227	-0.002599 ± 0.000389

The results in figure 5.14 show a change in trend from the reactions in which no metal ions were present at all (figures 5.12 and 5.13). This time, there is a decrease in absorbance in every experiment as opposed to the upwards trends seen before. This may indicate that the presence of metal ions is influencing enzyme stability. When comparing the gradients of the slope of each experiment (listed in table 5.4), when potassium is present, there is no significant difference between the reaction and the control reactions. When magnesium ions are present, there is a larger difference between reaction and control reactions, though again, this difference is not significant. For example, the reaction carried out with maltotriose (figure 5.14, A) has a steeper gradient (-0.0033) than that of the corresponding control reaction (-0.0007), however the difference is only 0.0026. If this is activity being observed, then the enzymatic activity for MtGlcT appears very weak. The trends observed in the maltoheptaose experiment (figure 5.14, C) appeared similar and so this experiment was therefore repeated with an increase in enzyme concentration to determine the effect this has on the rate of reaction.

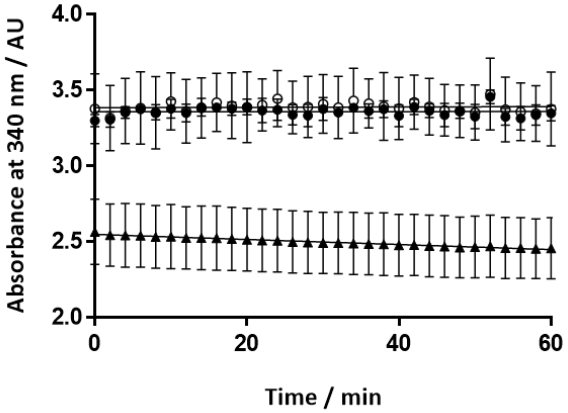


Figure 5. 15 PK/LDH coupled GT assay with MtGlcT and maltoheptaose (●). Reaction carried out in triplicate. Control reactions with no maltoheptaose present (○) or with no MtGlcT (▲) carried out in duplicate.

The results of the assay, shown in figure 5.15, show that the decrease in absorbance in the earlier assay was not due to GT activity. An increase in protein concentration did not lead to a steeper decrease in absorbance, but gave a more level trend (gradient of reaction = 0.000030 ± 0.000318). In addition, the slope of the reaction results is close to that of the control reaction in which no acceptor is present (0.000116 ± 0.000328). This result led to the conclusion that the glucans tested so far are not active substrates for MtGlcT.

In the above reactions, no positive control was used to validate the assay. For this reason, a known GT enzyme, OtsA, was expressed, purified and analysed in the PK/LDH assay as a positive control experiment.

5.4 Expression and purification of OtsA as a positive control for coupled glycosyl transferase assay

5.4.1 OtsA function and role

OtsA is a trehalose-6-phosphate synthase found throughout nature, from animals and insects (in which it is known as trehalose-phosphate synthase (TPS)) to plants and bacteria.[308-310] It works alongside OtsB (a trehalose-6-phosphate phosphatase) in one of the biosynthetic pathways in which trehalose is produced, forming trehalose-6-phosphate (**5.3**) from UDP-Glc (**1.33**) and Glucose-6-phosphate (G6P) (**5.2**) as depicted

in figure 5.16.[311] Trehalose is a non-reducing disaccharide widespread throughout nature and is prevalent in organisms that have evolved to survive environmental stresses.[312] OtsA is a retaining GT that has been categorised into the CAZy family GT20 (www.cazy.org) and has been mechanistically and structurally characterised.[308, 313]

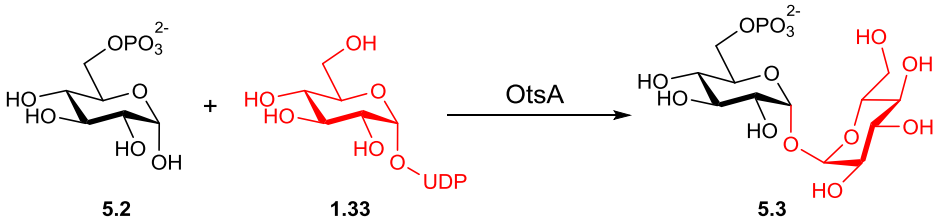


Figure 5. 16 Formation of trehalose-6-phosphate from UDP-Glc and G6P, the reaction catalysed by OtsA. Recreated from De Smet *et al*.[311]

OtsA uses a sugar nucleotide donor and therefore releases UDP during the GT reaction making it suitable for use in the PK/LDH coupled assay. Indeed, literature reports have used this assay to monitor GT activity of OtsA.[313]. For this reason, OtsA was expressed and purified for use in the coupled GT assay as a positive control.

5.4.2 Expression and purification of OtsA

The expression conditions for the OtsA protein were based on Gibson *et al.* who report its full expression, purification and characterisation.[308]

The *OtsA* gene was contained in a plasmid which also contained an ampicillin (Amp) resistance marker and a *lac* operon for induction of protein expression. The plasmid was transformed into *E. coli* BL21(DE3) chemically competent cells as per the heat shock method described in chapter 8, section 8.2.1.4. Expression was performed on a 1.0 L scale with 0.1 mM IPTG added for induction. Protein was expressed at 30 °C overnight. The lysate was purified using a Ni-NTA affinity column, the results of which were analysed using SDS-PAGE which is shown in figure 5.17.

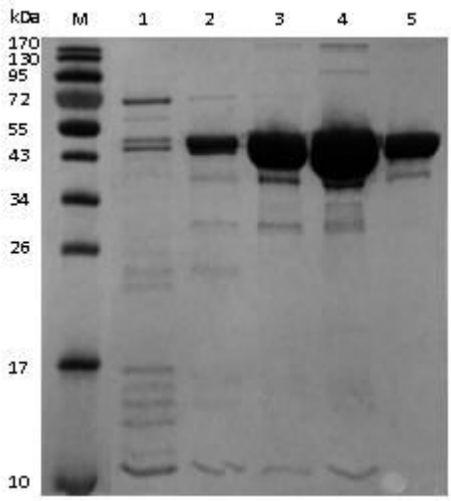


Figure 5. 17 SDS-PAGE analysis of Ni-NTA purification of OtsA. Lane 1: Flow through; Lane 2-5: eluted peak (fractions 27-30); M: marker. OtsA = 54.7 kDa.

The OtsA protein is 54.7 kDa in length[308], corresponding to the prominent band observed between the 43 and 55 kDa markers from the protein ladder. Upon dialysis into 5.0 mM HEPES buffer (pH 7) containing 60 mM $MgCl_2$ some precipitation was observed. However, sufficient soluble protein was retained and concentrated for use in future experiments.

The concentrated protein was then used in the PK/LDH coupled assay as a positive control experiment.

5.5 OtsA as a positive control in the PK/LDH coupled assay

As OtsA acts upon sugar nucleotides (in this case UDP-Glc), the PK/LDH coupled assay is suitable for monitoring its GT activity. As UDP-Glc is hydrolysed by the enzyme, the UDP that is released can act as a substrate for PK which produces the pyruvate substrate for LDH. This last reaction involves the oxidation of NADH to NAD^+, a process which is monitored by measuring the decrease in absorbance at 340 nm.

The assay solution used contained UDP-Glc, G6P, PEP, NADH, PK, LDH, $MgCl_2$, KCl and BSA. The final concentration of OtsA used was 0.1 mg/mL. The assay was initiated by the addition of OtsA and the reaction incubated at 37 °C for 1 hour. Full experimental details can be found in chapter 8, section 8.2.4.8. A negative control experiment was run in which no G6P was present. The results are shown in figure 5.18.

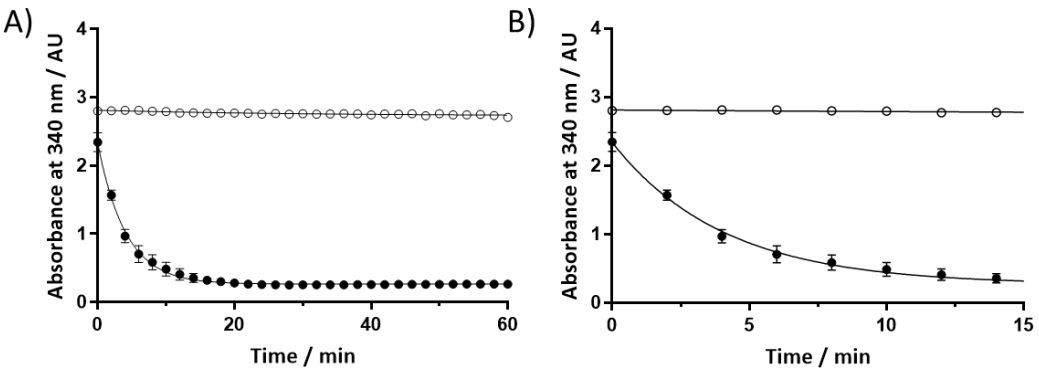


Figure 5. 18 A) PK-LD coupled transferase assay of OtsA and UDP-Glc and G6P (●). Control experiment run in which no Glc-6-P was present (○). Reaction carried out in triplicate. B) zoom region showing the first 15 minutes of the reaction. Linear and non-linear regression carried out on GraphPad Prism 7.0.

As seen in figure 5.18, the reaction was successful. As expected, the absorbance decreased rapidly within the first 10 minutes of the reaction (initial gradient = -0.56). After this, the decrease in absorbance slows down until the absorbance reading is a flat line and the reaction goes to completion. The control experiment in which no G6P was present confirms that this decrease was not due to the degradation of UDP-Glc by any other assay components, nor was OtsA hydrolysing UDP-Glc without an acceptor present. The small range of errors in the experiment suggests the results are reliable, thus confirming that the assay technique used is viable for the observation of GT activity.

5.6 Conclusions

Recombinant MtGlcT was located predominantly in inclusion bodies during over-expression of the protein. Attempts at co-expression, however, were unsuccessful at producing soluble protein for both sets of chaperone proteins used (GroEL/GroES system and TF chaperone protein). Though the TF chaperone protein appeared to yield soluble protein, Ni-NTA purification was unsuccessful. Future work could be carried out to investigate if FPLC could be optimised to obtain purified co-expressed protein.

MtGlcT was re-solubilised from IBs using 0.2 % sarkosyl (work carried out by Dr. Gu Yoo). The refolded protein was analysed and concentrated for use in the PK/LDH coupled assay.

The coupled PK/LDH assay was validated as a method for monitoring GT activity using OtsA as a positive control. OtsA was successfully expressed and purified using the methods previously published.[308, 313] When MtGlcT was tested in the PK/LDH assays, however, no transferase activity was observed for any of the following α-glucans: maltose, maltotriose, maltotetraose, maltoheptaose, amylose, amylopectin or glycogen. GT activity was also not observed when metal ions were present, suggesting that the lack of activity was not merely due to the requirement of a metal cofactor.

To test the hypothesis that MalG could be a donor for MtGBE and that this could be synthesised from UDP-Glc and GG, GG was tested as an acceptor as well as the α-glucans already mentioned. Again, no activity was observed when MtGlcT was incubated in the presence of UDP-Glc and GG, thus suggesting that this reaction either does not occur, or is not catalysed by MtGlcT.

The lack of activity may be due to the fact that the natural substrates for MtGlcT were not tested. It was noted that activity was observed with UDP-Glc and short chain glucans, but the glucans tested were not reported.[171, 290] In addition, other sugar nucleotides, such as ADP-Glc may act as substrates but were not tested in this work. It could also be that the glycerate moiety is essential for substrate recognition. If this were the case, activity may be observed if DGG were to be tested with MtGlcT.

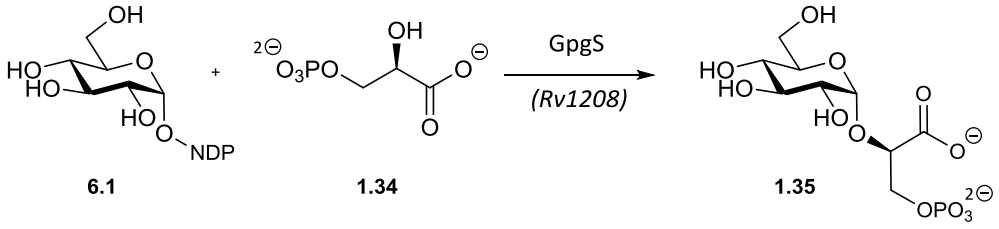
Due to the re-solubilisation of MtGlcT from IBs, the MtGlcT protein may have been incorrectly folded. Though there was no time to test this in this work, techniques such as circular dichroism (CD) could be employed to check for the folded state of the protein. Denatured protein would have to be used as a comparison as no known CD spectra of MtGlcT have been reported as far as the author is aware. In further work, difficulties in expressing soluble MtGlcT may be overcome through use of chaperones or other techniques, to obtain soluble recombinant protein.

In conclusion, activity was not observed when re-folded protein was used in the PK/LDH assay. This may be due to lack of correct folding or just down to the incorrect substrates being tested. To confirm this, further studies using protein that is confirmed as correctly folded and other glucan substrates or sugar nucleotides would need to be carried out.

6. Expression and purification of *Mycobacterium tuberculosis* glucosyl phosphoglycerate synthase (MtGpgS)

6.1 Role of MtGpgS in MGLP biosynthesis

The first step in MGLP biosynthesis in *Mycobacteria* is the formation of glucosyl 3-phosphoglycerate (GPG, **1.35**), a reaction catalysed by glucosyl phosphoglycerate synthase (GpgS) (figure 6.1).[171, 172] GPG is then converted to glucosyl glycerate (GG) by glucosyl-3-phosphoglycerate phosphatase (GpgP).[171, 178] GG, initially thought to be a rare biomolecule, is now acknowledged as a widespread metabolite found in many microbes.[314] Not only is GG a precursor to a range of biological molecules, but it is also found in its free form, the levels of which are found to increase during periods of environmental stress.[314, 315]



6.1 **1.34** **1.35**

Figure 6. 1 Reaction catalysed by GpgS: the formation of GPG from a sugar nucleotide donor (NDP-Glc, **6.1**) and 3-PGA (**1.34**).

The biosynthesis of GG was first reported in 2006 by Costa *et al.*[316] They report the use of recombinant GpgS and GpgP proteins from *Methanococcoides burtonii* in the synthesis of GG *via* the formation of GPG.[316] The authors found that GpgS from this archaeon was highly specific in terms of substrate recognition and the only sugar nucleotide that it would process was GDP-Glc.[316] The following year, a homologue of *M. burtonii* GpgS was identified in *Persephonella marina* and the recombinant protein expressed and studied biochemically.[317] This time, GpgS was found to show more flexibility in the sugar nucleotides utilised as donors, showing activity with UDP-Glc, GDP-Glc, ADP-Glc and TDP-Glc, in order of decreasing preference.[317] This was also the first report of the biosynthetic pathway of GG in a thermophilic organism.[317] As well as

this two-step pathway for GG biosynthesis, in 2007 Fernandes *et al.* discovered a one-step biosynthetic pathway to GG in, also in *P. marina*, in which they identified the enzyme glucosyl glycerate synthase (GgS), however this pathway to GG is less common.[318]

In 2008, the gene encoding the GpgS homologue in *Mycobacterium tuberculosis* was identified as *Rv1208*.[172] It was shown to be identical to the *Mycobacterium bovis* homologue and is believed essential to the survival of the bacterium.[169, 172] Though initially GpgS was classified as a family 2 glycosyl transferase, the expression, purification and characterisation of the recombinant forms of GpgS from *M. bovis* and *M. smegmatis* identified GpgS as a retaining GT belonging to CAZy family 81.[172, 174]

Crystallographic studies of GpgS began in 2008 when Gest *et al.* identified that although a monomer under the denaturing conditions of SDS-PAGE, GpgS appeared to form a dimer in solution.[173] The full 3D structure was reported the same year and residues involved in the binding of both donor and acceptor were identified.[174] Additional structural studies in 2012 were carried out, leading to a hypothesis regarding the enzymatic mechanism and highlighting the importance of a flexible loop in catalytic activity.[175]

The participation of GpgS in the biosynthesis of MGLPs was confirmed when Kaur *et al.* investigated the effects of disrupting the relevant gene (*MSMEG_5084*) in *M. smegmatis*.[177] When the group analysed extracts from mutant bacteria lacking *MSMEG_5084* they found an 80% reduction in the amount of MGLP when compared to the wild type cell extracts.[177] The authors concluded that their results proved that GpgS in both *M. smegmatis* and *M. tuberculosis* is involved in the biosynthesis of MGLPs. Furthermore, the authors postulated that other proteins with weak GpgS activity may have accounted for low levels of MGLPs detected in the mutant strain.[177]

More recently, detailed kinetic studies have been performed on MtGpgS which led to further insights as to its mechanism.[176] The study involved the use of dead end kinetic studies to determine that 3PGA is the first substrate to bind during catalysis, followed by the sugar nucleotide donor.[176] In addition, the group used site directed mutagenesis to identify key residues involved in catalysis and found a lack of acid/base residues in

the active site. This led to the conclusion that MtGpgS may undergo an S_Ni type mechanism possibly accompanied by a retaining mechanism.[176]

6.1.1 GPG as a potential donor substrate for MtGBE

Much work has been undertaken to characterise MtGpgS and its role in MGLP biosynthesis. In this project, the aim is not to study GpgS itself but to use it to synthesise GPG.

Substrate investigations discussed in chapter 4 gave no conclusive results as to potential substrates for the MtGBE protein. Therefore, other carbohydrate monomers thought to be involved in the MGLP biosynthetic pathway were investigated as potential substrates. It was hypothesised that GPG may act as a substrate of MtGBE to produce diglucosyl glycerate (DGG). To investigate this, a plasmid containing the *Rv1208* gene was received from the laboratory of Prof. Marcelo E. Guerin at Ikerbasque, Blibao in order to carry out the enzymatic synthesis of GPG. This was performed with the intention of subsequently reacting GPG with MtGBE to see if any products were produced.

6.2 Expression studies of GpgS

6.2.1 Transformation of pET29a(+)::*Rv1208* plasmid into *E. coli* BL21 cells

Rv1208 was contained in a pET29a(+) vector containing a kanamycin resistance marker, T7 promotor region and the *lac* operon. The *Rv1208* gene was designed to also include a His_6-tag in the recombinant protein to allow for purification using Ni-NTA affinity column.

pET29a(+)::*Rv1208* was transformed into *E. coli* BL21(DE3) chemically competent cells which were taken through to a plasmid mini prep kit. The results of the mini-prep for each colony were analysed *via* agarose gel electrophoresis (see chapter 8, procedure 8.2.1.2).

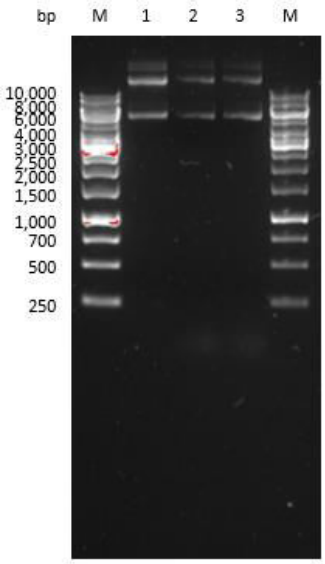


Figure 6. 2 Agarose gel analysis of the products from mini-prep experiments on pET28a(+)::Rv1208. Lane 1: Stock sample received from Guerin lab; Lane 2: Mini-prep product from colony 1; Lane 3: Mini-prep product from colony 2; M: Marker.

The samples of purified plasmid were compared to the sample of pET29a(+)::*Rv1208* obtained from the Guerin lab. Agarose gel analysis (figure 6.2) showed that the purified plasmids were of the same size as the control sample and it was concluded that the original sample of pET29a(+)::*Rv1208* had been successfully amplified.

6.2.2 Expression of MtGpgS

Literature reports on the expression of MtGpgS describe the use of *E. coli* BL21(DE3) pLysS cells.[173] However, this particular strain of competent cells were unavailable in this project and so *E. coli* BL21(DE3) cells were used.

Overnight growths and protein expression were carried out in 2x YT media as per the literature.[173] 0.5 mM IPTG was used in induction and the protein was expressed at 37 °C for 4 hours. Cells were then pelleted and a small sample lysed and analysed *via* SDS-PAGE. Results are shown in figure 6.3.

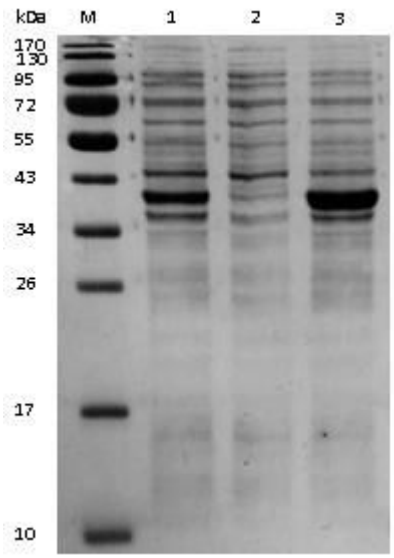


Figure 6. 3 SDS-PAGE analysis of expression of MtGpgS from E. coli BL21(DE3) cells. Lane 1: whole cell; Lane 2: lysate; Lane 3: insoluble; M: marker. MtGpgS = approx. 37 kDa.

MtGpgS is 37 kDa which corresponds to the thick band between 34-43 kDa in lanes 1 and 3 of the SDS-PAGE. This shows that MtGpgS has been overexpressed, though the majority of protein appears in inclusion bodies (IBs). As with MtGBE and MtGlcT, MtGpgS was then co-expressed using a system which promotes correct protein folding and the prevention of aggregation of the target protein. To determine the optimal chaperone system, both GroEL-GroES and the TF chaperone were tested in co-expression experiments in an attempt to express soluble MtGpgS.

6.3 Co-expression of MtGpgS with various chaperone proteins

6.3.1 Co-transformation of the pET29a(+)::*Rv1208* plasmid with pGro7 and pTf16 plasmids

Both the GroEL-GroES chaperone system and TF chaperone protein were tested for the ability to promote expression of soluble MtGpgS. In each case, co-transformation method 1 described in chapter 8, section 8.2.2.5 was employed to insert both plasmids into competent *E. coli* BL21(DE3) cells. Incubation overnight at 37 °C led to good colony growth on each plate, and thus the co-transformations were deemed successful.

To determine the best chaperone system for expression of soluble MtGpgS, the colonies grown in the co-transformations were used in small scale (15 mL) expression

experiments. For induction, 1.0 mM IPTG was added and protein was expressed at either 37 °C for 4 hours or 18 °C for 18 hours. Cells were lysed *via* sonication before being analysed by SDS-PAGE. The results are shown in figure 6.4.

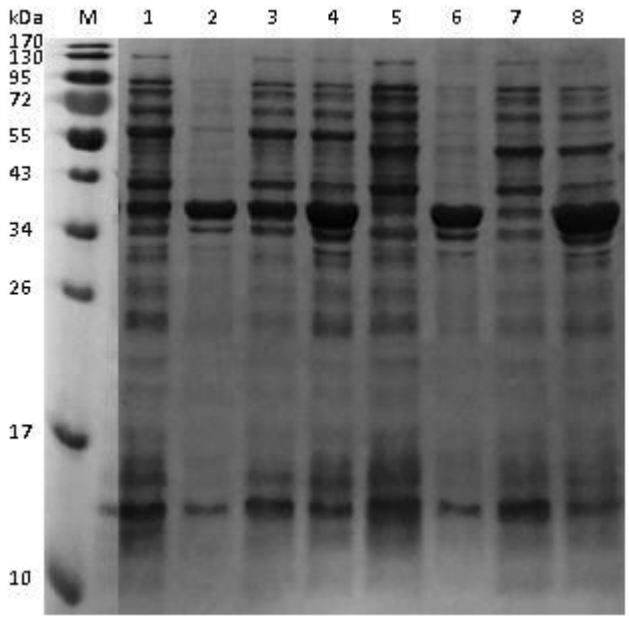


Figure 6. 4 SDS-PAGE analysis of various expression conditions for MtGpgS. Lanes 1-4: Co-expression with GroEL-GroES chaperone system; Lanes 1 + 2: 37 °C, 4 h expression, lysate, insoluble; Lanes 3 + 4: 18 °C, 18 h expression, lysate, insoluble. Lanes 5-8: Co-expression with TF chaperone protein; Lanes 5 + 6: 37 °C, 4 h expression, lysate, insoluble; Lanes 7 + 8: 18 °C, 18 h expression, lysate, insoluble; M: marker. MtGpgS = approx. 37 kDa.

SDS-PAGE analysis in figure 6.4 shows that the GroEL-GroES chaperone system was the most successful at producing soluble MtGpgS. This is indicated by the thick band produced between 34 and 43 kDa in the lane showing the soluble fraction for GroEL-GroES co-expression performed at 18 °C overnight (lane 3). In both cases, the chaperone proteins are different enough in mass so as not to be confused with the target protein, therefore the thick band of over-expressed protein is not due to the presence of either GroEL or TF which are 57 kDa and 48 kDa respectively. The respective chaperone bands can be seen separately in the gel, above the band for MtGpgS (lanes 1-4 for GroEL and lanes 5-8 for TF).

The expression studies led to the conclusion that co-expression with the GroEL-GroES chaperones at 18 °C for 18 hours was the most effective route to soluble MtGpgS

protein, therefore, a larger scale co-expression (1.0 L) was carried out in order to then purify enough protein for use in enzymatic reactions.

6.3.2 Co-expression and purification of MtGpgS with GroEL-GroES chaperone system

The optimised conditions from the experiments described in section 6.3.1 were applied to a 1.0 L scale growth and protein expression. The resulting lysate was then purified using Ni-NTA affinity column on an ÄKTA-prime purification system. MtGpgS protein was then eluted using a gradient of 0 – 0.25 M imidazole elution buffer over 18 column volumes and the eluate analysed by SDS-PAGE (see figure 6.5). During elution, two peaks were observed in the UV trace, suggesting non-specific binding of impurities to the resin. Fractions 10 and 11 were attributed to the first, smaller peak and fractions 13 and 14 to the second, much larger peak, suspected of containing the target protein.

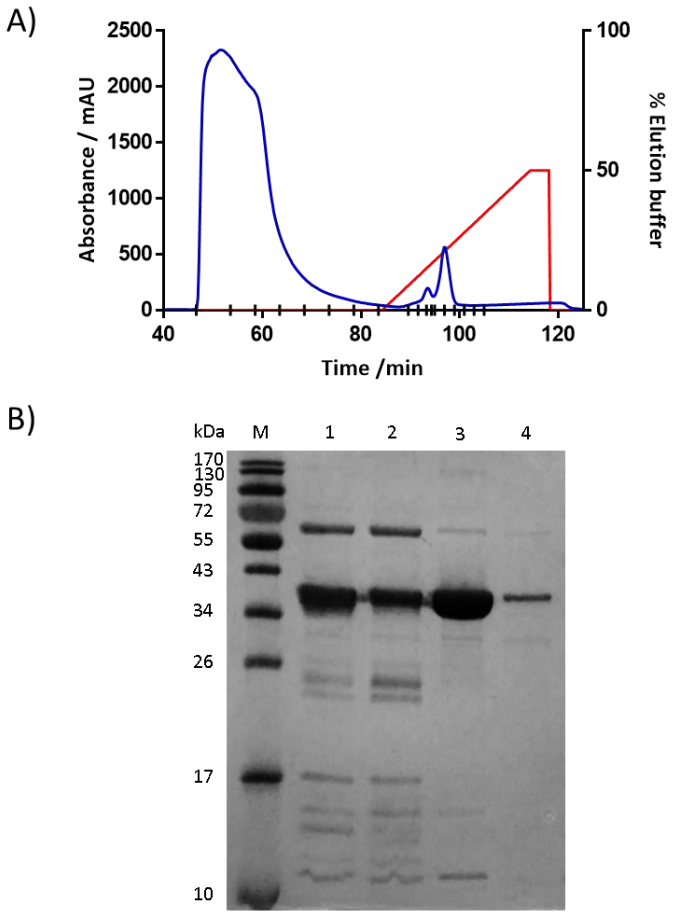


Figure 6. 5 A) UV trace (blue), % elution buffer (red) and fractions (black) from the Ni-NTA purification of co-expressed MtGpgS. B) SDS-PAGE analysis of the two peaks eluted from the resin. Lane 1: Fraction 10; Lane 2: Fraction 11; Lane 3: Fraction 13; Lane 4: Fraction 14; M: marker. MtGpgS = approx. 37 kDa.

SDS-PAGE analysis (figure 6.5, A) shows that MtGpgS was present in both eluted peaks, though in the initial, smaller peak, there were more impurities present, particularly a protein between 55 and 72 kDa in mass. This may correspond to the GroEL chaperone which is 57 kDa in mass, though its elution from the resin would be unusual given it does not contain a His-tag and experiments designed to test if GroEL was being eluted during FPLC (see chapter 3, section 3.2.2.4) suggested that GroEL was not co-eluted with target proteins. The larger peak contained higher purity MtGpgS protein and the relevant fractions were pooled and dialysed against 10 mM Tris-HCl containing 0.5 M NaCl.

Protein precipitation occurred during dialysis, though Bradford analysis suggested 0.2 mg/mL protein remained in solution. The supernatant was therefore concentrated to a final volume of 600 µL with a concentration of 5.0 mg/mL.

The concentrated protein solution was stored at -80 °C for future use in assays and enzymatically catalysed reactions.

6.4 Enzymatic synthesis of glucosyl-3-phosphoglycerate with MtGpgS

In the currently accepted order of reactions in MGLP biosynthesis (depicted in figure 6.6), UDP-Glc (**1.33**) and 3-PGA (**1.34**) are converted to glucosyl-3-phosphoglycerate (**1.35**) by GpgS. The phosphate group of GPG is then removed by glucosyl-3-phosphoglycerate phosphatase (GpgP) to produce glucosyl glycerate (**1.36**), which is believed to be subsequently converted to di-glucosyl glycerate (**1.37**) by MtGBE.[167, 171, 177]

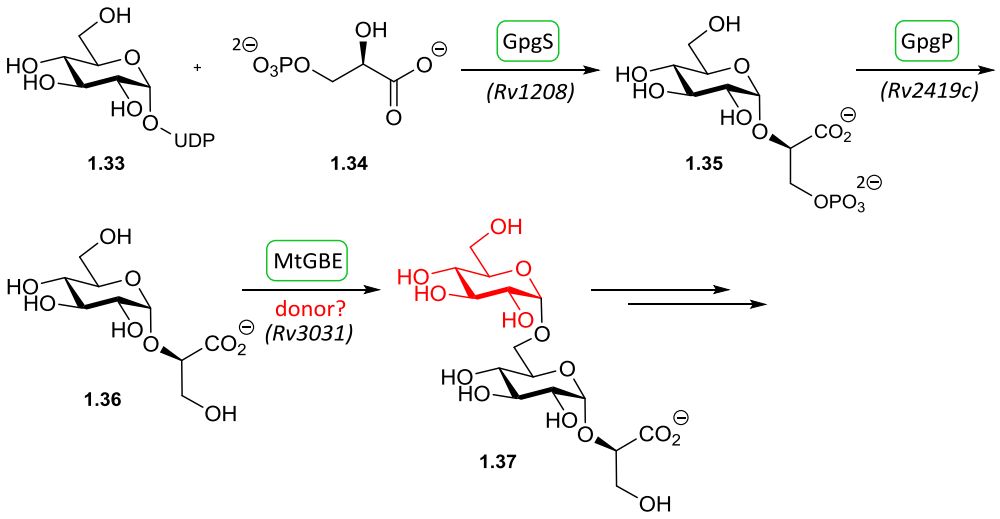


Figure 6. 6 Currently accepted order of reactions in the initial stage of MGLP biosynthesis. Figure adapted from Mendes et al.[167]

To investigate the possibility that **1.35** could act as a substrate for MtGBE and that dephosphorylation to yield **1.37** may occur later on, **1.35** was tested as a substrate for MtGBE. To produce **1.35**, the recombinant MtGpgS that had been expressed previously (see section 6.2.2) was used in a reaction with 3-PGA (**1.34**) and UDP-Glc (**1.33**).

6.4.1 PK/LDH coupled assay to track MtGpgS activity

To confirm that the recombinant MtGpgS was obtained in an active form and that the reaction was proceeding as expected, the PK/LDH coupled assay described in chapter 5 (section 5.3.1) was employed to monitor reaction progress. Reaction conditions were chosen based on reported data concerning the optimal reaction conditions for the GpgS proteins from *M. smegmatis* and *M. bovis*.[172] These data showed that the enzyme required the presence of $MgCl_2$ for any activity to be observed and that optimal concentrations were around 20 mM.[172] Therefore, the assay was carried out with the following components: UDP-Glc (1.0 mM), 3-PGA (1.0 mM) , PEP, NADH, PK, LDH and $MgCl_2$.The final concentration of MtGpgS used was 0.1 mg/mL. A more detailed assay composition can be found in chapter 8, section 8.2.5.4. The assay was carried out at 37 °C and the absorbance at 340 nm measured over time. The results are shown in figure 6.7.

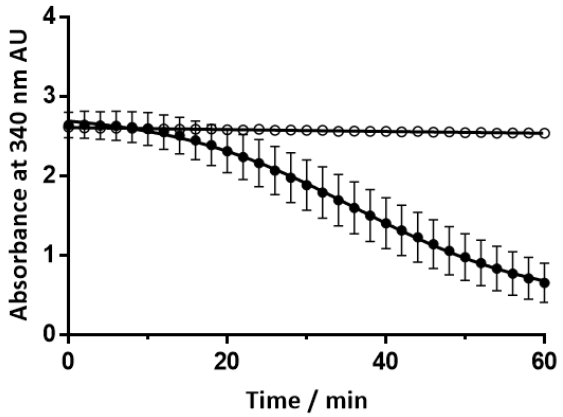


Figure 6. 7 PK/LDH coupled assay following the reaction between 3-PGA and UDP-Glc catalysed by MtGpgS (●). Control experiment contained no 3-PGA (○). Reaction carried out in triplicate. Lines of best fit generated using linear and non-linear regression tools in GraphPad Prism 7.0.

The results shown in figure 6.7 display a decrease in absorbance at 340 nm, suggesting the hydrolysis of UDP-Glc over time. The control experiment was performed in the absence of 3-PGA, and, due to the flat line showing very little decrease in absorbance (gradient = -0.001198 ± 0.000037), it appears that MtGpgS is not hydrolysing UDP-Glc in the absence of the acceptor. This also shows that degradation of UDP-Glc is not occurring spontaneously. The results show that the formation of GPG has not

completed before 1 hour. This is in agreement with the literature reports which say that optimal activity is seen at a pH level nearer pH 8 and at a temperature of 45 °C.[172] In this case, however, the results confirm that the MtGpgS protein expressed earlier in the project was active and able to produce GPG.

6.4.2 Scale up of MtGpgS catalysed reaction to produce GPG

The coupled assay confirmed that the reaction was successfully occurring and therefore the synthesis was performed on a larger scale with the aim to isolate GPG. The reaction was repeated with minor modifications to the earlier method. The concentrations of UDP-Glc and 3-PGA were increased to 10 mM and the components required for the coupled assay (PK, LDH, NADH and PEP) were not present. $MgCl_2$ and MtGpgS were used in the same concentrations as the PK/LDH assay described previously. HEPES pH 8 (20 mM) was used as the reaction buffer and total reaction volume was increased to 500 µL. The reaction components were mixed and incubated at 37 °C for 18 hours. The reaction products were then visualised using TLC. The reaction mixture was compared to the two substrates as well as to a control reaction in which no 3-PGA was present. The results of the TLC experiments are shown in figure 6.8.

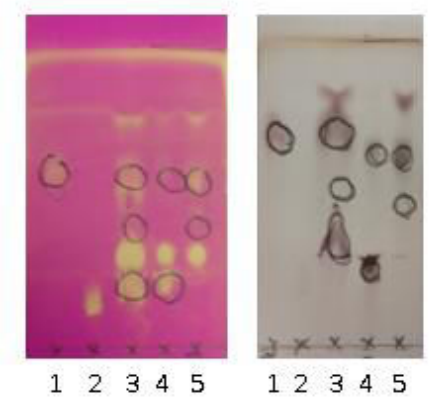


Figure 6. 8 TLC analysis of reaction between 3-PGA and UDP-Glc in the presence of MtGpgS. Lane 1: UDP-Glc standard; Lane 2: 3-PGA standard; Lane 3: mixed spot; Lane 4: reaction products: Lane 5: control (no 3-PGA present) reaction products. Eluent used: ACN/EtOAc/^{i}PA: H_2O in a ratio of 85:20:50:50. Plates visualised with UV irradiation (spots circled in pencil) and either $KMnO_4$ stain (left) or 10 % H_2SO_4 in MeOH stain (right).

The TLC results show that there are still small amounts of both substrates present in the reaction mixture as there are spots that correspond to both 3-PGA and UDP-Glc in lane 4 of each plate. However, the use of both stains shows that 3-PGA is not detected by either UV irradiation or by the acid stain. Therefore, the lower spot that occurs in lane 4 of the acid stained TLC plate is likely to be a new reaction product, different from the substrates. This is also evidenced by the fact that this spot does not appear in the control reaction in lane 5. This spot may either be due to the GPG that is produced in the reaction, or could also be due to the UDP which is released. Without a standard of UDP with which to compare to, it is unclear. However, as the coupled assay showed the reaction was progressing and the fact that this new spot is visualised by the acid stain which is typically used to identify sugars, it was concluded that this spot corresponded to GPG.

The next step in obtaining GPG would be to isolate it from the reaction mixture and subsequent analysis to confirm the structure of the desired compound. GPG has been separated and characterised from reaction mixtures as reported in the literature *via* ion exchange chromatography and NMR spectroscopy.[316] Due to time constraints, this was not possible as part of this work. However, the production and purification of GPG was attempted as part of an undergraduate project carried out by James Pearce, though ion exchange chromatography proved unsuccessful.

6.5 Conclusions

The aim of the work described in this chapter was to obtain GPG to test as a potential substrate of MtGBE. To achieve this, recombinant MtGpgS was to be expressed and purified for use in the enzymatic conversion of 3-PGA and UDP-Glc to GPG.

The aim was partly achieved as MtGpgS protein was successfully expressed and purified, despite initial difficulties in producing soluble MtGpgS. Following experiments involving co-expression with either the GroEL-GroES system or the TF chaperone protein, it was determined that co-expression with the GroEL-GroES chaperone proteins was the most efficient way of obtaining soluble protein. Purification using Ni-NTA FPLC led to protein of acceptable purity, sufficient for use in the planned enzymatic reaction.

The recombinant MtGpgS displayed activity in the PK/LDH coupled assay with UDP-Glc as a donor and 3-PGA as an acceptor. This activity was observed as a decrease in absorbance at 340 nm. Though the reaction progress was slow compared to literature reports, this was expected as not all conditions of the assay met the optimal conditions reported in the literature.[172] TLC analysis of the scaled up enzymatic reaction showed the formation of a new product, GPG, as new spots were identified using the acid based TLC stain.

The aim was not fully met as a lack of time meant that GPG could not be tested as a substrate for MtGBE. It was intended that the supernatant from the reaction producing GPG could be used directly in an experiment with MtGBE after the removal of MtGpgS. It was also planned that GPG be characterised to confirm that is was indeed successfully synthesised enzymatically. Purification would have been carried out by ion exchange chromatography and characterisation by NMR spectroscopy and or mass spectrometry.

7. Conclusions and future work

7.1 Conclusions on the chemical synthesis of various carbohydrate derivatives for use in probing MtGBE activity

The target compound deemed highest priority was glucosyl glycerate (**1.36**). This compound is the precursor to diglucosyl glycerate (**1.37**), the formation of which is believed to be catalysed by MtGBE. As part of this project the full chemical synthesis of GG was achieved over 8 steps, as determined by NMR and MS characterisation. Though the synthetic route to GG was not novel, adjustments were made compared to the procedure reported in the literature. The synthetic steps to the thioglucoside donor were successfully carried out as per the literature reports.[186] The glycosylation step during which the thioglucoside donor is coupled to the glycerate acceptor was carried out under the same conditions as the literature, however the purification did not separate anomers in this case. The synthesis was continued and it was found that after removal of TBDPS protecting group, the α anomer could be separated from the β. The deprotection steps involving the removal of the 6-O-acetyl group and the TBDPS groups were successfully completed. The removal of the benzyl ether protecting groups proved more difficult. The reported conditions of palladium on carbon under 35 psi in a solvent system of ethyl acetate and ethanol were attempted in this project, however with no success.[186] Once an adjustment was made to the setup of the reaction, the benzyl groups were sucessfully removed. The final step to produce GG was to remove the methyl ester which was achieved using LiOH. Neutralisation of the reaction with acidic resin avoided the need for removal of salt impurities *via* ion exchange chromatography.

The glycerate acceptor was successfully synthesised from D-mannitol over a five step synthesis. The only issue was the formation of the methyl ester from the protected glyceraldehyde. Initially, oxidation of the aldehyde to an acid was attempted, though this proved unsuccessful. A one-pot synthesis of the methyl ester using bromine and methanol reported by Ladame *et al.* was repeated successfully, yielding the desired methyl ester.[197] Following the deprotection of the acetal with acidic resin, selective

protection of the primary hydroxyl group with a TBDPS ether was straightforward and produced the acceptor compound.

The next compound of focus was maltosyl glycerate (MalG). The synthesis of this compound had not been reported and so a novel synthesis was developed. Initial experiments involved global protection of the hydroxyl groups with benzyl ethers, however once the selective removal of the anomeric benzyl ether proved inefficient, the strategy was changed to start with global acetylation instead. The per-*O*-acetylated maltose was then easily coupled with *p*-thiocresol to form a thiomaltoside donor. Before the glycosylation reaction was attempted, the acetyl groups were removed and replaced with benzyl ether groups in order to produce a 'ring-activated' donor and to remove the possibility of any NGP from the acetyl group protecting the 2-hydroxy group. The benzyl protected donor was then successfully coupled with the glycerate acceptor using the same conditions as those used in the GG synthesis. Following glycosylation, separation of the anomers proved difficult. Though some anomerically (α) pure product was obtained after column chromatography, the majority of the sample remained mixed. Given the time constraints of the project, the mix of anomers was taken through the rest of the synthesis, as well as the smaller amount of anomerically pure material. The remaining steps to MalG were then deprotection of the TBDPS and benzyl ether groups. Whilst the former was readily achieved, though with lower yield than expected, the removal of the benzyl ethers required optimisation. It was found that increasing the amount of catalyst to 1.5 molar equivalents facilitated the removal of all 7 benzyl ethers, allowing for the final step of methyl ester deprotection to give the target compound. Full characterisation of MalG was achieved using NMR spectroscopy as well as HRMS.

In addition to the syntheses of GG and MalG, two fluorinated compounds were also prepared. The first, 6FpNPαGlc was successfully prepared in one step by Laura Jowett as part of an undergraduate degree project. pNPαGlc was reacted with DAST which led to the successful, selective fluorination of the 6 position to give 6FpNPαGlc with a yield of 57 %. The product was stable enough for characterisation by ^{1}H, ^{13}C and ^{19}F NMR and MS.

Another fluorinated compound successfully synthesised was 5FβIdoF. The original plan was to produce 5FαGlcF, however, due to low yields and low amounts of material being produced, the synthesis was stopped at the 5F idosyl fluoride stage. This was considered useful as the Ido compound might act as an inhibitor for crystallisation of MtGBE, work which was being carried out by collaborators in the Kim group at SKKU, South Korea. The synthesis of 5FIdoF was successfully achieved over four steps, starting with the treatment of per-*O*-acetyl glucose with HF-pyridine. This reaction led to the production of the α-glucosyl fluoride in a 56 % yield. The bromination of position 5 was attempted using light irradiation to heat a solution of the starting material in CCl_4 in the presence of NBS to reflux. Despite literature precedence showing the success of this reaction[234, 235], the highest yield obtained in this project was 31%. The following step was also low yielding at 45%. This involved the replacement of the newly installed bromine with fluorine through the employment of AgF. It was also at this point at which the sugar took the idosyl conformation, as confirmed by the coupling constants in the ^{1}H NMR matching those of the reported idosyl structure.[235] The final step was the removal of the acetate protecting groups using a methanolic solution of ammonia which led to the final compound with a yield of 57%. Overall the target compound was successfully synthesised and characterised using NMR spectroscopy.

7.1.1 Future work proposed for the chemical synthesis of various carbohydrate derivatives for use in probing MtGBE activity

In the synthesis of GG, the 6-OAc protecting group was suspected of being involved in promoting formation of the α anomer by electronically shielding the top face of the sugar donor.[194] Future work could be carried out to probe the extent of this effect, and the potential influence on the stereochemical outcome of the reaction. To investigate this, future work could include the synthesis of various donors with different protecting groups at the 6 position, including one with an acetyl group. The use of these in glycosylation reactions with the glycerate acceptor and the subsequent determination of the α:β ratio of the reaction products may provide some insight as to the effect of the protecting group at the 6 position. Additionally, reactions with different acceptors may also prove interesting.

In the synthesis of MalG, further work would be required in order to fully separate the anomers in the mix of reaction products. Column chromatography did not prove successful enough in isolating the α anomer and so other separation techniques may prove more helpful. For example, HPLC.

If the synthesis of fluorinated compounds as part of this project were further investigated, the conversion of the 5-fluoro-β-L-idopyranosyl fluoride (**2.38**) to 5-fluoro-α-D-glucopyranosyl fluoride (**2.31**) as per the original aims would be carried out. This has been described in the literature by using HF-pyridine in an overnight reaction or through the use of BF_3, though reported yields were low.[234, 235]

In addition to the already synthesised compounds, other compounds may act as potential inhibitors. For example, it is generally believed that MtGBE acts upon GG and therefore, an inactive form of this compound may act as an inhibitor. 6-fluoro GG may behave as one such inhibitor due to the lack of a 6-OH and could possibly be produced using DAST to yield the 6F glucoside. This could then be converted to a suitable donor for glycosylation with the glycerate acceptor. Similarly, 2-deoxy-2,2-difluoro-*p*-nitrophenyl-α-D-glucopyranoside (2dF-pNPGlc) is proposed as another general inhibitor that could possibly be used in crystallographic studies.

7.2 Conclusions on the expression, purification and characterisation of MtGBE

Whilst the pET28a::*Rv3031* plasmid was successfully transformed and amplified into chemically competent cells, the over-expression of MtGBE led to the insoluble form of the protein. Therefore, two strategies were used to try and obtain soluble MtGBE. The first was solubilisation from IBs using sarkosyl, followed by re-folding of the protein facilitated by removal of the surfactant during dialysis. A subsequent purification attempt proved unsuccessful, suggesting that the protein had not refolded in a correct manner. This may have resulted in the His_6-tag not being available for binding to the nickel resin and thus not eluting in a pure fraction. Co-expression with the GroEL-GroES chaperone system was more successful and led to the production of soluble MtGBE. Although there was always still protein present in the insoluble fractions, increasing the amount of L-Ara in the culture media led to increased amounts of soluble protein.

This led to the purification of MtGBE by Ni-NTA FPLC. Subsequent dialysis and concentration led to protein solution pure enough for use in enzymatic assays.

Due to the lack of a His_6-tag, the chaperone proteins used in the co-expression of MtGBE would be removed from solution following FPLC, however concern was raised at the potential of chaperone proteins being present in solutions of MtGBE as GroEL was the same mass (57 kDa) and could not be separated on SDS-PAGE. However, experiments carried out in which MtGBE was expressed alone, co-expressed with chaperone proteins and the chaperone proteins expressed alone, confirmed that the chaperones were not being co-eluted off the nickel resin. MS further confirmed that the protein purified in MtGBE expression was not the same as any protein obtained after chaperones were expressed alone. The MS results were interesting as the mass of MtGBE expected (59,979 Da, calculated using ExPASy peptide mass tool) was not the same as that found (57,195 Da). Though the observed mass differs from the calculated mass, this is only a 3.3% discrepancy.

Initial assays carried out with MtGBE were designed to test for potential hydrolysis activity on small substrates; behaviour that would be novel for a GH57 branching enzyme. The assays were based on using *p*-nitrophenyl glucose (pNPGlc) as a substrate and measuring the absorbance at 400 nm to observe any *p*-nitrophenolate that may be produced. Despite many changes in conditions and attempts at optimisation of the assay no quantifiable hydrolysis activity was observed. This result is not surprising as GH57 branching enzymes do not usually display this kind of activity on such substrates and this behaviour would have been novel to this class of enzymes.

As no hydrolysis activity was observed, the next set of assays investigated the presence of branching activity from MtGBE. A BLAST search showed that MtGBE had the highest homology with GH57 branching enzymes, therefore this behaviour was investigated. Literature reports of iodine staining assays and branching assays involving subsequent debranching with isoamylase were found[180, 181] and attempts were made to repeat them with MtGBE. Substrates tested included amylose, amylopectin, glycogen and maltodextrin though no significant levels of activity were observed. It did appear that MtGBE showed very low activity with amylopectin, however, it was not possible to gain kinetic data with this substrate. The main difference between amylopectin and

amylose is that amylopectin is heavily branched whereas amylose is not. If amylopectin is a substrate of MtGBE, the presence of existing branch points may be significant.

In addition to the branching assays discussed, TLC of reaction products with various glycans and MtGBE was also carried out. Polysaccharides tested were those already used in the branching assays as well as pullulan. None of these compounds acted as substrates of MtGBE. Shorter di- or oligo- saccharides were also tested; maltose, maltotriose, maltotetraose, maltoheptaose and melibiose. Of these, only maltoheptaose could conclusively be ruled out as a potential donor. The other potential donors tested led to inconclusive results as the TLC spots were not clear. Reaction of MtGBE with MalG was also analysed by TLC but no activity was apparent.

Finally, with no conclusive evidence of any of the substances tested showing activity with MtGBE, attention was turned to the possibility of other donors. Glucose-1-phosphate was proposed as a possible substrate and so malachite green assays were employed to test this theory. No activity in the release of inorganic phosphate was observed, leading to the conclusion that G1P was not a substrate for MtGBE.

7.2.1 Future work proposed for the expression, purification and characterisation of MtGBE

With regards to the expression of MtGBE, co-expression with the GroEL-GroES chaperone system was successful. However, the purity of the enzyme, though deemed acceptable for assays could have been improved upon. Further purification by techniques such as size exclusion chromatography, would perhaps achieve this. Additionally, due to the MS results appearing to show the protein as lighter than expected, amino acid sequencing could be used to investigate whether or not the protein had been degraded during expression. Sequencing results suggest this is not due to an error at DNA level and so there may be a problem with the protein sequence itself. If that were to still suggest the protein were in-tact, alternative MS methods could be investigated, such as MALDI should there be any issues caused by the electrospray ionisation technique.

Future work with enzymatic assays could involve more quantitative analysis of substrates tested by TLC in this project. For example, HPLC could provide more

conclusive results by monitoring more accurately the consumption of substrates or production of products. Additionally, it may be possible that a sugar nucleotide is used as a donor to transfer a glucose unit to GG. This would be unusual activity for a branching enzyme, but given the widespread use of these donors throughout nature, it may be a possibility, especially as this is the substrate for the glucosyl glycerate synthase protein found in *P.marina*.[318]

7.3 Conclusions on the expression and purification of MtGlcT and MtGpgS

7.3.1 MtGlcT

In an attempt to investigate the hypothesis that MtGlcT could work in conjunction with MtGBE in the biosynthesis of MGLPs, the protein was expressed, purified and tested in a PK/LDH coupled assay.

The expression of MtGlcT led to the formation of IBs, making the protein inaccessible. Attempts at co-expression with the GroEL-GroES chaperone system and the TF chaperone initially indicated that the TF chaperone led to the production of soluble MtGlcT. Upon purification, however, no protein was eluted from the Ni-NTA column. This may have been due to a number of reasons. The protein may not have been stable or may not have folded correctly leading to insufficient binding to the Ni resin. Further attempts at isolating soluble protein from a large scale (8.0 L) protein expression were also unsuccessful. After FPLC purification, precipitation of the protein meant concentration was not possible.

As co-expression did not lead to soluble protein, re-solubilisation of MtGlcT from IBs was investigated. Solubilisation and refolding of MtGlcT was performed by Dr Gu Yoo using 0.2% sarkosyl. In this project, the refolded MtGlcT was concentrated and used in the PK/LDH assay.

The coupled PK/LDH assays relied on the degradation of NADH, the decrease of which was monitored by measuring the absorbance at 340 nm. MtGlcT was tested with a number of potential acceptors, including short chain α-glucans , glucose, GG and various polysaccharides; none showed any activity. When tested without metal ions present, an increase in absorbance was shown which contrasted to the result expected for GT activity. When metal ions were added to the reaction mixture, however, the

increase in absorbance no longer appeared and the reaction trends were the same as those for the negative controls. Though an acceptor was not identified, it would appear that metal ions are important to the stability of the MtGlcT protein. The negative results may also have been due to incorrect re-folding of the protein during the removal of the sarkosyl.

To verify that the PK/LDH coupled assay was set up correctly and was working as expected, a positive control was carried out with OtsA. Soluble OtsA protein was successfully expressed and purified for use in the coupled assay as per literature reports.[308] When tested, OtsA led to a decrease in absorbance at 340 nm when reacted with UDP-Glc and G6P, thus confirming the assay used worked for GT activity.

7.3.2 MtGpgS

Investigations into the possibility of GPG, a carbohydrate present in the initial stages of MGLP biosynthesis, being a potential substrate of MtGBE required the expression and purification of MtGpgS.

Initial attempts to express MtGpgS led to the formation of IBs, thus rendering the protein insoluble and not useful for enzymatic reactions. Expression studies were carried out to explore the effects of co-expression with chaperone proteins. Both the GroEL-GroES system and the TF chaperone protein were trialled. Co-expression with GroEL and GroES at 18 °C overnight led to the production of soluble protein. This method of expression followed by Ni-NTA purification led to soluble MtGpgS, suitable for use in the enzymatic synthesis of GPG.

Before GPG was synthesised, MtGpgS was used in the PK/LDH assay to confirm the protein was obtained in an active form. The assay results illustrated a decrease in absorbance at 340 nm with the control experiment showing no significant decrease in comparison. These results confirmed that MtGpgS was active upon its substrates, UDP-Glc and 3-PGA.

Once activity was confirmed, MtGpgS was used in the enzymatic synthesis of GPG. Analysis by TLC suggested that a new product was being formed, believed to be GPG. Isolation and subsequent characterisation of the reaction product was not possible due

to time constraints, nor was the reaction between the crude reaction mix and MtGBE able to be performed.

7.3.3 Future work proposed for the expression and purification of MtGlcT and MtGpgS

7.3.3.1 MtGlcT

The MtGlcT protein used in the experiments investigating potential acceptor substrates was obtained after re-folding following solubilisation from IBs. The re-folding of MtGlcT may not have been successful, which could account for lack of activity seen in the PK/LDH assay. To test this, experiments investigating the tertiary structure of the MtGlcT protein obtained in this work could be performed. For example, CD spectrometry (discussed in chapter 3) could provide insights into the folding state of the protein.

Prior to re-solubilisation, an 8.0 L protein expression and purification was performed, the results of which were hampered by protein precipitation following dialysis. On reflection, the dialysis buffer used was simple (50 mM HEPES pH 7.4) and future attempts could involve more complex buffers, more suited to stabilising the protein. For example, a buffer containing $MgCl_2$ could be used as the presence of this salt appeared to stabilise the protein in the coupled assays.

If MtGlcT were isolated in its native state, assays involving short chain α-glucans could be repeated. To investigate any involvement in the synthesis of DGG, the reaction between MtGlcT and GG could be repeated.

7.3.3.2 MtGpgS

Enzymatic synthesis of GPG was performed in this project but the isolation and analysis of the product was not. Future work, therefore, would include the purification of GPG from the reaction mixture and literature reports describe achieving this *via* ion exchange chromatography.[316] Isolated GPG could then be analysed using NMR spectroscopy and mass spectrometry.

If purification of GPG did not prove possible, the crude reaction mixture could be added to MtGBE and monitored by techniques such as TLC or HPLC so as to investigate the potential of GPG as a donor for MtGBE, as originally planned.

7.4 Overall concluding remarks

The main aim of this project was to identify the activity and donor substrate of MtGBE, however, despite the successful expression and purification of the protein, this was not achieved. This raises the question of why this main objective was not met and what could be done in future to further investigations into MtGBE activity.

Expression and purification of MtGBE was hindered by the formation of IBs rendering the protein inaccessible. To counter this issue, MtGBE was co-expressed with the GroEL-GoES chaperone system to promote correct protein folding of MtGBE.[251, 253, 319] This chaperone system was also used in the co-expression of MtGpgS and the purified protein displayed activity in the PK/LDH coupled assay, suggesting the chaperones had the desired effect and promoted correct protein folding. It therefore seems unlikely that the co-expression of MtGBE with GroEL and GroES produced an inactive form of the protein, though it could still be a possibility. CD experiments of MtGBE suggested that the protein did fold correctly though a denatured sample of MtGBE for CD analysis would confirm this. Additionally, protein NMR experiments may also be able to provide insights into the 3-D conformation of the protein and help assess if it is folded correctly or not.

Sequencing of the plasmid isolated from the bacterial cells used in expression confirmed no alterations to the DNA sequence of the original plasmid, suggesting no alterations to the protein were introduced at the DNA level. Mass spectrometry of the protein, however, led to a mass for MtGBE different to the predicted mass. Though this discrepancy was small, amino acid sequencing experiments could be performed to confirm no major alterations to the primary structure have occurred and are affecting protein activity.

The crystal structure of the enzyme, if successfully obtained, could provide important insights into the function of MtGBE. Comparison to the known structures of GH57 proteins[181, 182, 264, 266, 274] may confirm whether or not MtGBE acts on similar substrates – i.e. long chain polysaccharides or on shorter glucan chains. In addition, any metal ions that crystallise with the proteins could identify essential co-factors. In this work,

only potassium ions were investigated as a potential co-factor. It could be that divalent cations such as Mg^{2+} or Mn^{2+} are essential for activity or structural stability.

As well as essential co-factors, it may be that the donor for MtGBE was not tested in this project and therefore not identified. UDP-Glc is utilised as a substrate in the synthesis of GPG, the precursor to GG[174-176] and it may also be a substrate of GG, though this possibility is yet to be explored.

It should also be highlighted that the involvement of MtGBE in MGLP biosynthesis is so far speculative and no conclusive evidence has been reported to confirm this hypothesis.[167, 171, 184, 314] The *Rv3031* gene belongs to the same gene cluster as *Rv3032*, the product of which has been shown to directly affect MGLP production.[90] It is purely the location of *Rv3031* on this gene cluster that has led to the proposal of its involvement in MGLP synthesis. If *Rv3031* deficient mutant strains of *M. tb* could be grown, or if *Rv3031* were overexpressed in *M. tb* (as *Rv3032* has been[90]), the effect, or lack of effect, on MGLP production could be observed. If MtGBE is involved in the MGLP pathway, the accumulation of precursor compounds could help confirm the substrates of this protein.

Overall, though this project did not identify the substrates and mechanism of MtGBE, investigations have shown the need for other hypotheses to be explored regarding this enzyme.

8. Materials and Methods

8.1 Chemical synthesis of potential MtGBE substrates and inhibitors

All reagents were commercially sourced and used as received from the suppliers (Sigma-Aldrich®, Fisher Scientific, Acros Organics, VWR® and Carbosynth).

Anhydrous solvents were obtained commercially from Sigma-Aldrich.

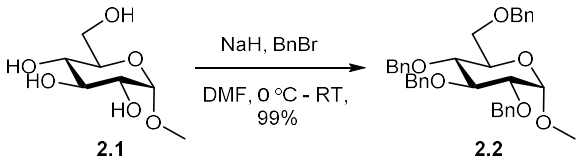
Reactions were monitored by thin layer chromatography using aluminium foil supported silica plates (MERCK 60 F254) and run in the solvents stated. Plates were developed using either a 5% H_2SO_4 in MeOH solution, 5% H_2SO_4 in MeOH solution with 0.3% *N*-(1-naphthyl)ethylenediamine $KMnO_4$ solution or under UV light.

Chromatography columns were prepared using Merck Millipore Geduran® Si 60 (40-63 µm) silica gel.

Nuclear magnetic resonance spectra were recorded using a Bruker AVIIIHD400 FT-NMR Spectrometer with COSY, HSQC, HMBC, PSYCHE, DEPT 45 and DEPT 135 used to assign spectra.

Mass spectra were recorded using a WATERS ZMD single quadrapole system and Bruker Apex III FT-ICR-MS spectrometers.

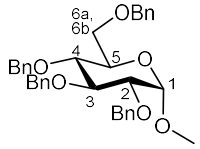
8.1.1 Methyl-2,3,4,6-tetra-*O*-benzyl-α-D-glucopyranoside (2.2)



2.1 2.2

Methyl-α-D-glucopyranoside (488 mg, 2.5 mmol) was dissolved in DMF (25 mL) and the resulting solution cooled to 0 °C. Sodium hydride (60% in paraffin, 520 mg, 13.0 mmol) was then added and the reaction mixture stirred for 30 minutes. Benzyl bromide (1.5 mL, 13.0 mmol) was then added and the reaction mixture stirred at room temperature overnight. Methanol (20 mL) was added to quench the reaction and the volume of the reaction mixture was reduced *in vacuo*. The solution was then washed with water (10 x

200 mL) and the organic layers combined and dried over $MgSO_4$ before being concentrated under reduced pressure to give the product in the form of a pale yellow oil (1.37 g, 99%). NMR and mass spectrometry determined that the product did not require further purification.



^{1}H NMR (400 MHz, $CDCl_3$): δ 7.4-7.3 (18H, m, H_{Ar}), 7.2-7.1 (2H,m, H_{Ar}), 5.0 (1H, dd, *J* = 11.0 Hz, ArC**H**$_2$),4.9-4.7 (1H, m, ArC**H**$_2$), 4.7-4.6 (1H, m, ArC**H**$_2$), 4.6 (1H, d, *J* = 3.6 Hz, H1), 4.6 (1H, m, ArC**H**$_2$), 4.5 (2H, m, ArC**H**$_2$), 4.0 (1H, t, *J* = 9.2 Hz, H3), 3.8-3.6 (4H, m, H4, H5, H6a, H6b), 3.6 (1H, dd, *J* = 3.5, 9.6 Hz, H2), 3.4 (3H, s, C**H**$_3$) ppm;

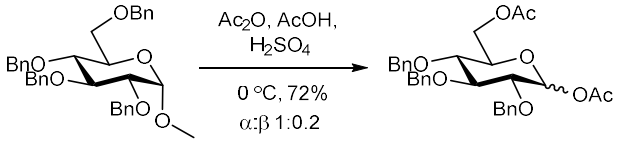
^{13}C NMR (101 MHz, $CDCl_3$): δ 138.8 (C_{Ar}), 138.3 (C_{Ar}), 138.2 (C_{Ar}), 138.0 (C_{Ar}), 128.4-127.6 (C_{Ar}), 98.2 (C1), 82.2 (C3), 79.9 (C2)l 77.7 (C4), 75.8 (Ar**C**H$_2$), 75.0 (Ar**C**H$_2$), 73.5 (Ar**C**H$_2$), 73.4 (Ar**C**H$_2$), 70.1 (C5), 68.5 (C6), 55.2 (**C**H$_3$) ppm;

LC-MS (ESI+) m/z: 577.4 $[M + Na]^+$, 593.6 $[M + K]^+$.

Data matches that of literature.[320]

8.1.2 1,6-di-*O*-acetyl-2,3,4-tri-*O*-benzyl-α/β-D-glucopyranoside (2.3α/β)

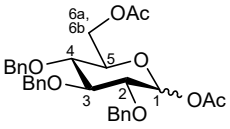
(Work carried out by Laura Jowett).



2.2 **2.3α** **2.3β**

Methyl-2,3,4,6-tetra-*O*-benzyl-α-D-glucopyranoside (500 mg, 0.90 mmol) was dissolved in a mix of acetic acid (5.0 mL) and acetic anhydride (5.0 mL) and the resulting solution cooled to 0 °C. Sulphuric acid (conc. 0.1 mL) was then added in a dropwise manner and the reaction stirred for 6 hours before the addition of saturated sodium bicarbonate solution (18 mL). The resulting solution was then washed with ethyl acetate (3 x 30 mL) and the organic layers were combined and dried over $MgSO_4$. After removal of the

solvent under reduced pressure, the product was purified via column chromatography (20% hexane in ethyl acetate) and a colourless oil was obtained (346 mg, 72%).



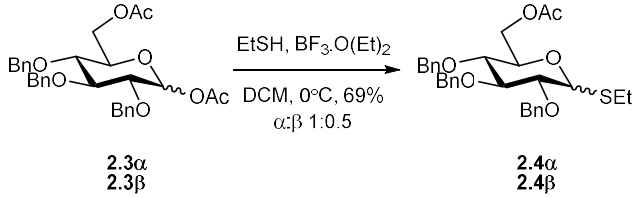
^{1}H NMR (400 MHz, $CDCl_3$): (α:β 1:0.2) δ 7.4-7.3 (17H, m, H_{Ar}α+β), 6.3 (1H, d, *J* = 3.6 Hz, H1 α), 5.6 (0.2H, d, *J* = 8.1 Hz, H1 β), 5.0 (1H, d, *J* = 10.8 Hz, ArC**H**$_2$α), 4.9 (0.2H, d, 10.8 Hz, ArC**H**$_2$β), 4.9 (0.2H, d, 10.9 Hz, ArC**H**$_2$β), 4.9 (1H, d, *J* = 10.8 Hz, ArC**H**$_2$α), 4.9 (1H, d, *J* = 10.8 Hz, ArC**H**$_2$α), 4.8 (0.2H, d, *J* = 11.4 Hz, ArC**H**$_2$β), 4.8 (0.2H, d, *J* = 11.4 Hz, ArC**H**$_2$β), 4.7 (1H, d, *J* = 11.5 Hz, ArC**H**$_2$α), 4.7 (1H, d, *J* = 11.4 Hz, ArC**H**$_2$α), 4.6 (1H, d, *J* = 10.6 Hz, ArC**H**$_2$α), 4.3 (1H, dd, *J* = 3.9, 12.1 Hz, H6a/6bα), 4.3, (1H, dd, *J* = 2.3, 12.2 Hz, H6a/6bα), 4.3, (0.4H, m, H6a+6bβ), 4.0 (2H, m, H3α, H5α), 3.8 (0.2H, t, *J* = 8.8 Hz, H3β), 3.7 (1H, dd, *J* = 3.6, 9.7 Hz, H2α), 3.7 (0.2H, m, H5β), 3.6 (0.2H, m, H2β), 2.2 (3H, s, C**H**$_3$α), 2.1 (0.6H, s, C**H**$_3$β), 2.0 (3H, s, C**H**$_3$α), 2.0 (0.6H, s, CH_3β) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 170.7 (Cq), 169.3 (Cq), 138.4 (C_{Ar}), 137.6 (C_{Ar}), 137.6 (C_{Ar}), 137.5 (C_{Ar}) 128.6 – 127.9 (C_{Ar}), 89.7 (C1), 81.6 (C3), 78.9 (C2), 76.6 (C4), 75.8 (Ar**C**H$_2$), 75.3 (Ar**C**H$_2$), 73.2 (Ar**C**H$_2$), 71.1 (C5), 62.7 (C6), 21.1 (**C**H$_3$), 20.8 (**C**H$_3$) ppm;

LC-MS (ESI+) m/z: 552.3 $[M + NH_4]^+$, 557.3 $[M + Na]^+$, 573.3 $[M + K]^+$.

Data matches that of literature.[186]

8.1.3 Ethyl 6-*O*-acetyl-2,3,4-tri-*O*-benzyl-1-thio-α/β-D-glucopyranoside (2.4α/β)

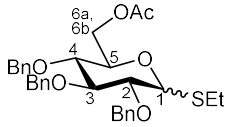


2.3α
2.3β

2.4α
2.4β

1,6-di-*O*-acetyl-2,3,4-tri-*O*-benzyl-α/β-D-glucopyranoside (867 mg, 1.6 mmol) was dissolved in DCM (10 mL) and the solution cooled to 0 °C. Ethanethiol (0.36 mL, 4.9 mmol) and boron trifluoride diethyl etherate (0.30 mL, 2.5 mmol) were then added and the solution stirred for 2 hours over an ice bath. Saturated sodium bicarbonate solution (20 mL) was added and the resulting aqueous layer was extracted with DCM (3

x 20 mL). The combined organic layers were dried over sodium sulphate before being concentrated under reduced pressure. Following purification with column chromatography (100 % DCM) the product was isolated as a pale yellow oil (599 mg, 69%).



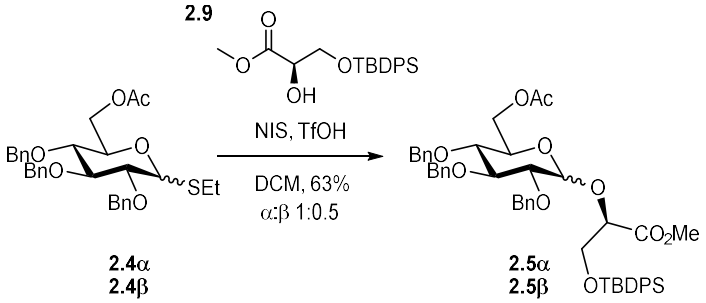
^{1}H NMR (400 MHz, $CDCl_3$): (α:β 1:0.5) δ 7.4-7.3 (20H, m, H_{Ar}α+β), 5.4 (1H, d, *J* = 5.4 Hz, H1α), 5.0 (1H, d, *J* = 10.7 Hz, ArC**H**$_2$α), 4.9 (1H, d, *J* = 9.6 Hz, ArC**H**$_2$α), 4.9-4.8 (2H, m, ArC**H$_2$**α), 4.8-4.7 (2.5H, m, ArC**H**$_2$α+β), 4.7 (1H, d, *J* = 11.8 Hz, ArC**H**$_2$α), 4.6 (1.5H, m, ArC**H**$_2$α+β), 4.5 (0.5H, d, *J* = 9.8 Hz, H1β), 4.4-4.2(4H, m, H5 α, H6aα, H6bα, H5β), 3.9 (1H, m, H3α), 3.8 (1H, dd, *J* = 3.4, 9.5 Hz, H2α), 3.7 (0.5H, m, H3β), 3.5 (2.5H, m, H4α, H2β, H4β), 2.8-2.7 (1H, m, C**H$_2$**CH_3β), 2.6-2.5 (2H, m, C**H$_2$**CH_3α), 2.1 (1.5H, s, O-C**H$_3$**β), 2.0 (3H, s, O-C**H$_3$**α), 1.3 (1.5H, t, *J* = 7.5 Hz, CH_2C**H$_3$**β), 1.3 (3H, t, *J* = 7.4 Hz, CH_2C**H$_3$**α) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 170.7 (Cqβ), 170.7 (Cqα), 138.6-137.8 (Arβ), 128.5-127.7 (Arα), 86.6 (C3β), 85.3 (C1β), 83.0 (C1α), 82.5 (C3α), 81.7 (C2/C4/C5β), 79.5 (C2α), 77.7 (C2/C4/C5β), 77.2 (C4α), 76.9 (C2/C4/C5β), 75.8 (Ar**C**H_2β), 75.8 (Ar**C**H_2α), 75.1 (Ar**C**H_2β), 75.0 (Ar**C**H_2α), 72.4 (Ar**C**H_2α), 69.0 (C5α), 63.4 (C6β), 63.2 (C6α), 25.3 (**C**H_2CH_3β), 23.8 (**C**H_2CH_3α), 20.9 (O-**C**H_3β), 20.8 (O-**C**H_3α), 15.1 (CH_2**C**H_3β), 14.8 (CH_2**C**H_3α) ppm;

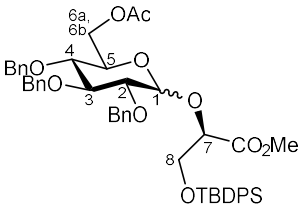
LC-MS (ESI+) m/z 554.3 $[M + NH4]^+$, 559.4 $[M + Na]^+$, 575.4 $[M + K]^+$.

Data matches that of literature.[186]

8.1.4 Methyl 3-*O*-*tert*-butyldiphenylsilyl-(2*R*)-2-*O*-(6-*O*-acetyl-2,3,4-tri-*O*-benzyl-α/β-D-glucopyranosyl)-2,3-dihydroxypropanoate (2.5α/β)



Ethyl 6-*O*-acetyl-2,3,4-tri-*O*-benzyl-1-thio-α/β-D-glucopyranoside (1.00 g, 1.86 mmol) and methyl (2*R*)-3-*O*-*tert*-butyldiphenylsilyl-2,3-dihydroxypropanoate (685 mg, 1.91 mmol) were dissolved in DCM (10 mL) and 4 Å molecular sieves were added. The solution was stirred at room temperature for 30 minutes. *N*-iodosuccinimide (556 mg, 2.47 mmol) was added before the reaction was cooled over ice. Triflic anhydride (24 µL, 0.13 mmol) was carefully added. The reaction was then stirred for 0.5 hours before saturated sodium thiosulphate solution (20 mL) and saturated sodium bicarbonate solution (10 mL) were added. The reaction mixture was then extracted with DCM (3 x 10 mL) and the organic layers combined, washed with brine (20 mL) and dried over $MgSO_4$. The solvent was then removed under reduced pressure and the crude product purified two successive times by column chromatography (10% ethyl acetate in hexane) to give the product as a pale yellow oil (979 mg, 63%).



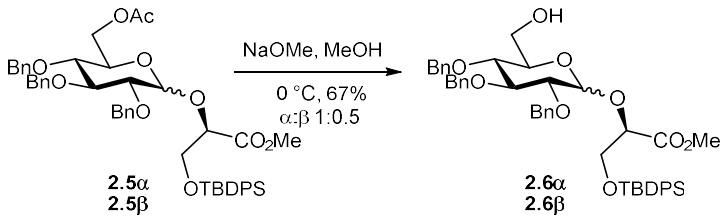
^{1}H NMR (400 MHz, $CDCl_3$): (α:β 1:0.5) δ 7.7-7.6 (7H, m, H_{Ar}α+β), 7.5-7.2 (30H, m, H_{Ar}α+β), 5.2 (1H, d, *J* = 3.6 Hz, H1α), 5.0 (1H, d, *J* = 10.7 Hz, ArC**H**$_2$α), 4.9 (1H, d, *J* = 11.7 Hz, ArC**H**$_2$α), 4.9 (1H, d, *J* = 11.1 Hz, ArC**H**$_2$α), 4.8 (1H, d, *J* = 10.7 Hz, ArC**H**$_2$α), 4.7 (1H, d, *J* = 11.7 Hz, ArC**H**$_2$α), 4.6 (1H, d, *J* = 11.1 Hz, ArC**H**$_2$α), 4.5 (1H, dd, *J* = 4.0, 6.5 Hz, H7α), 4.3 (0.5H, dt, *J* = 3.0, 8.1 Hz, H2β), 4.2 (2H, d, *J* = 3.2 Hz, H6a,6bα), 4.1 (1H, t, *J* = 9.3 Hz, H3α), 4.0 (4.5H, m, H5α, H8a,8bα, H3β, H4β), 3.8 (1.6H, s, C**H**$_3$β), 3.8 (3H, s, -OC**H**$_3$α), 3.6 (1H, dd, *J* = 3.6, 9.6 Hz, H2α), 3.5 (1H, m, H4α), 3.2 (0.5H, d, *J* = 8.0 Hz, H1β) 2.4 (1.5H, s, C**H**$_3$β), 2.0 (3H, s, O-C**H**$_3$α), 1.1 (4.5H, s, 3 x C**H**$_3$β), 1.1 (9H, s, 3 x C**H**$_3$α) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ173.3 (Cqβ), 170.7 (Cqα), 170.2 (Cqα), 138.7 (C_{Ar}), 138.2 (C_{Ar}), 138.1 (C_{Ar}), 135.6-135.5 (C_{Ar}), 133.0-132.9 (C_{Ar}), 129.9 (C_{Ar}), 129.8 (C_{Ar}), 128.4-127.7 (C_{Ar}), 125.3 (C_{Ar}), 94.8 (C1α), 81.7 (C3α), 79.2 (C2α), 77.1 (C4α), 75.8 (Ar**C**H$_2$α) 74.9 (C7α), 74.9 (Ar**C**H$_2$α), 72.1 (Ar**C**H$_2$α), 72.0 (C2β), 69.0 (C5), 65.8 (**C**H$_2$β), 64.7 (C8), 62.9 (C6), 52.4 (**C**H$_3$β), 52.0 (-O**C**H$_3$α), 26.8 (**C**H$_3$ x 3α), 26.7 (**C**H$_3$ x 3β), 21.5 (**C**H$_3$β), 20.8 (O-**C**H$_3$α), 19.3 (^{t}Buβ), 19.2 (^{t}Buα) ppm;

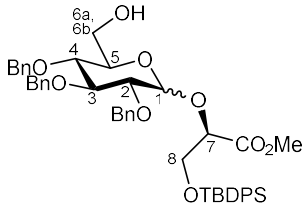
LC-MS (ESI+) m/z: 850.7 $[M + NH_4]^+$, 855.7 $[M + Na]^+$, 871.6 $[M + K]^+$.

Data matched that of literature.[186]

8.1.5 Methyl 3-*O*-*tert*-butyldiphenylsilyl-(2*R*)-2-*O*-(2,3,4-tri-*O*-benzyl-α/β-D-glucopyranosyl)-2,3-dihydroxypropanoate (2.6α/β)



Methyl 3-*O*-*tert*-butyldiphenylsilyl-(2*R*)-2-*O*-(6-*O*-acetyl-2,3,4-tri-*O*-benzyl-α/β-D-glucopyranosyl)-2,3-dihydroxypropanoate (508 mg, 0.60 mmol) was dissolved in methanol (5.0 mL) and the resulting solution cooled to 0 °C. Sodium methoxide solution in methanol (25%, 0.11 mL, 0.48 mmol) was added dropwise and the pH of the solution reached 10-11. The reaction was then stirred for 5.5 hours after which the solution was neutralised with saturated ammonium chloride solution (added until pH reached 7). The aqueous layer was then extracted with ethyl acetate (3 x 20 mL) after which the organic layers were combined and dried over Na_2SO_4. The solvent was then removed under reduced pressure to give the product as a yellow oil that was subsequently deemed clean enough to not warrant further purification, (318 mg, 67%).



^{1}H NMR (400 MHz, $CDCl_3$): (α:β 1:0.5) δ 7.7-7.6 (7H, m, H_{Ar}α+β), 7.4-7.2 (21H, m, H_{Ar}α+β), 5.1 (1H, d, *J* = 3.6 Hz, H1α), 5.0 (1H, d, *J* = 10.8 Hz, ArC**H**$_2$α), 4.9 (2H, m, ArC**H**$_2$α x 2), 4.8 (1H, d, *J* = 10.8 Hz, ArC**H**$_2$α), 4.7 (1H, d, *J* = 11.7 Hz, ArC**H**$_2$α), 4.6 (1H, d, *J* = 11.3 Hz, ArC**H**$_2$α), 4.5 (1H, dd, *J* = 4.2, 6.2 Hz, H7α), 4.2 (0.5H, m, Hβ), 4.0 (4.5H, m, H3α, H8a, 8bα, Hβ), 3.8 (1H, m, H5α), 3.8 (1.5H, s, C**H**$_3$β), 3.7 (3H, s, C**H**$_3$α), 3.7 (2H, m, H6a, 6bα), 3.5 (2H, m, H2α, H4α), 3.1 (0.5H, d, *J* = 8.0 Hz, Hβ), 1.0 (4H, s, ^{t}Buβ), 1.0 (9H, s, ^{t}Buα) ppm;

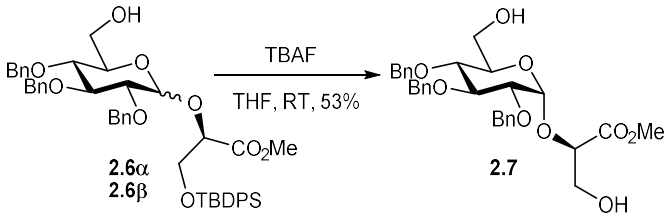
^{13}C NMR (101 MHz, $CDCl_3$): δ 170.4 (Cq), 138.8 (C_{Ar}α), 138.4 (C_{Ar}α), 138.2 (C_{Ar}α), 135.7-135.6 (C_{Ar}α), 133.1-132.9 (C_{Ar}β), 129.9 (C_{Ar}α/β), 129.8 (C_{Ar}α/β), 128.4-127.6 (C_{Ar}α/β), 95.0 (C1α), 81.6 (C3α), 79.4 (C2α), 77.0 (C4α), 75.7 (Ar**C**H$_2$α), 74.9 (C7α), 74.9 (Ar**C**H$_2$α),

72.2 (Ar**C**H_2α), 72.0 (Cβ), 71.2 (C5α), 65.8 (**C**H_2β), 64.7 (C8α), 61.7 (C6α), 52.4 (**C**H_3β), 26.8 (**C**H_3α x 3), 26.7 (**C**H_3β x 3), 19.3 (^{t}Buβ), 19.2 (^{t}Buα) ppm;

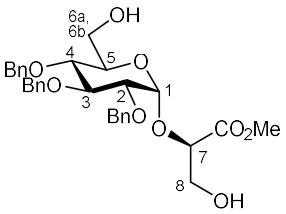
LC-MS (ESI+) m/z: 808.7 $[M + NH_4]^+$, 813.6 $[M + Na]^+$, 829.6 $[M + K]^+$.

Data matches that of literature.[186]

8.1.6 Methyl (2*R*)-2-*O*-(2,3,4-tri-*O*-benzyl-α-D-glucopyranosyl)-2,3-dihydroxypropanoate (2.7)



Methyl 3-*O*-*tert*-butyldiphenylsilyl-(2*R*)-2-*O*-(2,3,4-tri-*O*-benzyl-α/β-D-glucopyranosyl)-2,3-dihydroxypropanoate (242 mg, 0.31 mmol) was dissolved in THF (5.0 mL) and then TBAF solution in THF (1M, 0.37 mL, 0.37 mmol) was added. The reaction mixture was left to stir at room temperature for 3.5 hours after which water (5 mL) was added. The product was extracted from the reaction mixture using ethyl acetate (3 x 20 mL) and the combined organic layers were combined and dried over sodium sulphate. The solvent was then removed under reduced pressure and the resulting yellow oil purified by column chromatography (10% methanol in DCM) to give the product as a waxy solid (90 mg, 53%).



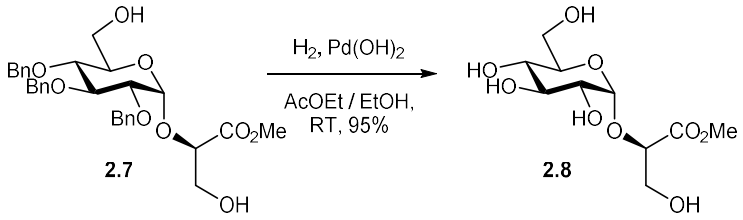
^{1}H NMR (400 MHz, $CDCl_3$): δ 7.4-7.3 (15H, m, H_{Ar}), 5.2 (1H, d, *J* = 3.7 Hz, H1), 5.0 (1H, d, *J* = 10.9 Hz, ArC**H**$_2$), 4.9 (2H, m, ArC**H**$_2$ x2), 4.8 (1H, d, *J* = 10.9 Hz, ArC**H**$_2$), 4.7 (1H, d, *J* = 11.5 Hz, ArC**H**$_2$), 4.6 (1H, d, *J* =11.0 Hz, ArC**H**$_2$), 4.4 (1H, t, *J* = 4.5 Hz, H7) 4.1 (1H, t, *J* = 9.3 Hz, H3), 4.0 (2H, d, *J* = 4.3 Hz, H8a, 8b), 3.8 (3H, s, C**H**$_3$), 3.8-3.7 (3H, m, H5, H6a, 6b), 3.6 (1H, dd, *J* = 3.7, 9.6 Hz, H2), 3.5 (1H, t, *J* = 9.3 Hz, H4) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 170.1 (Cq), 138.7 (C_{Ar}), 137.9 (C_{Ar}), 137.8 (C_{Ar}), 128.5-128 (C_{Ar}), 95.1 (C1), 81.4 (C3), 79.4 (C2), 77.1 (C4), 75.7 (Ar**C**H_2), 75.2 (Ar**C**H_2), 74.6 (C7), 72.5 (Ar**C**H_2), 71.7 (C5), 63.4 (C8), 61.7 (C6), 52.2 (**C**H_3) ppm;

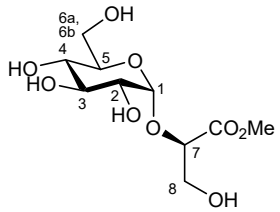
LC-MS (ESI+) m/z: 570.4 $[M + NH_4]^+$, 575.5 $[M + Na]^+$, 591.4 $[M + K]^+$.

Data matches that of literature.[186]

8.1.7 Methyl (2*R*)-2-*O*-(α-D-glucopyranosyl)-2,3-dihydroxypropanoate (2.8)



Methyl (2*R*)-2-*O*-(2,3,4-tri-*O*-benzyl-α-D-glucopyranosyl)-2,3-dihydroxypropanoate (123 mg, 0.22 mmol) was dissolved in a mix of ethyl acetate (4.0 mL) and ethanol (2.0 mL) and to the resulting solution was added palladium hydroxide on carbon (20%, 25 mg). The system was intermittently exposed to a vacuum, replacing the atmosphere in the vessel with hydrogen gas. A hydrogen filled balloon was used to maintain a hydrogen saturated atmosphere within the reaction vessel. The reaction was left to stir at room temperature overnight after which the reaction mixture was filtered over Celite® and washed with methanol. The filtrate was then washed with DCM (3 x 20 mL) after which the aqueous layer was concentrated and the solvent removed under reduced pressure. This led to the product being obtained as a colourless oil which did not require further purification (59 mg, 95%).

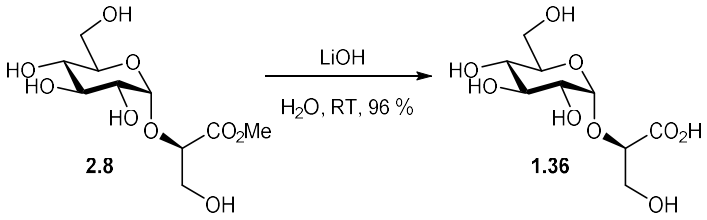


^{1}H NMR (400 MHz, CD_3OD): δ 5.0 (1H, d, *J* = 3.8 Hz, H1), 4.4 (1H, t, *J* = 3.6 Hz, H7), 3.9 (2H, dd, *J* = 2.4, 3.4 Hz, H8a, H8b), 3.8 (2H, m, H6a,6b), 3.8 (3H, s, C**H**$_3$), 3.7 (1H, m, H3), 3.6 (1H, m, H5), 3.4 (1H, dd, *J* = 3.8, 9.7 Hz, H2), 3.3 (1H, m, H4) ppm;

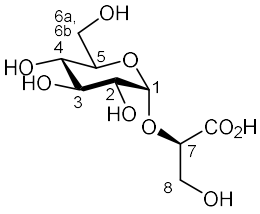
^{13}C NMR (101 MHz, CD_3OD): δ 170.9 (Cq), 97.5 (C1), 75.1 (C7), 73.6 (C3), 73.0 (C5), 72.2 (C2), 70.2 (C4), 62.8 (C8), 61.2 (C6), 51.1 (**C**H_3) ppm;

Data matches that of literature.[186]

8.1.8 (2*R*)-2-*O*-(α-D-glucopyranosyl)-2,3-dihydroxypropanoate (1.36)



Methyl (2*R*)-2-*O*-(α-D-glucopyranosyl)-2,3-dihydroxypropanoate (62 mg, 0.22 mmol) was dissolved in H_2O (10.0 mL) and a solution of LiOH (1 M in H_2O, 180 µL) was added. The reaction was left to stir overnight at room temperature. Reaction was neutralised with Dowex®-H^+ resin which was subsequently removed via filtration. The solvent was removed under reduced pressure to yield a colourless oil (56 mg, 96%).



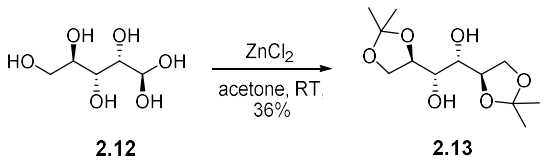
^{1}H NMR (400 MHz, D_2O): δ 5.00 (1H, d, *J* = 3.8 Hz, H1), 4.2 (1H, dd, *J* = 3.2, 6.1 Hz, H7), 3.9-3.7 (6H, m, H3, H5, H6a, H6b, H8a, H8b), 3.5 (1H, dd, *J* = 3.8, 9.8 Hz, H2), 3.4 (1H, m, H4) ppm;

^{13}C NMR (101 MHz, D_2O): δ 177.1 (Cq), 97.5 (C1), 79.1 (C7), 73.2 (C3/C5), 72.2 (C3/C5), 71.6 (C2), 69.4 (C4), 63.1 (C6/C8), 60.5 (C6/C8) ppm;

LC-MS (ESI-) m/z: 267.7 $[M - H]^-$.

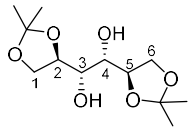
Data matches that of literature.[186]

8.1.9 1,2:5,6-Di-*O*-isopropylidene-D-mannitol (2.13)



Zinc chloride (47.0 g, 0.34 mol) was stirred in acetone (300 mL) and the reaction mixture cooled to 0 °C. D-mannitol (30.0 g, 0.16 mol) was then added and the reaction

left to stir overnight at room temperature. The reaction was then once again cooled to 0 °C and a solution of potassium carbonate (47.8 g) in water (60 mL) was added. The reaction was stirred at room temperature for 1 hour before being filtered under vacuum. The remaining solid was washed with ethyl acetate (3 x 50 mL) and to the filtrate was added concentrated NH_4OH (1.0 mL). The solvent was then removed under reduced pressure and the resulting white residue dissolved in water (50 mL). This was then extracted with ethyl acetate (2 x 200 mL), the organic layers combined and dried over sodium sulphate then the solvent removed under reduced pressure. The resulting white solid was then recrystallized from the minimum amount of boiling ethyl acetate to give the product in the form of white cotton like crystals (15.0 g, 36%).



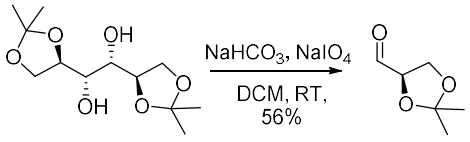
^{1}H NMR (400 MHz, $CDCl_3$): δ 4.2 (2H, dd, *J* = 6.3, 12.4, H5, H2), 4.1 (2H, m, H6a, H1a), 4.0 (2H, m, H6b, H1b), 3.8 (2H, m, H4, H3), 2.7 (2H, d, *J* = 6.7 Hz, -OH x 2), 1.4 (6H, s, C**H**$_3$), 1.4 (6H, s, C**H**$_3$) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 109.3 (Cq x 2), 76.3 (C5, C2), 71.2 (C4, C3), 66.7 (C6, C1), 26.7 (**C**H_3), 25.2 (**C**H_3) ppm;

LC-MS (ESI+) m/z: 263.4 $[M + H]^+$, 285.3 $[M + Na]^+$, 547.4 $[2M + Na]^+$.

Data matches that of literature.[195]

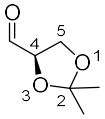
8.1.10 (*R*)-(+)-2,2-dimethyl-1,3-dioxolane-4-carboxaldehyde (2.14)



2.13 **2.14**

1,2:5,6-Di-*O*-isopropylidene-D-mannitol (12.25 g, 46.7 mmol) was dissolved in DCM (120 mL) and saturated sodium bicarbonate solution (5.4 mL) was added followed by sodium periodate (19.98 g, 93.4 mmol). The reaction was stirred at room temperature for 4 hours before the sodium periodate was removed via vacuum filtration. The DCM was removed under reduced pressure and the remaining aqueous solution was extracted with DCM (3 x 50 mL). The organic layers were then washed with water (50

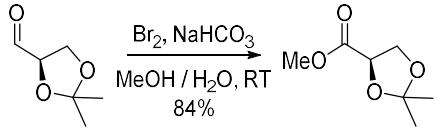
mL) followed by brine (50 mL) before being combined and dried over $MgSO_4$. The solvent was removed under reduced pressure to give the product as a colourless oil (3.42 g, 56%). The crude product was purified via vacuum distillation, fractions collected at 60-65 °C.



^{1}H NMR (400 MHz, $CDCl_3$): δ 9.7 (1H, d, *J* = 1.7 Hz, C**H**O), 4.4 (1H, ddd, *J* = 1.9, 4.9, 7.2 Hz, H4), 4.1 (2H, m, H5a, H5b), 1.5 (3H, s, C**H**$_3$), 1.4 (3H, s, C**H**$_3$) ppm;

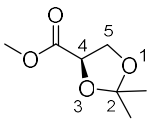
Data matches that of literature.[196]

8.1.11 Methyl (*R*)-(+)-2,2-dimethyl-1,3-dioxolane-4-carboxylate (2.15)



2.14 **2.15**

(*R*)-(+)-2,2-dimethyl-1,3-dioxolane-4-carboxaldehyde (116 mg, 0.80 mmol) was dissolved in a mix of methanol (9.0 mL) and water (1.0 mL) and to the resulting solution was slowly added sodium bicarbonate (491 mg, 5.8 mmol). Bromine (5 drops) was then added and the reaction stirred overnight at room temperature. Saturated sodium thiosulphate solution was then added until the yellow colour of the reaction mixture had disappeared. The product was extracted with DCM (3 x 20 mL) and the organic layers combined and dried over magnesium sulphate. The solvent was removed under reduced pressure and the residue filtered to give the product as a colourless oil (108 mg, 84%).



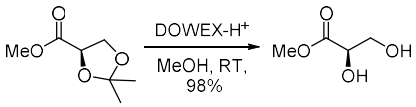
^{1}H NMR (400 MHz, $CDCl_3$): 4.6 (1H, dd, *J* = 5.3, 7.2 Hz, H4), 4.2 (1H, dd, *J* = 7.3, 8.6 Hz, H5a), 4.1 (1H, dd, *J* = 5.2, 8.6 Hz, H5b), 3.8 (3H, s, OC**H**$_3$), 1.5 (3H, s, C**H**$_3$), 1.4 (3H, s, C**H**$_3$) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 171.6 (Cq), 111.4 (Cq), 74.1 (C4), 67.2 (C5), 52.4 (O**C**H$_3$), 25.8 (**C**H$_3$), 25.5 (**C**H$_3$) ppm;

[α]$_D$ +19.0 (*c* 1.0, $CHCl_3$, 22 °C).

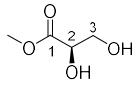
Data matches that in literature.[197, 321]

8.1.12 (*R*)-methyl-2,3-dihydroxyl-propionate (2.16)



2.15 2.16

(*R*)-(+)-2,2-dimethyl-1,3-dioxolane-4-carboxylate (547 mg, 3.43 mmol) was dissolved in methanol (20 mL) and to the resulting solution DOWEX®-H^+ resin (6.36 g) was added. The reaction was left to stir at room temperature overnight before the DOWEX®-H^+ resin was removed via vacuum filtration. The solvent was removed under reduced pressure to give the product as a yellow oil (403 mg, 98%).



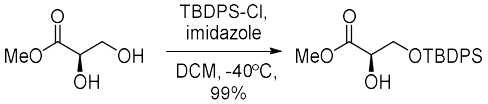
^{1}H NMR (400 MHz, $CDCl_3$): δ 4.3 (1H, t, *J* = 3.6 Hz, H2), 3.9 (1H, m, H3a), 3.8 (1H, m, H3b), 3.8 (3H, s, C**H**$_3$) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 173.5 (Cq), 71.8 (C2), 64.1 (C3), 52.7 (**C**H_3) ppm;

[α]$_D$ +11.0 (*c* 0.5, $CHCl_3$, 23 °C).

Data matches that of literature.[322]

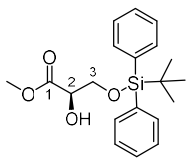
8.1.13 Methyl (2*R*)-3-*O*-tert-butyldiphenylsilyl-2,3-dihydroxypropanoate (2.9)



2.16 2.9

(*R*)-methyl-2,3-dihydroxyl-propionate (403 mg, 3.4 mmol) was dissolved in DCM (5 mL) and to the resulting solution was added imidazole (464 mg, 6.8 mmol). The reaction mixture was then cooled to -40 °C and TBDPS-Cl (1.05 mL, 4.02 mmol) was added. The reaction was left to stir for 1.5 hours before quenching with a saturated solution of NH_4Cl (20 mL). The organic and aqueous layers were separated and the aqueous layer

was extracted with DCM (3 x 20 mL). The combined organic layers were combined and dried over $MgSO_4$ before the solvent was removed under reduced pressure to give the product in the form of a yellow oil (1.09g, 91%).



^{1}H NMR (400 MHz, $CDCl_3$): δ 7.8-7.6 (5H, m, Ar), 7.5-7.4 (5H, m, Ar), 4.3 (1H, t, *J* = 2.9 Hz, H2), 4.0 (1H, dd, *J* = 2.9, 10.5 Hz, H3a), 3.9 (1H, dd, *J* = 2.9, 10.4 Hz, H3b), 3.8 (3H, s, -OC**H**$_3$), 1.1 (9H, s, C**H**$_3$ x 3) ppm;

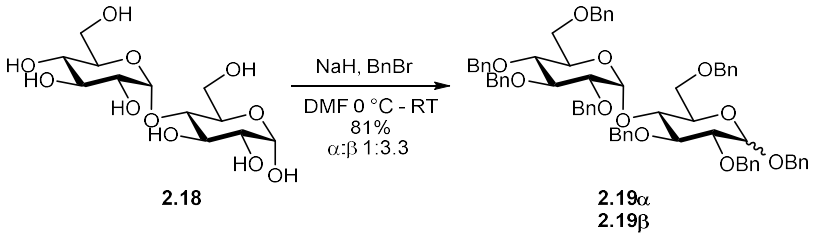
^{13}C NMR (101 MHz, $CDCl_3$): δ 173.1 (Cq), 135.8-127.3 (C_{Ar}), 71.8 (C2), 65.6 (C3), 52.2 (**C**H_3), 26.5 (C**H**$_3$ x 3), 19.1 (^{t}Bu) ppm;

[α]$_D$ -22.0 (*c* 1.0, $CHCl_3$, 21 °C);

LC-MS (ESI+) m/z: 376.2 $[M + NH_4]^+$, 381.2 $[M + Na]^+$, 397.3 $[M + K]^+$.

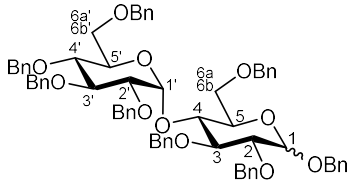
Data matches that of literature.[186, 323]

8.1.14 Octa-*O*-benzyl- α/β-D-maltopyranoside (2.19α/β)



2.18 **2.19α** **2.19β**

D-maltopyranose monohydrate (1.04 g, 2.9 mmol) was dissolved in dimethyl formamide (20 mL) and the resulting solution cooled over ice. Sodium hydride (60% in paraffin, 1.79 g, 44.75 mmol) was slowly added and the reaction stirred for 30 minutes before benzyl bromide (3.4 mL, 29 mmol) was added dropwise. The reaction was then slowly allowed to return to room temperature and stirred for a further 3 hours at which point it was cooled and water added slowly until gas evolution ceased. Product was then extracted with DCM (20 mL) and washed with water (5 x 10 mL) before being dried over $MgSO_4$. Solvent was removed under reduced pressure and the resultant oil was purified by column chromatography (30% ethyl acetate in hexane) to yield a viscous oil (2.48 g, 81%).



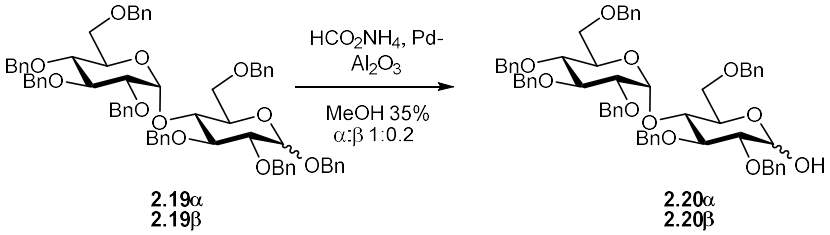
^{1}H NMR (400 MHz, $CDCl_3$): δ (α:β 1:3.3) 7.5-7.1 (50H, m, Ar), 5.7 (0.3H, d, *J* = 3.6 Hz, H1'α), 5.7 (1H, d, *J* = 3.7 Hz, H1'β), 5.1 (0.3H, d, *J* = 11.6 Hz, ArC**H**$_2$α), 5.0-4.3 (24H, m, ArC**H**$_2$, H1β, H1α), 4.2-4.1 (3H, m), 3.9-3.8 (6.6H, m, H3'β, H5'β, H6a'/H6aβ, H6b'/H6bβ), 3.7-3.4 (8.3H, m, H2'β, H4'β, H6a'/H6aβ, H6b'/H6bβ, H2β, H2α) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 138.8-137.5 (Cq_{Ar}), 128.5-125.3 (C_{Ar}), 102.4 (C1β), 96.8 (C1'β), 84.8, 82.3, 82.0, 79.4, 77.8, 75.5 (Bn-**C**H$_2$), 75.0 (Bn-**C**H$_2$), 74.7 (Bn-**C**H$_2$), 74.6, 73.9 (Bn-**C**H$_2$), 73.5 (Bn-**C**H$_2$), 73.4 (Bn-**C**H$_2$), 73.2 (CH_2α), 72.8, 72.1 (Bn-**C**H$_2$), 71.1, 71.0 (Bn-**C**H$_2$), 69.2 (C6/C6'β), 68.3 (C6/C6'β) ppm;

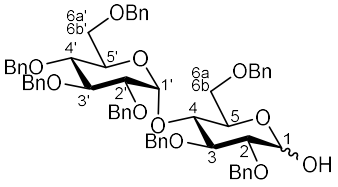
LR-MS (ESI+) m/z: 1085.8 $[M + Na]^+$, 1101.9 $[M + K]^+$.

Data matches that of literature.[324]

8.1.15 2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranose (2.20α/β)



Octa-*O*-benzyl- α/β-D-maltopyranoside (115 mg, 0.11 mmol) was dissolved in methanol (5 mL) and palladium on alumina (10%, 144 mg) was added, followed by ammonium formate (132 mg, 2.09 mmol). The reaction was left to stir at room temperature and monitored by TLC (2:1 diethyl ether: ethyl acetate). Upon complete consumption of the starting material the catalyst was removed via filtration over Celite® which was washed with ethyl acetate. The solvent was then removed in vacuo and the crude product purified by column chromatography (0-20% diethyl ether in toluene) to give a viscous oil (37mg, 35%).



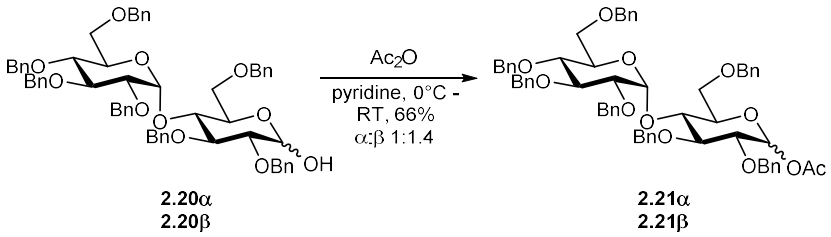
^{1}H NMR (400 MHz, $CDCl_3$): δ (α:β 1:0.2) 7.4-7.1 (34H, m, H_{Ar}), 5.6 (1H, d, *J* = 3.6 Hz, H1'α), 5.6 (0.2H, d, *J* = 3.6 Hz, H1'β), 5.2 (1H, d, *J* = 3.6 Hz, H1α), 5.0-4.8 (8.0H, m, H1β), 4.7-4.4 (11.4H, m), 4.3 (1.4H, m), 4.2-4.0 (5.2H, m), 3.9-3.6 (9.2H, m, H3'α, H4'α, H5'α, H2α, H6aα, H6bα), 3.6-3.5 (3.6H, m, H2'α, H6a'α), 3.4 (1.4H, m, H6b'α, H2β) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 138.9-137.7 (Cq_{Ar}), 129.1-125.3 (C_{Ar}), 97.4 (C1β), 97.0 (C1'α), 96.9 (C1'β), 90.8 (C1α), 84.4, 83.1, 82.0, 81.3, 80.0, 79.5, 79.3, 77.7, 75.5 (Bn-**C**H_2α), 75.0 (Bn-**C**H_2β), 75.0 (Bn-**C**H_2α), 74.5 (Bn-**C**H_2β), 74.4, 74.2 (Bn-**C**H_2α), 73.9 (Bn-**C**H_2β), 73.5 (Bn-**C**H_2β), 73.5 (Bn-**C**H_2α), 73.3 (Bn-**C**H_2β), 73.3 (Bn-**C**H_2α), 73.1 (Bn-**C**H_2α), 72.9, 72.9, 71.2 ,71.1, 69.9, 69.4 (**C**H_2β), 69.1 (C6/C6'α), 68.2 (C6/C6'α) ppm;

LR-MS (ESI+) m/z: 995.7 $[M + Na]^+$, 1011.7 $[M + K]^+$.

Data matches that of literature.[325]

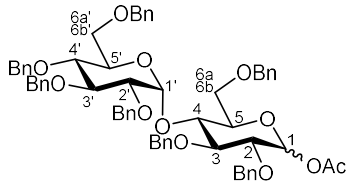
8.1.16 1-*O*-acetyl-2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranoside (2.21α/β)



2.20α
2.20β

2.21α
2.21β

2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranose (315 mg, 0.32 mmol) was dissolved in pyridine (5.0 mL) and acetic anhydride (0.05 mL) was added. Reaction was then stirred at room temperature overnight. Methanol (5.0 mL) was added to the reaction mixture and the solvents removed under reduced pressure. Remaining pyridine was co-evaporated with toluene then chloroform. Crude material was purified *via* column chromatography (0-10% diethyl ether in toluene) to yield a viscous oil (212 mg, 66%).



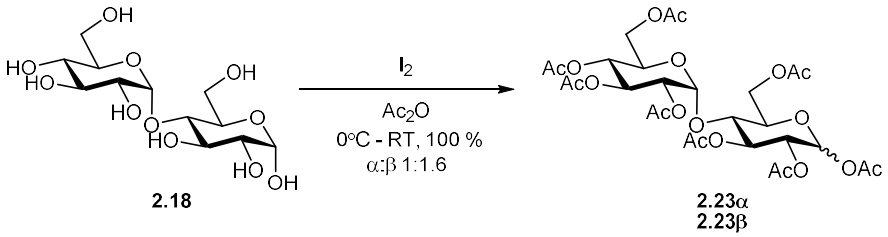
^{1}H NMR (400 MHz, $CDCl_3$): δ (α:β 1:1.4) 7.4-7.1 (45H, m, H_{Ar}), 6.4 (0.7H, d, *J* = 3.6 Hz, H1'α), 5.7 (0.7H, d, *J* = 3.7 Hz, H1α), 5.7 (1H, d, *J* = 8.0 Hz, H1β), 5.6 (1H, d, *J* = 3.7 Hz, H1'β), 5.0 (0.7H, d, *J* = 11.9 Hz, ArC**H**$_2$α), 4.9-4.8 (9H, m), 4.7-4.4 (15H, m), 4.3 (1.7H, m), 4.2 (1.7H, m), 4.1 (0.7H, m, H3'α), 4.0 (0.7H, m), 4.0-3.6 (13.7H, m, H2'α, H2β), 3.5 (3.7H, m, H2α, H2'β, H6'/H6β), 3.4 (2H, m, H6'/H6β), 2.2 (2.1H, s, C**H**$_3$α), 2.1 (3H, s, C**H**$_3$β) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 169.6 (Cq), 169.3 (Cq), 138.7-137.4 (Cq_{Ar}), 128.5-126.7 (C_{Ar}), 97.0 (C1'β), 96.9 (C1'α), 93.9 (C1β), 89.6 (H1α), 84.8, 82.0, 82.0, 81.6, 81.0, 80.5, 79.4, 79.4, 79.1, 77.7, 75.6 (Bn-**C**H$_2$), 75.5 (Bn-**C**H$_2$), 75.4, 75.0 (Bn-**C**H$_2$), 75.0 (Bn-**C**H$_2$), 74.8 (Bn-**C**H$_2$), 74.3 (Bn-**C**H$_2$), 73.5 (Bn-**C**H$_2$), 73.5 (Bn-**C**H$_2$), 73.3 (Bn-**C**H$_2$), 73.3 (Bn-**C**H$_2$), 73.3 (Bn-**C**H$_2$), 73.2 (Bn-**C**H$_2$), 72.8, 72.5, 71.6, 71.1, 68.7 (C6/6'β), 68.5 (C6/6'α), 68.2 (C6/6'β), 68.2 (C6/6'α), 21.23 (**C**H$_3$α), 21.1 (**C**H$_3$β) ppm;

LR-MS (ESI+) m/z: 1037.9 $[M + Na]^+$, 1053.7 $[M + K]^+$.

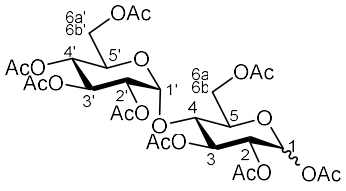
Data matches that of literature.[326]

8.1.17 Octa-*O*-acetyl-α/β-D-maltopyranoside (2.23α/β)



D-maltose monohydrate (2.00 g, 5.55 mmol) was dissolved in acetic anhydride (10 mL) and cooled over an ice bath. Iodine (0.1 g, 0.39 mmol) was then slowly added to the resultant solution. The reaction was warmed to room temperature and stirred for 2 hours. A saturated solution of sodium thiosulphate was added until the dark brown colour had gone leaving a yellow coloured solution. A saturated solution of sodium bicarbonate was then added until fizzing had subdued. Ethyl acetate (3 x 10 mL) was used to extract the product and resulting organic layers were washed with saturated

sodium bicarbonate (2 x 10 mL). Organic layers were dried over $MgSO_4$ and solvent removed under vacuum to give a white foam (2.85 g, 4.20 mmol, 76%).



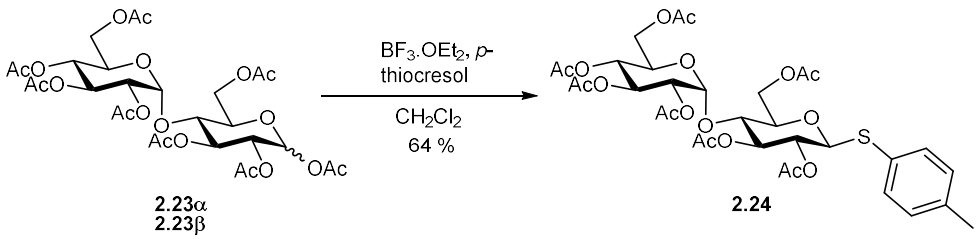
^{1}H NMR (400 MHz, $CDCl_3$): δ (α:β 0.6:1) 6.3 (0.6H, d, *J* = 3.8 Hz, H1α), 5.8 (1H, d, *J* = 8.2 Hz, H1β), 5.5 (0.6H, dd, *J* = 8.7, 10.2 Hz, H3α), 5.4 (0.6H, d, *J* = 3.9 Hz, H1'α), 5.4 (1H, d, *J* = 4.2 Hz, H1'β), 5.4 (1.6H, m, H3'β, Hα), 5.3 (1H, t, *J* = 9.0 Hz, H3β), 5.1 (1.6H, q, *J* = 9.8 Hz, H4'β, Hα), 5.00 (1.6H, m, H2β, H2α), 4.9 (1.6H, m, H2'β, Hα), 4.5 (1.6H, dd, *J* = 2.5, 12.2 Hz, H6aβ, H6/6'α), 4.2 (3.2H, m, H6bβ, H6a'β, H6/6'α x 2), 4.1 (1H, m, Hα), 4.1 (3.2H, m, H4β, H6b'β, H4α, H6/6'α), 4.0 (1.6H, m, H5'β, Hα), 3.8 (1H, ddd, *J* = 2.6, 4.3, 9.5 Hz, H5β), 2.2 (1.8H, s, C**H**3 α), 2.2 (1.8H, s, C**H**$_3$α), 2.1 (3H, s, C**H**$_3$β), 2.1 (4.2H, s, C**H**$_3$α+β), 2.1 (3H, s, C**H**$_3$β) 2.1 (1.8H, s, C**H**$_3$α), 2.1 (3H, s, C**H**$_3$β), 2.0 (1.6H, s, C**H**$_3$α), 2.0 (3H, s, C**H**$_3$β), 2.0 (1.6H, s, C**H**$_3$α), 2.0 (3H, s, C**H**$_3$β), 2.0 (4.2H, s, C**H**$_3$α+β), 2.0 (3H, s, C**H**$_3$β), 2.0 (1.8H, s, C**H**$_3$α) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 170.6 (Cq), 170.6 (Cq), 170.5 (Cq), 170.5 (Cq), 170.1 (Cq), 170.0 (Cq), 170.0 (Cq), 169.9 (Cq), 169.8 (Cq), 169.6 (Cq), 169.5 (Cq), 169.4 (Cq), 168.9 (Cq), 168.8 (Cq), 95.8 (C1'α), 95.7 (C1'β), 91.3 (C1β), 88.9 (C1α), 75.3 (C3β), 73.0 (C5β), 72.4 (C4β), 72.3 (Cα), 72.3 (Cα), 70.9 (C2β), 70.1 (Cα), 70.1 (Cα), 70.0 (C2'), 69.7 (C2α), 69.3 (C3'β), 69.3 (Cα), 68.6 (Cα), 68.6 (C5'β), 68.0 (C4'β), 67.9 (Cα), 62.5 (C6β), 62.4 (C6/6'α), 61.5 (C6' β), 61.4 (C6/6'α), 21.0 - 20.4 (**C**H$_3$α+ β) ppm;

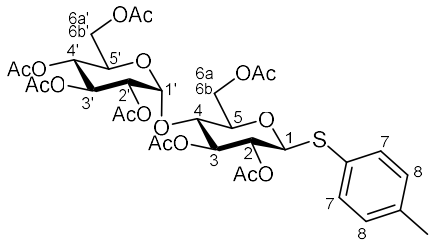
LR-MS (ESI+) m/z: 696.5 $[M + NH_4]^+$, 701.5 $[M + Na]^+$, 717.5 $[M + K]^+$.

Data matches that of literature.[327]

8.1.18 *p*-tolyl-2,3,6,2',3',4',6'-hepta-*O*-acetyl-1-thio-β-D-maltopyranoside (2.24)



Octa-*O*-acetyl-α/β-D-maltopyranoside (2.85 g, 4.2 mmol) was dissolved in anhydrous DCM (15 mL) a solution of *p*-thiocresol (1.07 g, 8.4 mmol) in anhydrous DCM (4 mL) was then added, followed by $BF_3.OEt_2$ (1.0 mL, 8.4 mmol). The reaction was left to stir at room temperature overnight. Reaction was cooled over ice and triethylamine was added until deep red colour changed to yellow. Reaction mixture was washed with saturated sodium bicarbonate (20 mL) followed by water (20 mL). Organic layers were dried over $MgSO_4$ and solvent removed under reduced pressure. The crude product was purified via column chromatography (40% ethyl acetate in hexane) to give a white foamy solid (1.98 g, 2.7 mmol, 64 %).



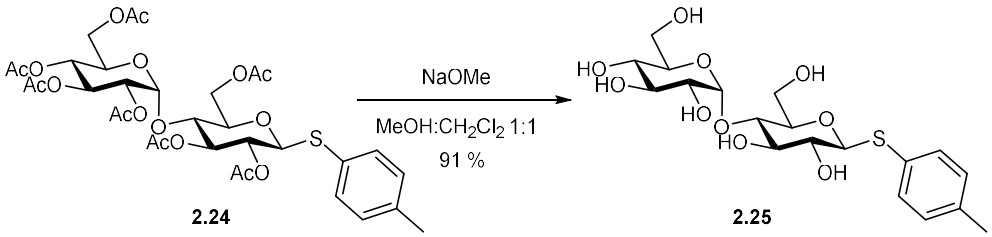
^{1}H NMR (400 MHz, $CDCl_3$): δ 7.4 (2H, d, *J* = 8.1 Hz, H7), 7.1 (2H, d, *J* = 8.0 Hz, H8), 5.4 (1H, d, *J* = 4.0 Hz, H1'), 5.3 (1H, dd, J = 9.7, 10.4 Hz, H3'), 5.3 (1H, t, *J* = 9.0 Hz, H3), 5.0 (1H, t, *J* = 9.9 Hz, H4'), 4.8 (1H, dd, *J* = 4.0, 10.5 Hz, H2'), 4.8 (1H, t, *J* = 9.7 Hz, H2), 4.7 (1H, d, *J* = 10.0 Hz, H1), 4.5 (1H, dd, *J* = 2.5, 12.0 Hz, H6a), 4.2 (2H, td, *J* = 4.4, 12.1 Hz, H6b, H6b'), 4.1 (1H, dd, *J* = 2.2, 12.4 Hz, H6a'), 3.9 (2H, m, H5', H4), 3.7 (1H, m, H5), 2.4 (3H, s, Ar-C**H**$_3$), 2.1 (3H, s, C**H**$_3$), 2.1 (3H, s, C**H**$_3$), 2.1 (3H, s, C**H**$_3$), 2.0 (3H, s, C**H**$_3$), 2.0 (3H, s, C**H**$_3$), 2.0 (3H, s, C**H**$_3$), 2.0 (3H, s, C**H**$_3$) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 170.5 (Cq), 170.3 (Cq), 170.2 (Cq), 169.9 (Cq), 169.5 (Cq), 169.4 (Cq), 134.1 (C7), 129.7 (C8), 95.5 (C1'), 85.1 (C1), 76.5 (C3), 76.1 (C5), 72.4 (C5'/C4), 70.7 (C2), 70.0 (C2'), 69.3 (C3'), 68.5 (C5'/C4), 68.0 (C4'), 62.8 (C6), 61.5 (C6'), 21.2 (Ar-**C**H$_3$), 20.9 (**C**H$_3$), 20.8 (**C**H$_3$), 20.8 (**C**H$_3$), 20.7 (**C**H$_3$), 20.6 (**C**H$_3$), 20.6 (**C**H$_3$), 20.6 (**C**H$_3$) ppm;

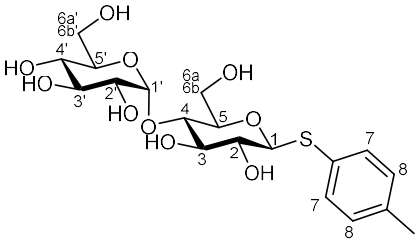
LR-MS (ESI+) m/z: 760.5 $[M + NH_4]^+$, 765.4 $[M + Na]^+$, 781.5 $[M + K]^+$.

Data matches that of literature.[216]

8.1.19 *p*-tolyl-1-thio-β-D-maltopyranoside (2.25)



p-tolyl-2,3,6,2',3',4',6'-hepta-*O*-acetyl-1-thio-β-D-maltopyranoside (17.93 g, 24.1 mol) was dissolved in anhydrous DCM (80 mL) and methanol (80 mL). Sodium methoxide (25% w/v in methanol) (10.5 mL, 48.3 mmol) was added and the reaction stirred at room temperature for 3 hours. Dowex®-H^+ resin was used to neutralise the solution and was then filtered off and washed with methanol. Solvent was removed under reduced pressure and crude product was purified via column chromatography (5-30% methanol in DCM) to give a white foamy solid (9.85 g, 21.96 mmol, 91 %).



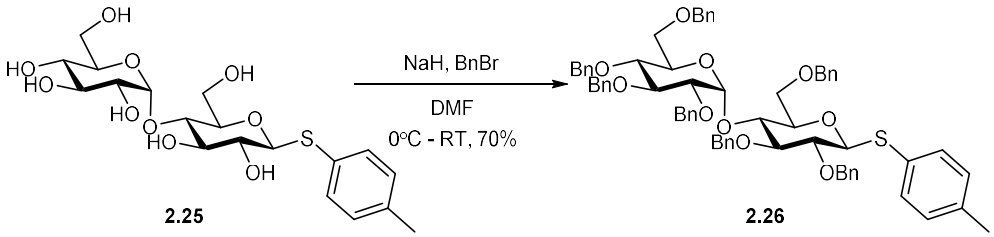
^{1}H NMR (400 MHz, CD_3OD): δ 7.5 (2H, d, *J* = 8.1 Hz, H7), 7.1 (2H, d, *J* = 8.0 Hz, H8), 5.2 (1H, d, *J* = 3.8 Hz, H1'), 4.5 (1H, d, *J* = 9.8 Hz, H1), 3.8 (3H, m, H6a', H6a, H6b), 3.6 (4H, m, H3', H5', H3, H6b'), 3.5 (1H, t, *J* = 9.3 Hz, H4), 3.4 (1H, dd, *J* = 3.7, 9.7 Hz, H2'), 3.4 (1H, m, H5), 3.2 (2H, m, H4', H2), 2.3 (3H, s, Ar-C**H**$_3$) ppm;

^{13}C NMR (101 MHz, CD_3OD): δ 137. 5 (Cq_{Ar}), 132.2 (C7), 129.7 (Cq_{Ar}), 129.2 (C8), 101.4 (C1'), 88.2 (C1), 79.4 (C4), 79.2 (C5), 78.0, 73.7, 73.4 (C3', C5', C3), 72.8 (C2'), 72.0, 70.1 (C4', C2), 61.3 (C6'), 61.0 (C6), 19.7 (Ar-**C**H_3) ppm;

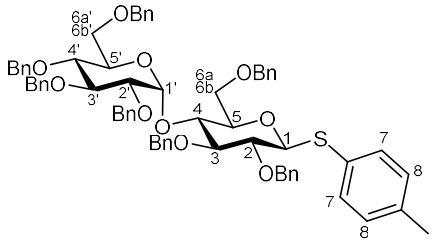
LR-MS (ESI+) m/z: 466.4 $[M + NH_4]^+$, 471.4 $[M + Na]^+$, 487.2 $[M + K]^+$, 897.5 $[2M + H]^+$, 919.4 $[2M + Na]^+$.

Data matches that of literature.[216]

8.1.20 *p*-tolyl-2,3,6,2',3',4',6'-hepta-*O*-benzyl-1-thio-β-D-maltopyranoside (2.26)



p-tolyl-1-thio-β-D-maltopyranoside (0.95 g, 2.12 mmol) was dissolved in anhydrous dimethylformamide (20 mL) and cooled over ice. NaH (60% in paraffin, 682 mg, 17.04 mmol), was added and the reaction stirred for 1 hour over ice. Benzyl bromide (2.00 mL, 16.82 mmol) was then added and the reaction allowed to warm to room temperature and stirred overnight. Reaction was quenched with MeOH until fizzing subdued then H_2O until no further gas evolution. Product was extracted with diethyl ether (3 x 20 mL) and organic layers washed with H_2O (20 mL) followed by brine (20 mL) before being dried over $MgSO_4$. Solvent was removed under vacuum and the crude oil was purified via column chromatography (5-20% ethyl acetate in hexane) to give a pale yellow, viscous oil (1.61 g, 1.49 mmol, 70%).



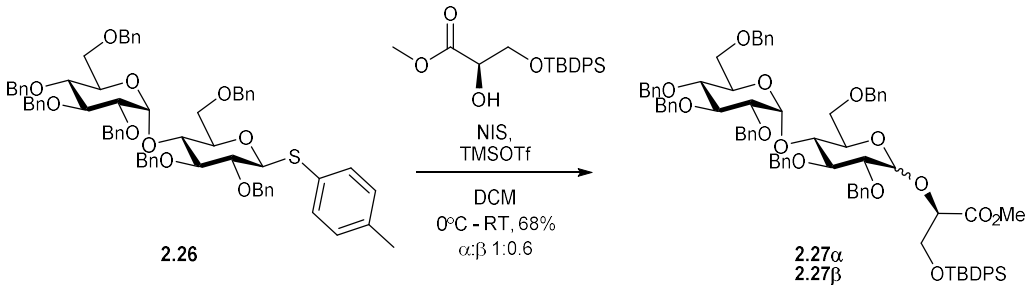
^{1}H NMR (400 MHz, $CDCl_3$): δ 7.5 (2H, d, *J* = 8.1 Hz, H7), 7.21-7.3 (7H, m, H_{Ar}), 7.3-7.1 (27H, m, H_{Ar}), 7.0 (2H, d, *J* = 8.1 Hz, H8), 5.6 (1H, d, *J* = 3.6 Hz, H1'), 4.9-4.8 (4H, m, ArC**H**$_2$), 4.8 (2H, m, ArC**H**$_2$), 4.6 (1H, d, *J* = 9.8 Hz, H1), 4.6-4.4 (8H, m, ArC**H**$_2$), 4.3 (1H, d, *J* = 12.2 Hz, ArC**H**$_2$), 4.1 (1H, t, *J* = 9.1 Hz, H5), 3.9-3.8 (5H, m, H3, H3', H5', H6a, H6b), 3.6 (1H, m, H4'), 3.6 (3H, m, H2, H4, H6a'), 3.5 (1H, dd, *J* = 3.6, 10.0 Hz, H2'), 3.4 (1H, dd, *J* = 0.9, 10.3 Hz, H6b') 2.3 (3H, s, Ar-C**H**$_3$) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 138.7-137.7 (Cq_{Ar}), 132.8 (C7), 129.7 (C8), 129.6 (Cq_{Ar}), 128.4-126.5 (C_{Ar}) 97.1 (C1'), 87.4 (C1), 86.8 (C3/C3'/C5'), 82.0 (C3/C3'/C5'), 80.9 (C2/C4), 79.4 (C2'), 78.8 (C2/C4), 77.7 (C4'), 75.5 (Bn-**C**H$_2$), 75.2 (Bn-**C**H$_2$), 75.0 (Bn-**C**H$_2$), 74.3 (Bn-**C**H$_2$), 73.5 (Bn-**C**H$_2$), 73.4 (Bn-**C**H$_2$), 73.3 (Bn-**C**H$_2$), 72.7 (C5), 71.1 (C3/C3'/C5'), 69.2 (C6), 68.2 (C6'), 21.2 (Ar-CH_3) ppm;

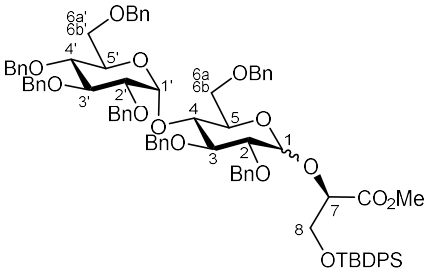
LR-MS (ESI+) m/z: 1096.8 $[M+NH_4]^+$, 1101.8 $[M+Na]^+$, 1117.8 $[M+K]^+$.

Data matches that of literature.[217]

8.1.21 Methyl 3-*O*-*tert*-butyldiphenylsilyl-(2*R*)-2-*O*-(2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (2.2 α/β)



p-tolyl-2,3,6,2',3',4',6'-hepta-*O*-benzyl-1-thio-β-D-maltopyranoside (9.93 g, 9.2 mmol) and methyl (2*R*)-3-*O*-tert-butyldiphenylsilyl-2,3-dihydroxypropanoate (3.96 g, 11.06 mmol) were dissolved in anhydrous DCM (90 mL) and stirred over 4 Å molecular sieves for 30 minutes. *N*-iodosuccinimide (2.71 g, 12.05 mmol) was added and the reaction cooled over ice. TMS-OTf (120 µL, 0.64 mmol) was added and the reaction stirred for 30 minutes. Saturated aqueous solutions of $NaHCO_3$ and $Na_2S_2O_3$ were added until deep red colour of reaction turned yellow. Aqueous layer was extracted with DCM (2 x 50 mL) and the combined organic layers were washed with saturated $NaHCO_3$ solution then dried over $MgSO_4$. Solvent was removed *in vacuo* to give a yellow oil which was purified by column chromatography (10-15% ethyl acetate in hexane) to yield a pale yellow oil (8.27 g, 6.3 mmol, 68%).



^{1}H NMR (400 MHz, $CDCl_3$): δ (α:β 1:0.6) 7.7 (7H, m, H_{Ar}), 7.5-7.1 (64H, m, H_{Ar}), 5.7 (1H, d, *J* = 3.6 Hz, H1'α), 5.7 (0.6H, d, *J* = 3.7 Hz, H1'β), 5.2 (1H, d, *J* = 3.7 Hz, H1α), 5.1 (1H, d, *J* = 11.5 Hz, ArC**H**$_2$α), 5.0-4.8 (7H, m, ArC**H**$_2$α x 4), 4.7 (1H, d, *J* = 10.8 Hz, ArC**H**$_2$α), 4.6 (1H, d, *J* = 11.5 Hz, ArC**H**$_2$α), 4.6 (0.6H, d, *J* = 11.1 Hz, ArC**H**$_2$β), 4.6-4.5 (11H, m, ArC**H**$_2$α, H7α, H1β), 4.4 (1H, d, J = 10.8 Hz, ArC**H**$_2$α), 4.3 (2.4H, m, Ar-C**H**$_2$α), 4.2 (1H, t, *J* = 9.1 Hz, H3α), 4.1-3.8 (9.5H, m, H4α, H5α, H8a+bα, H3'α), 3.8-3.7 (5.5H, m, H6aα, -OC**H**$_3$α), 3.7 (1H, dd, *J* = 3.8, 9.4 Hz, H2α), 3.7 (1.8H, s, -OC**H**$_3$β), 3.7-3.5 (5.5H, m, 4'α, H5'α, 6bα, H2β), 3.5-3.4 (3.8H, m, H6a'α, H2'α, H2'β), 3.4 (0.6H, dd, *J*

= 1.4, 10.7 Hz, H6/H6'β), 3.4 (1H, dd, *J* = 1.0, 10.7 Hz, 6b'α), 1.1 (9H, s, ^{t}Buα), 1.0 (5H, s, ^{t}Buβ) ppm;

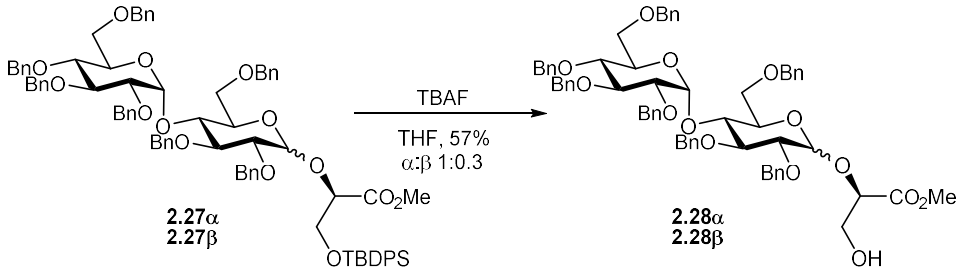
^{13}C NMR (101 MHz, $CDCl_3$): δ 171.0 (Cq), 170.5 (Cq), 139.1 (Cq), 138.8 (Cq), 138.7 (Cq), 138.6 (Cq), 138.5 (Cq), 138.2 (Cq), 138.2 (Cq), 138.1 (Cq), 138.1 (Cq), 138.0 (Cq), 135.7 (C_{Ar}), 135.6 (C_{Ar}), 135.6 (C_{Ar}), 133.1 (Cq), 133.0 (Cq), 132.9 (Cq), 132.9 (Cq), 129.9 (C_{Ar}), 129.9 (C_{Ar}), 129.8 (C_{Ar}), 128.3-126.6 (C_{Ar}), 103.7 (C1β), 97.1 (C1'α), 96.7 (C1'β), 94.7 (C1α), 84.7, 82.0 (C3'α), 81.8, 81.7 (C3α), 79.6 (C2'α), 79.4 (C2α), 79.3, 79.3, 77.7, 77.6 (C4'/C5'α), 75.5 (Bn-**C**H_2α), 75.0 (C7α), 74.9 (Bn-**C**H_2α), 74.6 (Bn-**C**H_2β), 74.3 (Bn-**C**H_2α), 73.5 (Bn-**C**H_2β), 73.5 (Bn-**C**H_2α), 73.2 (Bn-**C**H_2α), 73.1 (Bn-**C**H_2α), 72.4 (C4/C5α), 72.1, 72.0 (Bn-**C**H_2α), 71.0 (C4'/C5'α), 71.0, 70.2 (C4/C5α), 68.9 (C6/6'β), 68.9 (C6α), 68.2 (C6/6'β), 68.2 (C6'α), 64.7 (C8α), 63.8 (C8β), 52.0 (CH_3α), 51.9 (CH_3β), 26.8 (^{t}Buα), 26.7 (^{t}Buβ), 19.2 (Cq-^{t}Bu) ppm;

HR-MS (ESI+) m/z: for $C_{81}H_{88}O_{14}SiNa$ $[M+Na]^+$ calcd. 1335.5841; observed: 1335.58;

[α]$_D$ (α:β 1:0.7) +52.7 (*c* 0.5, $CHCl_3$, 23 °C);

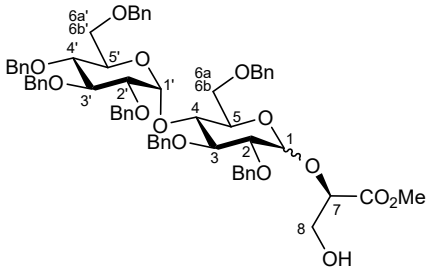
IR (cm^{-1}) 3030 (Ar C-H), 2857 (C-H), 1752 (C=O), 1497-1428 (C=C), 1360 (C-H), 1027 (C-O).

8.1.22 Methyl (2*R*)-2-*O*-(2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (2.28α/β)



Methyl 3-*O*-*tert*-butyldiphenylsilyl-(2*R*)-2-*O*-(2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (578 mg, 0.44 mmol) was dissolved in anhydrous THF (7 mL) and tetrabutylammonium fluoride (TBAF) (1M solution in THF, 530 µL, 0.53 mmol) was added. The reaction was stirred at room temperature for 3

hours. Water (15 mL) was added to the reaction mixture following which the product was extracted with ethyl acetate (3 x 15 mL). The combined organic layers were washed with brine (20 mL) then dried over $MgSO_4$. Solvent was removed and resulting crude oil was purified via column chromatography (20-30% ethyl acetate in petroleum ether) to yield a colourless oil (263 mg, 0.25 mmol, 57%).



^{1}H NMR (400 MHz, $CDCl_3$): δ (α:β 1:0.3) 7.4-7.1 (38H, m, H_{Ar}), 5.7 (1H, d, *J* = 3.6 Hz, H1'α), 5.6 (0.3H, d, *J* = 3.7 Hz, H1'β), 5.2 (1H, d, *J* = 3.7 Hz, H1α), 5.1 (1H, d, *J* = 11.6 Hz, ArC**H**$_2$α), 5.0-4.8 (7H, m, ArC**H**$_2$α+β), 4.7 (0.3H, m, H1β), 4.7-4.4 (9.3H, m, ArC**H**$_2$β), 4.4 (1H, dd, *J* = 3.6, 5.4 Hz, H7α), 4.4 (0.6H, m, Hβ), 4.3 (1H, d, *J* = 12.2 Hz, ArC**H**$_2$α), 4.1 (1H, dd, *J* = 8.3, 9.4 Hz, H3α), 4.0-3.9 (5H, m, H4α, H5α, H8aα, H8bα), 3.9 (2H, m, H3'α, H8aβ, H8bβ), 3.9-3.8 (0.3H, m, Hβ), 3.80 (3H, s, -OC**H**$_3$α), 3.8-3.7 (2H, m, H6aα), 3.7 (0.9H, s, -C**H**$_3$β), 3.7-3.6 (5H, m, H2α, H4'α, H5'α, H6bα), 3.6 (0.6H, m, H2β), 3.5 (2.3H, m, H2'α, H6a'α, H2β), 3.4 (1.3H, m, H6b'α) ppm;

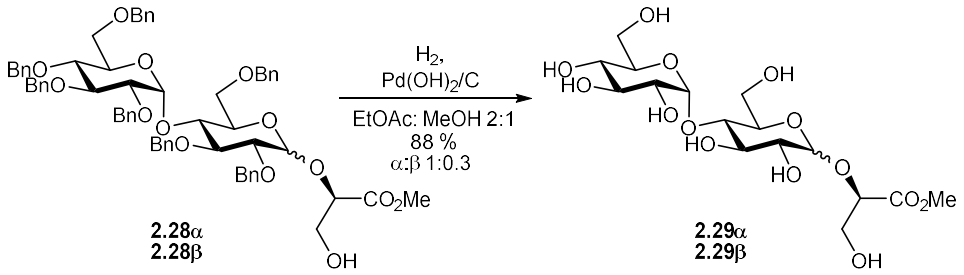
^{13}C NMR (101 MHz, $CDCl_3$): δ (β signals too weak) 170.2 (Cq), 138.9 (Cq), 138.7 (Cq), 138.4 (Cq), 137.9 (Cq), 137.8 (Cq), 137.7 (Cq), 128.4-127.4 (Cq), 127.1 (Cq), 126.7 (Cq), 126.6 (Cq), 96.8 (C1'α), 95.1 (C1α), 81.9 (C3'α), 81.4 (C3α), 79.5 (C2α /C4'α/C5'α), 79.4 (C2'α), 77.7 (C2α /C4'α/C5'α), 75.5 (Bn-**C**H$_2$α), 75.5 (C7α), 75.0 (Bn-**C**H$_2$α), 74.2 (Bn-**C**H$_2$α), 73.5 (Bn-**C**H$_2$α), 73.4 (Bn-**C**H$_2$α), 73.3 (Bn-**C**H$_2$α), 72.4 (Bn-**C**H$_2$α), 72.3 (C4α/C5α), 71.1 (C2α/C4'α/C5'α), 70.3 (C4α/C5α), 69.0 (C6α), 68.2 (C6'α), 63.5 (C8α), 52.2 (**C**H$_3$α) ppm;

HR-MS (ESI+) m/z: for $C_{65}H_{70}O_{14}Na$ $[M + Na]^+$ calcd. 1097.4663; observed: 1097.47;

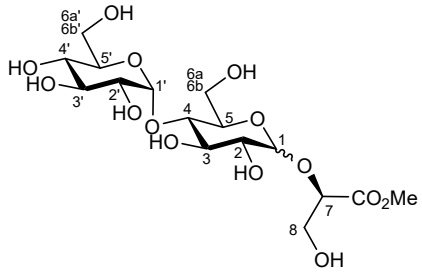
[α]$_D$ (α:β 1:0.7) +59.9 (*c* 0.5, $CHCl_3$, 21 °C);

IR (cm^{-1}) 3500 (O-H), 3030 (Ar C-H), 2867 (C-H), 1738 (C=O), 1497-1453 (C=C), 1360 (C-H), 1027 (C-O).

8.1.23 Methyl (2*R*)-2-*O*-(α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (2.29α/β)



Methyl (2*R*)-2-*O*-(2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (273 mg, 0.25 mmol) was dissolved in a mixture of ethyl acetate and methanol (36 mL, 2:1) and palladium hydroxide on activated charcoal (20% wt, 258 mg, 0.37 mmol) was added. The reaction was purged with hydrogen gas which was then used to maintain a saturated atmosphere. The reaction was left to stir at room temperature for 18 hours. The palladium catalyst was removed via filtration over Celite® which was washed with methanol. The solvent was then removed *in vacuo* to yield a white solid foam (99 mg, 0.22 mmol, 88%)



^{1}H NMR (400 MHz, CD_3OD): δ (α:β 1:0.3) 5.2 (1H, d, *J* = 3.8 Hz, H1'α), 5.2 (0.3H, d, *J* = 4.0 Hz, H1'β), 5.0 (1H, d, *J* = 3.8 Hz, H1α), 4.5 (0.3H, d, *J* = 7.7 Hz, H1β), 4.4 (1H, t, *J* = 3.7 Hz, H7α), 4.4 (0.3H, t, *J* = 4.0, 5.1 Hz, Hβ), 4.0 (1H, t, *J* = 9.3 Hz, H3α), 3.9 (2H, d, *J* = 3.7 Hz, H8a, H8b), 3.9-3.8 (4H, m, H6a'α, H6aα, H6b/H6b'α), 3.8 (3H, s, C**H**$_3$α), 3.8 (0.9H, s, C**H**$_3$β), 3.8-3.7 (3H, m, H5'α, H5α, H6b/H6b'α), 3.6 (2H, m, H3'α), 3.5 (1H, t, *J* = 9.4 Hz, H4α), 3.5-3.4 (2.3H, m, H2'α, H2α, H2'β), 3.4 (0.3H, m, H2β), 3.3 (1H, m, H4'α) ppm;

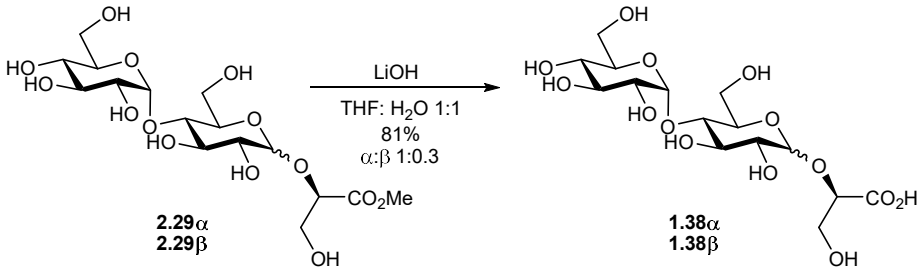
^{13}C NMR (101 MHz, CD_3OD): δ 171.6 (Cq), 170.9 (Cq), 102.6 (C1β), 101.5 (C1'β), 101.5 (C1'α), 97.5 (C1α), 79.9 (C4α), 79.5 (Cβ), 79.0 (Cβ), 76.1 (Cβ), 75.4 (Cβ), 75.4 (C7α), 73.7 (C3'α), 73.7 (Cβ), 73.5 (C3α/C5α/C5'α), 73.4 (C3α/C5α/C5'α), 73.1 (Cβ), 72.8 (C2α/C2'α), 72.7 (Cβ), 71.8 (C2α/C2'α), 71.6, (C5α/C5'α) 70.1 (C4'α), 62.7 (C8α), 61.7 (Cβ), 61.3 (C6α/C6'α), 60.6 (C6α/C6'α), 51.3 (**C**H$_3$β), 51.2 (**C**H$_3$α) ppm;

HR-MS (ESI+) m/z: for $C_{16}H_{28}O_{14}Na$ $[M + Na]^+$ calcd. 467.1377; observed: 467.14;

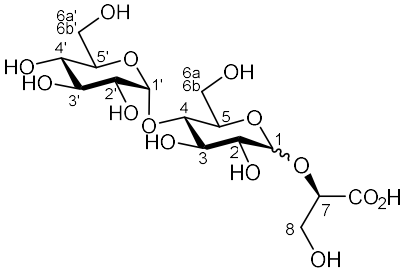
[α]$_D$ (α:β 1:0.7) +102.0 (*c* 0.25, CH_3OH, 21 °C);

IR (cm^{-1}) 3313 (O-H), 2929 (C-H), 1735 (C=O), 1376 (C-H), 1016 (C-O).

8.1.24 (2*R*)-2-*O*-(α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (1.38α/β)



Methyl (2*R*)-2-*O*-(α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (99 mg, 0.22 mmol) was dissolved in a mixture of water and THF (5 mL, 1:1) to which a solution of lithium hydroxide (1 M in H_2O, 300 µL) was added. The reaction was left to stir at room temperature overnight and TLC (7:2:1 EtOAc: MeOH: H_2O) showed complete consumption of the starting material. Dowex-H^+ resin was used to neutralise the solution before being filtered over cotton wool. Solvent was removed *in vacuo* to yield a colourless oil (76 mg, 0.18 mmol, 81%).



^{1}H NMR (400 MHz, D_2O): δ (α:β 1:0.3) 5.4 (1H, d, *J* = 3.8 Hz, H1'α), 5.4 (0.3H, d, *J* = 3.9 Hz, H1'β), 5.0 (1H, d, *J* = 3.8 Hz, H1α), 4.5 (0.3H, d, *J* = 7.8 Hz, H1β), 4.2 (1H, dd, *J* = 3.1, 6.2 Hz, H7α), 4.1 (0.3H, dd, *J* = 3.2, 6.6 Hz, Hβ), 4.1 (1H, m, H3α/H3'α), 3.9 (1H, ddd, *J* = 2.1, 4.4, 10.0 Hz, H5α/H5'α), 3.9-3.6 (12H, m, H3α/H3'α, H6aα, H6a'α, H6bα, H6b'α, H8aα, H8bα, H4α/H4'α), 3.6 (1.3H, dd, *J* = 0.8, 3.9 Hz, H2α/H2'α, Hβ), 3.5 (1H, dd, *J* = 0.7, 3.8 Hz, H2α/H2'α), 3.5 (0.3H, m, Hβ), 3.4 (1.3H, m, Hα, H2β) ppm;

^{13}C NMR (101 MHz, D_2O): δ 177.8 (Cqβ), 177.1 (Cq), 102.2 (Cβ), 99.6 (C1'α), 99.6 (Cβ), 97.3 (C1α), 81.8 (Cβ), 79.2 (C7α), 76.7 (Cα), 76.5 (Cβ), 76.0 (Cβ), 74.5 (Cβ), 73.7 (C3α/C3'α), 72.9 (Cβ), 72.9 (Cα), 72.6 (Cα), 72.6 (Cβ), 71.7 (C2α/C2'α), 71.7 (Cβ), 71.4 (C2α/C2'α), 70.7 (C5α/C5'α), 69.3 (Cα), 63.1 (Cα), 62.6 (Cβ), 60.5 (Cα), 60.4 (Cα) ppm;

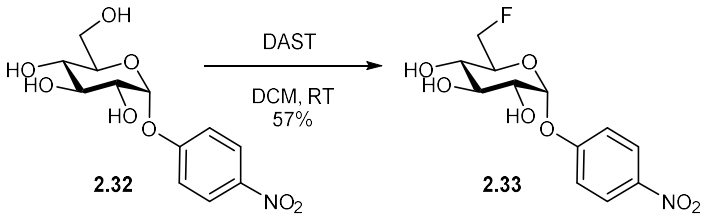
HR-MS (ESI+) m/z: for $C_{15}H_{26}O_{14}Na$ $[M + Na]^+$ calcd. 453.1220; observed: 453.12;

[α]$_D$ (α-anomer only) +98.0 (*c* 0.25, H_2O, 21 °C);

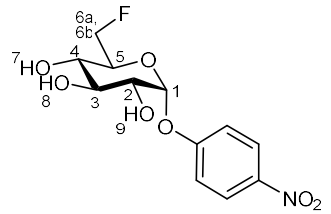
IR (cm^{-1}) 3191 (O-H), 1421 (C-H), 1006 (C-O).

8.1.25 *p*-Nitrophenyl 6-deoxy-6-fluoro-α-D-glucopyranoside (2.33)

(Work carried out by Laura Jowett).



p-nitrophenyl-α-D-glucopyranoside (200 mg, 0.66 mmol) was dissolved in DCM (8.00 mL) and the resulting solution cooled to -40°C. DAST (0.5 mL, 3.78 mmol) was added and the solution left to warm to room temperature. The reaction was stirred for 3 hours before being cooled to -20 °C at which point methanol (8.00 mL) was added. The solvent was removed under reduced pressure and the crude product purified via column chromatography (25% methanol in ethyl acetate) to afford a white solid (115 mg, 57%).



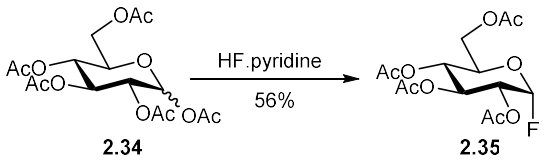
^{1}H NMR (400 MHz, d6-DMSO): δ 8.2 (2H, m, H_{Ar}), 7.3 (2H, m, H_{Ar}), 5.7 (1H, d, *J* = 3.6 Hz, H1), 5.4 (1H, d, *J*= 6.0 Hz, H7), 5.3 (1H, d, *J*= 6.2 Hz, H9), 5.2 (1, d, *J*= 5.1 Hz, H8), 4.5 (2H, m, H6a, H6b), 3.7-3.6 (1H, m, H3), 3.6-3.5 (1H, m, H5), 3.5 (1H, ddd, *J*= 3.7, 6.2, 9.7 Hz, H2), 3.3 (1H, ddd, *J* = 6.0, 8.8, 10.2 Hz, H4) ppm;

^{13}C NMR (101 MHz, d6-DMSO): δ 162.4 (C_{Ar}), 142.2 (C_{Ar}), 126.2 (C_{Ar}), 117.3 (C_{Ar}), 97.9 (C1), 82.7 (C6), 73.2 (C3), 72.6 (C5), 71.5 (C2), 69.0 (C4) ppm;

^{19}F NMR (376 MHz, d6-DMSO): δ -232.2 (1F, td, *J* = 26.9, 48.1, Hz) ppm;

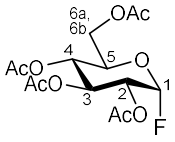
LC-MS (ESI$^-$) m/z: 302.1 $[M - H]^-$, 338.1 $[M + Cl^-]^-$, 348.1 $[M + HCOO^-]^-$, 605.2 $[2M – H]^-$, 651.2 $[2M + HCOO^-]^-$.

8.1.26 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl fluoride (2.35)



2.34 **2.35**

1,2,3,4,6-penta-*O*-acetyl-D-glucopyranose (2.00 g, 5.1 mmol) was dissolved in HF-pyridine (70 %, 15 mL) and the reaction stirred overnight at room temperature. TLC (6:4 hexane: acetone) showed complete consumption of starting material. DCM (15 mL) was added to the reaction mixture and the resulting solution neutralised with saturated aqueous sodium bicarbonate. Product then extracted with DCM (4 x 200 mL) and the organic layers combined and washed with saturated sodium bicarbonate (200 mL) then HCl (2 M, 200 mL) before being dried over $MgSO_4$. Solvent was removed *in vacuo* to give a yellow oil which was purified by column chromatography (30% acetone in hexane) to give a white waxy solid upon drying (1.00 g, 2.86 mmol, 56%).



^{1}H NMR (400 MHz, $CDCl_3$): δ 5.8 (1H, dd, *J* = 2.8, 52.9 Hz, H1), 5.5 (1H, t, *J* = 9.9 Hz, H3), 5.2 (1H, t, *J* = 9.8 Hz, H4), 5.0 (1H, ddd, *J* = 2.8, 10.3, 24.2 Hz, H2), 4.3 (1H, dd, *J* = 3.8, 12.1 Hz, H6a), 4.2 (2H, m, H5, H6b), 2.1 (3H, s, C**H**$_3$), 2.1 (3H, s, C**H**$_3$), 2.1 (3H, s, C**H**$_3$), 2.0 (3H, s, C**H**$_3$) ppm;

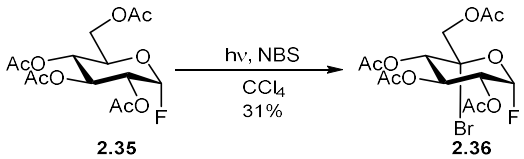
^{13}C NMR (101 MHz, $CDCl_3$): δ 170.5 (Cq), 170.0 (Cq), 170.0 (Cq), 169.4 (Cq), 104.1 (d, C1), 70.2 (d, C2), 69.8 (d, C5), 69.4 (C3), 67.3 (C4), 61.2 (C6), 20.7 (**C**H$_3$), 20.6 (**C**H$_3$), 20.6 (**C**H$_3$ x 2) ppm;

^{19}F NMR (376 MHz, $CDCl_3$): δ -149.8 (1F, dd, *J* = 24.3, 53.8 Hz) ppm;

LR-MS (ESI+) m/z: 368.4 $[M + NH_4]^+$, 373.4 $[M + Na]^+$, 389.3 $[M + K]^+$.

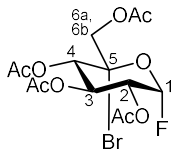
Data matches that of literature.[234]

8.1.27 2,3,4,6-tetra-*O*-acetyl-5-bromo-α-D-glucopyranosyl fluoride (2.36)



2.35 **2.36**

2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl fluoride (507 mg, 1.45 mmol) was dissolved in CCl_4 and N-bromosuccinimide (1.33g, 7.47 mmol) was added. The reaction was then refluxed using a 400 W tungsten bulb. After 9 hours the reaction was allowed to cool and then quenched with saturated sodium thiosulphate solution. Reaction mixture was filtered then organic layer washed with water (200 mL) then saturated sodium bicarbonate (200 mL) before being dried over $MgSO_4$. The solvent was removed under reduced pressure and the crude material purified via column chromatography (1 % ethyl acetate in DCM) to give yellow foam (196 mg, 0.45 mmol, 31%).



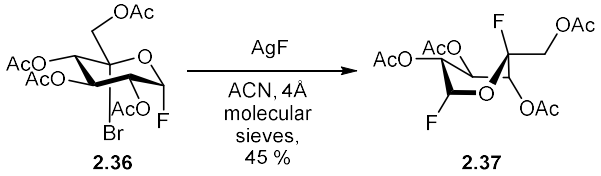
^{1}H NMR (400 MHz, CD_3OD): δ 6.0 (1H, dd, *J* = 3.3, 52.9 Hz, H1), 5.8 (1H, t, *J* = 10.27 Hz, H3), 5.3 (1H, d, *J* = 10.0 Hz, H4), 5.2 (1H, ddd, *J* = 3.4, 10.5, 24.0 Hz, H2), 4.4 (2H, dd, *J* = 12.4, 29.2 Hz, H6a+H6b), 2.1 (3H, s, C**H**$_3$), 2.1 (3H, s, C**H**$_3$), 2.1 (3H, s, C**H**$_3$), 2.0 (3H, s, C**H**$_3$) ppm;

^{13}C NMR (101 MHz, CD_3OD): δ 169.8 (Cq), 169.7 (Cq), 169.2 (Cq), 104.4 (d, C1), 92.9, 69.4, 69.2, 68.0, 66.9, 66.2, 19.0 (**C**H$_3$), 19.0 (**C**H$_3$), 18.8 (**C**H$_3$) ppm;

^{19}F NMR (376 MHz, CD_3OD): δ -147.0 (1F, dd, *J* = 24.3, 53.8 Hz) ppm.

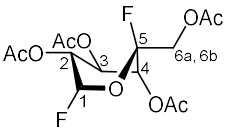
Data matches that of literature.[234]

8.1.28 2,3,4,6-tetra-*O*-acetyl-5-fluoro-β-L-idopyranosyl fluoride (2.37)



2.36 **2.37**

2,3,4,6-tetra-*O*-acetyl-5-bromo-α-D-glucopyranosyl fluoride (85 mg, 0.2 mmol) was dissolved in anhydrous acetonitrile (3 mL) and AgF (37 mg, 0.28 mmol) was added. The reaction was stirred in the dark over 4 Å molecular sieves overnight at room temperature. Reaction was filtered over celite which was washed with DCM. The organic filtrate was then washed with water (10 mL) then brine (10 mL) before being dried over $MgSO_4$. The solvent was removed under reduced pressure to give a yellow oil which was then purified by column chromatography (50% diethyl ether in petroleum ether) to yield a colourless oil (33 mg, 0.09 mmol, 45%).



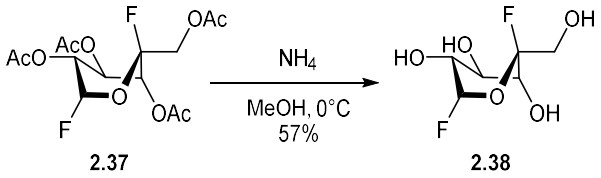
^{1}H NMR (400 MHz, $CDCl_3$): δ 5.8 (1H, dd, *J* = 2.1, 58.2 Hz, H1), 5.4 (1H, dddd, *J* = 1.3, 1.8, 8.4, 24.8 Hz, H2), 5.3 (1H, m, H3), 5.2 (1H, dd, *J* = 1.7, 5.2 Hz, H4), 4.5 (1H, dd, *J* = 12.5, 24.3 Hz, H6a), 4.2 (1H, t, *J* = 12.7 Hz, H6b), 2.1 (3H, s, C**H**$_3$), 2.1 (6H, s, C**H**$_3$ x2), 2.1 (3H, s, C**H**$_3$) ppm;

^{13}C NMR (101 MHz, $CDCl_3$): δ 169.9 (Cq), 169.8 (Cq), 169.4 (Cq), 168.6 (Cq), 110.0 (d, C1), 103.8, 101.5, 69.4, 69.3, 69.3, 69.0, 68.3 (m), 62.4, 62.2, 20.7 (**C**H$_3$), 20.6 (**C**H$_3$), 20.5 (**C**H$_3$) ppm;

^{19}F NMR (376 MHz, $CDCl_3$): δ -106.3 (1F, m, F5), -138.6 (1F, ddd, *J* = 20.8, 24.3, 57.2 Hz, F1) ppm.

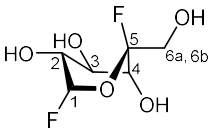
Data matches that of literature.[234]

8.1.29 5-fluoro-β-L-idopyranosyl fluoride (2.38)



2.37 2.38

2,3,4,6-tetra-*O*-acetyl-5-fluoro-β-L-idopyranosyl fluoride (86 mg, 0.23 mmol) was dissolved in a solution of ammonia in methanol (7 N, 10 mL) at 0°C and the reaction stirred for 3 hours when TLC (9:1 ethyl acetate: hexane) showed complete consumption of starting material. The solvent was then removed under reduced pressure and the resulting yellow oil purified by column chromatography (27:2:1 ethyl acetate: methanol: water) to give a yellow oil (25 mg, 0.13 mmol, 57%).



^{1}H NMR (400 MHz, CD_3OD): δ 5.7 (1H, dd, *J* = 1.7, 56.0 Hz, H1), 3.9 (1H, dd, *J* = 12.7, 17.6 Hz), 3.8 (3H, m), 3.7 (1H, ddd, *J* = 0.6, 12.6, 22.6 Hz) ppm;

^{19}F NMR (376 MHz, CD_3OD): δ -116.8 (1F, m, F5), -143.9 (1F, ddd, *J* = 13.9, 22.5, 57.2 Hz, F1) ppm.

Data matched that of literature.[234]

8.2 Materials and methods for biological experiments

Luria Broth (Miller's) for culture media and the Invitrogen™ ChargeSwitch® -Pro plasmid mini prep kit were purchased from Invitrogen. IPTG, kanamycin, agarose, ampicillin and Tris-HCl were all purchased from Melford Laboratories Ltd. Pre-made SDS-PAGE (4-20%) gels were bought in from Bio-Rad, as was concentrated (5x) Bradford reagent. The pre-made SDS running buffer used in SDS-PAGE was bought as a 20x solution from Novex and the stain used for visualisation was Instant Blue, purchased from Expedeon. Pre-packed nickel columns were purchased from GE Healthcare. Pierce™ BCA assay kit and 6x DNA loading dye were purchased from Thermo Scientific. All other buffer components and reagents were purchased from Sigma-Aldrich®, Fisher Scientific or Acros Organics.

The bacterial strains used as chemically competent cells were either *E. coli* BL21 (DE3) purchased from Invitrogen Life Technologies or *E. coli* JM109 which were obtained from the lab of Prof Peter Roach at the University of Southampton. Both strains of competent cells were then made in house from these stocks.

Equipment used: ÄKTAPrime liquid chromatography system, POLARstar® Omega micro plate reader (BMG Labtech), Tecan infinite M200 Pro, A Bio-Rad Mini-PROTEAN® Tetra handcast system for SDS-PAGE, BioRad Gel Doc™ EZ imager, A Bio-Rad Mini-Sub® Cell GT Cell for agarose gels, BioMate™ 3 spectrometer, Agilent Technologies Cary Series UV-Vis spectrophotometer, MSE Soniprep 150.

A Sorvall Evolution RC was used for large scale or high speed centrifugation with SLA-1500, SLA-6000 and SS-34 rotors with all runs kept at 4°C, a VWR® microstar 17 centrifuge was used for samples of 1 mL or less and all other sample sizes were centrifuged in a Heraeus™ Contrifuge™ Stratos™.

8.2.1 General Laboratory Protocols

8.2.1.1 Plasmid purification

A ChargeSwitch®-Pro plasmid mini-prep kit (Invitrogen) was used to isolate and purify plasmids. When the kit was first used, RNase A was added to the resuspension buffer which was subsequently stored at 4 °C. Cells from 10 mL overnight culture were harvested by centrifugation (8000 rpm, 4 °C, 30 min) and resuspended in resuspension

buffer (250 µL) *via* pipetting. The centrifugation protocol contained within the kit was then followed.

8.2.1.2 Agarose Gel Electrophoresis

1% agarose gels were prepared by adding agarose (0.6 g) to 1x TAE buffer (See section 8.2.6.1, 60 mL) and heating *via* microwave until the agarose fully dissolved. Once the solution had cooled enough to handle, Nancy-520 (3 µL) was added and gently mixed without the introduction of bubbles. The gel was poured into the cast and bubbles were removed using a pipette tip. A comb was then added and the gel left to set at 4 °C.

Samples were prepared by adding 5 µL DNA sample to 5 µL sterile water and 2 µL 6x DNA loading dye (Thermo Scientific). Samples (5 µL) were then loaded into the gel which was covered with 1 x TAE buffer before being run at 90 V for 40-60 minutes. Gels were then imaged using a Bio-Rad Gel Doc™ EZ imager.

8.2.1.3 Chemically competent cells

LB media (10 mL) was inoculated with stock competent cells (Prof Roach group, University of Southampton or Invitrogen Life Technologies) using a loop and incubated at 37 °C, 180 rpm overnight. LB media (100 mL) was inoculated with overnight culture (1.0 mL) and incubated at 37 °C, 180 rpm until the $O.D_{600}$ reached 0.6. The cells were then harvested by centrifugation (4000 rpm, 4 °C, 10 min) and the supernatant discarded. The cell pellet was resuspended by pipetting into 40 mL calcium chloride buffer (see section 8.2.6.2) and incubated on ice for 30 minutes. Cells were harvested once more (4000 rpm, 4 °C, 10 min) and re-suspended in cold calcium chloride buffer (10 mL). Sterile 75% glycerol solution (2.0 mL) was then added and mixed *via* pipetting before the cells were divided into 150 µL aliquots and stored at -80 °C.

8.2.1.4 Heat shock transformation

The relevant aliquots of competent cells were thawed over ice and re-suspended gently. Plasmid solution (2-5 µL) was then transferred and gently mixed with the pipette before being incubated on ice for 30 minutes. The solution was incubated in a water bath (42 °C, 35 seconds) and immediately placed back on ice for 5 minutes incubation. SOC media (see section 8.2.6.3, 250µL) was then added and the resulting

suspension gently mixed before being incubated for 1 hour at 37 °C whilst shaking at 180 rpm. Once the cells had been incubated, 50 µL were plated onto agar plates containing the appropriate antibiotics (see section 8.2.6.4) and incubated upside down overnight at 37 °C.

8.2.1.5 Overnight bacterial growths

A loop of either the relevant glycerol stock of bacteria or of bacteria from a colony grown on an agar plate was used to inoculate LB (see section 8.2.6.6) or 2x YT (see section 8.2.6.7) media (10 mL for small scale or 50 mL for large scale) containing the appropriate antibiotics. The cultures were incubated at 37 °C overnight whilst shaking at 180 rpm.

Table 8.1 concentrations of antibiotics used in overnight cultures

Antibiotic composition	**Concentrations of antibiotics in media (unless otherwise stated) / µg/mL**
Kanamycin	50
Ampicillin	100
Chloramphenicol	30
Kanamycin + Chloramphenicol	30 (Kan), 20 (Chlor)

8.2.1.6 Bacterial glycerol stocks

Relevant cells from an overnight culture (500 µL) were mixed gently and thoroughly with sterile 75% glycerol solution (500 µL) ensuring no distinct layers were visible. Stocks were then flash frozen and stored at -80 °C.

8.2.1.7 Protein expression

Fresh overnight culture was used to inoculate a larger volume (1.0 L, 5.0 L or 8.0 L) of culture media containing the relevant antibiotics and/or supplements (1% inoculate used). The cells were then grown at 37 °C, 180 rpm until the desired $O.D_{600}$ was met. At this point, the relevant inducing agent was added to the desired concentration and protein expression allowed to occur at the desired temperature and time period for

the experiment. Once expression was complete, cells were harvested *via* centrifugation (8,000 rpm, 4 °C, 30 min).

8.2.1.8 Cell Lysis *via* Sonication

Cell pellets were re-suspended in the appropriate volume of lysis buffer (4 x mass of pellet) (for lysis buffer see section 8.2.6.8). Sonication was then carried out with the sample cooled in ice for the desired amount of time. Cycles of intermittent sonication and rest periods were determined depending upon the volume of the sample. Samples less than or equal to 1.0 mL were sonicated for 10 cycles of 1 s on, 1 s rest. Samples greater in volume than this were sonicated for 6 cycles of 90 s on, 60 s rest. Once lysis was complete, the cell debris was pelleted *via* centrifugation (12,000 rpm, 4 °C, 30 min) leaving behind the soluble cell lysate.

8.2.1.9 SDS-PAGE

Where pre-packed gels were not used, SDS-PAGE gels were prepared as per the following:

Table 8.2 Composition of 12.5% and 7.5% acrylamide resolving gels and stacking gel

Component	12.5 % Acrylamide Resolving Gel	Stacking Gel
30 % Acrylamide /Bis-Acrylamide / mL	3.15	0.33
4x Appropriate Gel Buffer / mL	1.88	0.62
Sterile Water / mL	2.4	1.55
10% Ammonium per sulphate (APS) / µL	75	15

Table 8.3 Components of 4x Resolving and 4x Stacking Gel Buffers

Component (in 100 mL Sterile Water)	4x Resolving Gel Buffer	4x Stacking Gel Buffer
Tris-Base / g	18.17	6.06
Sodium Dodecyl Sulphate (SDS) / g	0.4	0.4
Tetramethylethylenediamine (TEMED) / g	0.4	0.4
Adjust with HCl to pH:	8.8	6.8

During hand casting, 10% APS was added to the resolving gel before it was poured first and allowed to set. During setting, the resolving gel was topped with isopropanol to prevent drying. This was followed by pouring the stacking gel (following addition of 10% APS), after which the relevant size comb (10 or 15 well) was placed in the mould before the stacking gel was allowed to set. Once set, either the comb was removed and the gels rinsed with deionised water before use or wrapped in damp tissue and stored at 4 °C for no longer than 3 days.

Samples for SDS-PAGE were prepared by adding either a pre-determined (by Bradford assay) or equal volume of 2x gel loading dye (section 8.2.6.11) and heating to 95 °C for 15 minutes. The samples were then loaded onto the relevant gel and run at 200 V for 30-40 minutes in 1x running buffer (see section 8.2.6.10). Once run, the gels were removed, rinsed with deionised water and stained with Expedeon Instant Blue stain which was left to cover the gel for 1 hour before rinsing with deionised water.

8.2.1.10 Ni affinity Fast Protein Liquid Chromatography

Ni affinity FPLC was carried out using either a HisTrap™ HP histidine-tagged protein purification column (GE Healthcare) (5.0 mL) or a manually pre-packed column consisting of Ni-NTA resin and an ÄKTA Prime liquid chromatography system. The column was initially washed with binding buffer (buffer A, section 8.2.6.9) to remove any traces of storage solution (20 % ethanol). The cell lysate was then loaded onto the column and anything that did not bind was collected in fractions as the flow through. The column was then allowed to equilibrate before the bound proteins were eluted with elution buffer (buffer B, section 8.2.6.9) over a gradient of 0-100 %. Fractions were collected during elution and were later analysed using SDS-PAGE. The column was washed with elution buffer (5 x column volumes) to ensure no protein remained followed by 20% ethanol (5 x column volumes) before being disconnected and stored at 4 °C.

8.2.1.11 Bradford Assay

5x Bradford reagent was diluted to 1x with deionised water. A calibration curve was set using solutions of known concentrations of bovine serum albumin (BSA) in sterile water (concentrations used were 0 mg/mL, 0.125 mg/mL, 0.25 mg/mL, 0.5 mg/mL and 1.0 mg/mL).

When cuvettes were used with a UV-spectrophotometer, samples were prepared by adding 20 µL sample to 1x Bradford reagent (1.00 mL) and mixed *via* gentle pipetting. In the case of a 96-well plate in a plate reader, samples were prepared by adding 10 µL sample to 200 µL 1x Bradford reagent and mixed *via* gentle pipetting taking care not to introduce bubbles. In each case, a blank using the relevant buffer was also prepared. The absorbance was then measured at 595 nm and compared to the calibration curve.

GraphPad Prism 7.0 was used to create calibration curve, line of best fit and to interpolate unknown concentrations.

8.2.1.12 Concentration of protein solutions

Amicon® spin filtration (used in expressions below 5.0 L in volume):

Amicon® Ultra – 15 (Merk) centrifugal units (10 kDa MWCO) were used at 8,000 rpm at 4 °C in bursts of 10 minutes. Between each 10-minute cycle, the protein solution was mixed to prevent protein aggregation at the filter. Once the desired volume had been reached, protein solution was analysed *via* Bradford assay and stored in aliquots at -80 °C.

Amicon® stirred pressure cell filtration units (used in expression greater than 5.0 L in volume):

Amicon® stirred pressure cell filtration units were used at 4 °C with a 10 kDa MWCO PES filter (Sartorius) and kept under pressure using nitrogen gas. Samples were stirred throughout the entire process and once the desired volume was met, the solution was analysed *via* Bradford assay and stored in aliquots at -80 °C.

8.2.1.13 Dialysis of protein solutions

Dialysis was carried out using dialysis tubing with a 10 kDa MWCO (Sigma Aldrich). Tubing was first wet in the dialysis buffer before being secured at one end with a clip. Dialysis buffer was then used to check for leaks. Protein solution was then carefully added *via* a syringe and needle or, for smaller volumes, a 1000 µL pipette. Air bubbles were removed with care taken not to lose protein solution. A clip was used to secure the other end of the tubing which was then suspended in dialysis buffer by floats. The volume of buffer used was kept to at least 50 x that of the volume of protein solution.

Dialysis was generally carried out at 4 °C for 3 hours, with buffer being changed after the first 1.5 hours. For volumes under 1.0 mL, dialysis cassettes with 10 kDa MWCO were used instead of tubing (Spectra-Por® Float-A-Lyzer® G2 1.0 mL).

8.2.2 Experimental for Chapter 3

8.2.2.1 Digestion of pET28::*Rv3031* by NdeI and XhoI

The plasmid pET28a::*Rv3031* was subjected to a double digest with the restriction enzymes NdeI and XhoI. The reaction composition was as per table 8.4:

Table 8.4 composition of double digest reaction

Reaction:	***Negative control***	***NdeI only***	***XhoI only***	***Double digest***
Component	**Volume used /µL**			
Plasmid solution	5	5	5	5
10 x cut smart buffer (NEB)	1	1	1	1
NdeI (NEB)	0	1	0	1
XhoI (NEB)	0	0	1	1
Deionised water	4	3	3	2

Digests were carried out at 37 °C for 18 hours after which the crude product mixes were analysed *via* agarose gel electrophoresis.

8.2.2.2 Initial expression studies of MtGBE

The plasmid pET28a::*Rv3031* was transformed into *E. coli* BL21(DE3) chemically competent cells and successfully grown colonies were resuspended in 3 x 50 mL LB media containing 50 µg/mL Kan (one colony per flask). Once grown overnight at 37 °C, 180 rpm, a portion of each culture (5.0 mL) was used to inoculate 6 x 500 mL LB media containing Kan (50 µg/mL). The large scale growth was grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.6 was reached. At this point, protein expression was induced using IPTG (for concentrations used, see table 8.5).

Table 8.5 Concentrations of IPTG used in initial expression studies of MtGBE

Experiment	Final concentration of IPTG used /mM
JG-7399-35_A	0.1
JG-7399-35_B	0.5
JG-7399-35_C	1.0

Protein expression was carried out at 18 °C overnight as well as at 37 °C overnight. Cells were then harvested (8000 rpm, 4 °C, 30 min), lysed and analysed using SDS-PAGE.

8.2.2.3 Initial expression and purification of MtGBE

The plasmid pET28a::*Rv3031* was transformed into *E. coli* BL21(DE3) chemically competent cells and successfully grown colonies were resuspended in 100 mL 2 x YT media containing 50 µg/mL Kan. Once grown overnight at 37 °C, 180 rpm, a portion of overnight culture (10 mL) was used to inoculate 1.0 L LB media containing Kan (50 µg/mL). The large scale growth was grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.7 was reached. At this point, protein expression was induced using IPTG (final concentration of 0.1 mM) and was carried out at 18 °C overnight.

Cells were then harvested (8000 rpm, 4 °C, 30 min), lysed and the resulting lysate purified *via* FPLC using the Tris-HCl binding and elution buffers described in 8.2.6.9. During purification, resin-bound protein was eluted over a gradient of 0-100% elution buffer. Collected fractions were then analysed *via* SDS-PAGE.

8.2.2.4 Solubilisation of MtGBE from inclusion bodies

MtGBE protein was expressed as per procedure 8.2.2.3 using 1.0 mM IPTG. Expression was carried out at 37 °C overnight. 2.0% sarkosyl solution (final concentration 0.2%, see section 8.2.6.12) was either used to resuspend pelleted cells before cell lysis and sonication, or to resuspend the insoluble cell debris obtained after lysis and centrifugation. During the latter experiment, resuspension was carried out *via* stirring at 4 °C overnight. A control experiment in which no sarkosyl was added at all was also carried out. Samples from each experiment were then analysed using SDS-PAGE.

8.2.2.5 Co-transformation of pGro7 and pET28a::*Rv3031*

Method 1:

2.0 µL each of each plasmid were mixed with 150 µL chemically competent *E. coli* BL21(DE3) cells. Transformation was then performed as per the heat shock technique described in section 8.2.1.4. Transformants were then incubated on agar plates containing a mix of Kan (30 µg/mL) and Chlor (20 µg/mL) at 37 °C overnight.

Method 2:

pGro7 was transformed into BL21(DE3) cells as described in section 8.2.1.4 resulting in BL21 pGro7 cells. One colony of these cells was used to inoculate 10 mL LB media containing 30 µg/mL Chlor which was then incubated at 37 °C overnight. The culture was then used to make chemically competent cells as per procedure 8.2.1.3.

The BL21 pGro7 chemically competent cells were then transformed with the pET28a(+)::*Rv3031* plasmid *via* the heat shock method. As above, successful transformants were selected by incubation at 37 °C overnight on agar plates containing a mix of the antibiotics Kan (30 µg/mL) and Chlor (20 µg/mL).

8.2.2.6 Small scale co-expression and purification of GroEL-GroES chaperone system with MtGBE

One colony of BL21 pGro7,pET28a(+)::*Rv3031* was used to inoculate 100 mL LB containing Chlor (25 µg/mL) and Kan (30 µg/mL). The cells were grown overnight at 37 °C, 180 rpm. 10 mL of this overnight culture was used to inoculate 1.0 L of LB containing Chlor (25 µg/mL), Kan (30 µg/mL) and L-Ara (500 µg/mL). Cells were grown at 37 °C, 180 rpm to an $O.D_{600}$ of 0.6 at which point IPTG (final concentration 0.5 mM) was added. Expression was carried out at 18 °C, 180 rpm overnight. Cells were then harvested by centrifugation (8000 rpm, 4 °C, 30 min) before being lysed and analysed *via* SDS-PAGE.

8.2.2.7 Co-expression experiments with varying L-arabinose concentrations

One colony of BL21 pGro7,pET28a(+)::*Rv3031* was used to inoculate 5.0 mL LB containing Chlor (20 µg/mL) and Kan (30 µg/mL). The cells were grown overnight at 37 °C, 180 rpm. 1.0 mL of this overnight culture was used to inoculate 3 x 100 mL of LB

containing Chlor (25 µg/mL), Kan (30 µg/mL) and L-Ara (see table 8.6). Cells were grown at 37 °C, 180 rpm to an $O.D_{600}$ of 0.5 at which point IPTG (final concentration 0.5 mM) was added. Expression was carried out at 18 °C, 180 rpm overnight. Cells were then harvested by centrifugation (8000 rpm, 4 °C, 30 min) before being lysed and analysed *via* SDS-PAGE.

Table 8.6 Concentrations of L-Ara used in co-expression studies of MtGBE and GroEL-GroES.

Experiment	Final concentration of L-Ara used / µg/mL	Mass of cell pellet / g
Ara_600	600	0.47
Ara_700	700	0.52
Ara_800	800	0.47

8.2.2.8 Large scale co-expression and purification of MtGBE with GroEL-GroES chaperones

One colony of BL21 pGro7, pET28a(+)::*Rv3031* was used to inoculate 100 mL LB containing Chlor (25 µg/mL), Kan (30 µg/mL) and 800 µg/mL L-Ara. The cells were grown overnight at 37 °C, 180 rpm. 50 mL of this overnight culture was used to inoculate 5.0 L of LB containing Chlor (25 µg/mL), Kan (30 µg/mL) and L-Ara (800 µg/mL). Cells were grown at 37 °C, 170 rpm to an $O.D_{600}$ of 0.5 at which point IPTG (final concentration 0.5 mM) was added. Expression was carried out at 18 °C, 170 rpm overnight. Cells were then harvested by centrifugation (8000 rpm, 4 °C, 30 min) before being lysed and purified *via* FPLC (Tris-HCl binding and elution buffers from 8.2.6.9). Fractions containing protein were pooled and dialysed against 50 mM HEPES pH 7.4 before being concentrated to 4.0 mg/mL. The solution was then divided into aliquots to be stored at -80 °C.

8.2.2.9 Experiments to determine whether GroEL is co-eluted with MtGBE during FPLC

Heat shock transformation / co-transformation was carried out as per procedures 8.2.1.4 and 8.2.2.5 (method 1) respectively with the following:

Table 8. 7 Conditions used in heat shock transformation of various experiments to test effect of chaperone proteins

Experiment	Plasmid used	ABs in agar plates
MtGBE only	pET28a(+)::*Rv3031*	Kan – 50 µg/mL
GroEL-GroES only	pGro7	Chlor – 30 µg/mL
Co-expression	pGro7 + pET28a(+)::*Rv3031*	Kan – 30 µg/mL, Chlor – 25 µg/mL

For each experiment, one colony of the relevant bacteria was used to inoculate 100 mL LB containing the appropriate antibiotics (see table 8.8). The cells were grown overnight at 37 °C, 180 rpm. 20 mL of this overnight culture was used to inoculate 2.0 L of LB containing the appropriate antibiotics and L-Ara (see table 8.8). Cells were grown at 37 °C, 170 rpm to an $O.D_{600}$ of 0.5 at which point IPTG (final concentration 0.5 mM) was added. Expression was carried out at 18 °C, 170 rpm overnight. Cells were then harvested by centrifugation (8000 rpm, 4 °C, 30 min) before being lysed and analysed using SDS-PAGE. The lysate from each experiment was then purified *via* Ni-NTA FPLC (Tris-HCl binding and elution buffers from 8.2.6.9). Fractions containing protein were pooled and dialysed against either 50 mM HEPES pH 7.4 (for assays), 20 mM ammonium formate (for MS samples) or 20 mM potassium phosphate pH 7.0 (CD samples) before being concentrated or diluted to the desired concentration.

Table 8. 8 Conditions used in overnight cultures and protein expressions in various experiments carried out to show effects of chaperone proteins

Experiment	ABs in overnight culture (μg/mL)	large scale culture contained (μg/mL)
MtGBE only	Kan (30)	Kan (50) L-Ara (600)
GroEL-GroES only	Chlor (20)	Chlor (25) L-Ara (600)
Co-expression	Kan (30) Chlor (20)	Kan (30) Chlor (20) L-Ara (600)
BL21 cells (control)	none	none

8.2.2.10 LC-MS of MtGBE and protein fractions from chaperone expression and purification

Samples were prepared in ammonium formate buffer (20 mM). Samples were diluted with buffer to give a final concentration of 50 μg/mL. ESI-MS was performed using a WATERS ZMD single quadrapole system and Bruker Apex III FT-ICR-MS spectrometers.

8.2.2.11 Circular dichroism experiments of MtGBE and protein fractions from chaperone expression and purification

Samples were prepared in potassium phosphate buffer at pH 7.0 (see section 8.2.6.16). Samples were then diluted with buffer and the final concentrations of each sample (determined using Bradford assay) are given in table 8.9.

Table 8. 9 Concentration of protein samples submitted for CD analysis

Experiment	Concentration of sample / mg/mL
MtGBE only	0.1
GroEL-GroES (fraction 56)	0.5
GroEL-GroES (fraction 57)	0.5
Co-expression	0.5

CD experiments were carried out on a Jasco J720 CD Spectrophotometer covering the spectral range 180 - 800nm.

8.2.3 Experimental for Chapter 4

8.2.3.1 Absorbance spectrum of *p*-nitrophenol at varying pH levels

p-nitrophenol (139 mg, 1.0 mmol) was dissolved in 20 mM sodium phosphate buffer (section 8.2.6.16) of the relevant pH (1.0 mL) to give a final concentration of 1.0 M. Each solution was further diluted with the relevant buffer to give a 0.1 mM solution. 200 µL of each solution were pipetted into wells on a 96-well plate (Greiner, flat-bottomed, colourless, full volume) and an absorbance scan measured from 230 – 1000 nm in a Tecan infinite M200 Pro plate reader. pH levels tested were: 6.5, 7.0, 7.5 and 8.5.

8.2.3.2 Absorbance spectrum of *p*-nitrophenyl-α-D-glucopyranoside

p-nitrophenol-α-D-glucopyranoside (30 mg, 0.1 mmol) was dissolved in 20 mM sodium phosphate buffer (section 8.2.6.16) of the relevant pH (1.0 mL) to give a final concentration of 0.1 M. Each solution was further diluted with the relevant buffer to give a 0.1 mM solution. 200 µL of each solution were pipetted into wells on a 96-well plate (Greiner, flat-bottomed, colourless, full volume) and an absorbance scan measured from 230 – 1000 nm in a Tecan infinite M200 Pro plate reader. pH levels tested were: 6.5, 7.0, 7.5 and 8.5.

8.2.3.3 *p*-nitrophenol standard curve for absorbance assays

p-nitrophenol (139 mg, 1.0 mmol) was dissolved in deionised water (50.0 mL) to give a 20 mM solution. This was further diluted in water to give a 2.0 mM solution. The 2.0

mM pNP solution was mixed with an equal volume of 100 mM HEPES pH 7 containing 2.0 M KCl to give a 1 mM sample. This was then used in sequential dilutions to make the following standards: 1.0, 0.5, 0.25, 0.125, 0.0625, 0.03125, 0.01563, 0.007813 mM.

8.2.3.4 General method for MtGBE absorbance assays with *p*-nitrophenyl-α-D-glucopyranoside

Plate reader:

For absorbance assays, all components except for the enzyme were mixed together beforehand and left to warm to room temperature. Assays were initiated upon the addition of the enzyme and were mixed in the plate reader using orbital shaking (unless otherwise stated). Reactions were carried out in a Greiner 96-well plate (flat-bottom, clear, full volume) at 37 °C. The absorbance reading at 400 nm was taken at regular intervals for the time specified for each experiment in the main text (Chapter 4).

UV-vis spectrophotometer:

For absorbance assays, all components except for the enzyme were mixed together beforehand and left to warm to room temperature. Assays were initiated upon the addition of the enzyme and were gently mixed before readings taken. Reactions were carried out at 37 °C in a 10.00 mm quartz cuvette (Hellma®) to which enzyme was added *via* syringe. Absorbance was measured continuously at 400 nm over the time specified in the main text (Chapter 4). The cuvette was rinsed with water and acetone and dried with air between each experiment.

General composition of experiments:

Table 8.10 shows the general composition of the absorbance experiments, changes to the concentration, pH or presence of certain components are detailed in the main text (Chapter 4).

Table 8.10 General composition of absorbance assay reactions

Component	concentration / mM
HEPES buffer (pH 7.0)*	50.0
KCl	200.0
pNPGlc	1.0
MtGBE	1.0 mg/mL
Glucosyl glycerate (not always present)	1.0

* For assays in which pH was varied, MES buffer was used for pH levels below 6.8.

8.2.3.5 General procedure for iodine staining assays

Initial activity test:

All reaction components except for enzyme were pre-mixed and allowed to warm to room temperature. Reactions were initiated on addition of enzyme and incubated at 37 °C for 10 mins. Assay reaction mixtures contained 25 µL enzyme solution, 50 µL sample solution and 25 µL glucosyl glycerate (4.0 mM). Substrate solution was prepared using 100 µL of substrate stock (amylopectin and maltodextrin – 2.5% in DMSO, glycogen – 5% in DMSO, amylose - 1.2 % in DMSO), 200 µL HEPES (0.5 M, pH 7.0) and 700 µL H_2O. Enzyme solutions were prepared by diluting stock solution to appropriate concentrations using sterile water (stocks made: 8.0, 4.0, 2.0, 1.0, 0.5 mg/mL). Reactions were stopped by addition of 400 µL of 0.4 mM HCl and 400 µL iodine staining solution (see section 8.2.6.13) was added. Absorbance was measured at 530 nm for amylopectin, 395 nm for glycogen and maltodextrin and at 660 nm for amylose.

Table 8.11 shows the general composition of the iodine staining experiments, changes to the concentration, pH or presence of certain components are detailed in the main text (Chapter 4).

Table 8. 11 General composition of iodine staining assays

Assay component	Final concentration
HEPES (pH 7.0)	50.0 mM
DMSO	5.0%
MtGBE	2.0, 1.0, 0.5, 0.25, 0.125, 0 mg/mL
Glucosyl glycerate (not always present)	1 mM
Polysaccharide:	
Amylopectin	0.125 %
Glycogen	0.25 %
Maltodextrin	0.125 %
Amylose	0.06 %

End point assays:

Enzyme reaction mixtures contained 25 µL enzyme solution (1 mg/mL), 25 µL of either glucosyl glycerate (4.0 mM) or deionised H_2O and the relevant volume of substrate solution. Substrate solution was prepared using 100 µL of substrate stock (amylopectin and maltodextrin – 2.5% in DMSO, glycogen – 5% in DMSO), 200 µL HEPES (0.5 M, pH 7.0) and 700 µL H_2O. Enzyme solutions were prepared by diluting stock solution to 1.0 mg/mL using sterile water. Reaction mixtures were made up to 100 µL using deionised water if necessary. All assay components except for the enzyme solution were pre-mixed and incubated at 37 °C for 5 minutes. Enzyme solution was used to initiate the reaction which was then incubated for 10 minutes at 37 °C. At relevant time points, 10 µL aliquots were taken and added to 40 µL of 0.4 mM HCl. Iodine staining solution (40 µL) was then added and the absorbance measured at the relevant wavelength (as above).

Table 8.12 shows the general composition of the end point assays, changes to the concentration, pH or presence of certain components are detailed in the main text (Chapter 4).

Table 8. 12 composition of end point iodine staining assays

Assay component	Final concentration
HEPES (pH 7.0)	50.0 mM
DMSO	5.0%
MtGBE	0.25 mg/mL
Glucosyl glycerate (not always present)	1.0 mM
Polysaccharide (excluding amylose):	0.0625, 0.125, 0.25, 0.5 %

8.2.3.6 Maltose standard curve with BCA reagent

Maltotriose standards were made up in deionised water. Stock solutions of 500 µM, 400 µM and 300 µM were initially prepared with the 400 µM solution being used in further dilutions. The concentrations tested were: 500, 400, 300, 200, 100, 50, 25 and 10 µM. Standards (20 µL) were mixed with BCA reagent (120 µL) and heated to 70 °C for 30 minutes. After cooling to room temperature, the absorbance of each sample at 560 nm was measured. GraphPad Prism 7.0 was used to generate the standard curve and interpolate samples of unknown concentration.

8.2.3.7 General procedure for isoamylase branching assays

Assay reaction mixtures contained 25 µL enzyme solution, 50 µL sample solution and 25 µL glucosyl glycerate (4.0 mM). Substrate solution was prepared using 100 µL of substrate stock (amylopectin and maltodextrin – 2.5% in DMSO, glycogen – 5% in DMSO, amylose - 1.2 % in DMSO), 200 µL HEPES (0.5 M, pH 7.0) and 700 µL H_2O. Enzyme solutions were prepared by diluting stock solution to 1.0 mg/mL using sterile water. All components except for the enzyme solution were pre-mixed and allowed to warm to room temperature before the initiation of the reaction with the addition of the enzyme solution. Reactions were incubated at 37 °C for 25 minutes. Every 5 minutes, an aliquot (40 µL) was taken of the reaction which was heated to 99 °C for 5 minutes to denature the enzyme. 15 µL of the aliquot were used to determine reducing end concentration before isoamylase treatment. The remaining 25 µL of

aliquot was mixed with 75 µL debranching reaction master mix (table 8.14). Debranching was carried out at 40 °C for 20 hours after which point the reaction mixture was used to determine the number of reducing ends after debranching. Values were multiplied by 4 to account for the dilution of the sample when added to the debranching mixture.

To determine the number of reducing ends, sample (15 µL) was mixed with BCA reagent (90 µL, see section 8.2.6.14), heated to 70 °C for 30 minutes and cooled to room temperature before the absorbance was read at 560 nm.

Table 8.13 shows the general composition of the branching experiments, changes to the concentration, pH or presence of certain components are detailed in the main text (Chapter 4).

Table 8. 13 general composition of iodine branching assays (before debranching reaction)

Assay component	Final concentration
HEPES (pH 7.0)	50.0 mM
DMSO	5.0%
MtGBE	0.25 mg/mL
Glucosyl glycerate	1 mM
Polysaccharide:	
Amylopectin	0.5 %
Amylose	0.25 %

Table 8. 14 composition of de-branching reaction master mix

Component	Volume / µL
Isoamylase (500 U/mL)	0.35
25 mM Sodium phosphate pH 8	64.65
1.0 M Sodium acetate pH 3.8	10

8.2.3.8 Phosphate standard curve with malachite green dye

Solutions of potassium phosphate (mono basic) were made up in either deionised water or HEPES (50 mM, pH 7.0) to the following concentrations: 10, 20, 40, 50, 60, 80, 100 µM. Each phosphate solution (50 µL) was added to malachite green dye (50 µL, see section 8.2.6.18), mixed *via* pipetting and incubated at room temperature for 5 minutes. Absorbance scans from 600-750 nm were then taken of each solution. For each sample, scans were performed at 0, 10 and 20 minutes after incubation. The absorbance maxima were then plotted against concentration to create a standard curve, analysed in GraphPad Prism 7.0.

8.2.3.9 General procedure for malachite green assay

All assay components (see table 8.15) except for the enzyme were pre-mixed before initiation of the reaction was carried out upon addition of the enzyme. The reaction mixture was then incubated at 37 °C over 60 minutes, during which time aliquots (50 µL) were taken at 15, 30 and 60 minutes and added to the malachite green dye (50 µL, see section 8.2.6.18). The dyed sample solutions were left to incubate for 5 minutes before absorbance spectra were taken 3 times, each 2 minutes apart. Absorbance data was collected between 600-750 nm. The absorbance maxima at 630 nm was then plotted as a function of time using GraphPad Prism 7.0.

Table 8.15 shows the general composition of the malachite green assay, changes to the concentration, pH or presence of certain components are detailed in the main text (Chapter 4).

Table 8. 15 general composition of malachite green assays

Assay component	**Final concentration / mM**
HEPES pH 7.0	50.0
Glucose-1-phosphate	1.0
Glucosyl glycerate	1.0
MtGBE	1.0 mg/mL

8.2.3.10 TLC screening of potential glucose donors

All reaction components were mixed together and incubated at 37 °C for 6-18 hours. Substrates tested and the composition of the various reactions are summarised in tables 8.16 – 8.18:

Table 8. 16 general composition of substrate screening reactions - maltooligosaccharides

Assay component	Final concentration / mM
HEPES (pH 7.0)	25.0
KCl	50.0
MtGBE	0.5 mg/mL
Glucosyl glycerate	1.0
Potential substrate:	
Maltose	1.0
Maltotriose	1.0
Maltotetraose	1.0
Maltoheptaose	1.0

Table 8. 17 general composition of substrate screening reactions – melibiose and pullulan

Assay component	Final concentration / mM
HEPES (pH 7.0)	50.0
KCl (not always present)	100.0
MtGBE	0.1 mg/mL
Glucosyl glycerate	1.0
Potential substrate:	
Melibiose	1.0 mg/mL
Pullulan	1.0 mg/mL

Table 8. 18 general composition of substrate screening reactions – maltosyl glycerate

Assay component	Final concentration / mM
HEPES (pH 7.4)	25.0
MtGBE	1.0 mg/mL
Maltosyl glycerate	1.0

Following incubation of the reaction mixtures, crude samples were loaded onto aluminium foil supported silica plates (MERCK 60 F254). Samples of starting materials and blank reaction mixtures (no enzyme present) were also loaded onto the plates. Eluent details and visualisation techniques are described in table 8.19:

Table 8. 19 Eluent and visualisation techniques for each TLC experiment described

Donor being tested	Eluent	Stain used before heating
Maltooligosaccharides	ACN: EA: iPA: H_2O (85:20:50:50)	20% H_2SO_4 in ethanol
Melibiose and pullulan	iPA: acetone: H_2O (2:2:1)	20% H_2SO_4 in ethanol
Maltosyl glycerate	ACN: EA: iPA: H_2O (85:20:50:50)	5% H_2SO_4 + *N*-(1-naphthyl)ethylenediamine (0.3% w/v) in methanol

8.2.3.11 Site directed mutagenesis to create MtGBE mutants

The composition of the PCR reaction mixture is shown in table 8.20 (any modifications to the composition are detailed in chapter 4):

Table 8.20 general composition of PCR reactions

Component	Volume / µL
H_2O	38.5
10x buffer	5
dNTP mix	4
Forward primer (100 pmol)	0.5
Reverse primer (100 pmol)	0.5
Template	1
Taq polymerase	0.5

Unless otherwise stated, the PCR reaction cycle was as follows:

Table 8.21 PCR cycle used in SDM

Stage	Temperature / °C	Time of stage / min
1	95	5.0
2 (x18 cycles)	95	1.0
	55	1.5
	68	15.0
3	68	5.0

After completion of stage 3, reaction mixtures were held at 4 °C.

DpnI digestion reactions had the following composition (enzyme and buffer used were from the FastDigest DpnI kit from ThermoFisher):

Table 8.22 general composition of DpnI digestions

Component	Volume / µL
PCR product	50
10 x digestion buffer	6
DpnI enzyme	1
Sterile water	3

Digestion was carried out at 37 °C for 2 hours before products were analysed using agarose gel electrophoresis.

8.2.4 Experimental for Chapter 5

8.2.4.1 Initial expression studies on MtGlcT

The plasmid pET28a::*Rv3032* was transformed into *E. coli* BL21(DE3) chemically competent cells and successfully grown colonies were resuspended in 10 mL LB media containing 50 µg/mL Kan. Once grown overnight at 37 °C, 180 rpm, a portion of overnight culture (1.0 mL) was used to inoculate 5 x 100 mL LB media containing Kan (50 µg/mL). The large scale growth was grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.5 was reached. At this point, protein expression was induced using the relevant amount of IPTG (see table 8.23) and was carried out at either 37 or 18 °C for either 4 hours or overnight.

Table 8.23 Conditions used in expression studies of MtGlcT

Experiment	[IPTG] used to induce expression / mM	Temperature of expression /°C	Time of expression / h
1	1.0	18	18
2	0.5	18	18
3	0.1	18	18
4	1.0	37	18
5	1.0	37	4

Cells were then harvested (8000 rpm, 4 °C, 30 min), lysed and the resulting samples analysed *via* SDS-PAGE.

8.2.4.2 Studies on co-expression of MtGlcT with the GroEL-GroES chaperone system

The plasmids pET28a::*Rv3032* and pGro7 were co-transformed into *E. coli* BL21(DE3) chemically competent cells as per co-transformation method 2 described in section 8.2.2.5. Successfully grown colonies were resuspended in 5.0 mL LB media containing Kan (30 μg/mL) and Chlor (20 μg/mL). Once grown overnight at 37 °C, 180 rpm, a portion of overnight culture (1.0 mL) was used to inoculate 4 x 100 mL LB media containing Kan (30 μg/mL), Chlor (20 μg/mL) and L-Ara. The large scale growths were grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.5 was reached. At this point, IPTG (1.0 mM) was added, and the cultures grown either at 18 °C or 37 °C for either 4 or 18 hours. Cells were then harvested (8000 rpm, 4 °C, 30 min), lysed and the resulting samples analysed *via* SDS-PAGE.

8.2.4.3 Co-expression and purification of MtGlcT with the trigger factor chaperone protein

Co-expression of TF and MtGlcT:

The plasmids pET28a::*Rv3032* and pTf16 were co-transformed into *E. coli* BL21(DE3) chemically competent cells as per co-transformation method 1 described in section 8.2.2.5. Successfully grown colonies were resuspended in 100 mL LB media containing Kan (30 μg/mL), Chlor (20 μg/mL) and L-Ara (600 μg/mL). Once grown overnight at 37 °C, 180 rpm, a portion of overnight culture (10.0 mL) was used to inoculate 1.0 L LB media containing Kan (30 μg/mL), Chlor (20 μg/mL) and L-Ara (600 μg/mL). The large scale growth was grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.5 was reached. At this point, IPTG (0.5 mM) was added, and the culture grown at 18 °C for 18 hours. Cells were then harvested (8000 rpm, 4 °C, 30 min), lysed (see section 8.2.1.8) and the resulting samples analysed *via* SDS-PAGE. The resulting lysate was then purified by Ni-affinity FPLC using the Tris-HCl FPLC buffers as per section 8.2.6.9.

Expression of TF only:

The pTf16 plasmid was transformed into *E. coli* BL21(DE3) chemically competent cells as per the heat shock method described in section 8.2.1.4. Transformants were grown

on plates containing Chlor (20 μg/mL). One successful colony was used to inoculate 5.0 mL LB media containing Chlor (20 μg/mL) and L-Ara (600 μg/mL). Once grown overnight at 37 °C, 180 rpm, a portion of overnight culture (1.0 mL) was used to inoculate 100 mL LB media containing Chlor (20 μg/mL) and L-Ara (600 μg/mL). The large scale growth was grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.5 was reached. At this point, IPTG (0.5 mM) was added, and the culture grown at 18 °C for 18 hours. Cells were then harvested (8000 rpm, 4 °C, 30 min), lysed and the resulting samples analysed *via* SDS-PAGE.

8.2.4.4 Large scale expression and attempted purification of MtGlcT

The plasmid pET28a::*Rv3032* was transformed into *E. coli* BL21(DE3) chemically competent cells and successfully grown colonies were resuspended in 100 mL LB media containing 50 μg/mL Kan. Once grown overnight at 37 °C, 180 rpm, a portion of overnight culture (80 mL) was used to inoculate 8.0 L LB media containing Kan (50 μg/mL). The large scale growth was grown at 37 °C, 170 rpm until an $O.D_{600}$ of 0.5 was reached. At this point, protein expression was induced using IPTG (0.5 mM) and was carried out at 18 °C overnight. Cells were then harvested (8000 rpm, 4 °C, 30 min) and lysed before purification was carried out as per section 8.2.1.10 using the Tris-HCl FPLC buffers as per section 8.2.6.9. Fractions containing protein were pooled and dialysed against 50 mM HEPES pH 7.4.

8.2.4.5 Concentration of solubilised MtGlcT protein

The work carried out by Dr. Gu Yoo included the inoculation of an overnight culture of LB broth containing Kan (50μg/mL) with a colony of *E. coli* BL21 pET28a::*Rv3032*. The culture was grown overnight at 37 °C and a 1% inoculate was added to a large-scale growth. Expression was carried out at 37 °C for 4 hours. Once cells were harvested (8,000 rpm, 4 °C, 30 min), they were lysed and the insoluble cell debris was solubilised using buffer containing 0.2% sarkosyl (4 °C, 18 hours). Dialysis was then performed to remove the sarkosyl from the solubilised protein solution. Samples were then analysed using SDS-PAGE. This concludes the work carried out by Dr. Gu Yoo on this experiment.

The solution containing the re-solubilised MtGlcT was initially centrifuged to remove precipitate (8,000 rpm, 4 °C, 30 min) then analysed by Bradford assay and deemed to contain 2.0 mg/mL protein. The supernatant was then concentrated using Amicon®

Ultra – 15 (Merk) centrifugal units (10 kDa MWCO). The final volume of solution was 1.0 mL and was found to have a concentration of 40 mg/mL, as determined using the Bradford assay. The concentrated solution was then aliquoted and stored at -80 °C.

8.2.4.6 PK/LDH coupled transferase assay

Prior to assays being performed, a 5x master mix of assay components excluding the enzyme and potential acceptor was prepared. The composition of this 5x master mix is detailed in table 8.24.

Table 8.24 composition of LDH/PK 5 x assay master mix

Component	concentration / mM
UDP-Glc	5.0
PEP	5.0
NADH	1.5
PK	≈ 85 U/mL
LDH	≈ 120 U/mL
HEPES pH 7.0	100

All reaction components except for the enzyme were warmed to room temperature before initiation of the reaction upon the addition of the enzyme. Reaction mixtures contained 5 x master mix (40 µL), acceptor solution (50 µL) and enzyme solution (50 µL). Enzyme solution was prepared by diluting the stock volume with deionised water to give a final concentration of 4.0 mg/mL. Substrate solutions were prepared at the following concentrations; monosaccharides 4.0 mM, oligosaccharides 4.0 mM, amylose 0.12% amylopectin 1.0 % and glycogen 0.25 %. Polysaccharides were added as solutions in DMSO. Reactions were carried out at 37 °C over a period of 60 minutes. Assays were carried out in Greiner 96 well plates (flat-bottom, clear, full volume) and the absorbance at 340 nm was measured over regular periods of 2 minutes. Negative controls were carried out as above without the addition of MtGlcT. Data were plotted and analysed using GraphPad Prism 7.0.

8.2.4.7 Expression and purification of OtsA

The plasmid containing the *OtsA* gene was transformed into *E. coli* BL21(DE3) chemically competent cells and successfully grown colonies were resuspended in 50 mL LB media containing 100 µg/mL Amp. Once grown overnight at 37 °C, 180 rpm, a portion of overnight culture (10 mL) was used to inoculate 1.0 L LB media containing Amp (100 µg/mL). The large scale growth was grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.5 was reached. At this point, protein expression was induced using IPTG (0.1 mM) and was carried out at 30 °C, 170 rpm overnight. Cells were harvested (8,000 rpm, 4 °C, 30 min) and lysed as per section 8.2.1.8. Ni-affinity FPLC purification was then carried out using the HEPES buffers described in section 8.2.6.9. Fractions containing protein were pooled and dialysed against 5.0 mM HEPES pH 7.0, 60 mM $MgCl_2$ before being concentrated to 3.5 mg/mL. Aliquots were then stored at -80 °C.

8.2.4.8 PK/LDH assay with OtsA

Prior to assays being performed, a 5x master mix of assay components excluding the enzyme and potential acceptor was prepared. The composition of this 5x master mix is detailed in table 8.24, section 8.2.4.6. All reaction components except for the enzyme were warmed to room temperature before initiation of the reaction upon the addition of the enzyme. Reaction mixtures contained KCl (1.0 M, 40µL), $MgCl_2$ (1.0 M, 2.0 µL), BSA (10%, 2.0 µL), 5 x master mix (40 µL), G6P (50 mM, 4.0 µL) and enzyme solution (6.0 µL). The final volume of the reaction mixture was made up to 200 µL using deionised water. Enzyme was used as stored, at a concentration of 3.5 mg/mL. Reactions were carried out at 37 °C over a period of 60 minutes. Assays were carried out in Greiner 96 well plates (flat-bottom, clear, full volume) and the absorbance at 340 nm was measured over regular periods of 2 minutes. Negative control reactions were carried out in which no G6P was present. Data were plotted and analysed using GraphPad Prism 7.0.

8.2.5 Experimental for Chapter 6

8.2.5.1 Initial expression of MtGpgS

The plasmid pET29a(+)::*Rv1208* (Prof. Marcelo E. Guerin, Ikerbasque, Blibao) was transformed into *E. coli* BL21(DE3) chemically competent cells and successfully grown colonies were resuspended in 100 mL 2 x YT media containing 50 µg/mL Kan. Once

grown overnight at 37 °C, 180 rpm, a portion of overnight culture (10 mL) was used to inoculate 1.0 L 2 x YT media containing Kan (50 μg/mL). The large scale growth was grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.5 was reached. At this point, protein expression was induced using IPTG (0.5 mM) and was carried out at 37 °C, 180 rpm for 4 hours. Cells were pelleted (8,000 rpm, 4 °C, 30 min) and lysed before being analysed using SDS-PAGE.

8.2.5.2 Studies on the co-expression of MtGpgS with the GroEL-GroES chaperone system and the TF chaperone protein

The pET29a(+)::*Rv1208* plasmid and either the pGro7 or the pTf16 plasmid were co-transformed into *E. coli* BL21(DE3) chemically competent cells as per co-transformation method 1 in section 8.2.2.5. For each co-transformation experiment, a successful colony was used to inoculate 5.0 mL 2x YT media containing Kan (30 μg/mL) and Chlor (20 μg/mL). The cells were grown overnight at 37 °C, 180 rpm before 0.15 mL were used to inoculate 15 mL 2x YT media containing Kan (30 μg/mL), Chlor (20 μg/mL) and L-Ara (100 μg/mL). These larger cultures were grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.5 was reached. At that point, IPTG was added at a final concentration of 1.0 mM. Protein expression was then carried out either at 37 °C for 4 hours or 18 °C for 18 hours. Cells were harvested (8,000 rpm, 4 °C, 30 min) and lysed before being analysed *via* SDS-PAGE.

8.2.5.3 Co-expression of MtGpgS with the GroEL-GroES chaperone system and subsequent purification of MtGpgS

The pET29a(+)::*Rv1208* plasmid and pGro7 plasmid were co-transformed into *E. coli* BL21(DE3) chemically competent cells as per co-transformation method 1 in section 8.2.2.5. One resultant colony was used to inoculate 50 mL 2 x YT media containing Kan (30 μg/mL) and Chlor (20 μg/mL). The cells were grown overnight at 37 °C, 180 rpm before 10 mL were used to inoculate 1.0 L 2x YT media containing Kan (30 μg/mL), Chlor (20 μg/mL) and L-Ara (100 μg/mL). The larger culture was grown at 37 °C, 180 rpm until an $O.D_{600}$ of 0.5 was reached. At that point, IPTG was added at a final concentration of 1.0 mM. Protein expression was then carried out at 18 °C for 18 hours. Cells were harvested (8,000 rpm, 4 °C, 30 min) and lysed. Purification was then carried out as per section 8.2.1.10 using the HEPES buffers described in section 8.2.6.9.

Fractions containing protein were pooled and dialysed against 10 mM Tris-HCl pH 8.0, 0.5 M NaCl before being concentrated to 5.0 mg/mL. Solution was then aliquoted and stored at -80 °C.

8.2.5.4 PK/LDH coupled assay with MtGpgS

Prior to assays being performed, a 5x master mix of assay components excluding the enzyme and potential acceptor was prepared. The composition of this 5x master mix is detailed in table 8.24, section 8.2.4.6. All reaction components except for the enzyme were warmed to room temperature before initiation of the reaction upon the addition of the enzyme. Reaction mixtures contained $MgCl_2$ (1.0 M, 2.0 µL), 5 x master mix (40 µL), 3-PGA (50 mM, 4 µL) and enzyme solution (4 µL). Enzyme was used as stored, at a concentration of 5.0 mg/mL. Reaction volumes were made up to 200 µL with deionised water. Reactions were carried out at 37 °C over a period of 60 minutes. Assays were carried out in Greiner 96 well plates (flat-bottom, clear, full volume) and the absorbance at 340 nm was measured over regular periods of 2 minutes. Negative control reactions were carried out in which no 3-PGA was present. Data were plotted and analysed using GraphPad Prism 7.0.

8.2.5.5 Scaled up enzymatic reaction with MtGpgS

Reactions were carried out in 1.5 mL microcentrifuge tubes and total reaction volume was 500 µL. Each reaction was incubated at 37 °C for 18 hours and the crude products were then visualised using TLC. The reaction mixture was compared to the two substrates as well as to a control reaction in which no 3-PGA was present. TLC experiments were run in an eluent of ACN: EtOAc: ^{i}PA: H_2O in a ratio of 85:20:50:50. TLC analyses were run in duplicate and visualised with either $KMnO_4$ stain or 10 % H_2SO_4 in MeOH. The composition of the reaction mixture is given in table 8.25.

Table 8.25 composition of scaled up enzymatic synthesis of GPG

Component	concentration / mM
UDP-Glc	10.0
3-PGA	10.0
$MgCl_2$	20.0
HEPES pH 8.0	20.0
MtGpgS	0.1 mg/mL

8.2.6 Composition of buffers and other solutions

8.2.6.1 Tris-acetate-EDTA (TAE) buffer for agarose gel electrophoresis

Before being used to make agarose gels or as tank buffer, 50 x TAE solution was diluted to 1 x using deionised water.

Table 8.26 Composition of 50 x TAE buffer

Component	Amount in 1 L made up with deionised water
Tris-base	242 g
Glacial acetic acid	57.1 mL
0.5 M EDTA (pH 8.0)	100 mL

8.2.6.2 Calcium Chloride buffer used in preparation of chemically competent cells

Table 8.27 Composition of calcium chloride buffer used to make chemically competent cells

Component	Mass added to 100 mL water / mg	Final concentration / mM
$CaCl_2$	555	50
Tris-HCl pH 7.4	158	10

8.2.6.3 Super optimal broth with catabolite repression (SOC) media

Table 8.28 Composition of SOC media

Component	Volume added to total / mL
LB media	50
1 M $MgCl_2$	0.50
2 M $MgSO_4$	0.50
Glucose (20% w/v)	0.50

Once components were mixed, SOC media was filtered through a sterile syringe filter and divided into 3.0 mL aliquots before being stored at 4 ˚C.

8.2.6.4 LB-Agar plates

10 g of LB media and 6 g of agar were suspended in 400 mL deionised water before being sterilised *via* heating to 121 ˚C for 15 minutes in an autoclave. The solution was then allowed to cool before the appropriate volume of stock antibiotic solution was added and the plates poured. Once set, plates were wrapped with laboratory film and stored upside down at 4 ˚C.

Table 8.29 Amounts of antibiotics used in the preparation of agar plates

Antibiotic (stock concentration / mg/mL)	Volume added to 400 mL agar solution/ µL	Final Concentration / µg/mL
Kanamycin (50)	400	50
Ampicillin (100)	400	100
Chloramphenicol (30)	400	30
Kan (50) + Chlor (30) mix	Kan: 240 Chlor: 267	Kan: 30 Chlor: 20

8.2.6.5 Stock antibiotic solutions

Table 8.30 Composition of stock antibiotic solutions

Antibiotic	Amount to add to 10 mL water / mg	Stock concentration / mg/mL
Kanamycin	500	50
Ampicillin	1000	100
Chloramphenicol*	300*	30

*Chloramphenicol solutions prepared in ethanol

Once antibiotic solutions were prepared, they were then sterilised by being filtered through a sterile syringe filter. Solutions were stored in 1.0 mL aliquots at -20 ˚C.

8.2.6.6 Luria Broth (LB) media

LB powder was dissolved in deionised water (25 g/L) before being sterilised *via* heating to 121 ˚C for 15 minutes in an autoclave. Relevant antibiotic / inducer solutions were added once media had cooled below 55 °C.

Table 8.31 Amount of each component in 25 g LB (Miller) media

Component	Mass / g
Peptone 140	10.0
NaCl	10.0
Yeast extract	5.0

8.2.6.7 2 x YT media

The components of 2 x YT media were dissolved in water before being sterilised *via* heating to 121 ˚C for 15 minutes in an autoclave. Relevant antibiotic / inducer solutions were added once media had cooled below 55 °C.

Table 8.32 composition of 2 x YT media in 1 L

Component	Mass / g
Tryptone	16.0
NaCl	5.0
Yeast extract	10.0

8.2.6.8 Lysis buffer

Lysis buffer consisted of Ni affinity FPLC binding buffer (volume equal to 3 x the mass of cell pellet) to which was added lysozyme (1-5 mg), DNAseI (1-5 mg) and 1 cOmplete™ protease inhibitor cocktail tablet (Roche).

8.2.6.9 Ni-NTA FPLC buffers

Binding buffers:

Table 8.33 Composition of Tris based binding buffer (A buffer)

Component	Mass in 1 L water/ g	Final concentration / mM
Tris HCl	3.15	20
NaCl	29.2	500
Imidazole	0.34	5.0

Table 8.34 Composition of HEPES based binding buffer (A buffer)

Component	Mass in 1 L water /g	Final concentration / mM
HEPES	4.77	20
NaCl	29.2	500
Imidazole	0.34	5.0

Once buffer components are dissolved in most of the water, the pH is adjusted to 7.4 using HCl / NaOH. The solution is then made up to volume using deionised water.

Elution buffers:

Table 8.35 Composition of Tris based elution buffer (B buffer)

Component	Mass in 1 L water / g	Final concentration / mM
Tris HCl	3.15	20
NaCl	29.2	500
Imidazole	34	500

Table 8.36 Composition of HEPES based elution buffer (B buffer)

Component	Mass in 1 L water / g	Final concentration / mM
HEPES	4.77	20
NaCl	29.2	500
Imidazole	34	500

Once buffer components are dissolved in most of the water, the pH is adjusted to 7.4 using HCl / NaOH. The solution is then made up to volume using deionised water.

8.2.6.10 SDS-PAGE running buffer

Table 8.37 Composition of 5 x SDS running buffer

Component	Amount in 1 L*	Final concentration
Tris-base	15.1 g	0.12 M
Glycine	94 g	1.25 M
SDS (20% w/v)	25 mL	0.5 % w/v

*Total volume made up to 1.0 L using deionised water.

8.2.6.11 SDS-PAGE 2x loading dye

Upon use, 2x loading dye is diluted to 1x with the sample being analysed.

Table 8.38 composition of 2x SDS-PAGE loading dye

Component	Stock solution	Volume added* / mL	Final concentration
Tris HCl	1.0 M	2.5	100 mM
SDS	20% w/v	5.0	4%
Bromophenol Blue	1% w/v	1.25	0.05%
Glycerol	50% v/v	10	20%
DTT	1.0 M	1.25	50 mM

*Total volume made up to 25 mL with deionised water.

8.2.6.12 Sarkosyl solubilisation buffer

Solubilisation buffer consisted of 20 mM HEPES, pH 7.4, lysozyme (1-5 mg), DNAseI (1-5 mg) and 1 cOmplete™ protease inhibitor cocktail tablet (Roche) with the addition of 2.0 % sarkosyl solution in deionised water (final concentration 0.2%).

8.2.6.13 Iodine staining solution

Stock iodine/KI solution:

I_2 (260 mg, 2.04 mmol) and KI (2.6 g, 15.66 mmol) were dissolved in 10 mL deionised water. Solution was stored at 4 °C.

Freshly prepared solution used for assays:

I_2/KI stock (25 µL) (or 50 µL when used with glycogen or maltodextrin) was added to HCl (1.0 M, 100 µL) and the solution made up to 13.0 mL with deionised water. This solution was prepared fresh each day an assay was carried out.

8.2.6.14 BCA solution

BCA solution used was from the Pierce™ BCA protein assay kit (Thermo Scientific™).

BCA reagent A contained: sodium carbonate, sodium bicarbonate, bicinchoninic acid and sodium tartrate in sodium hydroxide (0.1 M).

BCA solution B contained: copper (II) sulphate (4%).

Before use in assay, BCA solution was prepared by mixing BCA reagent A and BCA reagent B in a 50:1 ratio as per kit instructions.

8.2.6.15 Malachite green dye solution

Malachite green carbinol base (13.5 mg, 0.04 mmol) was dissolved in deionised water (30 mL) to which HCl (4 M, 1.0 mL) was added. A blue green solution was formed. A solution of ammonium molybdate tetrahydrate (420 mg, 0.34 mmol) in HCl (4 M, 10 mL) was added slowly to the malachite green solution which then turned yellow once all molybdate was added. The resulting solution was stirred at room temperature for 30 minutes before any precipitate was removed *via* filtration. The dye solution was stored at -80 °C in 1.0 mL aliquots. Upon defrosting and prior to use, Triton-X100 (1.0%, 20 µL) was added and thoroughly mixed.

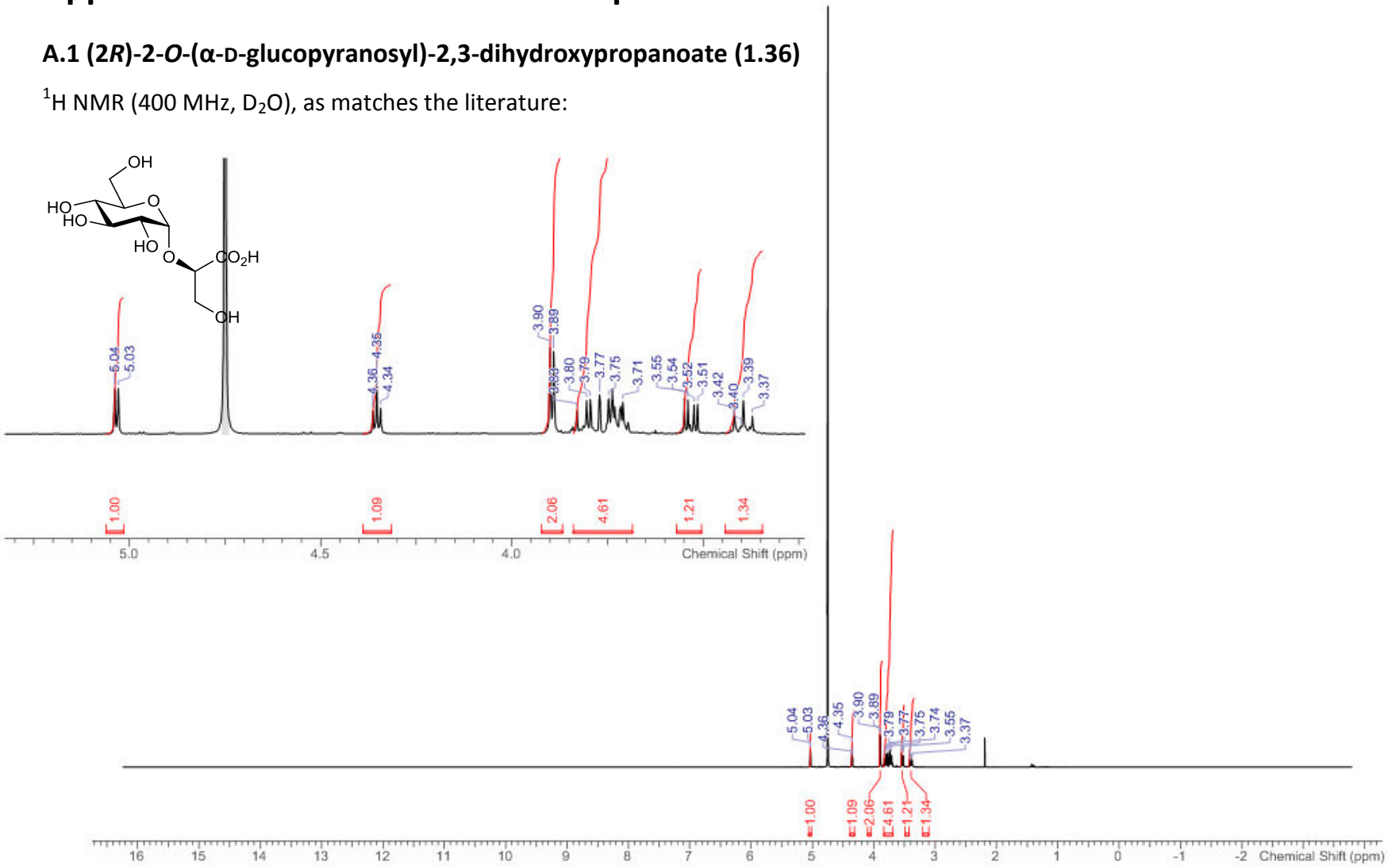
8.2.6.16 Preparation of phosphate buffer

Phosphate buffers of a desired pH were made by mixing solutions of monobasic potassium/sodium phosphate (0.2 M) and dibasic potassium/sodium phosphate (0.2 M) at a certain ratio. Ratios were taken from the buffer reference centre on the Sigma Aldrich website: (https://www.sigmaaldrich.com/life-science/core-bioreagents/biological-buffers/learning-center/buffer-reference-center.html#phosphate). If necessary, the final solution was diluted to the desired concentration.

Appendix A: Additional data for Chapter 2

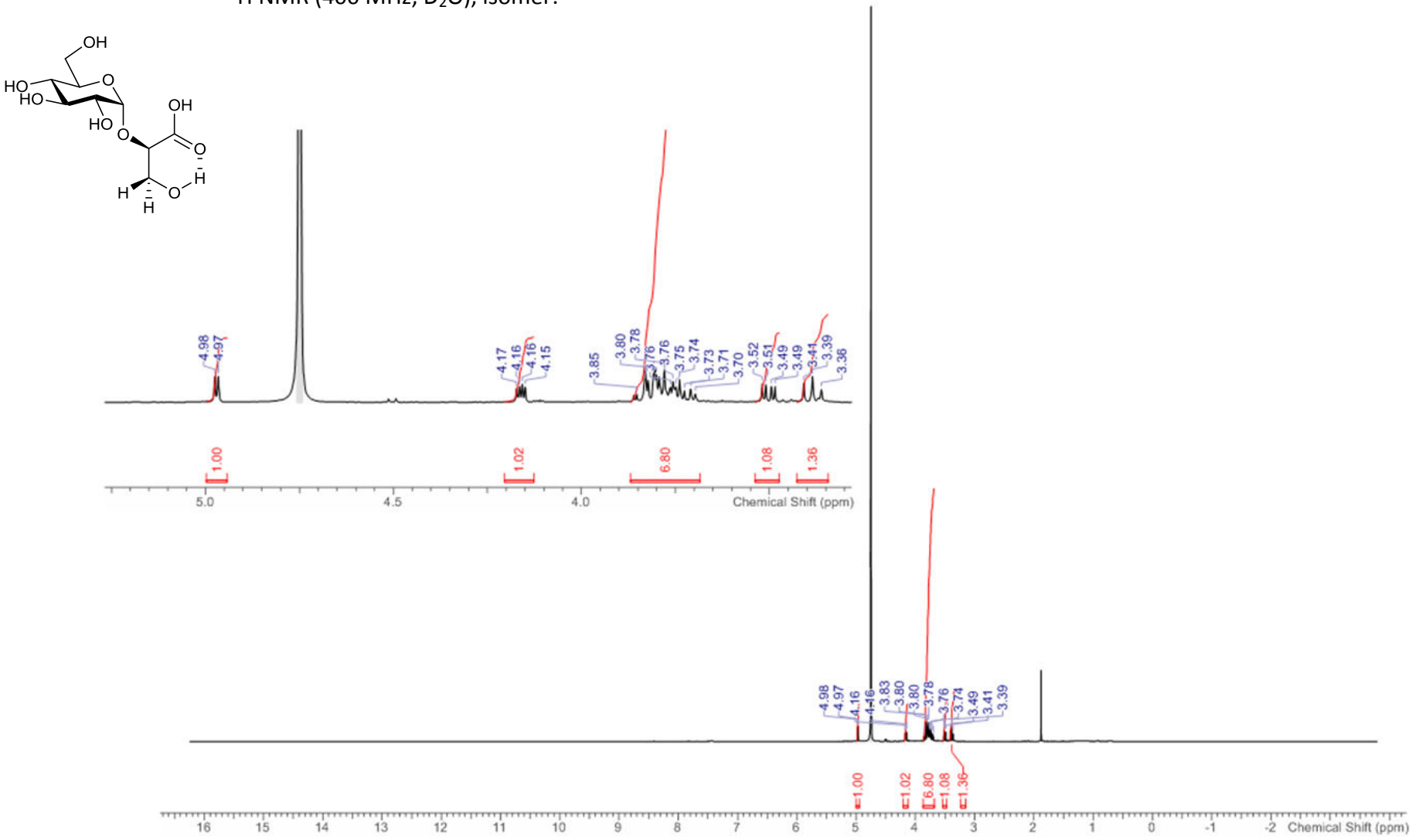
A.1 (2*R*)-2-*O*-(α-D-glucopyranosyl)-2,3-dihydroxypropanoate (1.36)

^{1}H NMR (400 MHz, D_2O), as matches the literature:



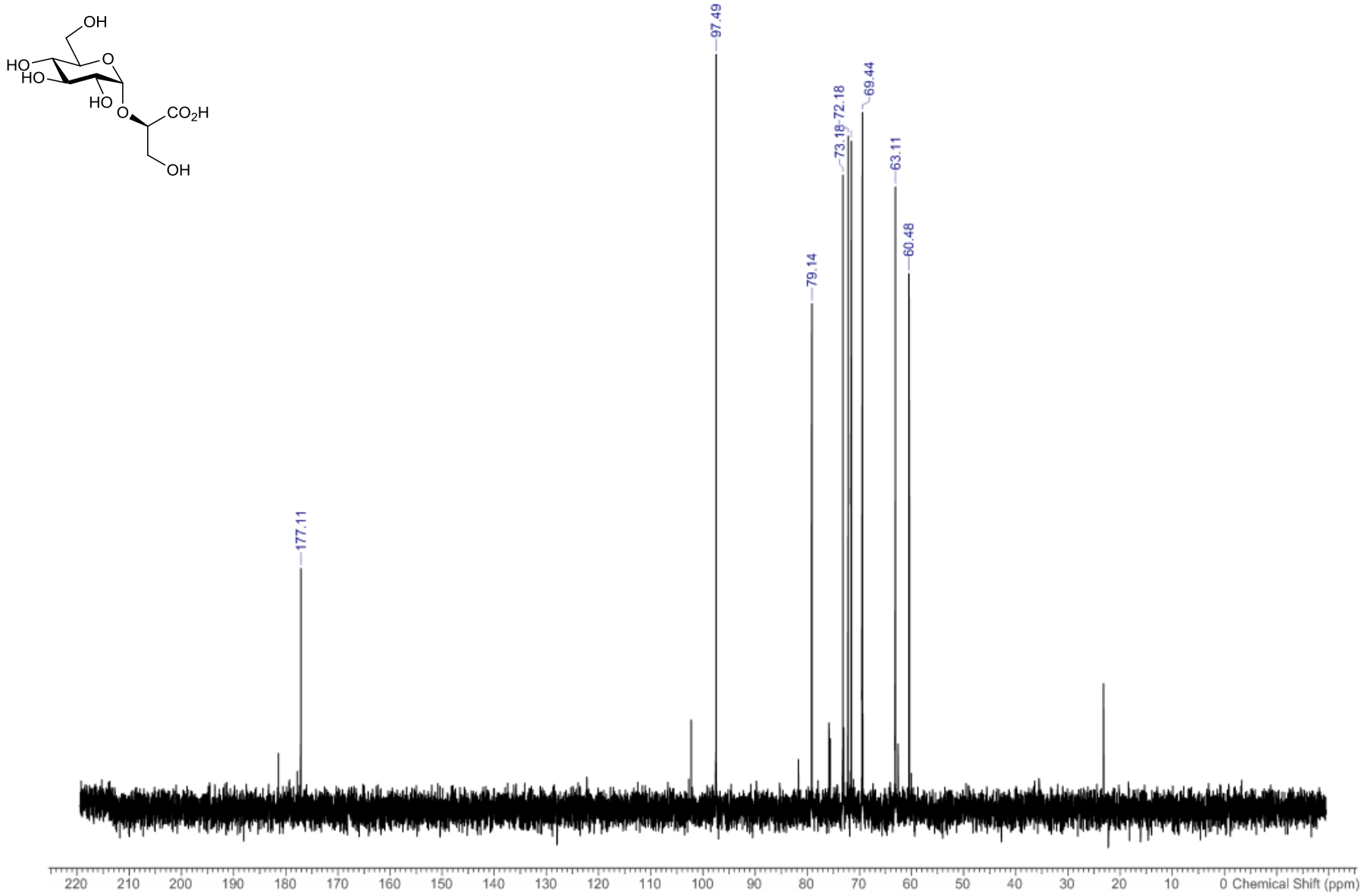
Appendix A

^{1}H NMR (400 MHz, D_2O), isomer:



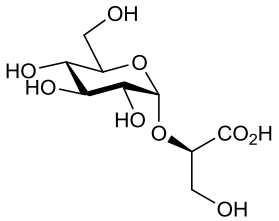
Appendix A

^{13}C NMR (101 MHz, D_2O):

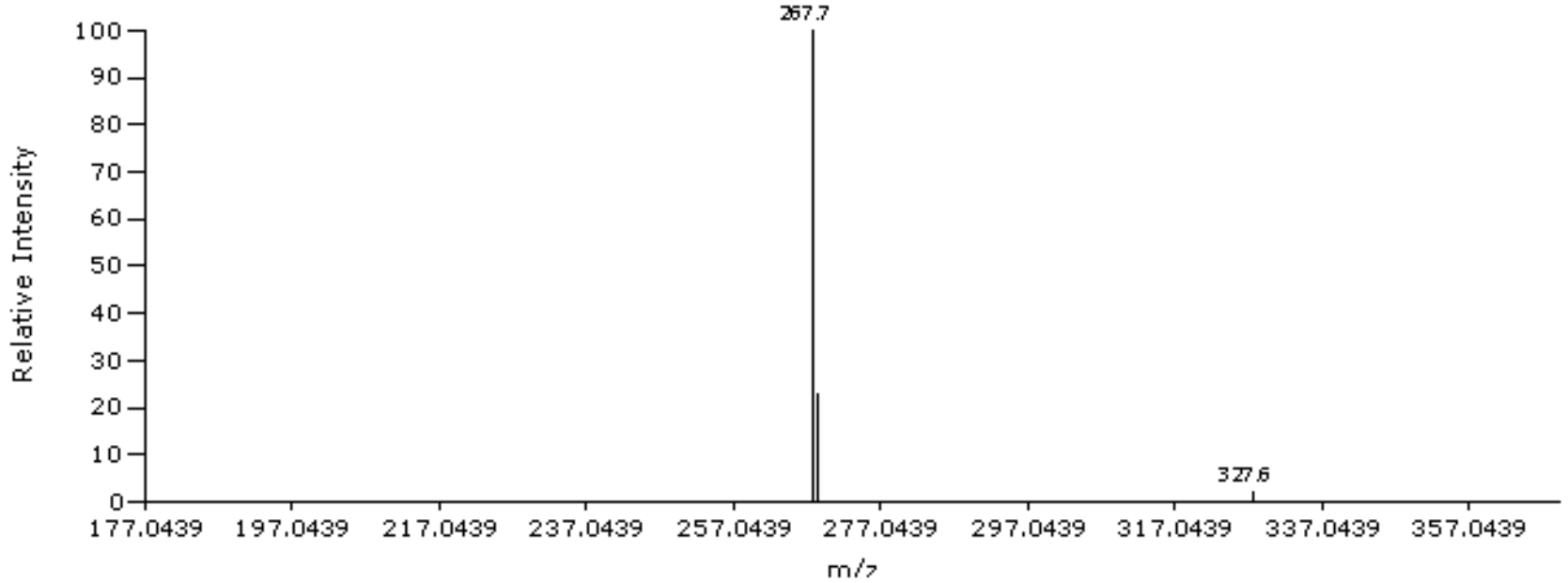


Appendix A

LC-MS (ESI-):

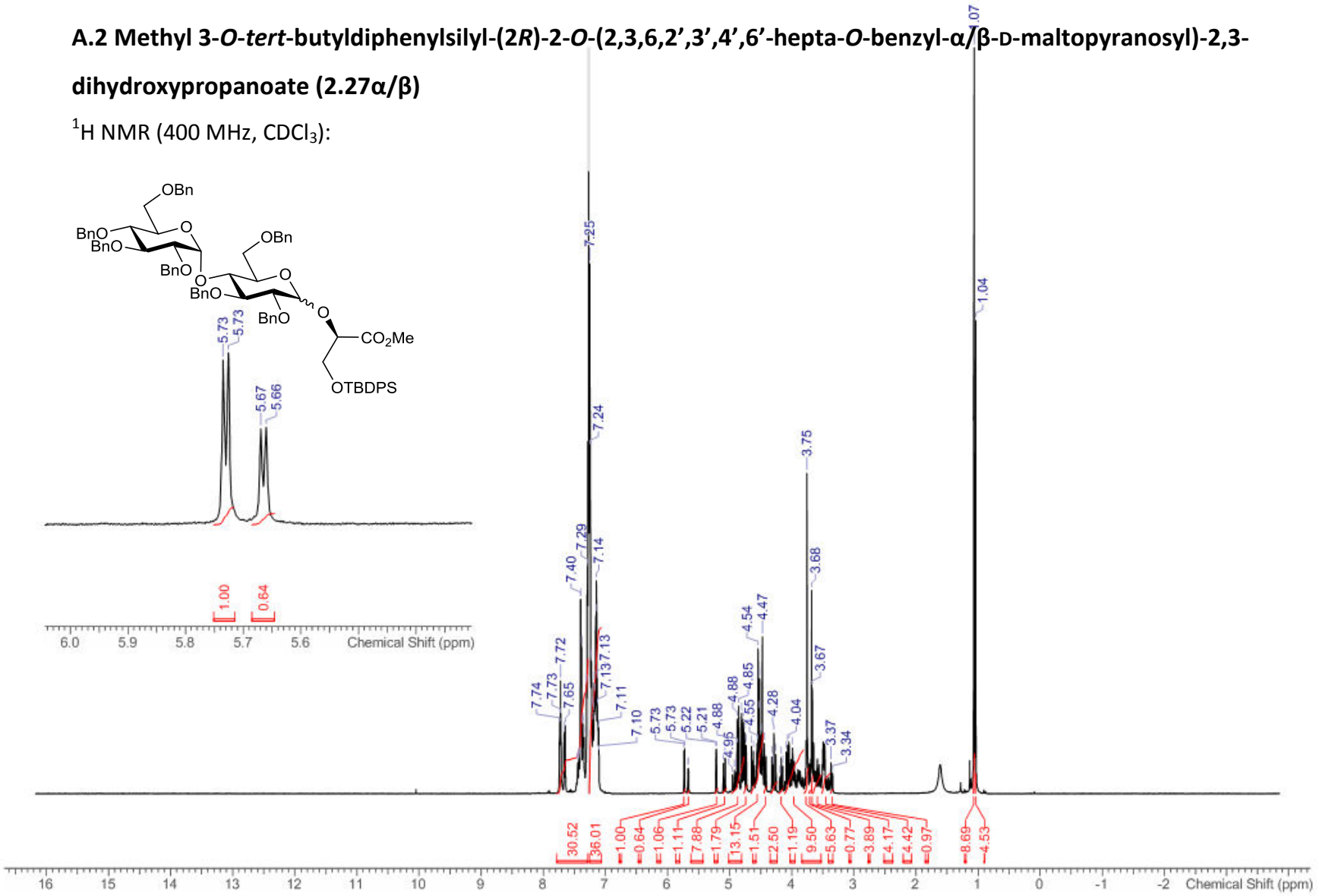


JG-8183-GG, BLUE ESINEG C18 5 min, RT 0.2486 mins, Scan# 70, NL 1.050E6, 14/07/2017 08:47, mz [176-1,417]

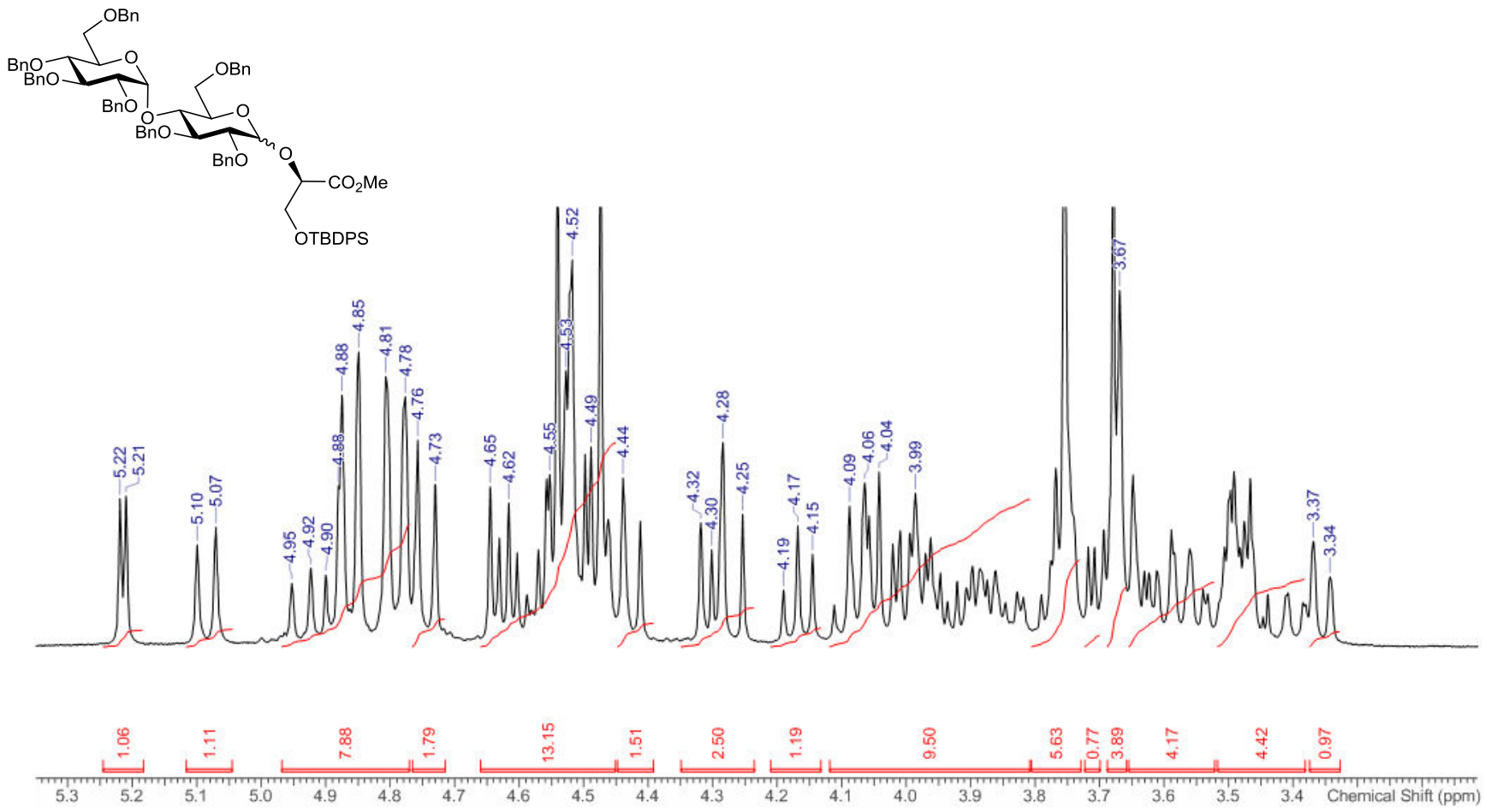


A.2 Methyl 3-*O*-*tert*-butyldiphenylsilyl-(2*R*)-2-*O*-(2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (2.27α/β)

^{1}H NMR (400 MHz, $CDCl_3$):

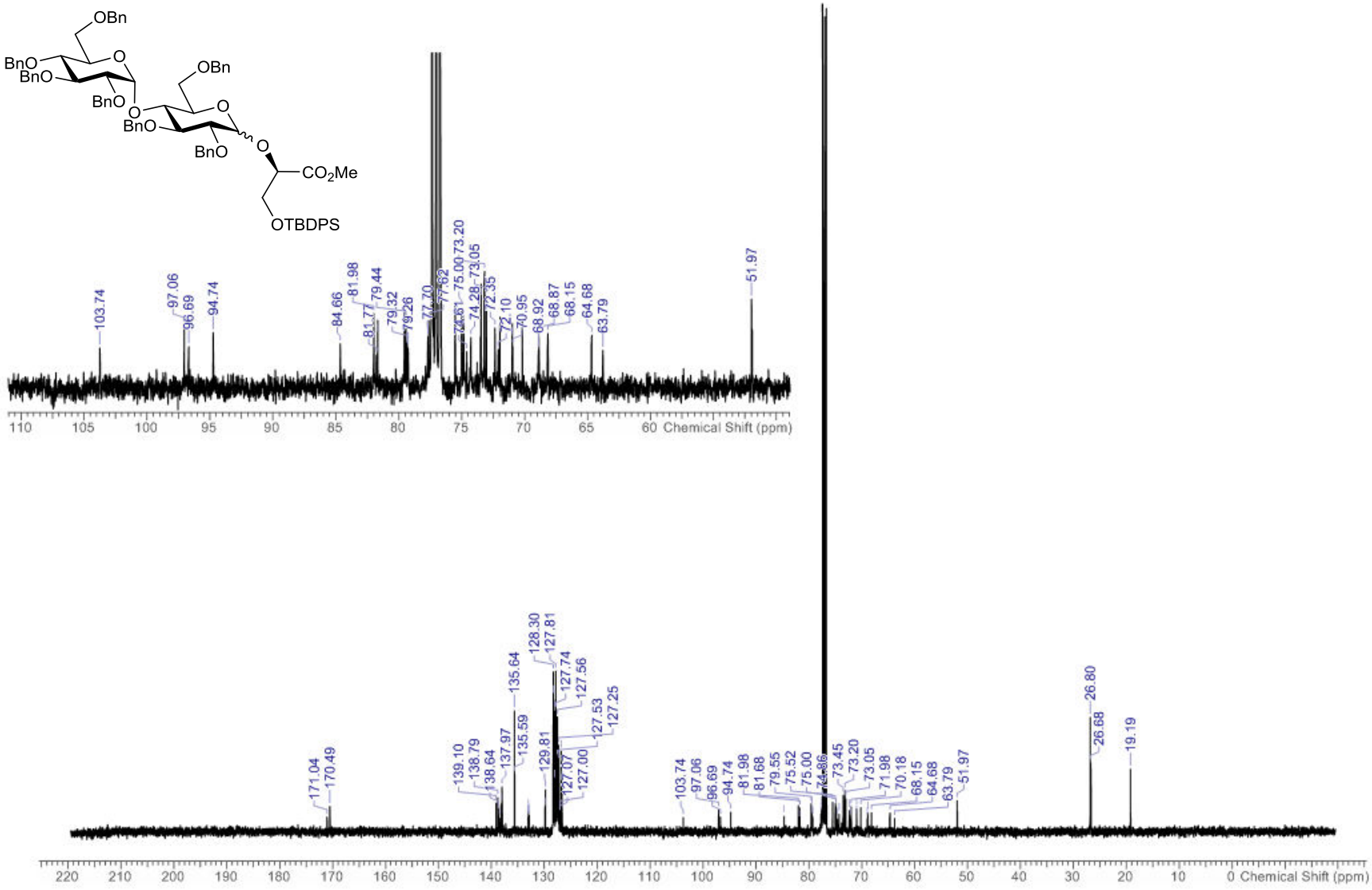


Appendix A


OBn
BnO
BnO
BnO
O
OBn
O
BnO
BnO
O
CO2Me
OTBDPS
5.22
5.21
5.10
5.07
4.95
4.92
4.90
4.88
4.88
4.85
4.81
4.78
4.76
4.73
4.65
4.62
4.55
4.53
4.52
4.49
4.44
4.32
4.30
4.28
4.25
4.19
4.17
4.15
4.09
4.06
4.04
3.99
3.67
3.37
3.34
1.06
1.11
7.88
1.79
13.15
1.51
2.50
1.19
9.50
5.63
0.77
3.89
4.17
4.42
0.97
5.3
5.2
5.1
5.0
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
3.9
3.8
3.7
3.6
3.5
3.4
Chemical Shift (ppm)

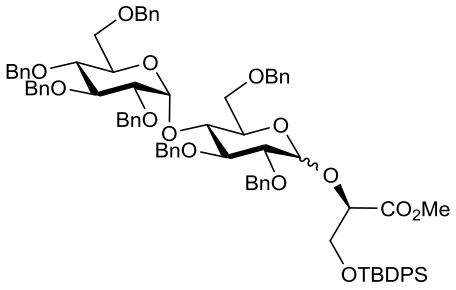
Appendix A

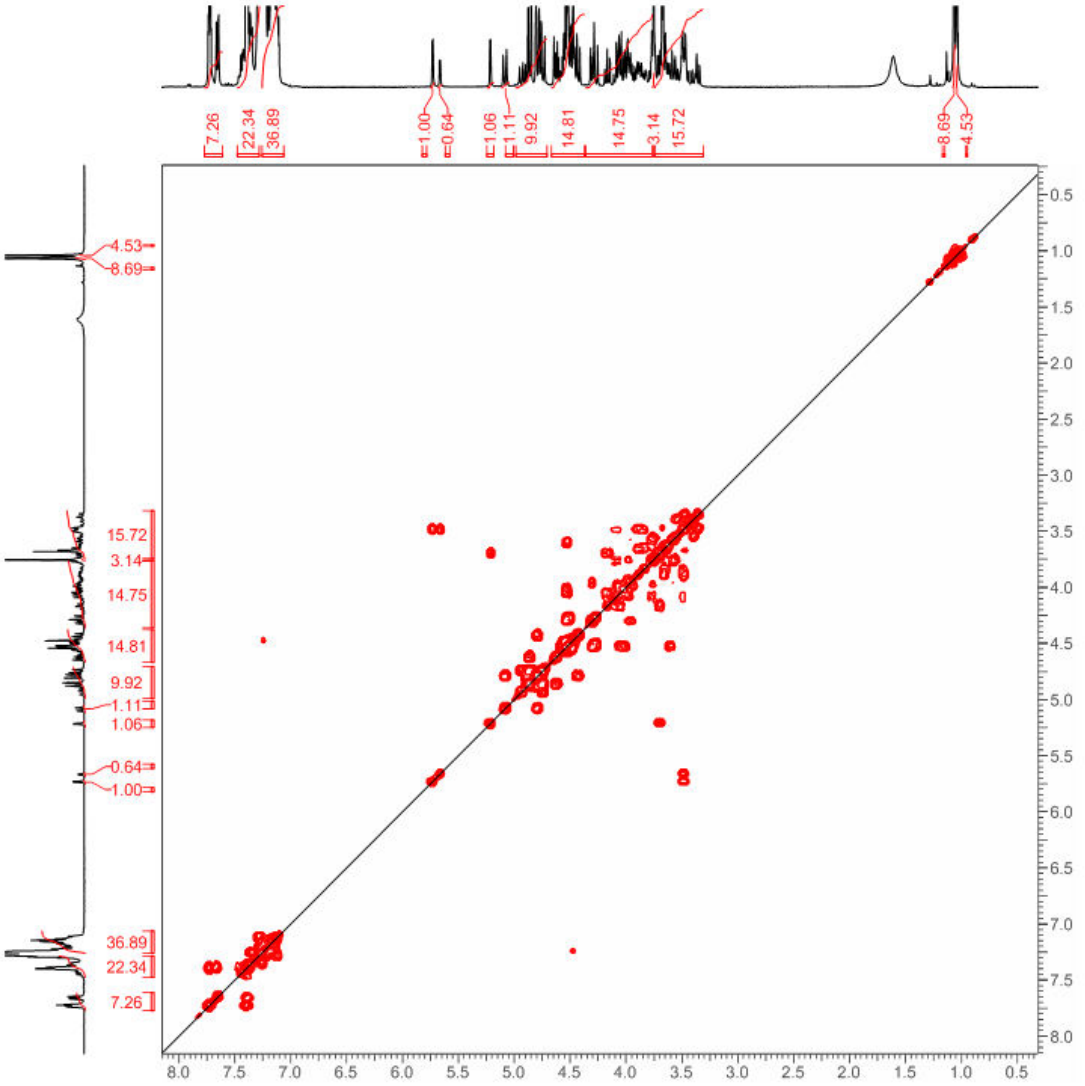
^{13}C NMR (101 MHz, $CDCl_3$):



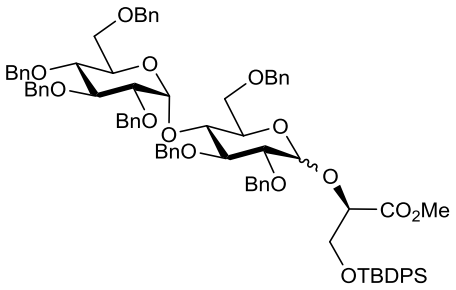
Appendix A

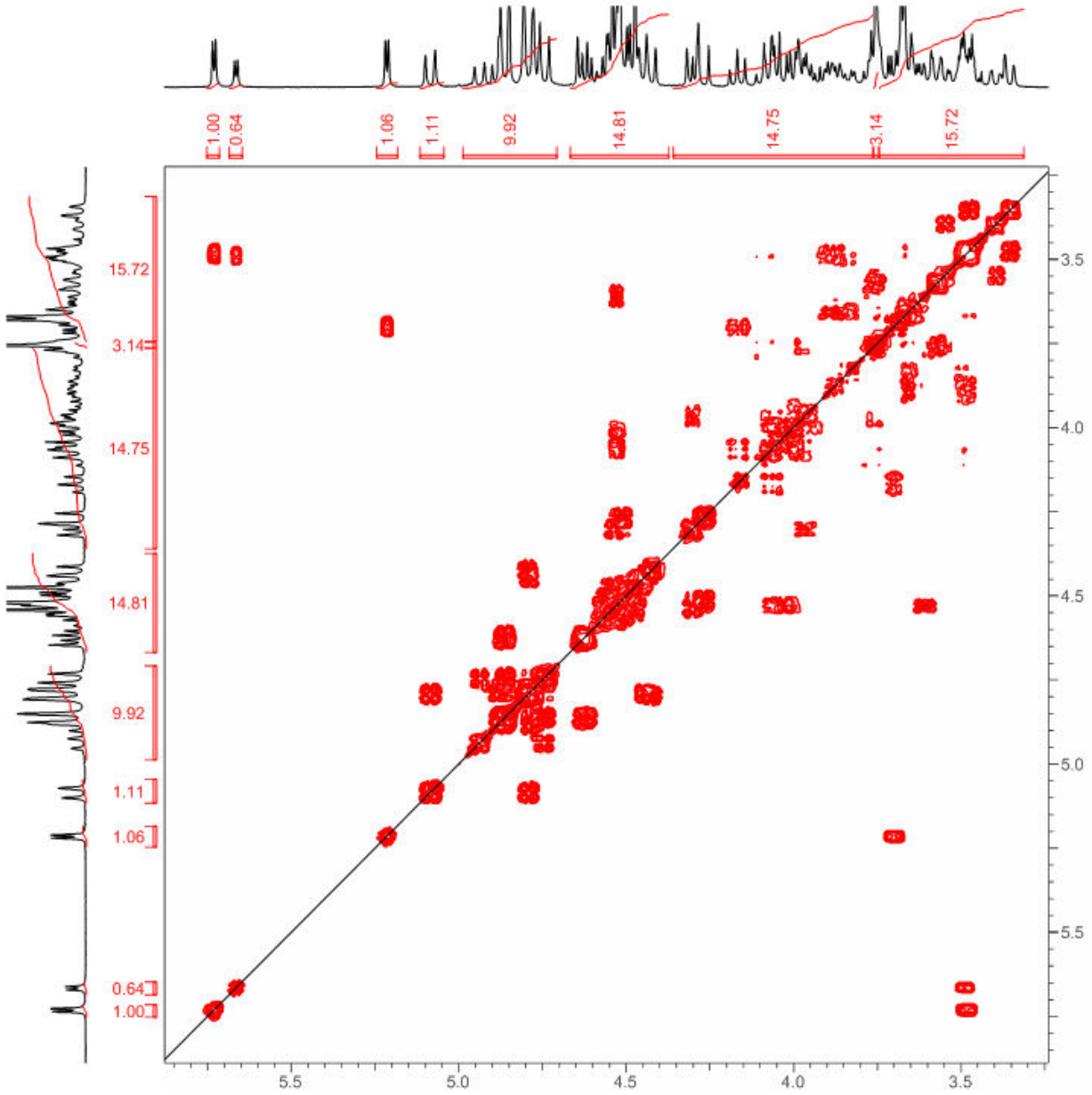
^{1}H-^{1}H COSY (400 - 400 MHz, $CDCl_3$):



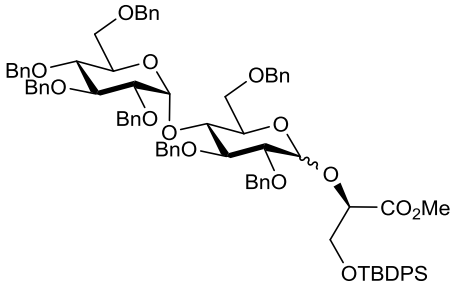


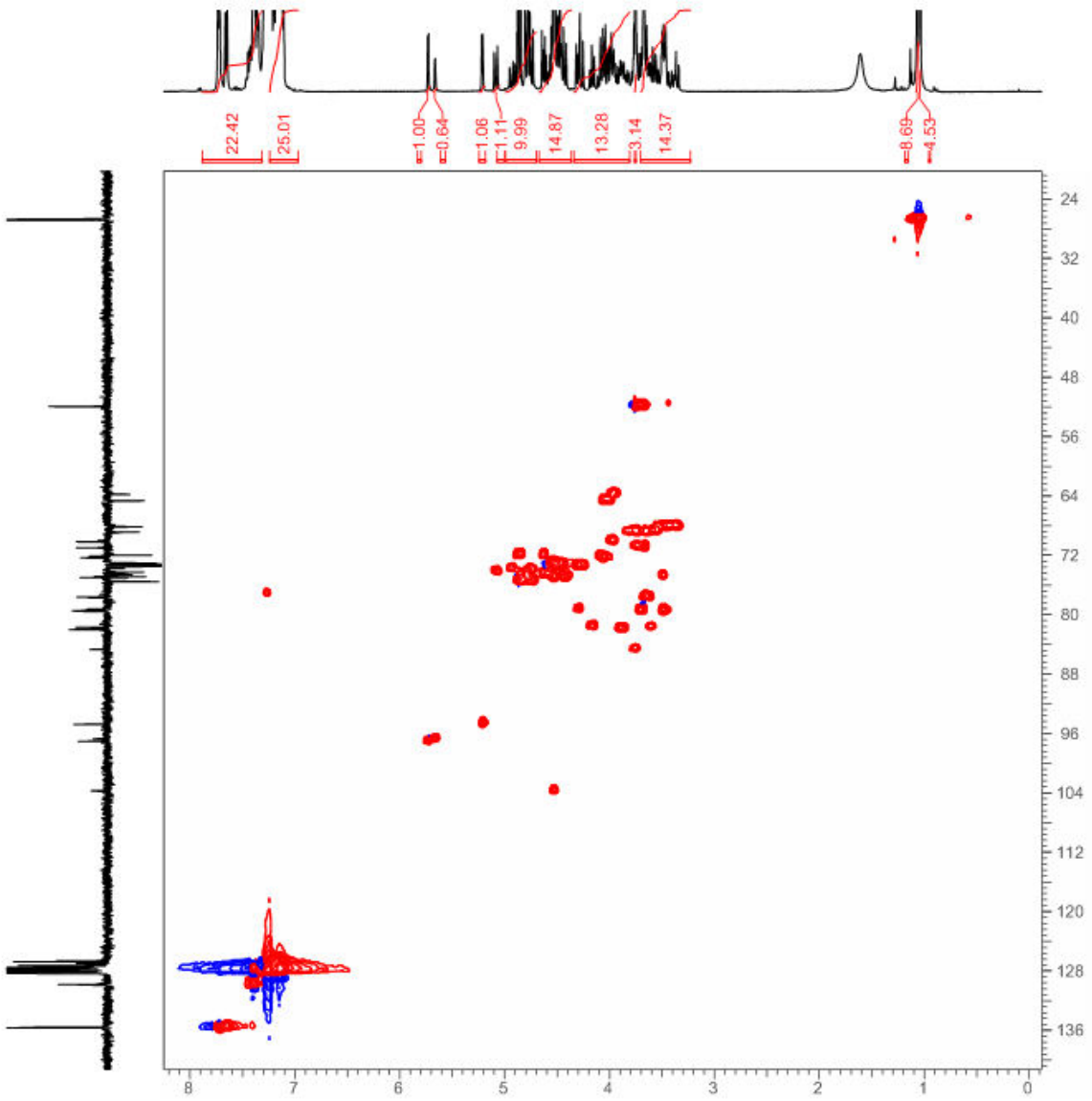
Appendix A


OBn
BnO
BnO
O
BnO
OBn
O
BnO
BnO
O
CO2Me
OTBDPS

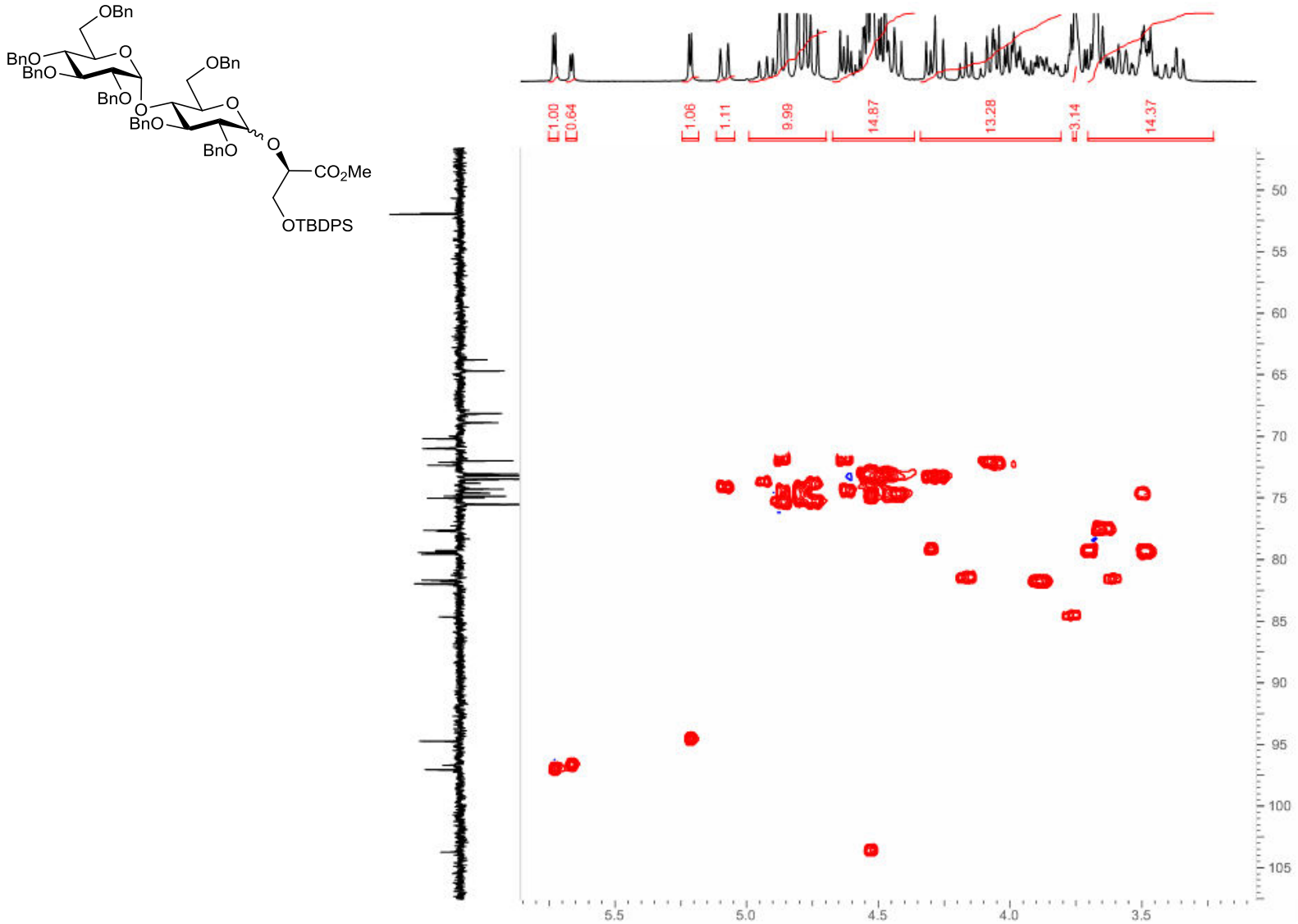

1.00
0.64
1.06
1.11
9.92
14.81
14.75
3.14
15.72
5.5
5.0
4.5
4.0
3.5

^{1}H-^{13}C HSQC (101 - 400 MHz, $CDCl_3$):



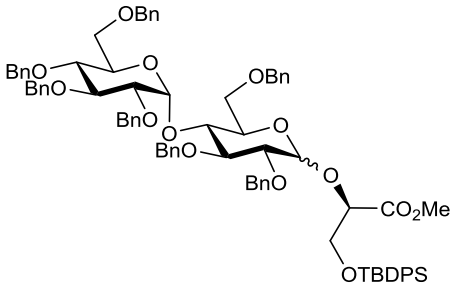


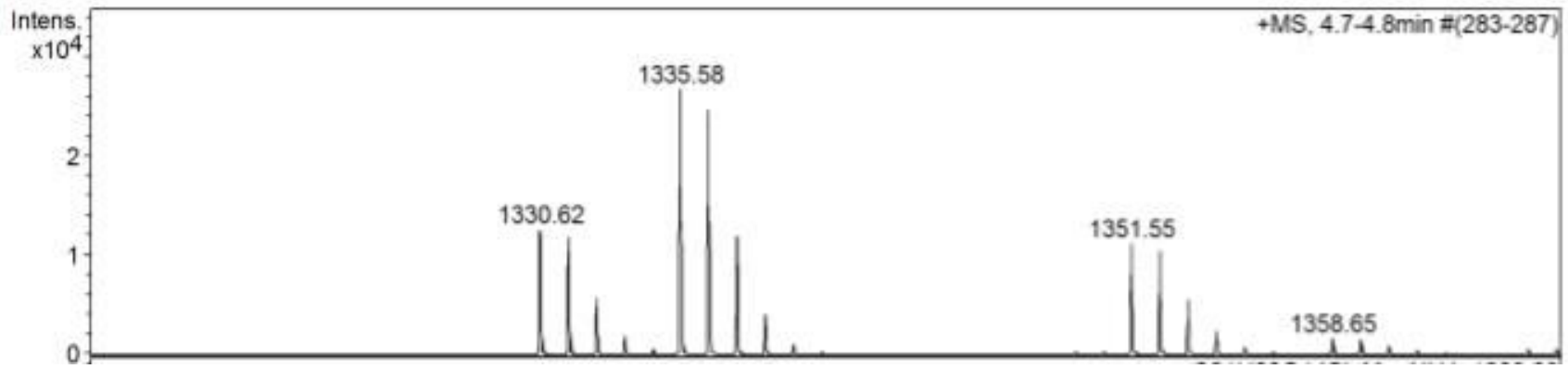
Appendix A


OBn
BnO
BnO
O
BnO
OBn
O
BnO
BnO
O
CO2Me
OTBDPS
1.00
0.64
1.06
1.11
9.99
14.87
13.28
3.14
14.37
50
55
60
65
70
75
80
85
90
95
100
105
5.5
5.0
4.5
4.0
3.5

Appendix A

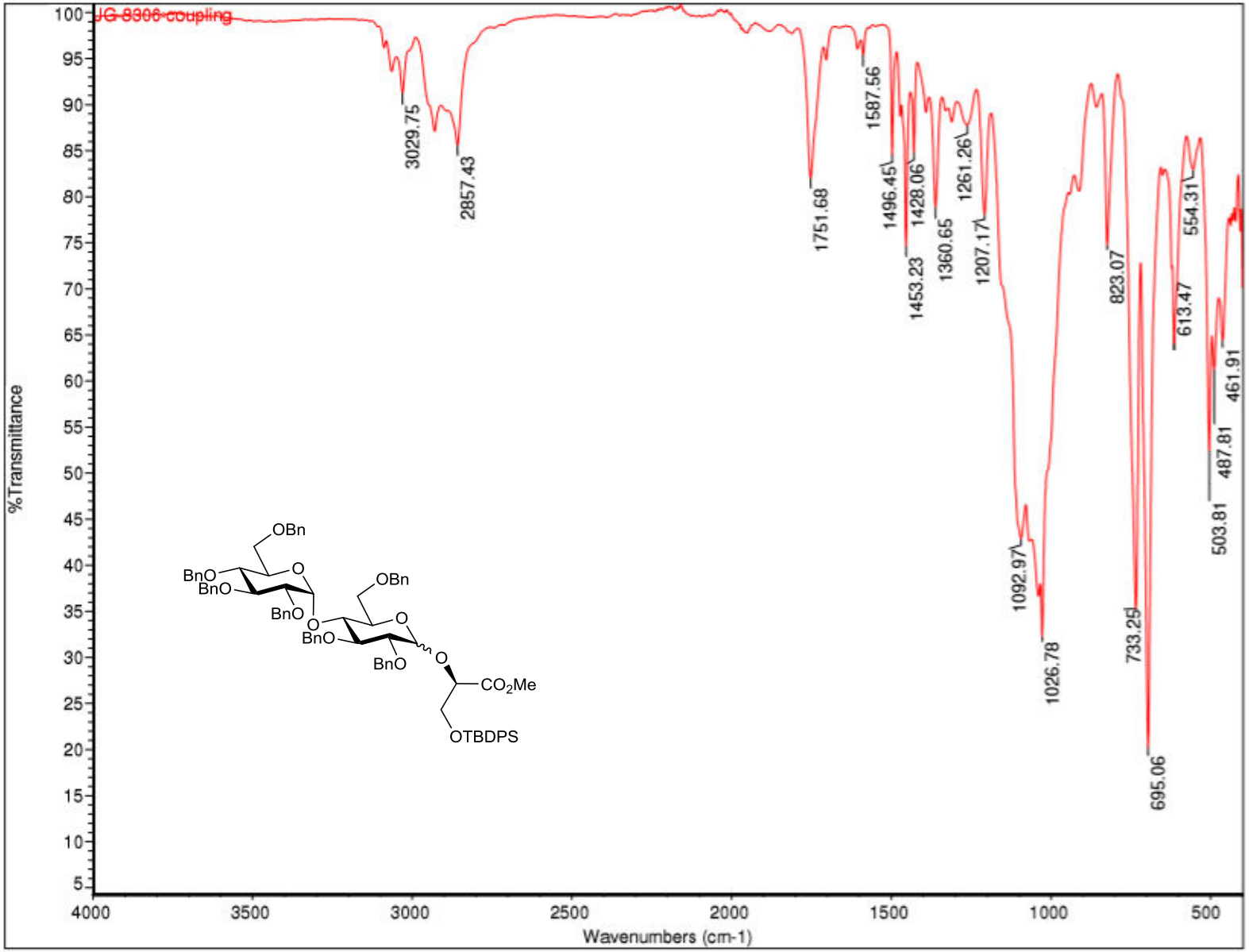
HR-MS (ESI+) m/z:





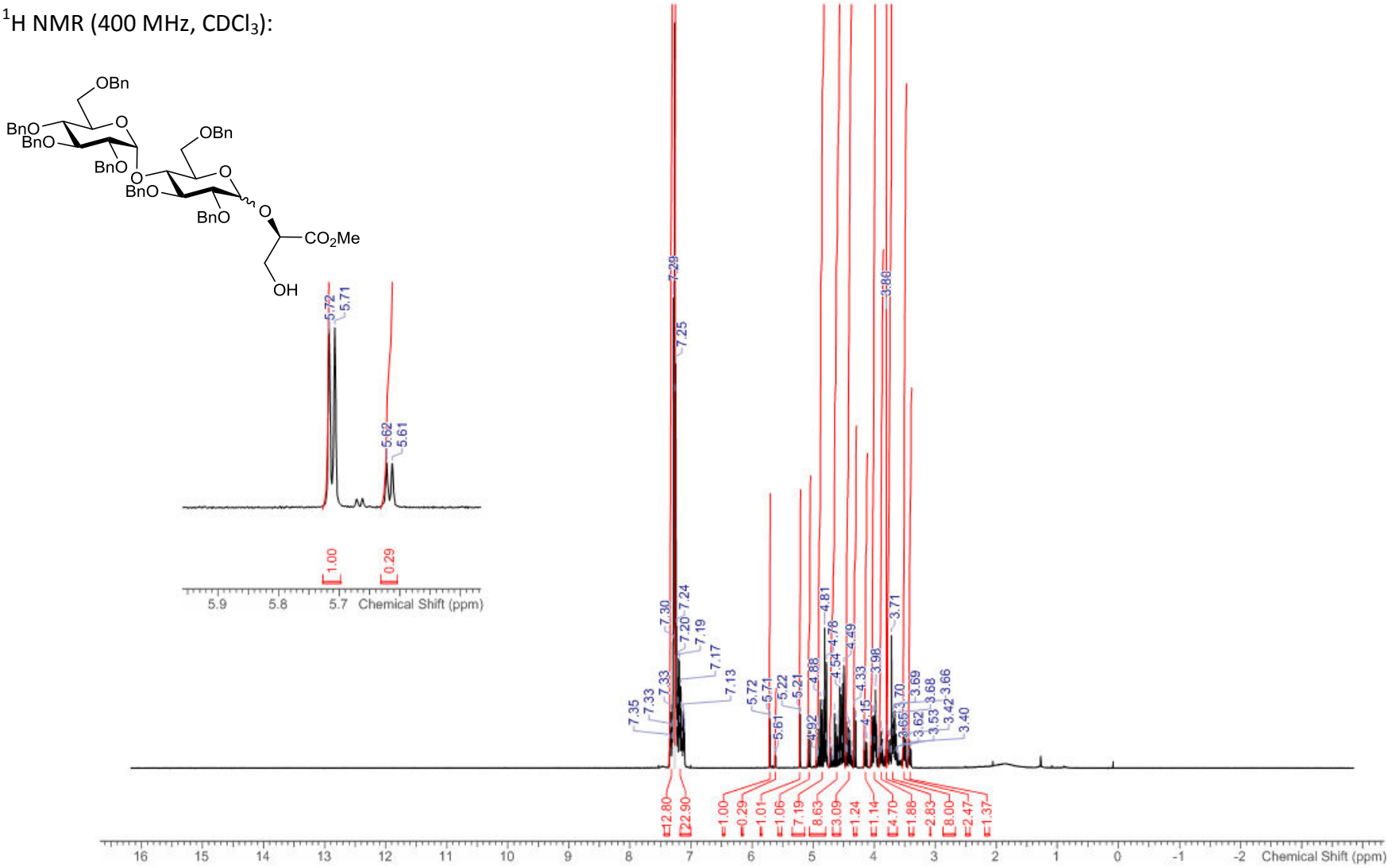
Appendix A

IR:

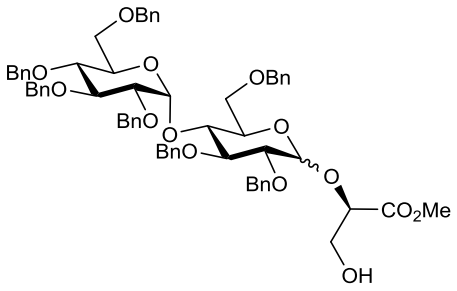


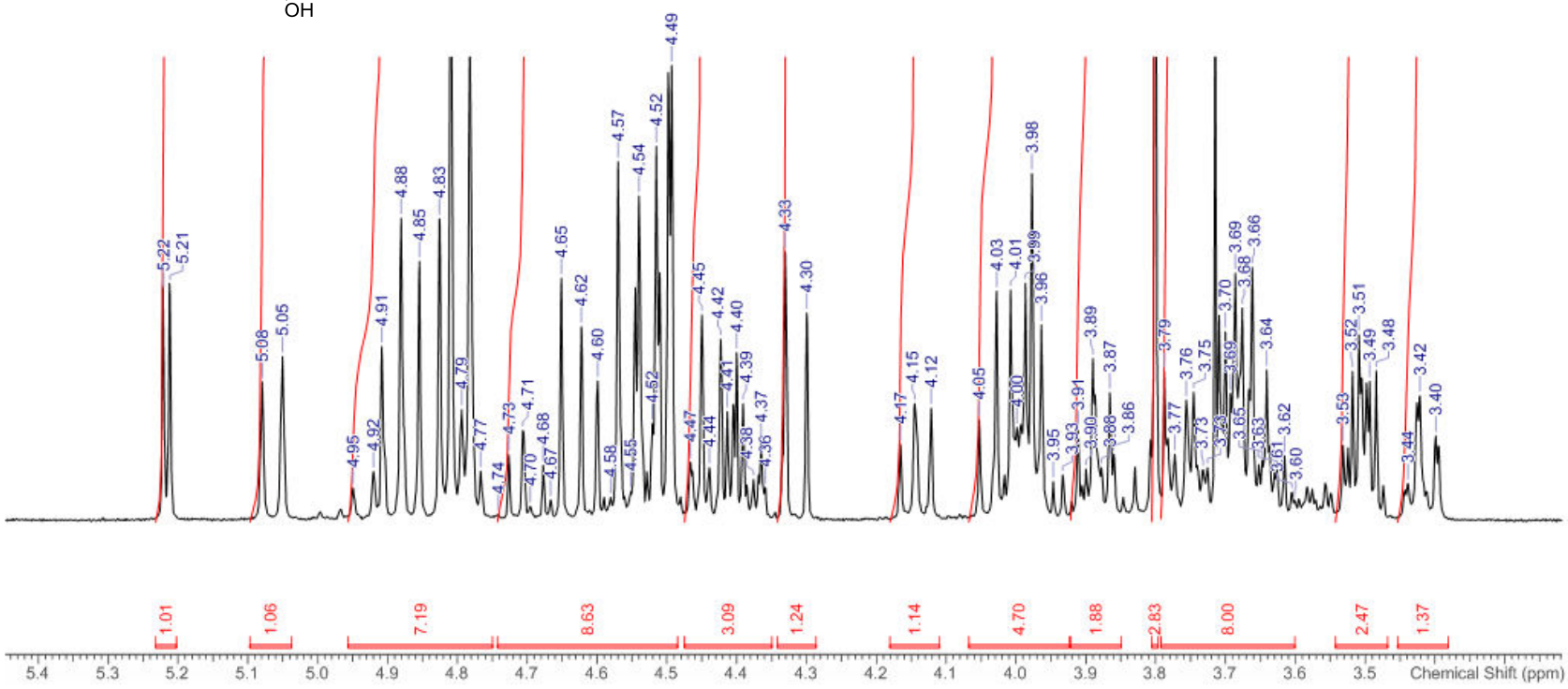
A.3 Methyl (2*R*)-2-*O*-(2,3,6,2',3',4',6'-hepta-*O*-benzyl-α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (2.28α/β)

^{1}H NMR (400 MHz, $CDCl_3$):



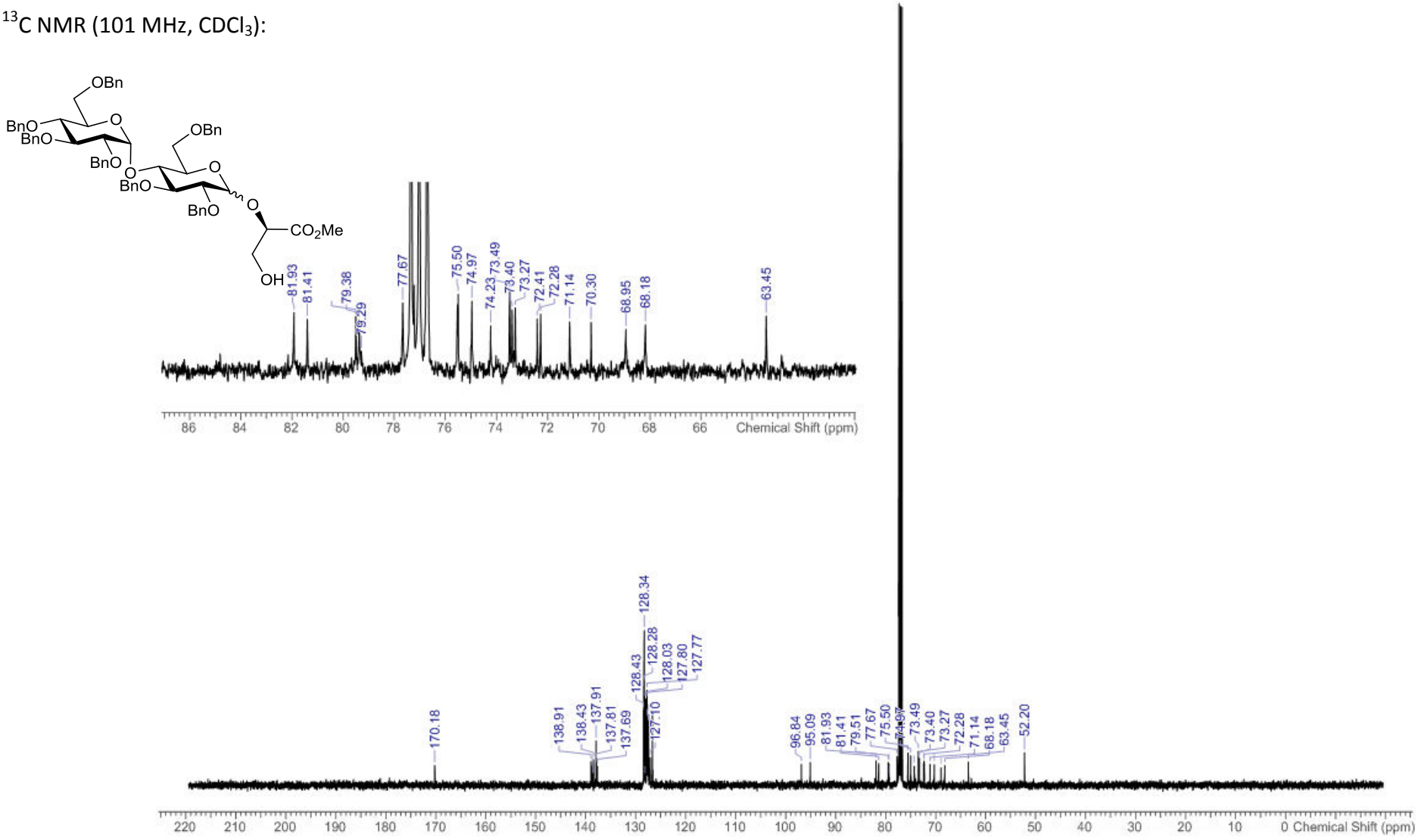
Appendix A


OBn
BnO
BnO
O
BnO
O
OBn
O
BnO
O
BnO
O
CO2Me
OH


5.22
5.21
5.08
5.05
4.95
4.92
4.91
4.88
4.85
4.83
4.79
4.77
4.74
4.73
4.71
4.70
4.68
4.67
4.65
4.62
4.60
4.58
4.57
4.55
4.54
4.52
4.62
4.49
4.47
4.45
4.44
4.42
4.41
4.40
4.39
4.38
4.37
4.36
4.33
4.30
4.17
4.15
4.12
4.05
4.03
4.01
4.00
3.99
3.98
3.96
3.95
3.93
3.91
3.90
3.89
3.88
3.87
3.86
3.79
3.77
3.76
3.75
3.73
3.70
3.69
3.68
3.66
3.65
3.64
3.63
3.62
3.61
3.60
3.53
3.52
3.51
3.49
3.48
3.44
3.42
3.40
1.01
1.06
7.19
8.63
3.09
1.24
1.14
4.70
1.88
2.83
8.00
2.47
1.37
5.4
5.3
5.2
5.1
5.0
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
3.9
3.8
3.7
3.6
3.5
Chemical Shift (ppm)

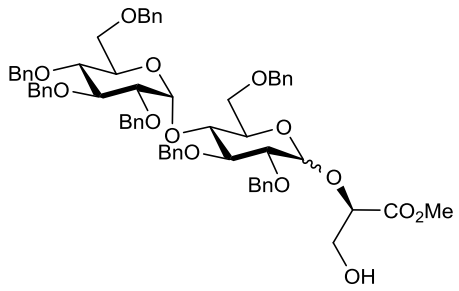
Appendix A

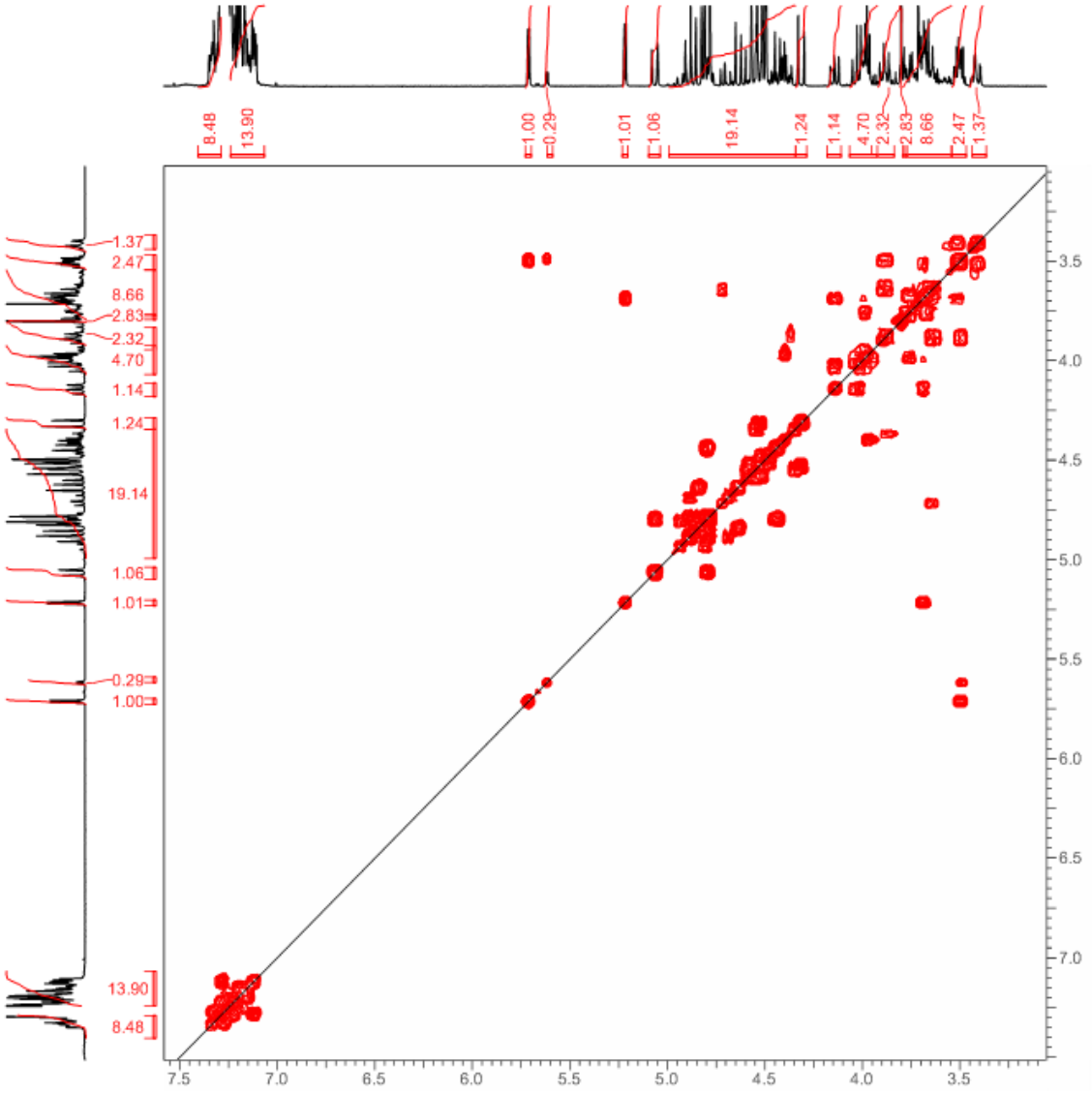
^{13}C NMR (101 MHz, $CDCl_3$):



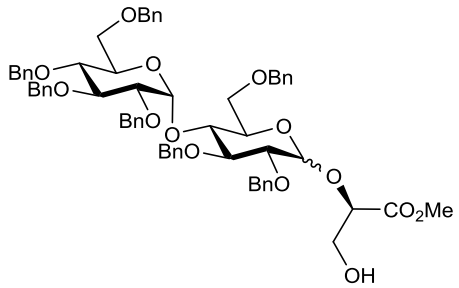
Appendix A

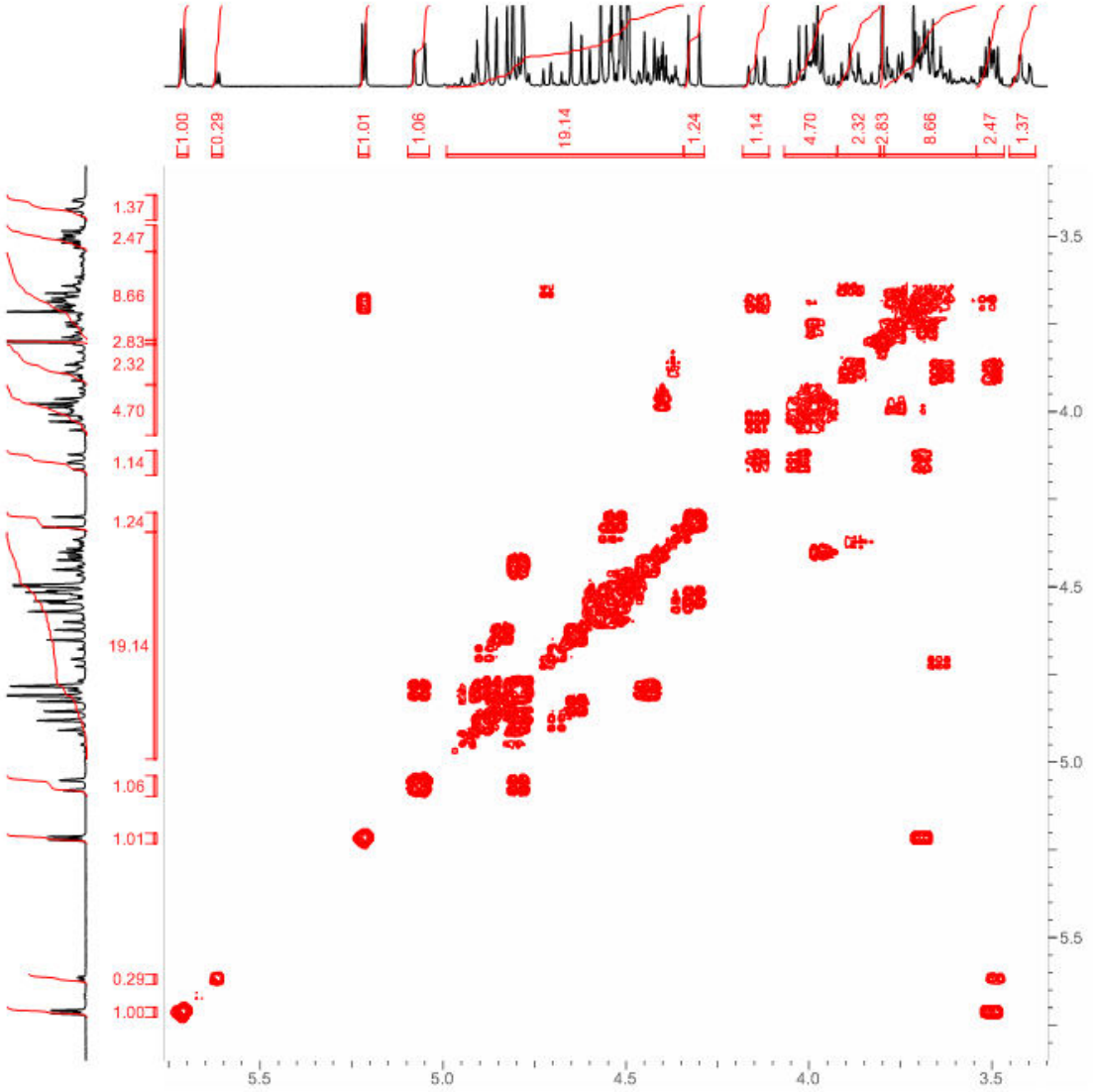
^{1}H-^{1}H COSY (400 - 400 MHz, $CDCl_3$):





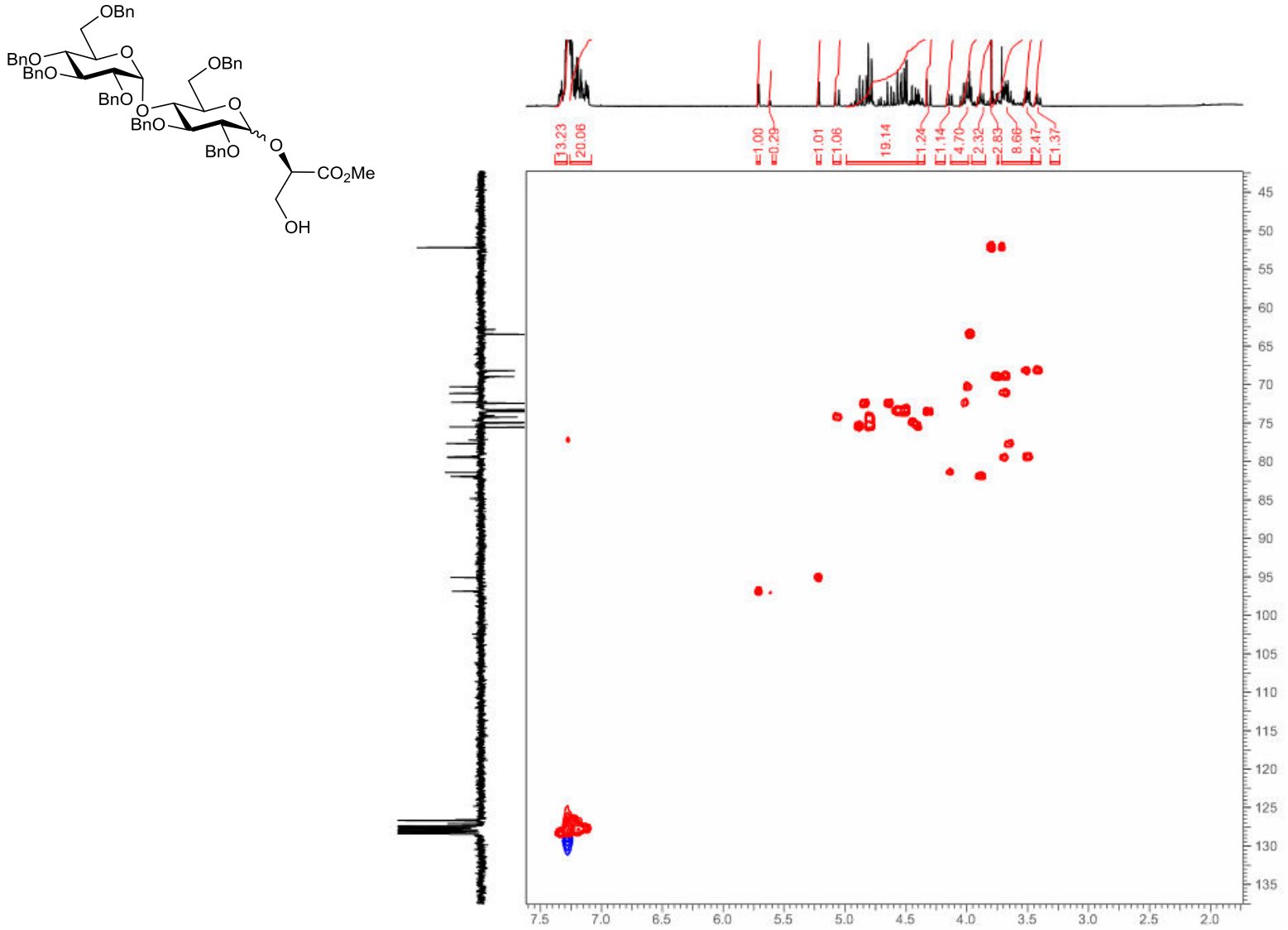
Appendix A


OBn
BnO
BnO
O
BnO
OBn
O
BnO
O
BnO
O
CO2Me
OH

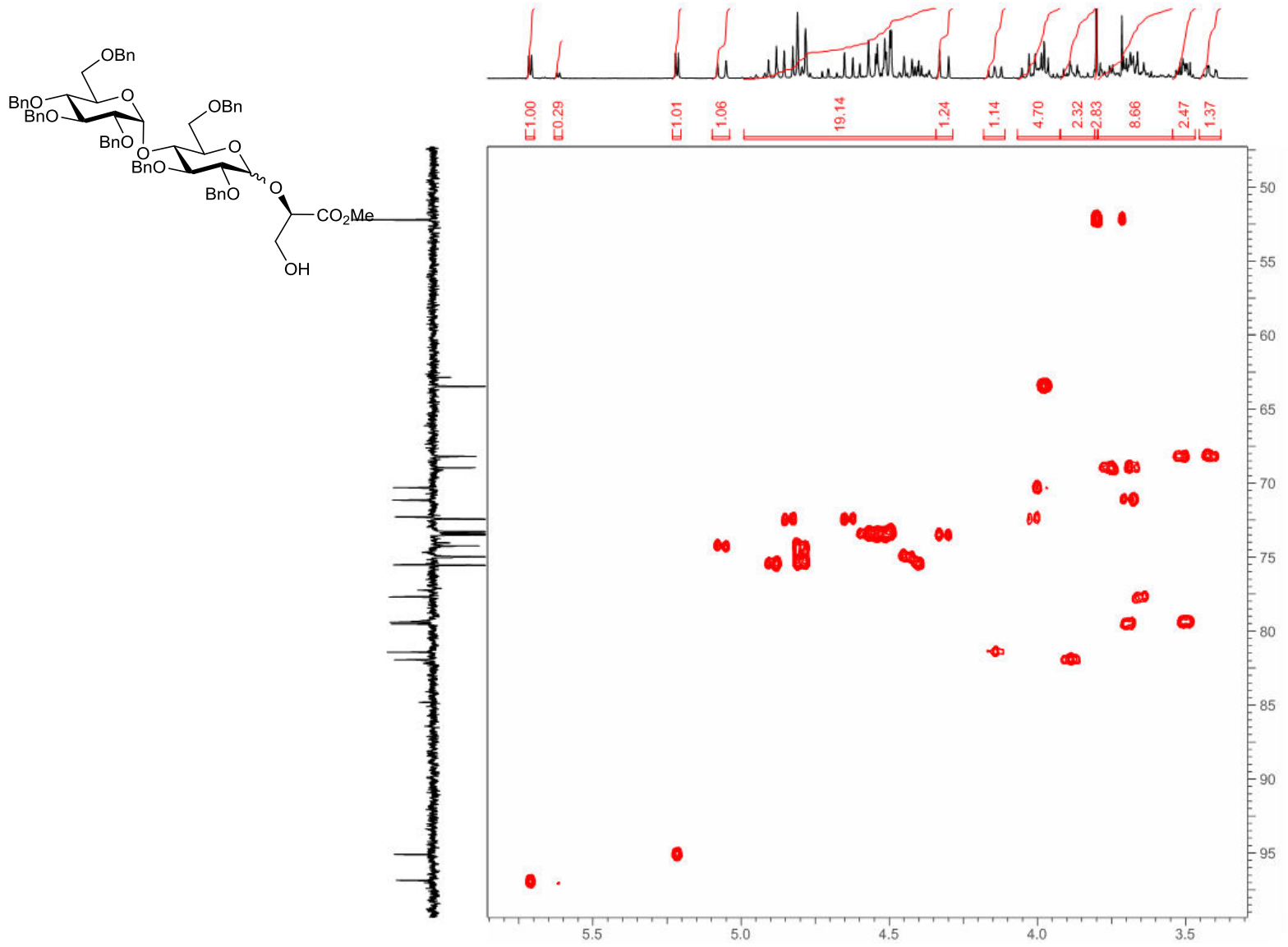

1.00
0.29
1.01
1.06
19.14
1.24
1.14
4.70
2.32
2.83
8.66
2.47
1.37
5.5
5.0
4.5
4.0
3.5

Appendix A

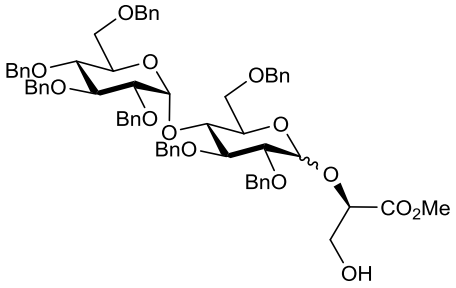
^{1}H-^{13}C HSQC (100 - 400 MHz, $CDCl_3$):

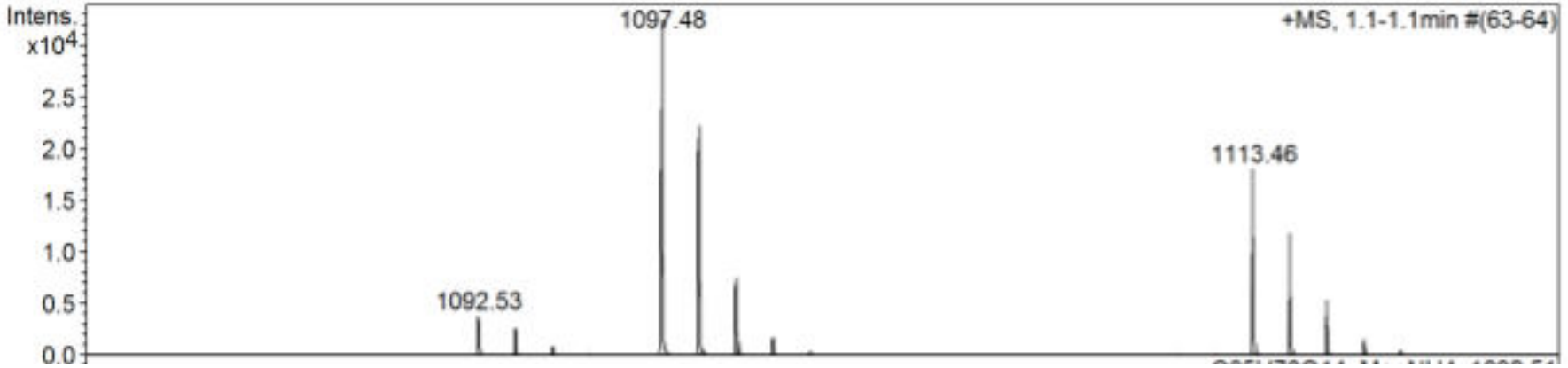


Appendix A


OBn
BnO
BnO
BnO
O
OBn
O
BnO
BnO
O
CO2Me
OH
1.00
0.29
1.01
1.06
19.14
1.24
1.14
4.70
2.32
2.83
8.66
2.47
1.37
50
55
60
65
70
75
80
85
90
95
5.5
5.0
4.5
4.0
3.5

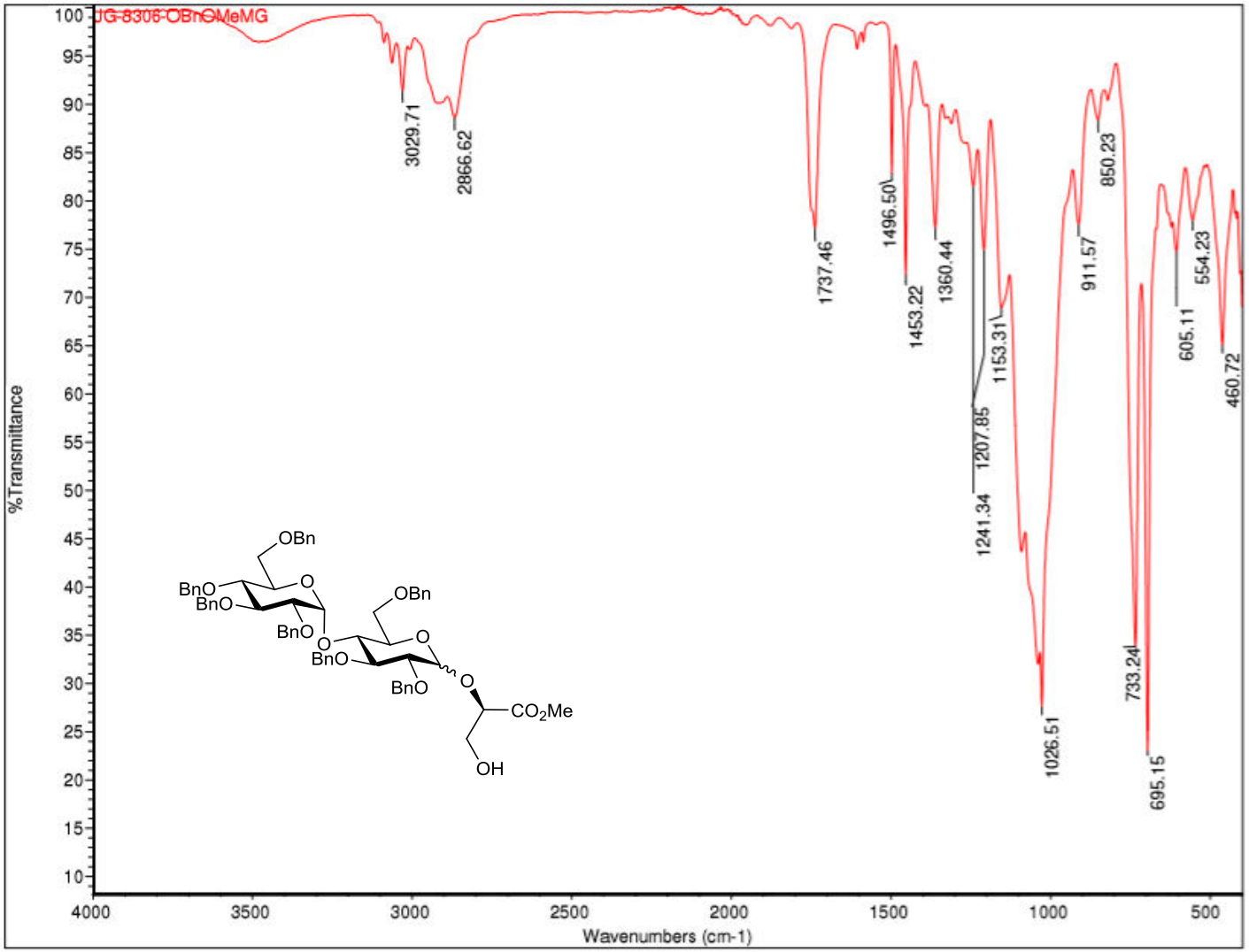
HR-MS (ESI+) m/z:


OBn
BnO
BnO
O
BnO
OBn
O
BnO
O
BnO
O
CO2Me
OH


Intens.
x10^4
2.5
2.0
1.5
1.0
0.5
0.0
1097.48
+MS, 1.1-1.1min #(63-64)
1113.46
1092.53

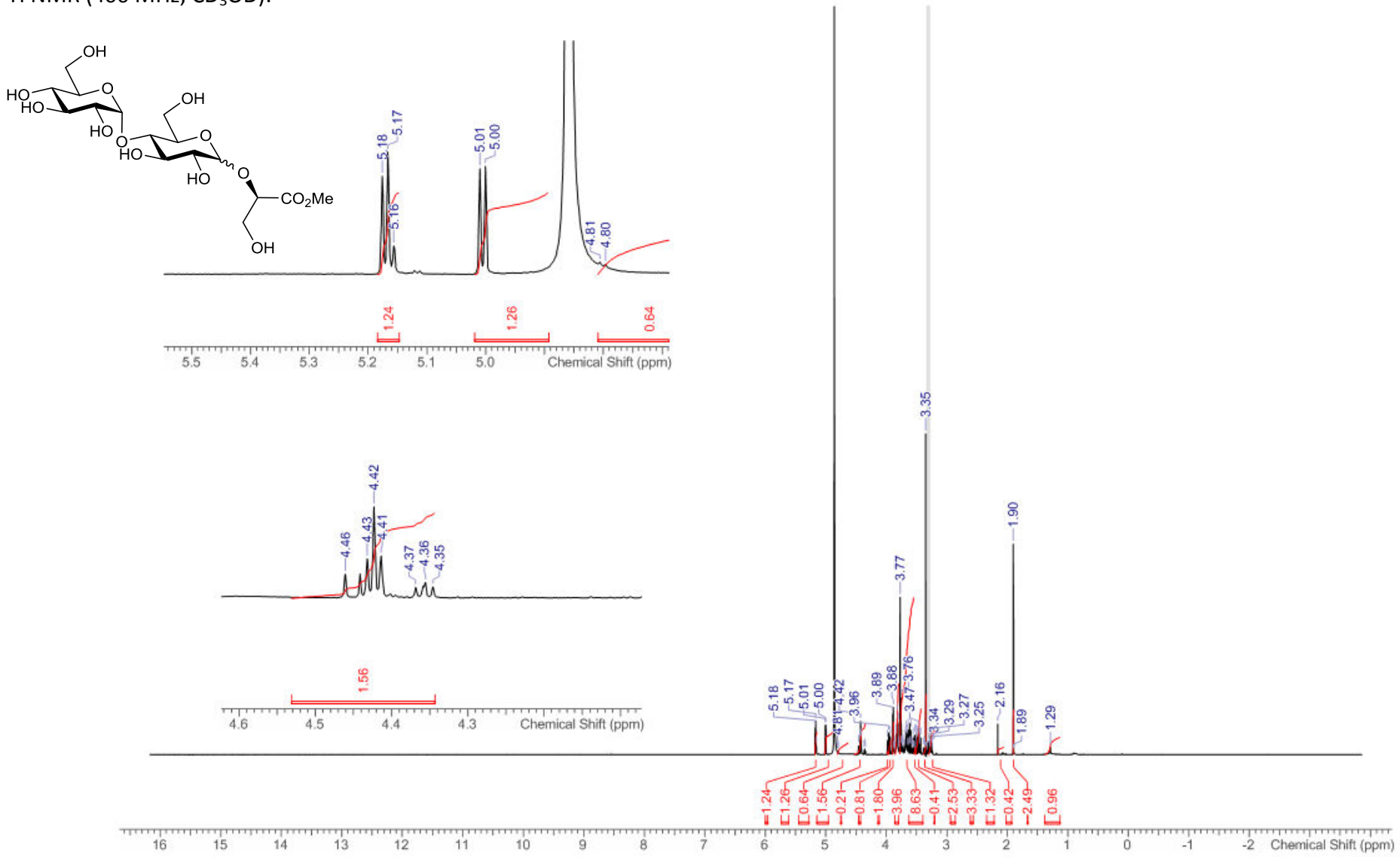
Appendix A

IR:

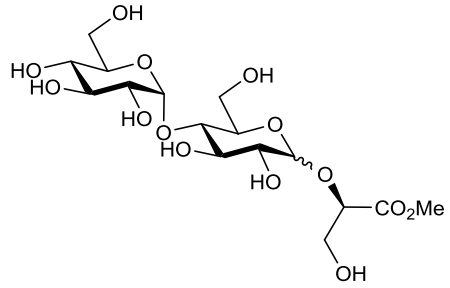


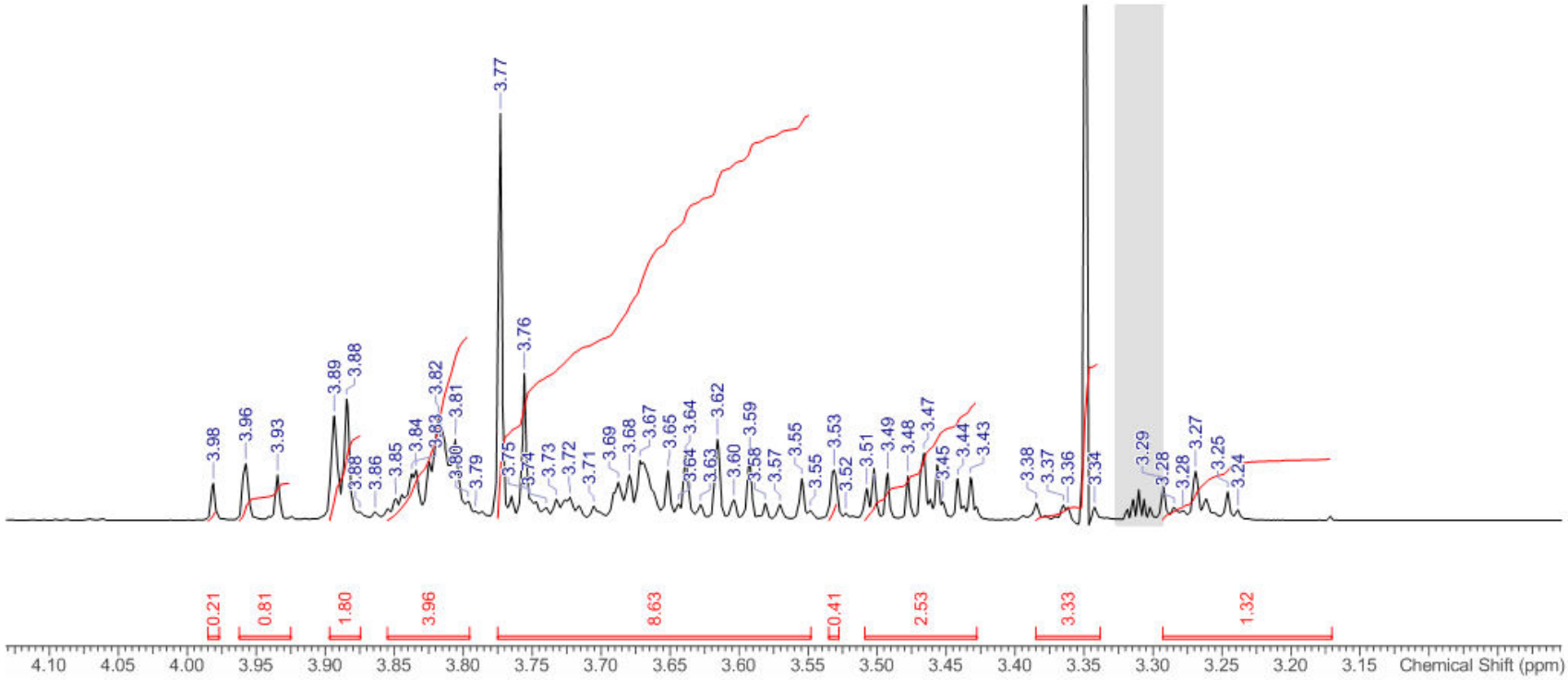
A.4 Methyl (2*R*)-2-*O*-(α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (2.29α/β)

^{1}H NMR (400 MHz, CD_3OD):



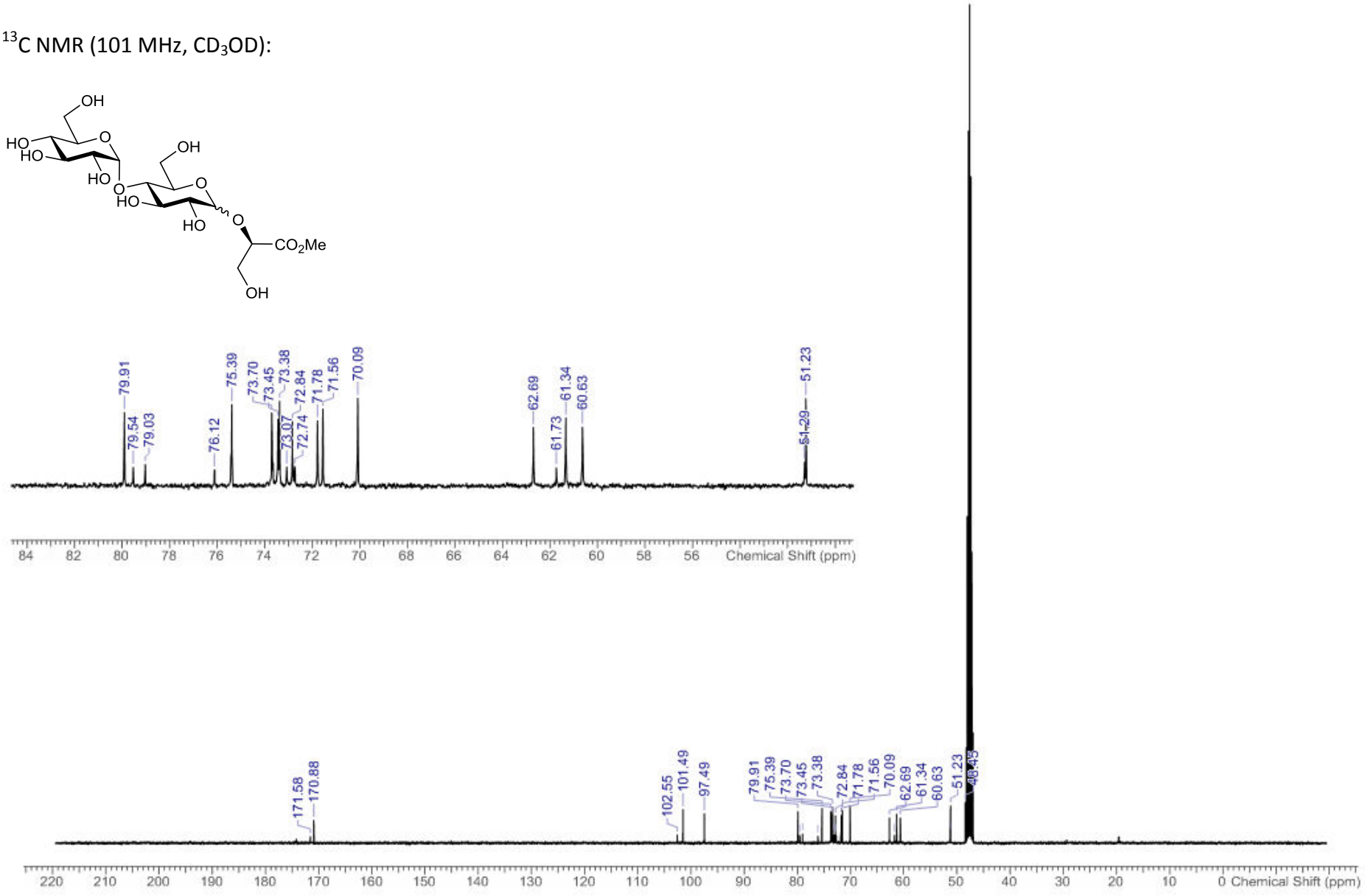
Appendix A


OH
HO
HO
HO
O
OH
O
HO
HO
O
CO2Me
OH


3.98
3.96
3.93
3.89
3.88
3.88
3.86
3.85
3.84
3.83
3.82
3.81
3.80
3.79
3.77
3.76
3.75
3.74
3.73
3.72
3.71
3.69
3.68
3.67
3.65
3.64
3.63
3.62
3.60
3.59
3.58
3.57
3.55
3.55
3.53
3.52
3.51
3.49
3.48
3.47
3.45
3.44
3.43
3.38
3.37
3.36
3.34
3.29
3.28
3.28
3.27
3.25
3.24
0.21
0.81
1.80
3.96
8.63
0.41
2.53
3.33
1.32
4.10
4.05
4.00
3.95
3.90
3.85
3.80
3.75
3.70
3.65
3.60
3.55
3.50
3.45
3.40
3.35
3.30
3.25
3.20
3.15
Chemical Shift (ppm)

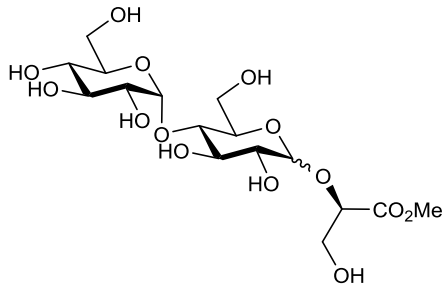
Appendix A

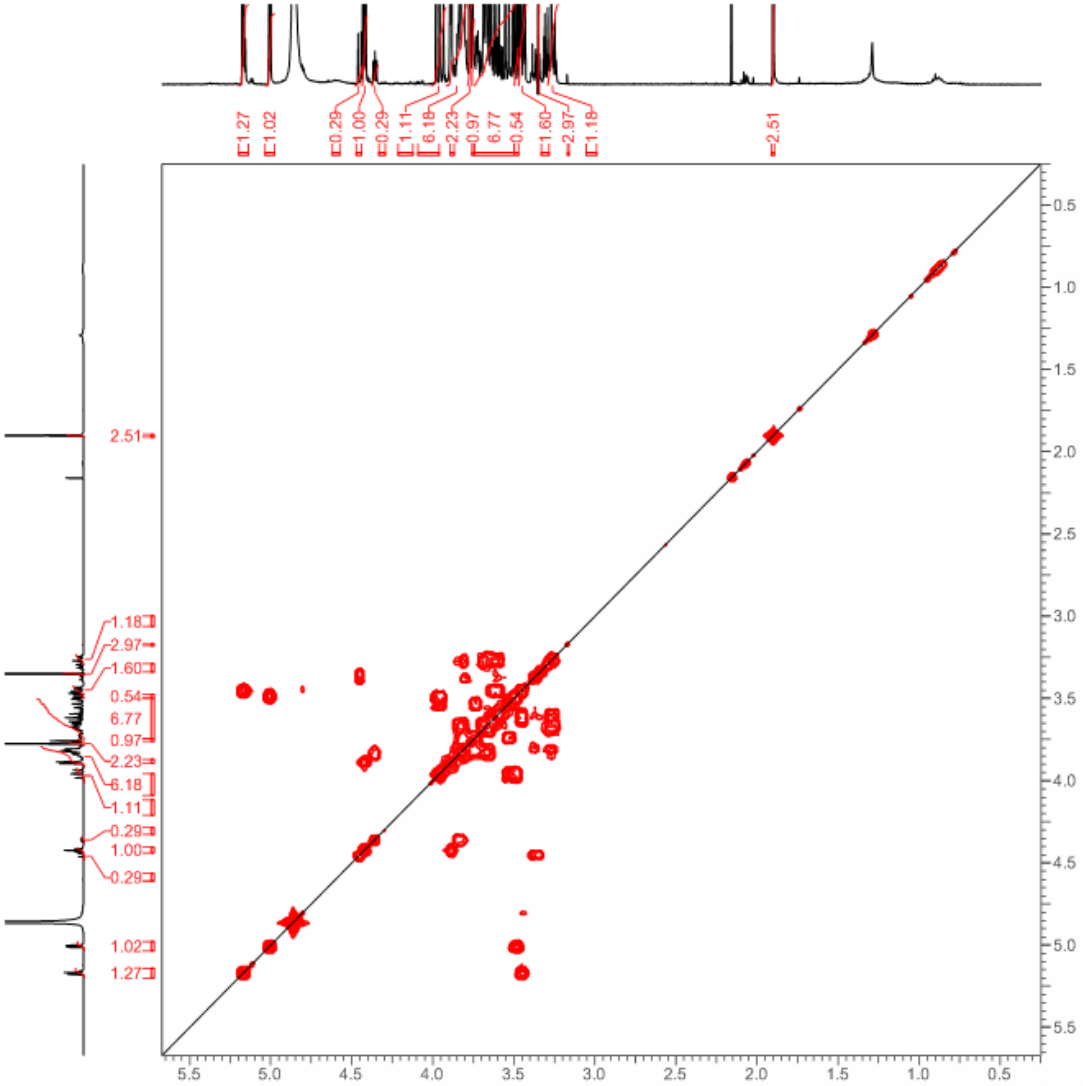
^{13}C NMR (101 MHz, CD_3OD):



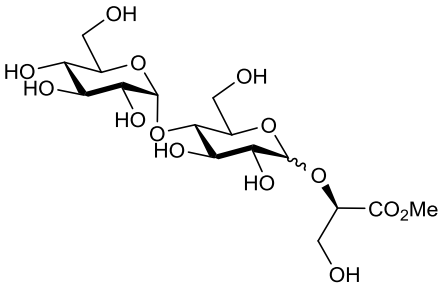
Appendix A

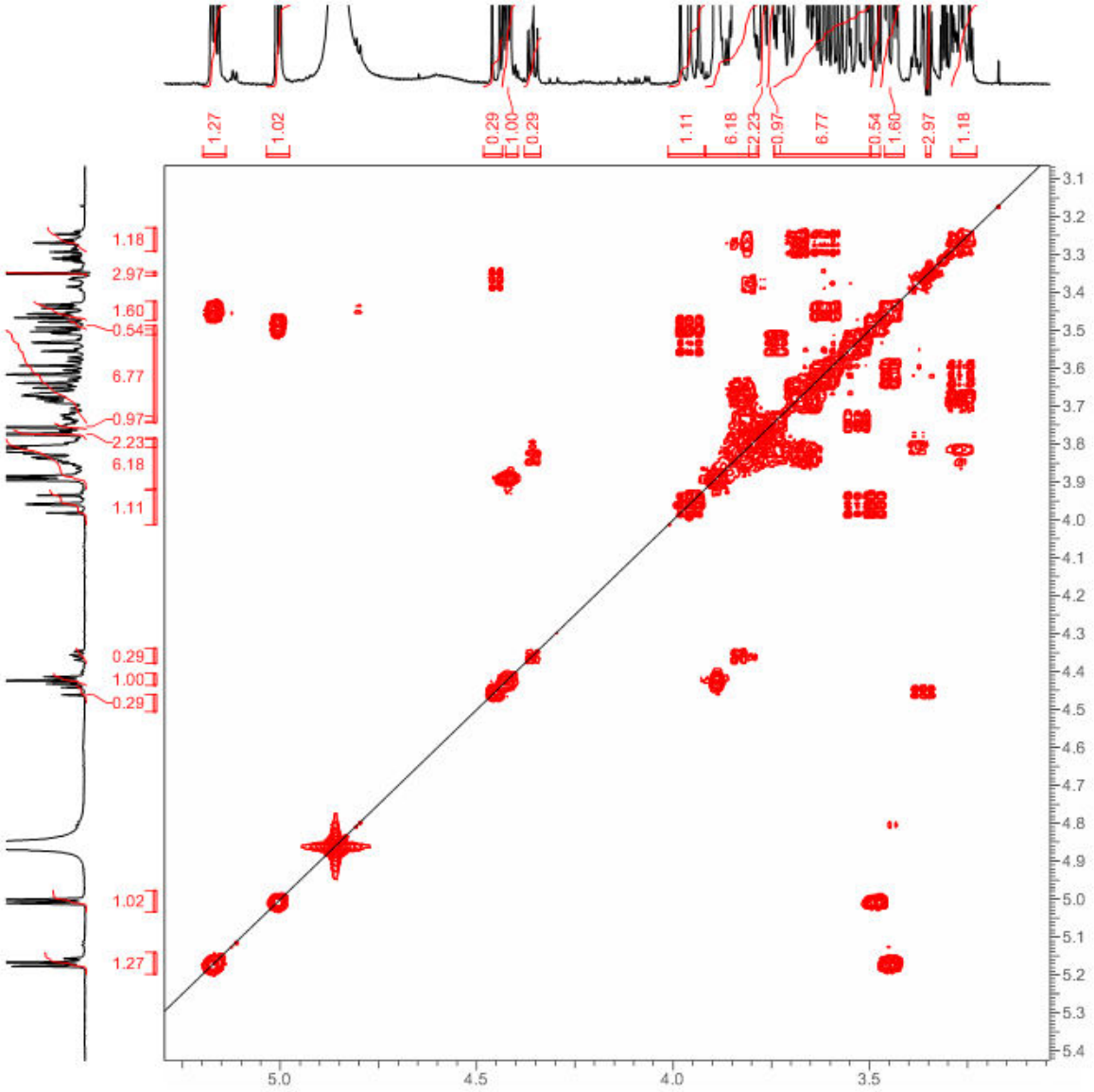
^{1}H-^{1}H COSY (400 - 400 MHz, CD_3OD):





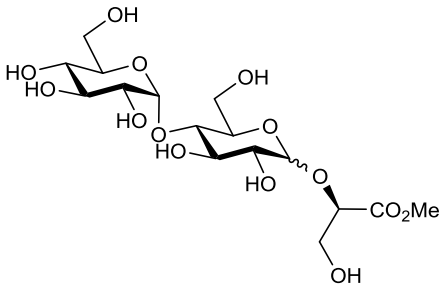
Appendix A

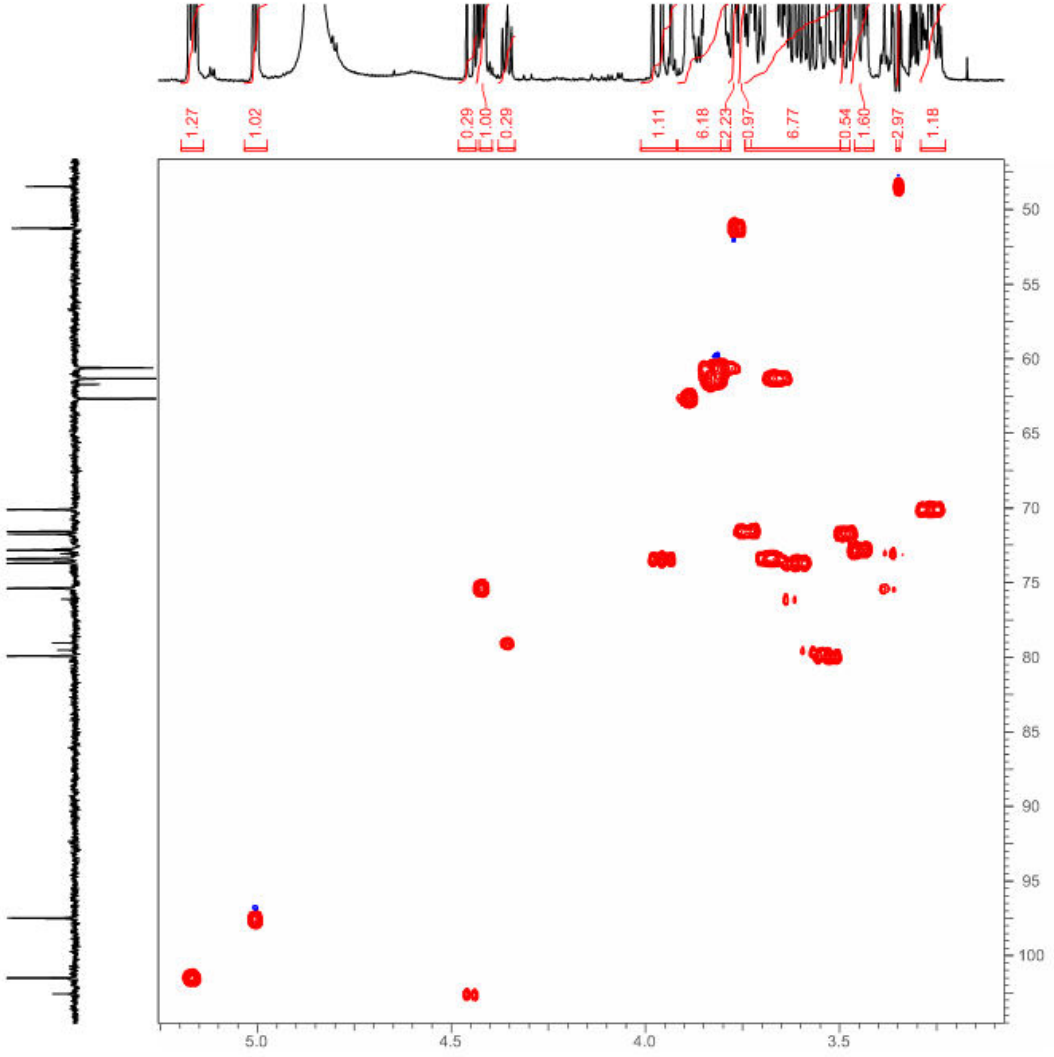

OH
HO
HO
O
HO
OH
O
HO
HO
O
CO2Me
OH


1.27
1.02
0.29
1.00
0.29
1.11
6.18
2.23
0.97
6.77
0.54
1.60
2.97
1.18
5.0
4.5
4.0
3.5

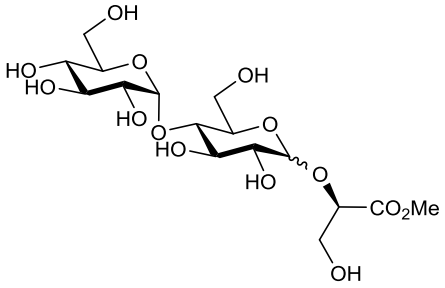
Appendix A

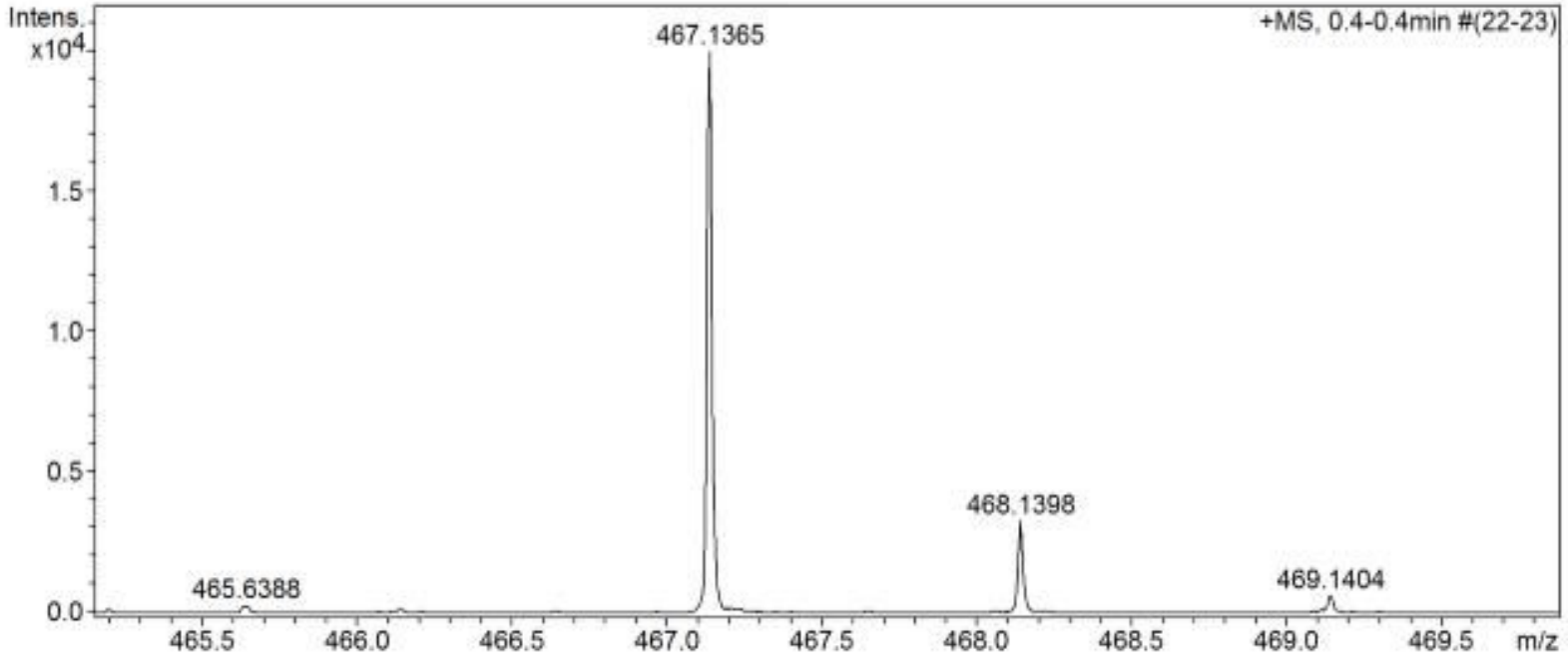
^{1}H-^{13}C HSQC (100 - 400 MHz, CD_3OD):





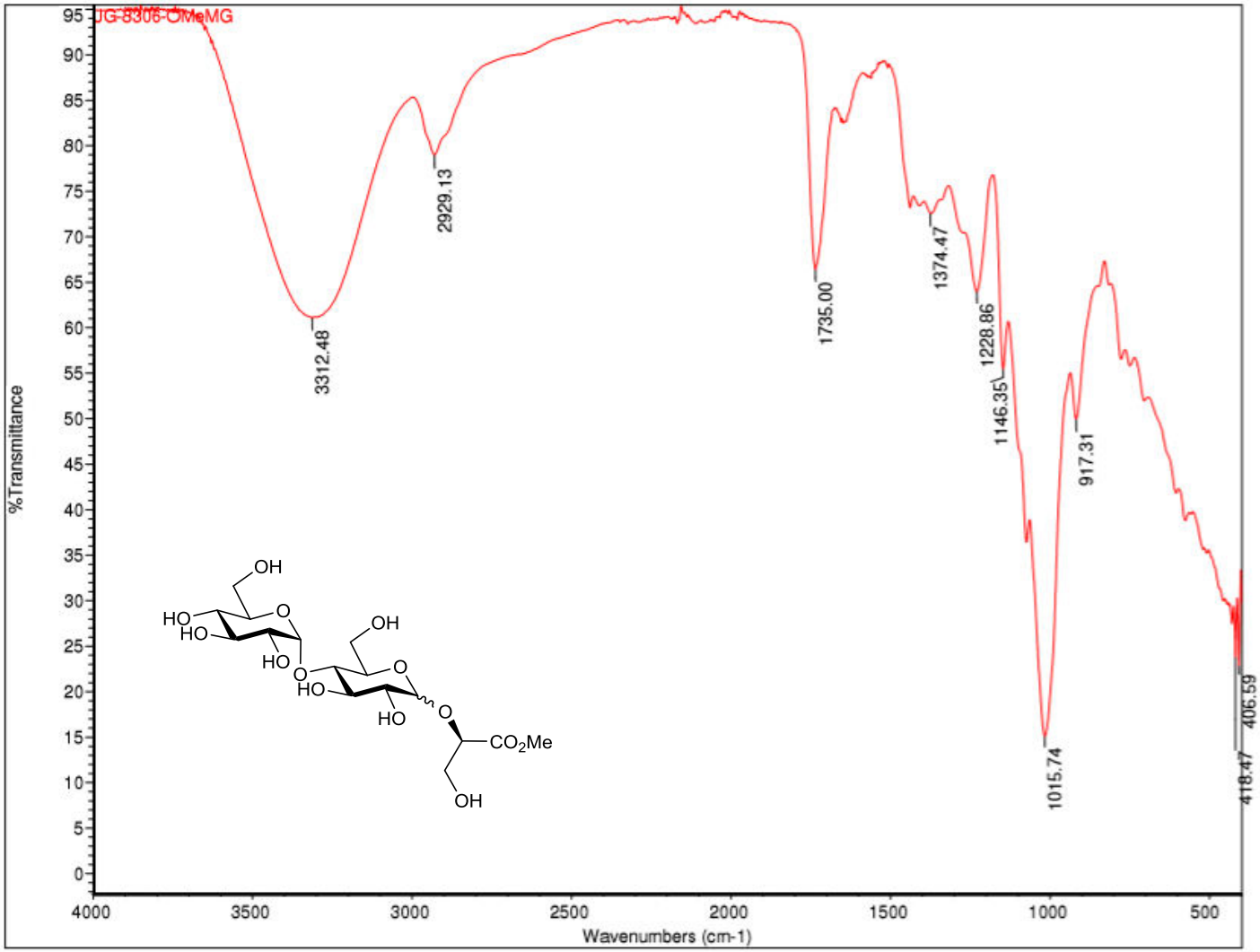
HR-MS (ESI+) m/z:


OH
HO
HO
O
HO
O
OH
O
HO
HO
O
CO2Me
OH


Intens.
x10^4
+MS, 0.4-0.4min #(22-23)
467.1365
468.1398
469.1404
465.6388
1.5
1.0
0.5
0.0
465.5
466.0
466.5
467.0
467.5
468.0
468.5
469.0
469.5
m/z

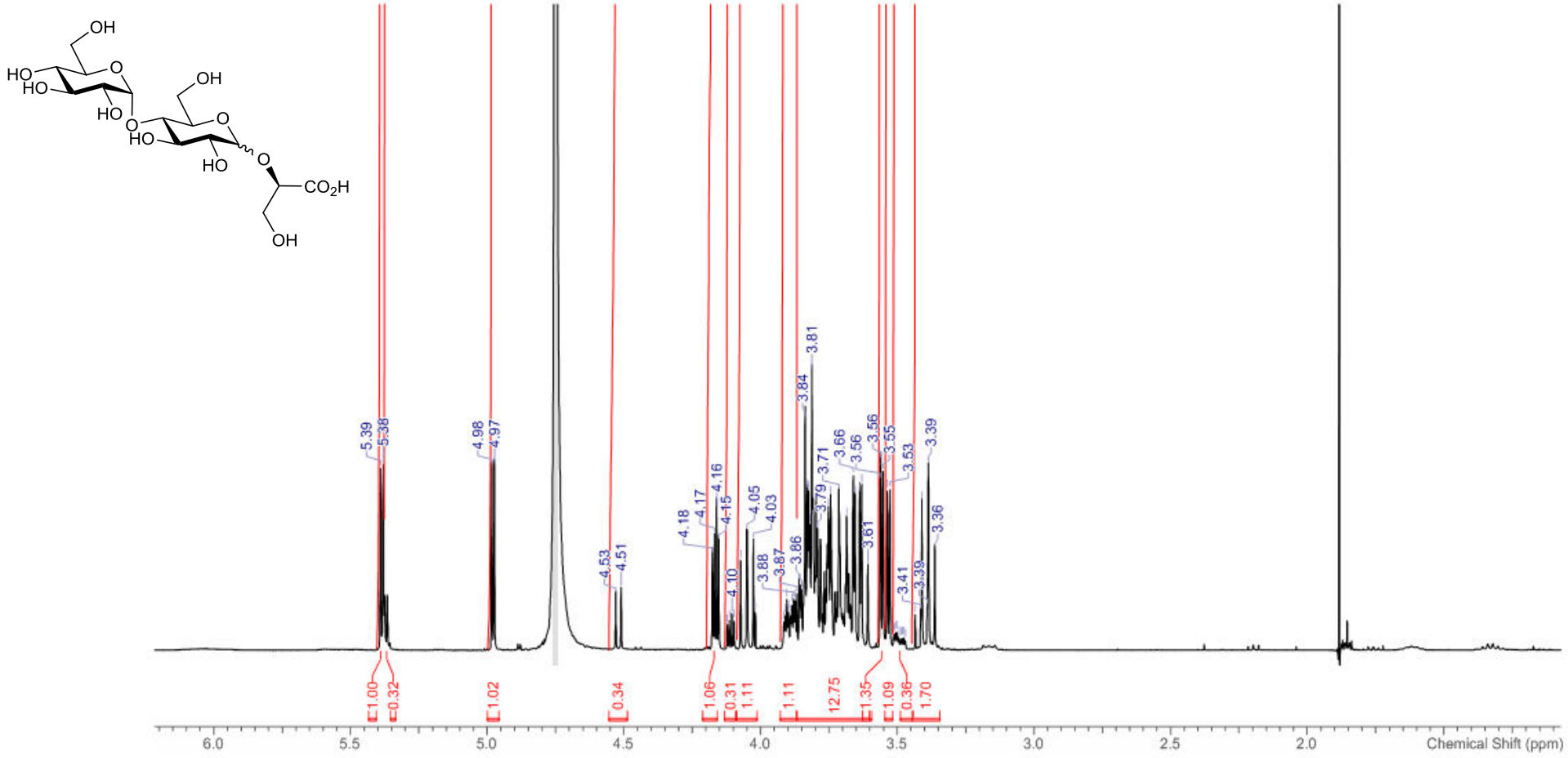
Appendix A

IR:

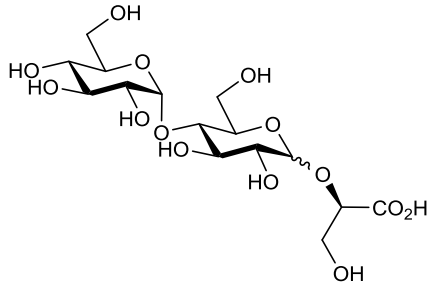


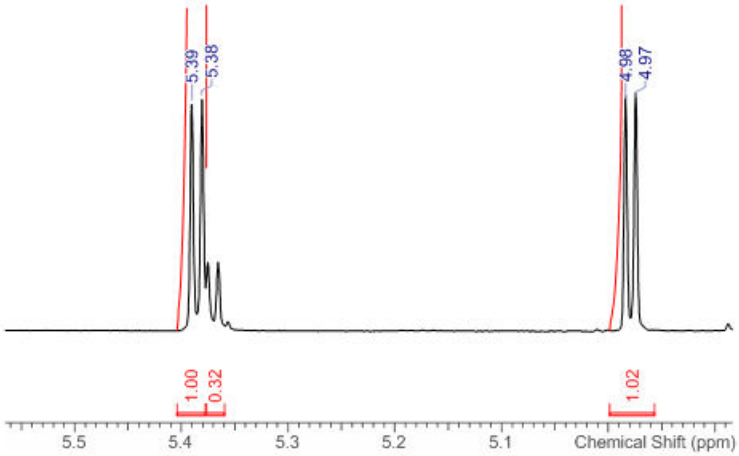
A.5 (2*R*)-2-*O*-(α/β-D-maltopyranosyl)-2,3-dihydroxypropanoate (1.38α/β)

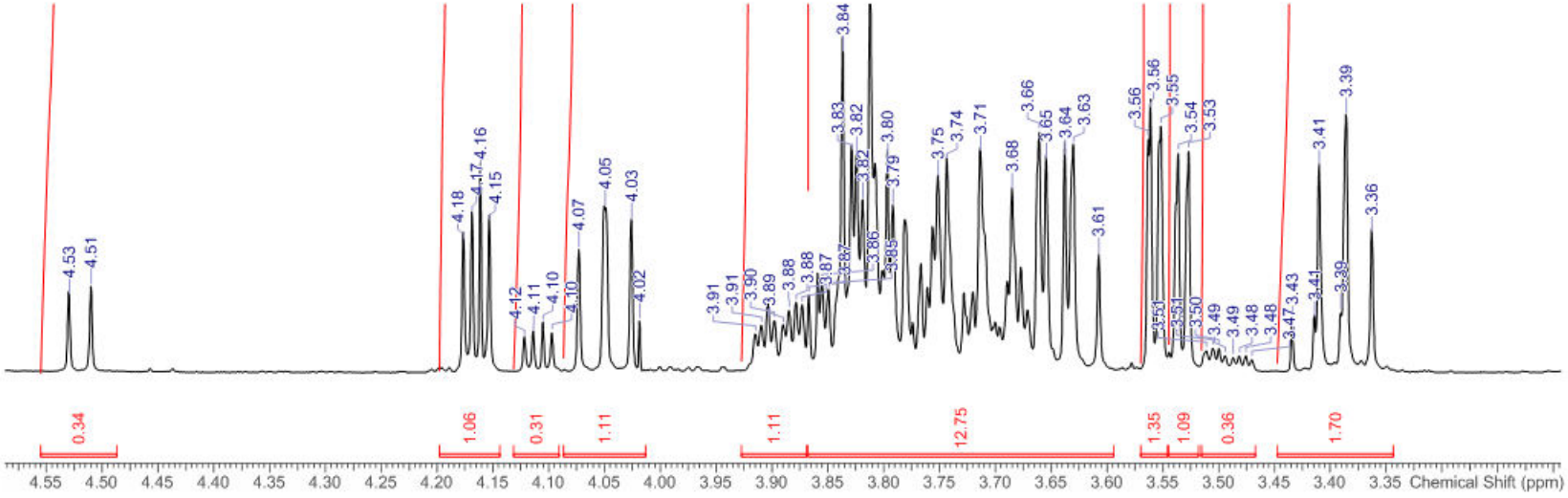
^{1}H NMR (400 MHz, D_2O):



Appendix A

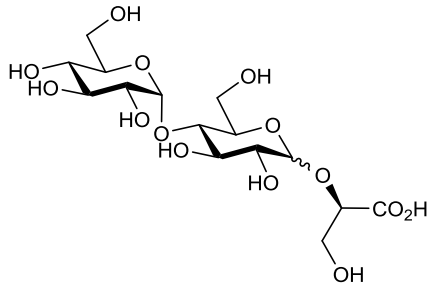

OH
HO
HO
O
HO
OH
O
HO
O
HO
O
CO2H
OH

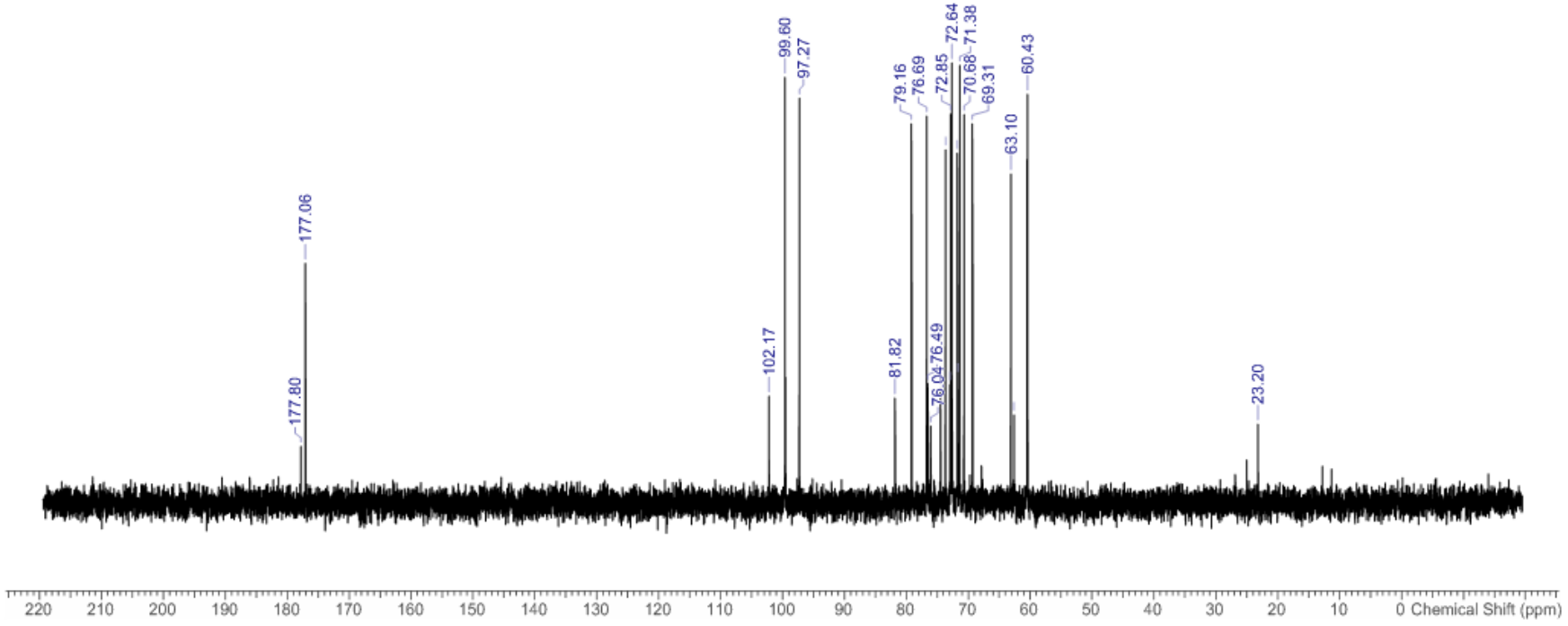

5.39
5.38
4.98
4.97
1.00
0.32
1.02
5.5
5.4
5.3
5.2
5.1
Chemical Shift (ppm)


4.53
4.51
4.18
4.17
4.16
4.15
4.12
4.11
4.10
4.10
4.07
4.05
4.03
4.02
3.91
3.91
3.90
3.89
3.88
3.88
3.87
3.84
3.83
3.82
3.82
3.80
3.79
3.75
3.74
3.71
3.68
3.66
3.65
3.64
3.63
3.61
3.56
3.56
3.55
3.54
3.53
3.51
3.50
3.49
3.49
3.48
3.48
3.47
3.43
3.41
3.41
3.39
3.39
3.36
0.34
1.06
0.31
1.11
1.11
12.75
1.35
1.09
0.36
1.70
4.55
4.50
4.45
4.40
4.35
4.30
4.25
4.20
4.15
4.10
4.05
4.00
3.95
3.90
3.85
3.80
3.75
3.70
3.65
3.60
3.55
3.50
3.45
3.40
3.35
Chemical Shift (ppm)

Appendix A

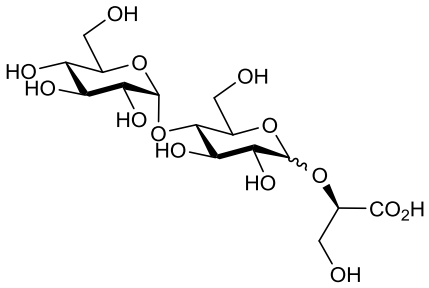
^{13}C NMR (101 MHz, D_2O):

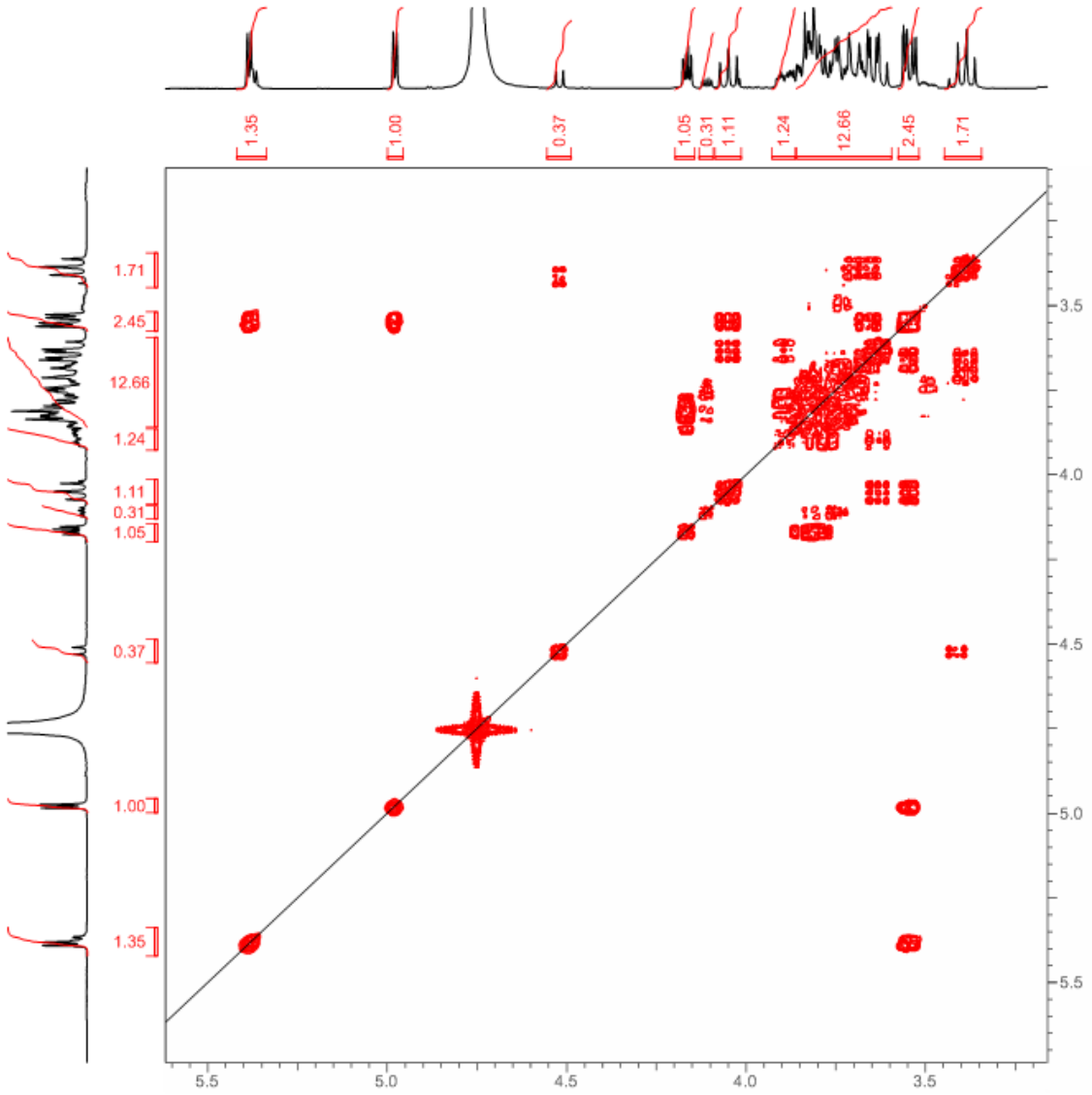




Appendix A

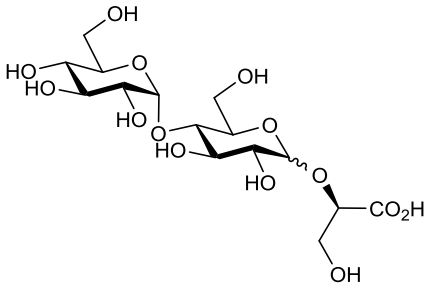
^{1}H-^{1}H COSY (400 - 400 MHz, D_2O):

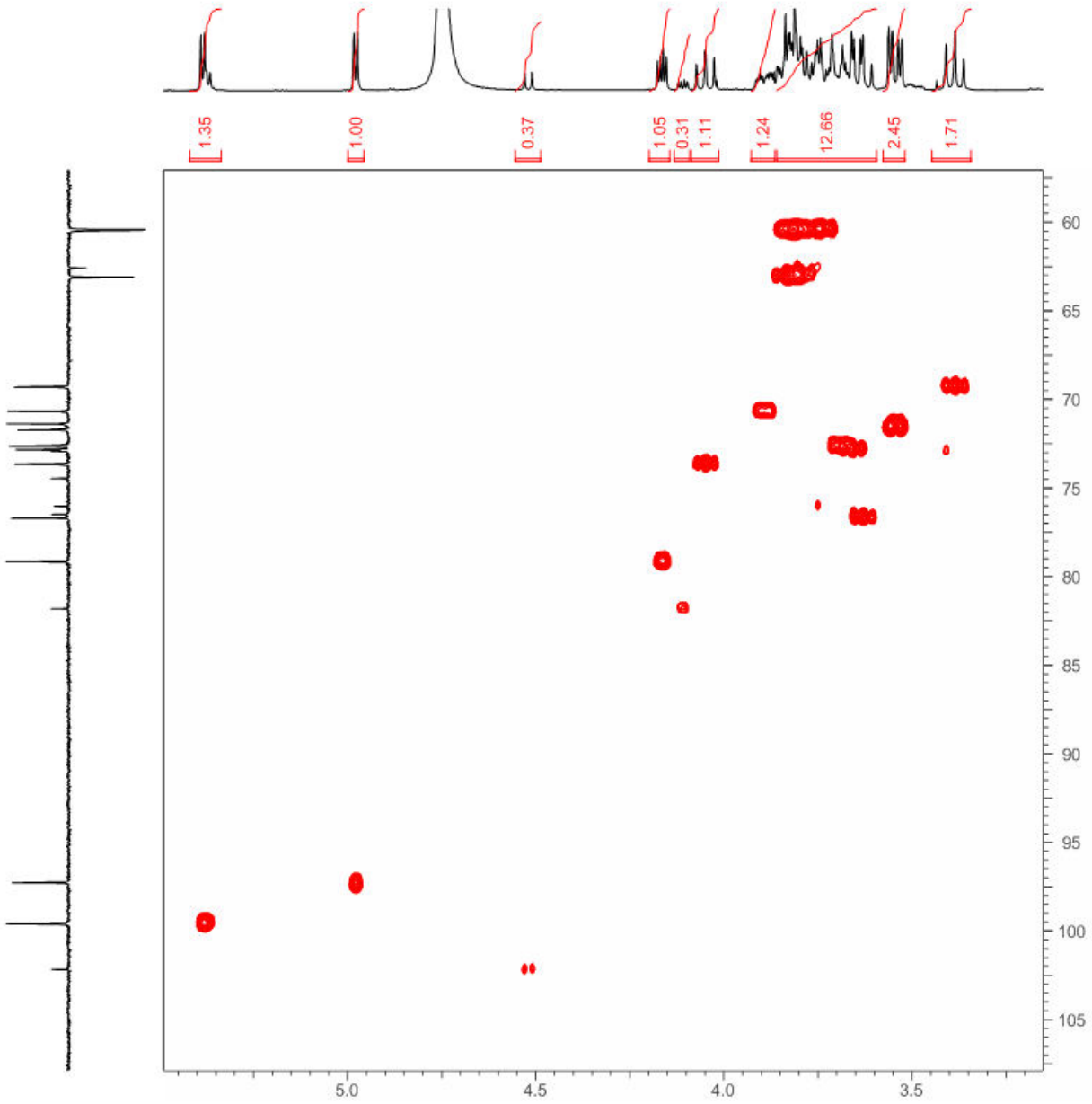




Appendix A

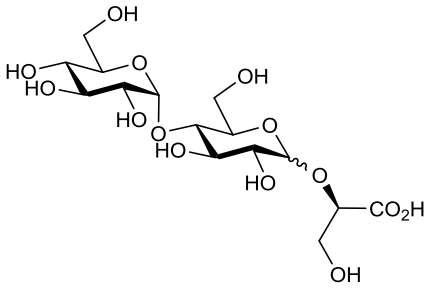
^{1}H-^{13}C HSQC (100 - 400 MHz, D_2O):

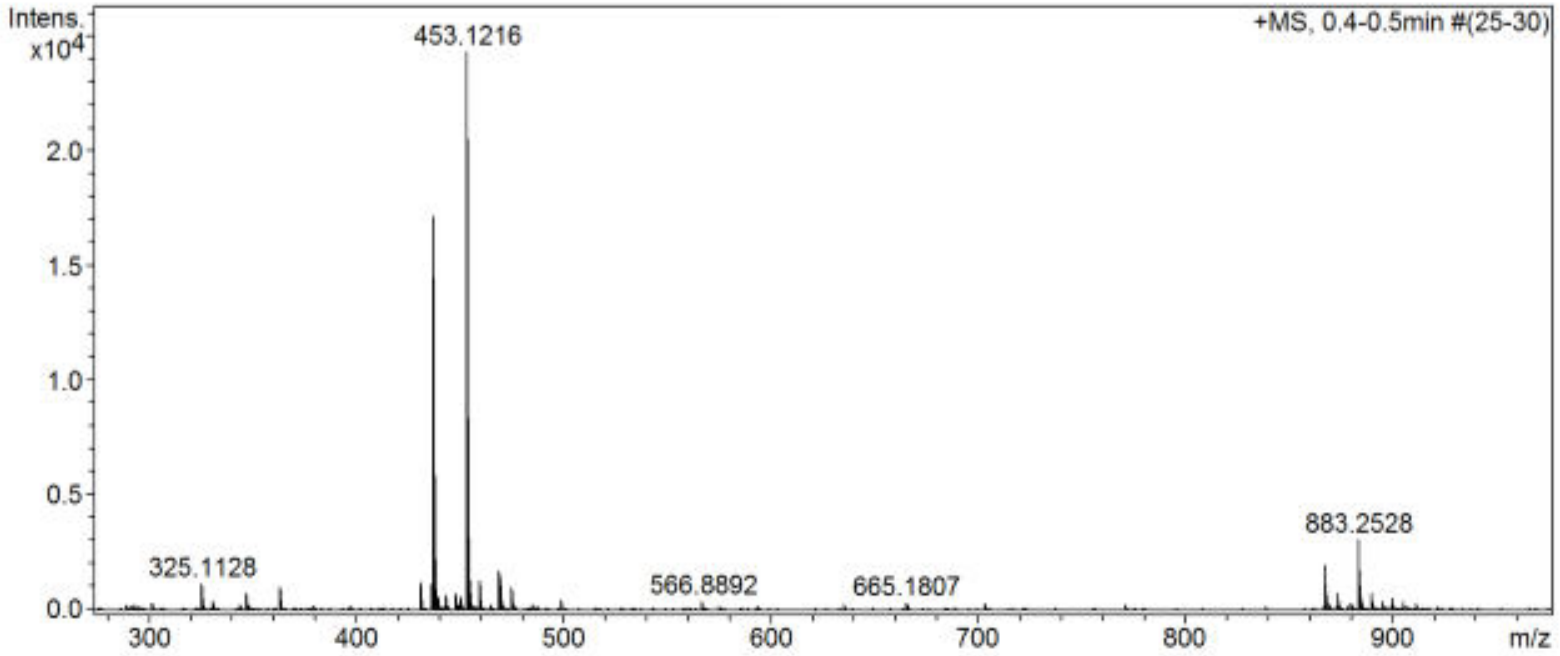




Appendix A

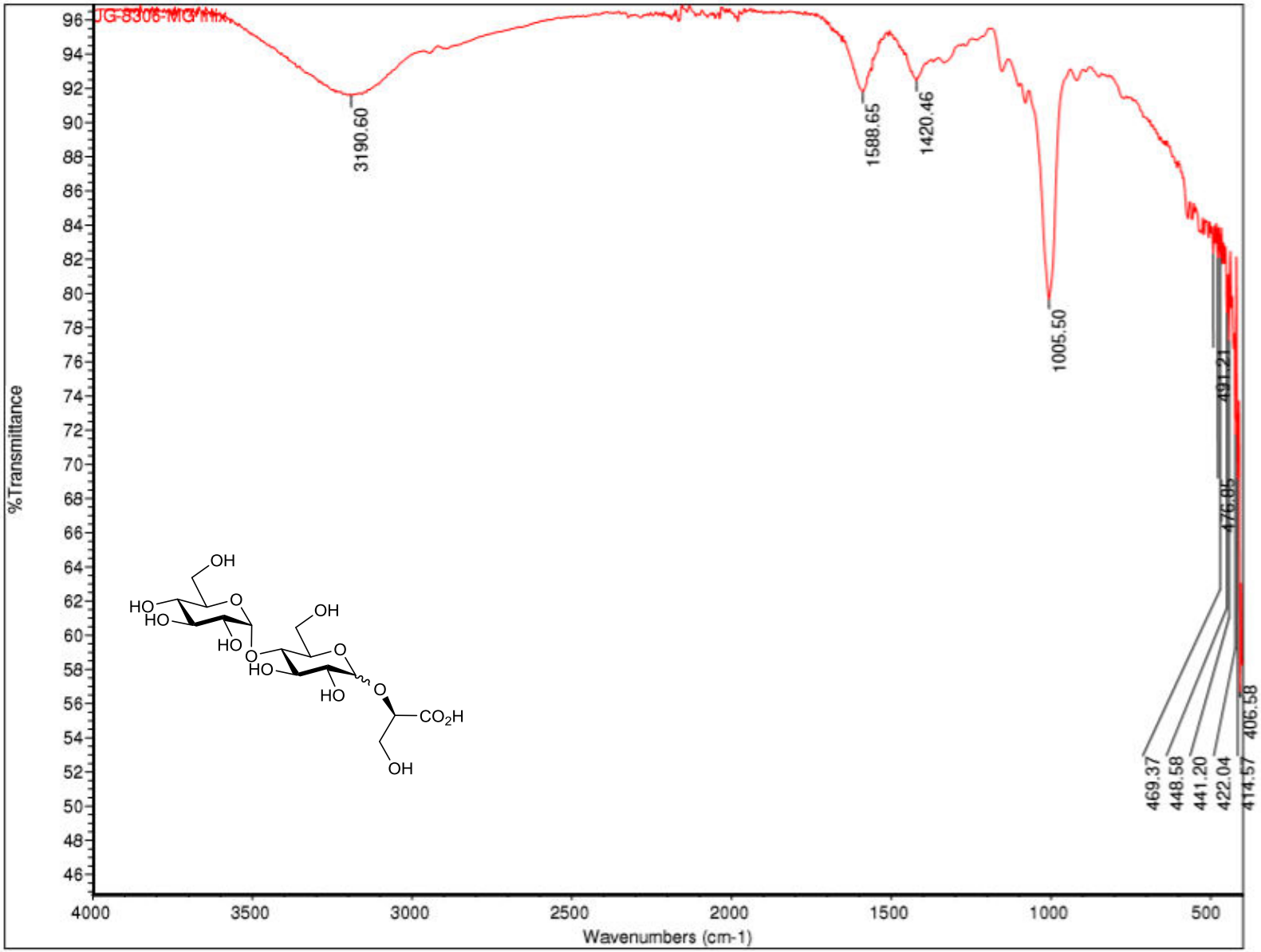
HR-MS (ESI+) m/z:





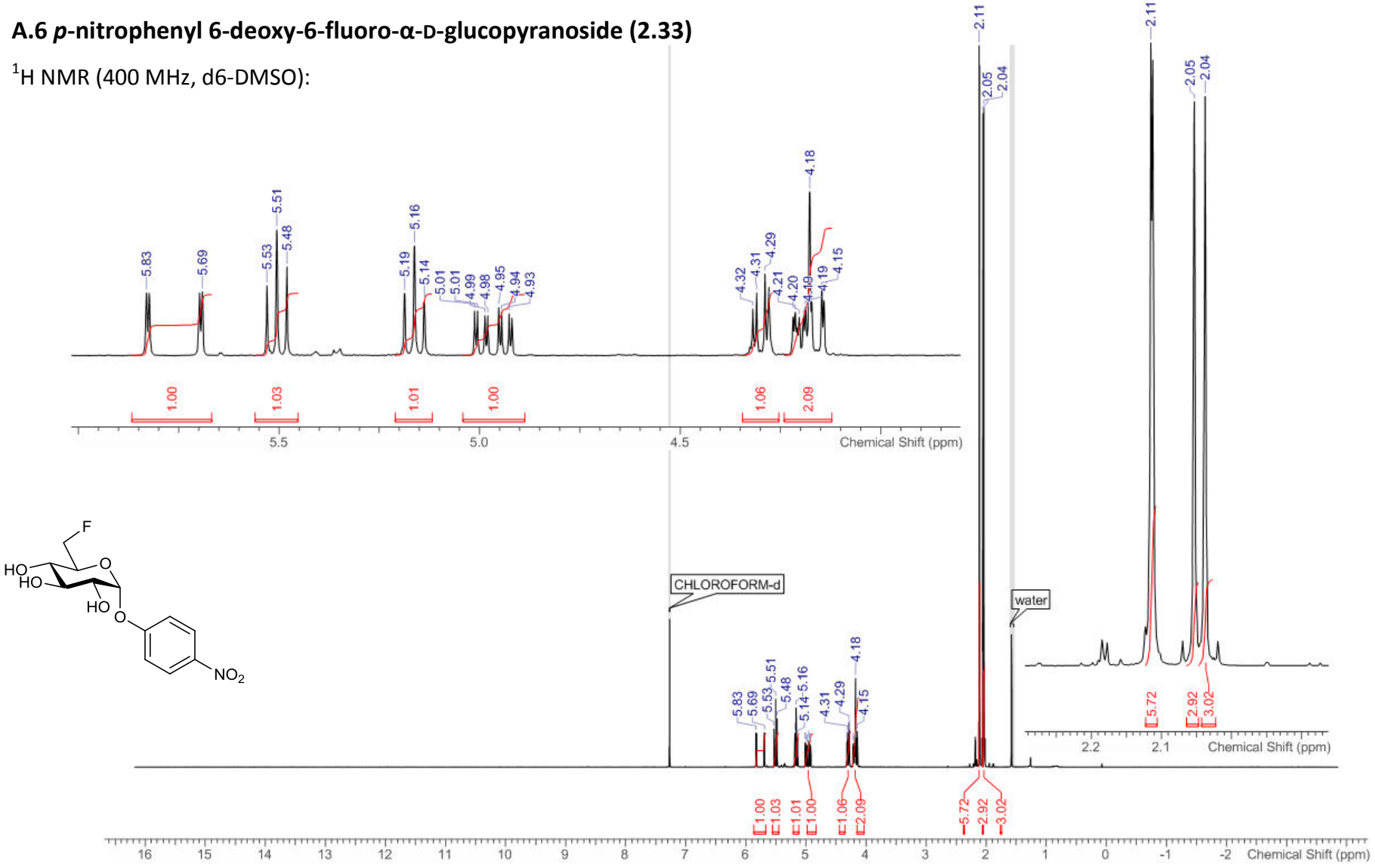
Appendix A

IR:

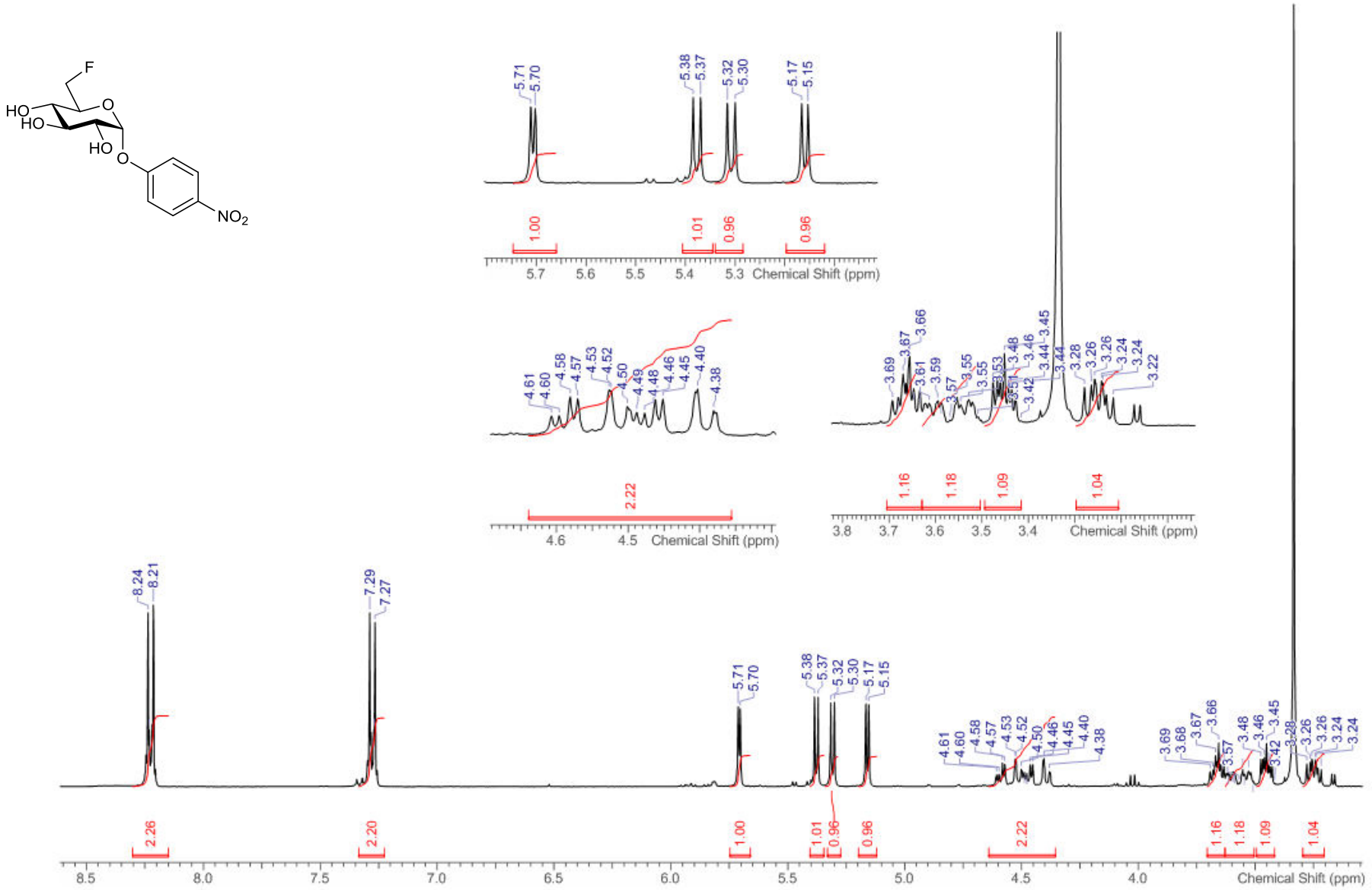


A.6 *p*-nitrophenyl 6-deoxy-6-fluoro-α-D-glucopyranoside (2.33)

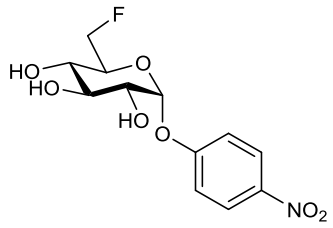
^{1}H NMR (400 MHz, d6-DMSO):



Appendix A

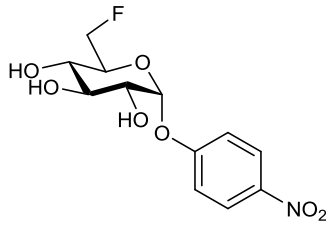

F
HO
HO
HO
O
O
NO2
5.71
5.70
5.38
5.37
5.32
5.30
5.17
5.15
1.00
1.01
0.96
0.96
Chemical Shift (ppm)
4.61
4.60
4.58
4.57
4.53
4.52
4.50
4.49
4.48
4.46
4.45
4.40
4.38
2.22
Chemical Shift (ppm)
3.69
3.67
3.66
3.61
3.59
3.57
3.55
3.55
3.53
3.51
3.48
3.46
3.44
3.45
3.44
3.28
3.26
3.26
3.24
3.24
3.22
1.16
1.18
1.09
1.04
Chemical Shift (ppm)
8.24
8.21
7.29
7.27
2.26
2.20
1.00
1.01
0.96
0.96
2.22
1.16
1.18
1.09
1.04
Chemical Shift (ppm)

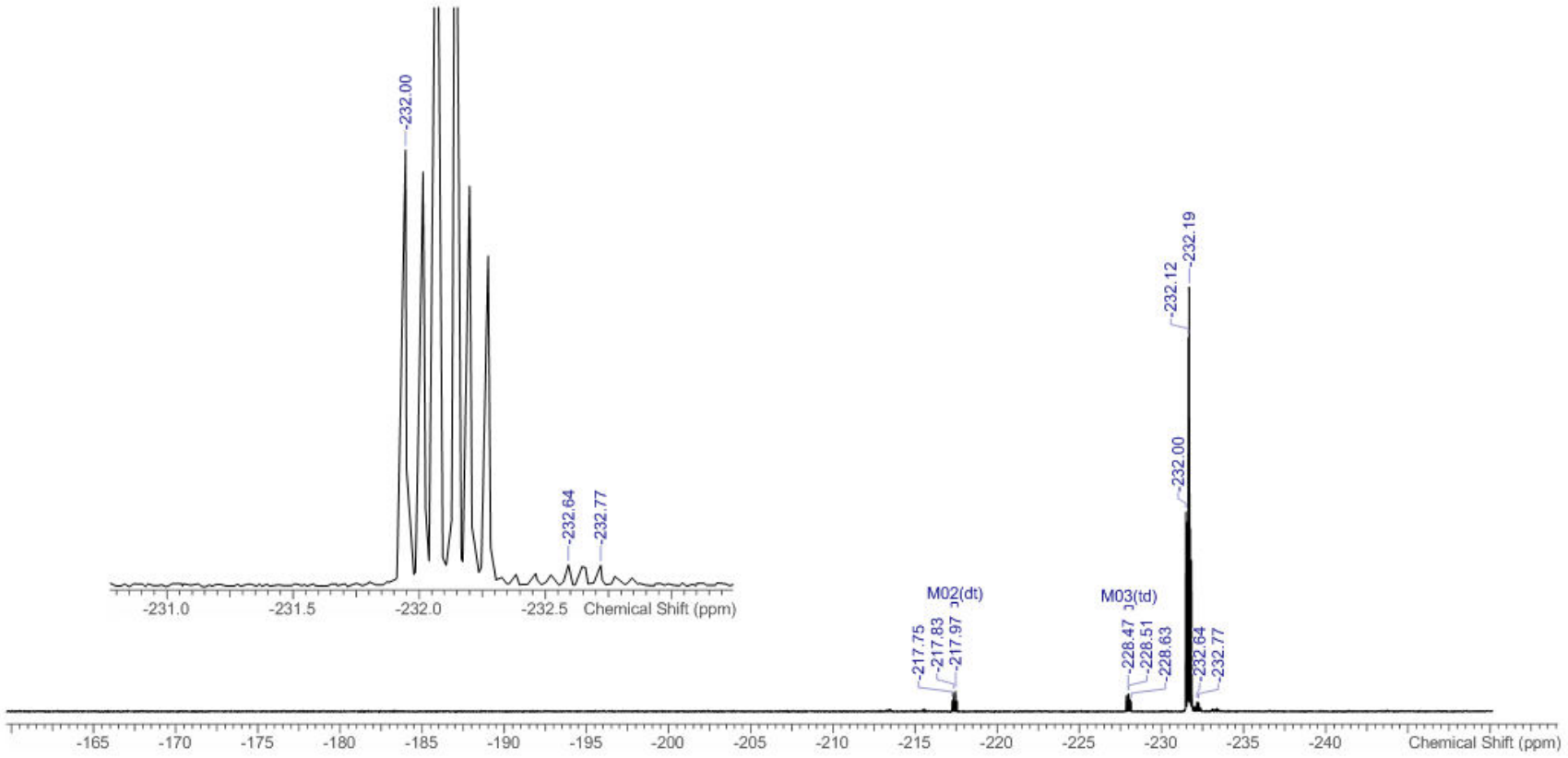
^{13}C NMR (101 MHz, d6-DMSO):



Appendix A

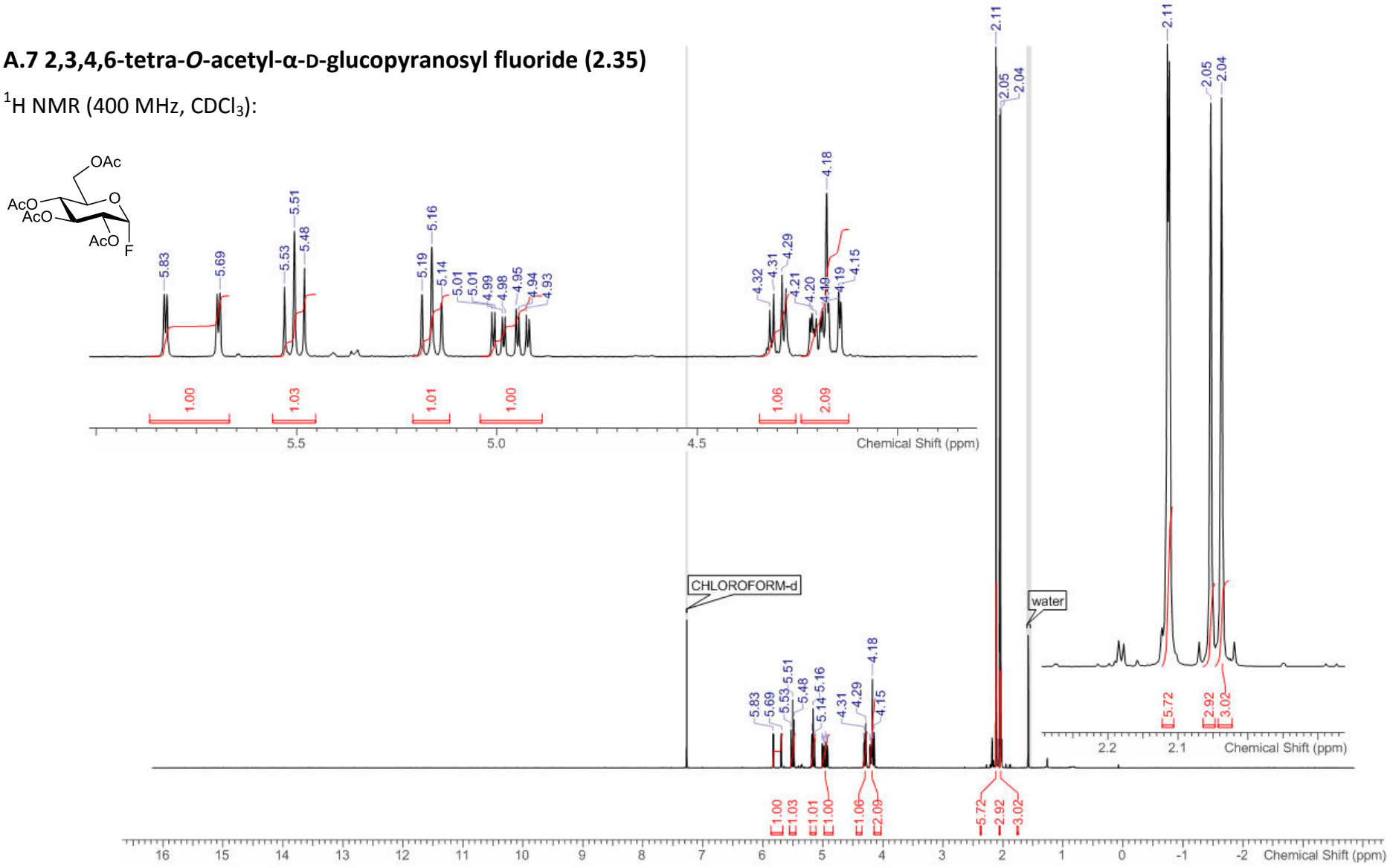
^{19}F NMR (376 MHz, d6-DMSO):





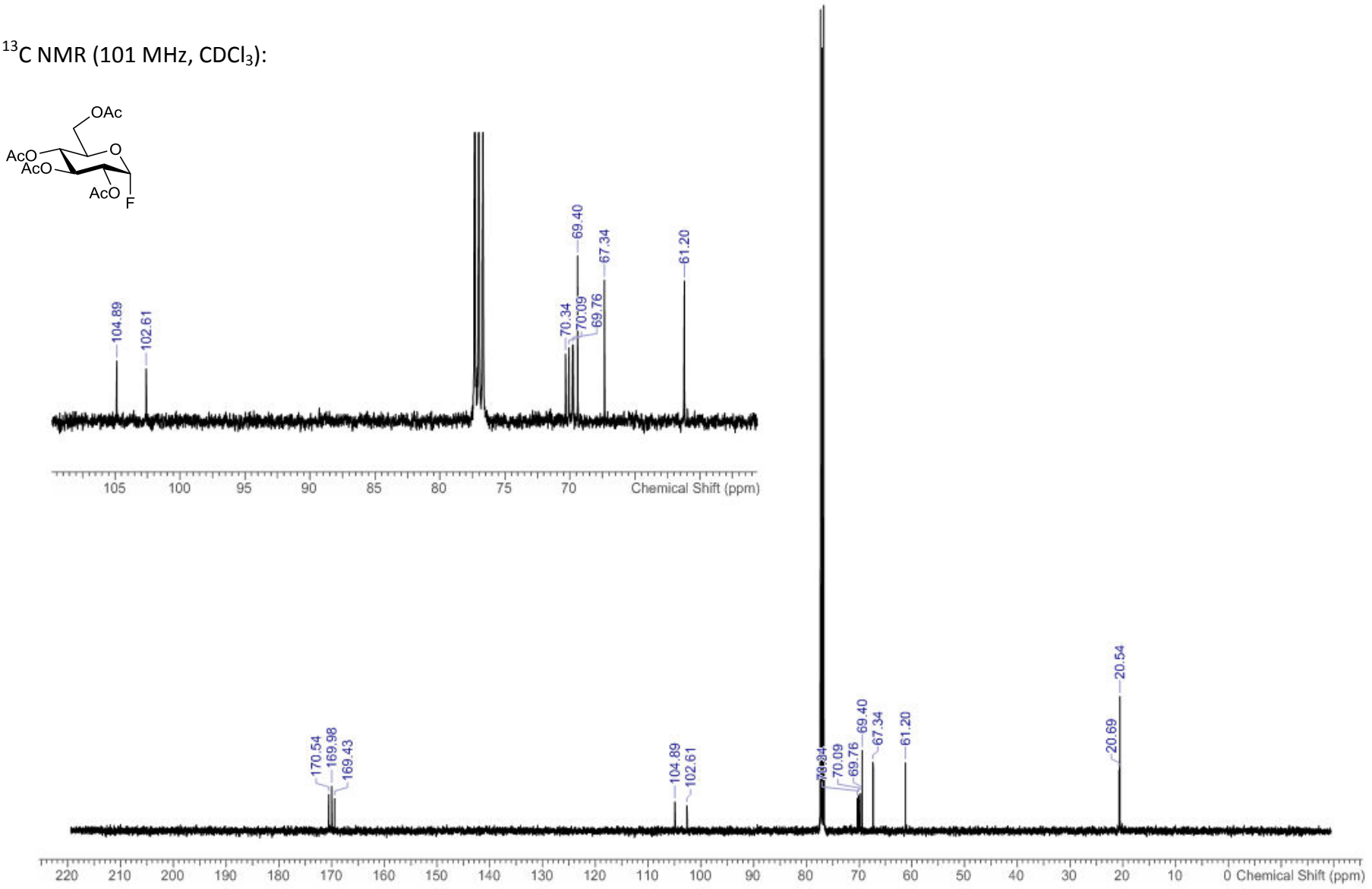
A.7 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl fluoride (2.35)

^{1}H NMR (400 MHz, $CDCl_3$):



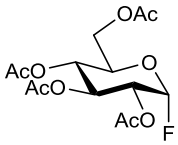
Appendix A

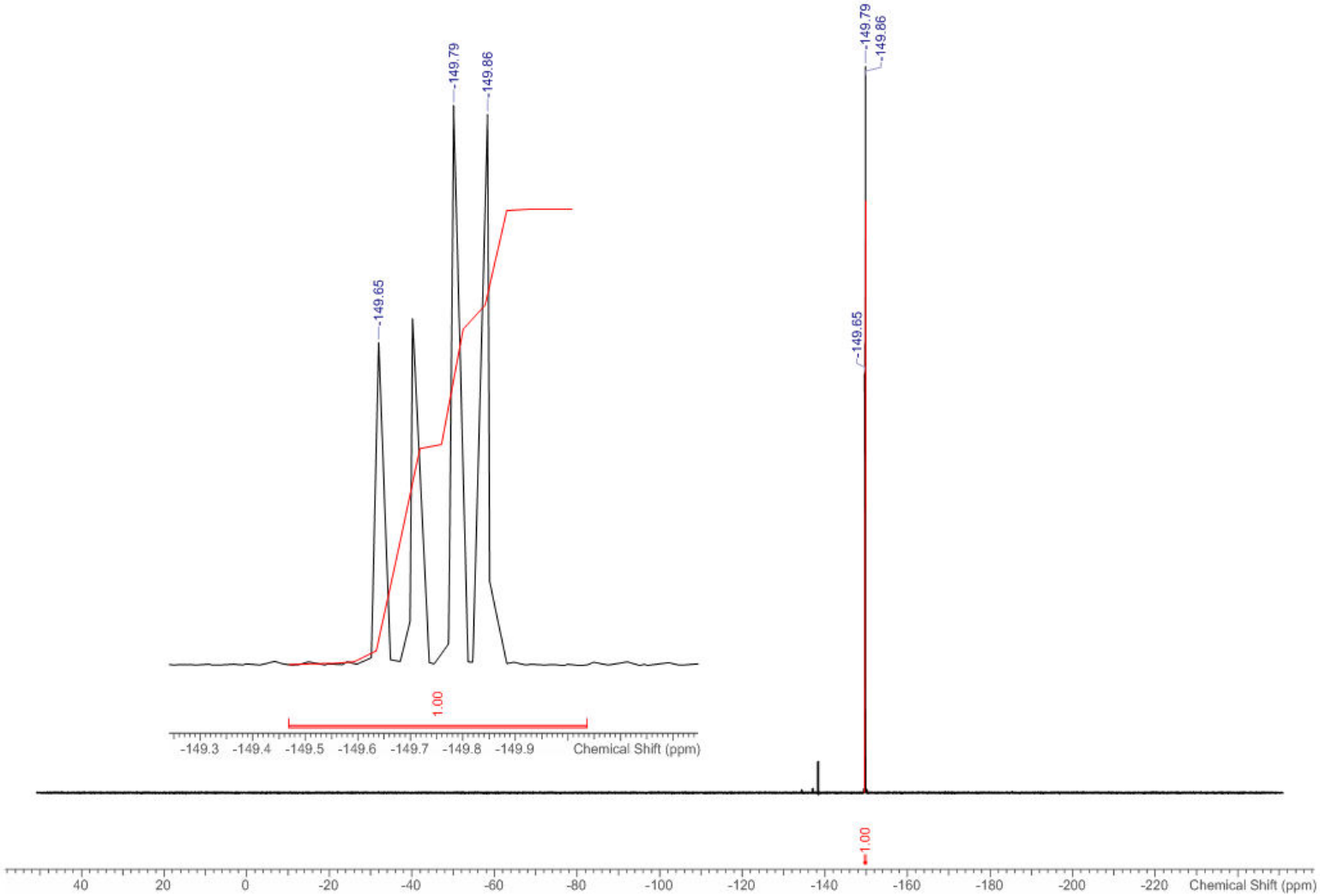
^{13}C NMR (101 MHz, $CDCl_3$):



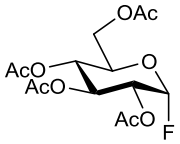
Appendix A

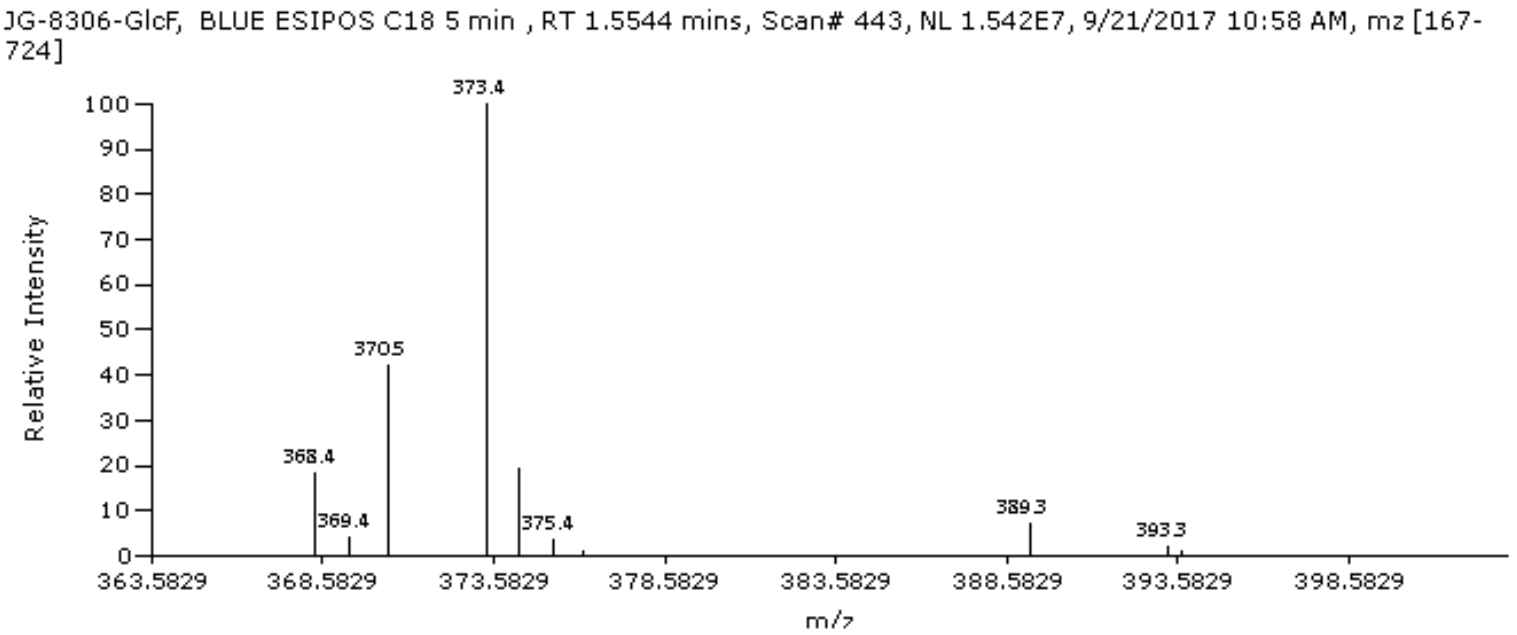
^{19}F NMR (376 MHz, $CDCl_3$):





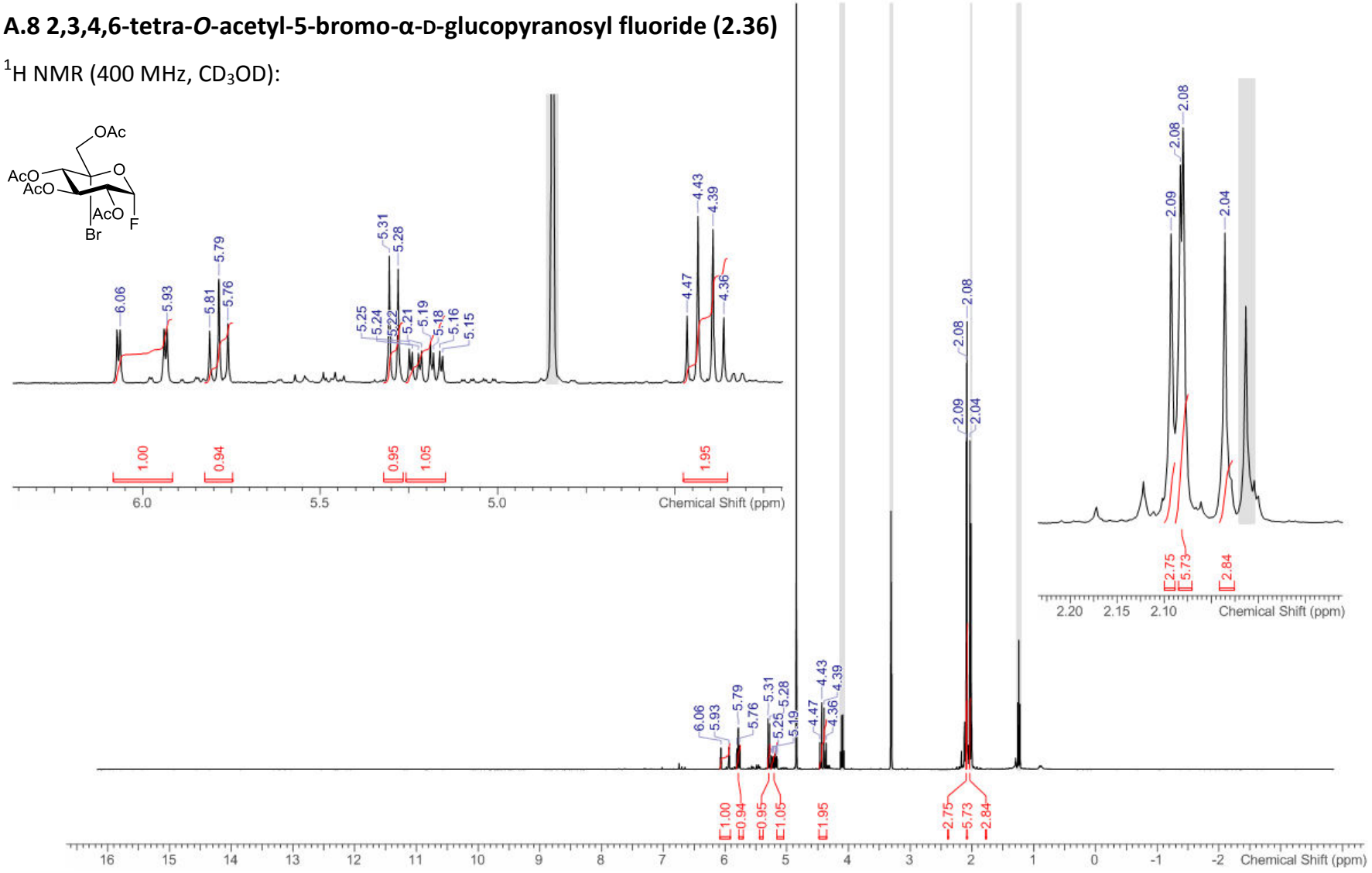
LR-MS (ESI+) m/z:


OAc
AcO
AcO
O
AcO
F


JG-8306-GlcF, BLUE ESIPOS C18 5 min , RT 1.5544 mins, Scan# 443, NL 1.542E7, 9/21/2017 10:58 AM, mz [167-724]
Relative Intensity
m/z
373.4
370.5
368.4
369.4
375.4
389.3
393.3
100
90
80
70
60
50
40
30
20
10
0
363.5829
368.5829
373.5829
378.5829
383.5829
388.5829
393.5829
398.5829

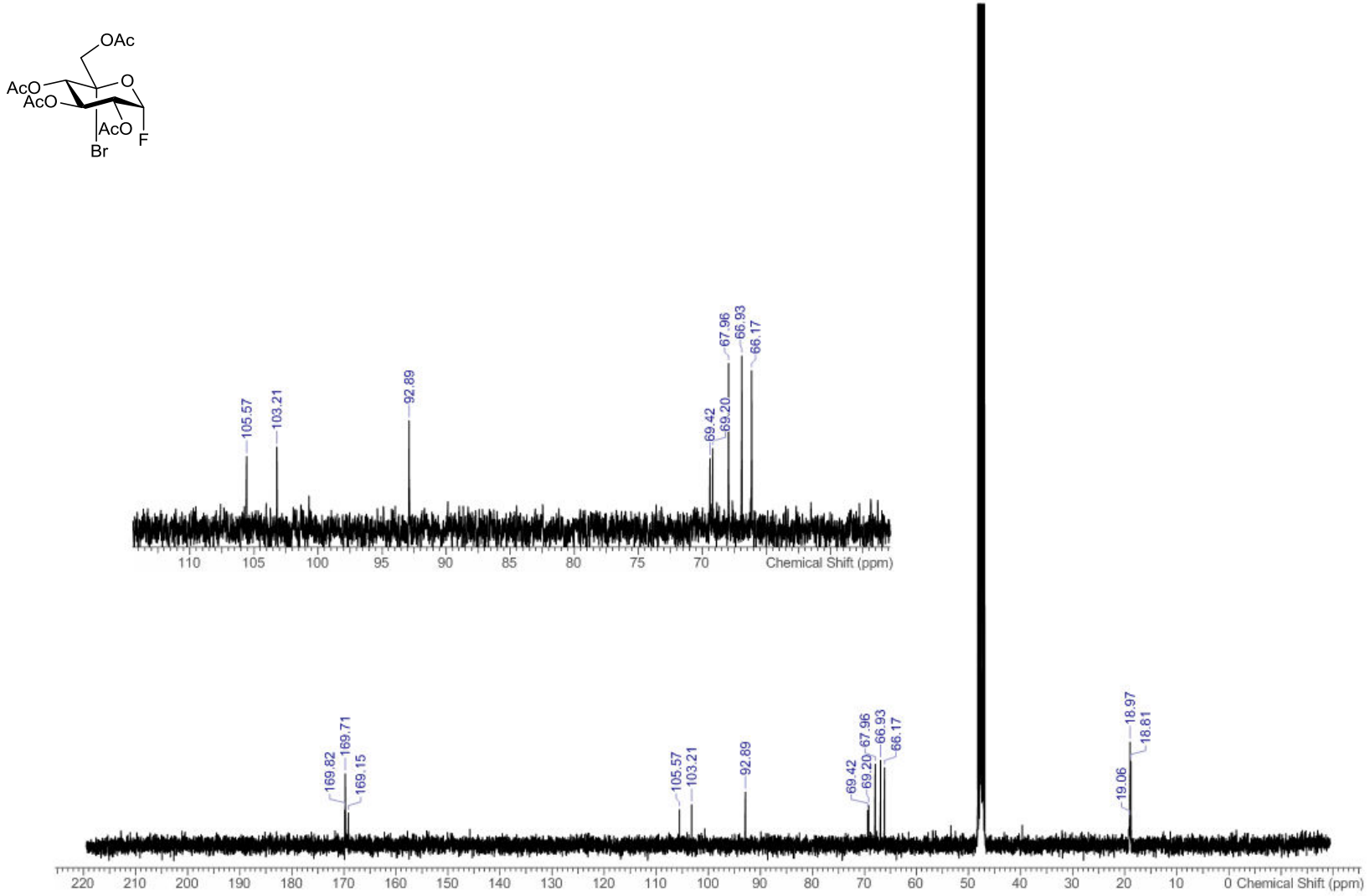
A.8 2,3,4,6-tetra-*O*-acetyl-5-bromo-α-D-glucopyranosyl fluoride (2.36)

^{1}H NMR (400 MHz, CD_3OD):

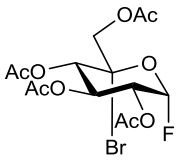


Appendix A

^{13}C NMR (101 MHz, CD_3OD):

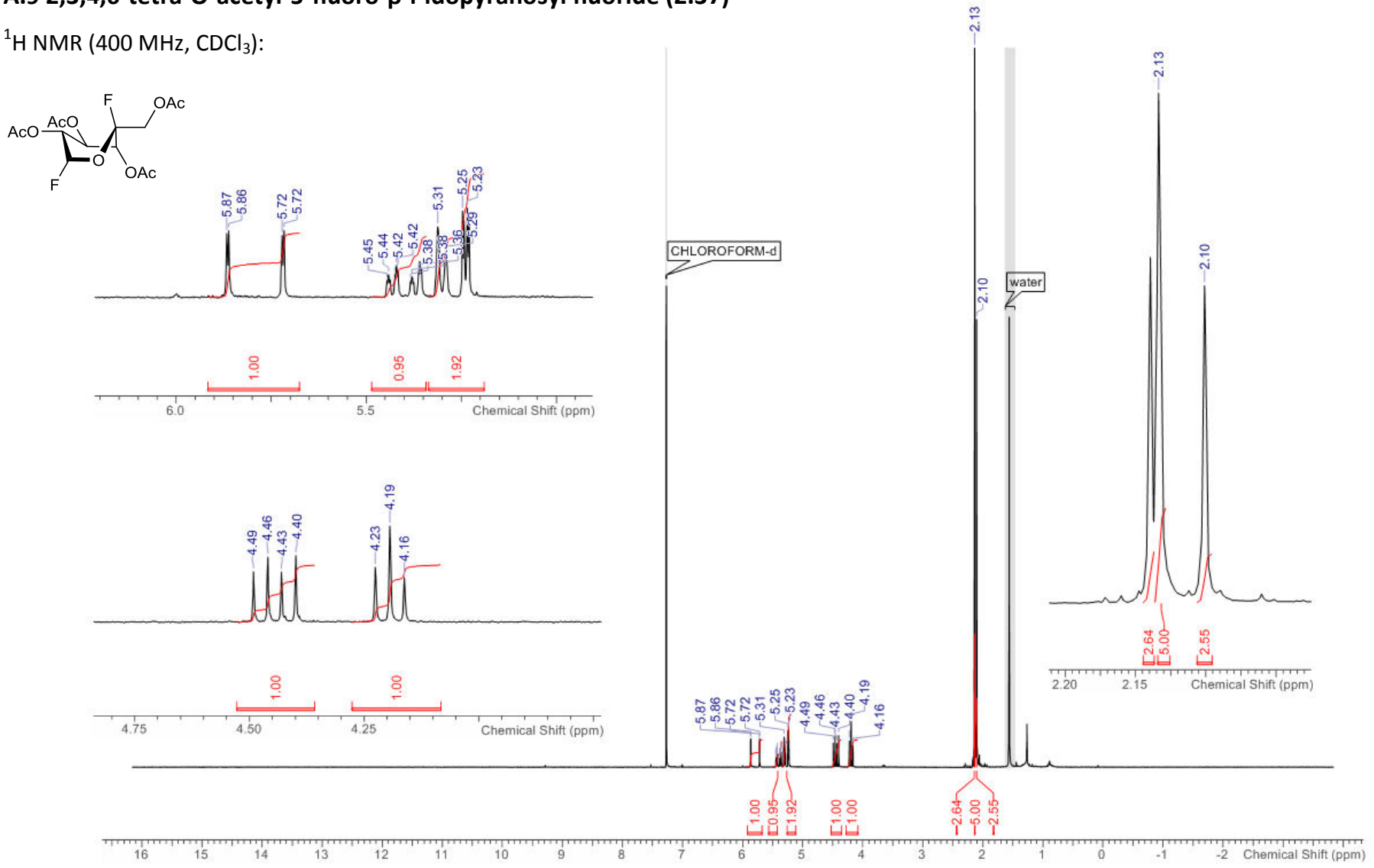


^{19}F NMR (376 MHz, CD_3OD):

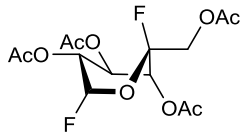


A.9 2,3,4,6-tetra-O-acetyl-5-fluoro-β-l-idopyranosyl fluoride (2.37)

^{1}H NMR (400 MHz, $CDCl_3$):

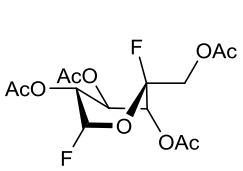


^{13}C NMR (101 MHz, $CDCl_3$):



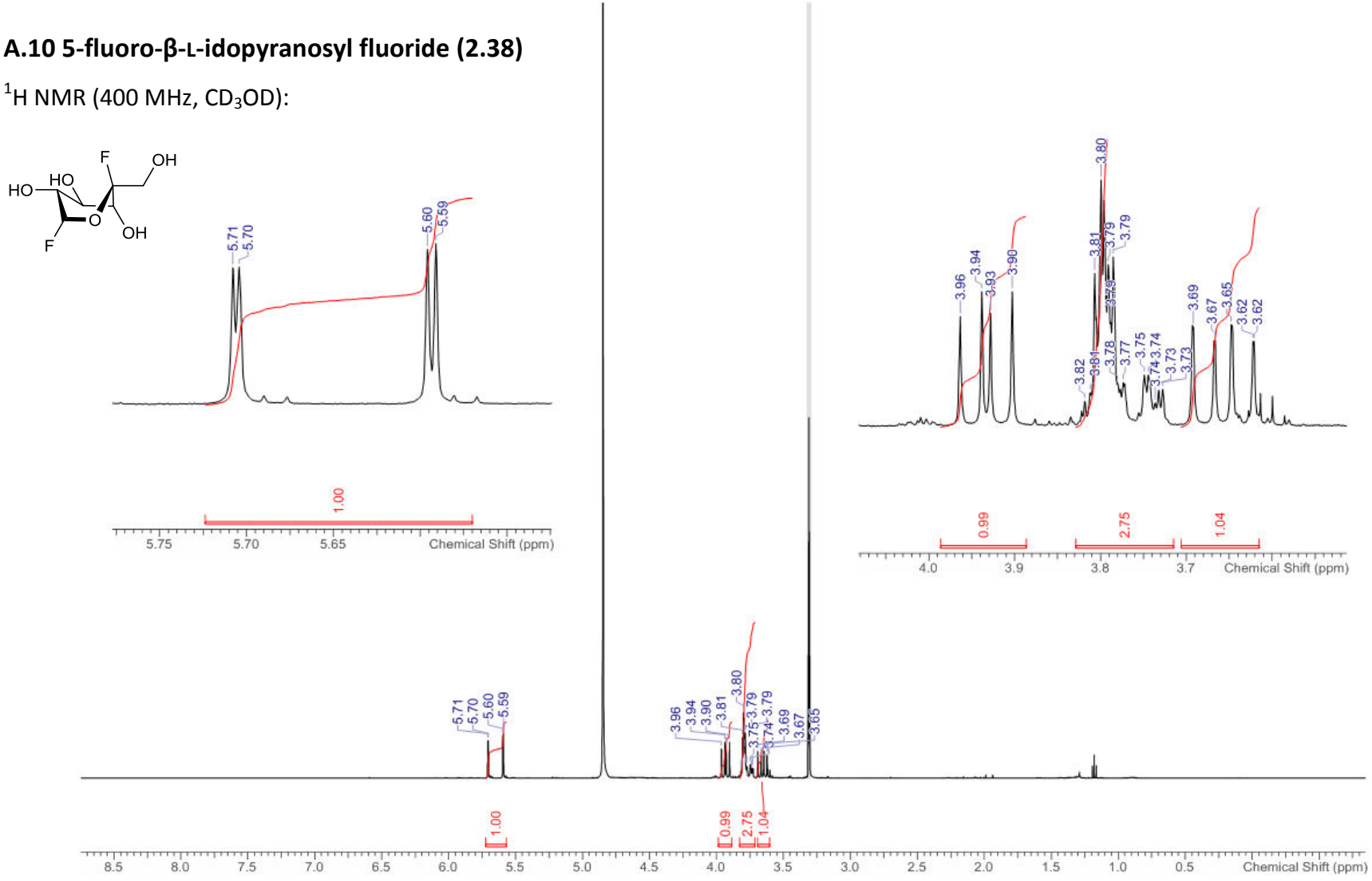
Appendix A

^{19}F NMR (376 MHz, $CDCl_3$):



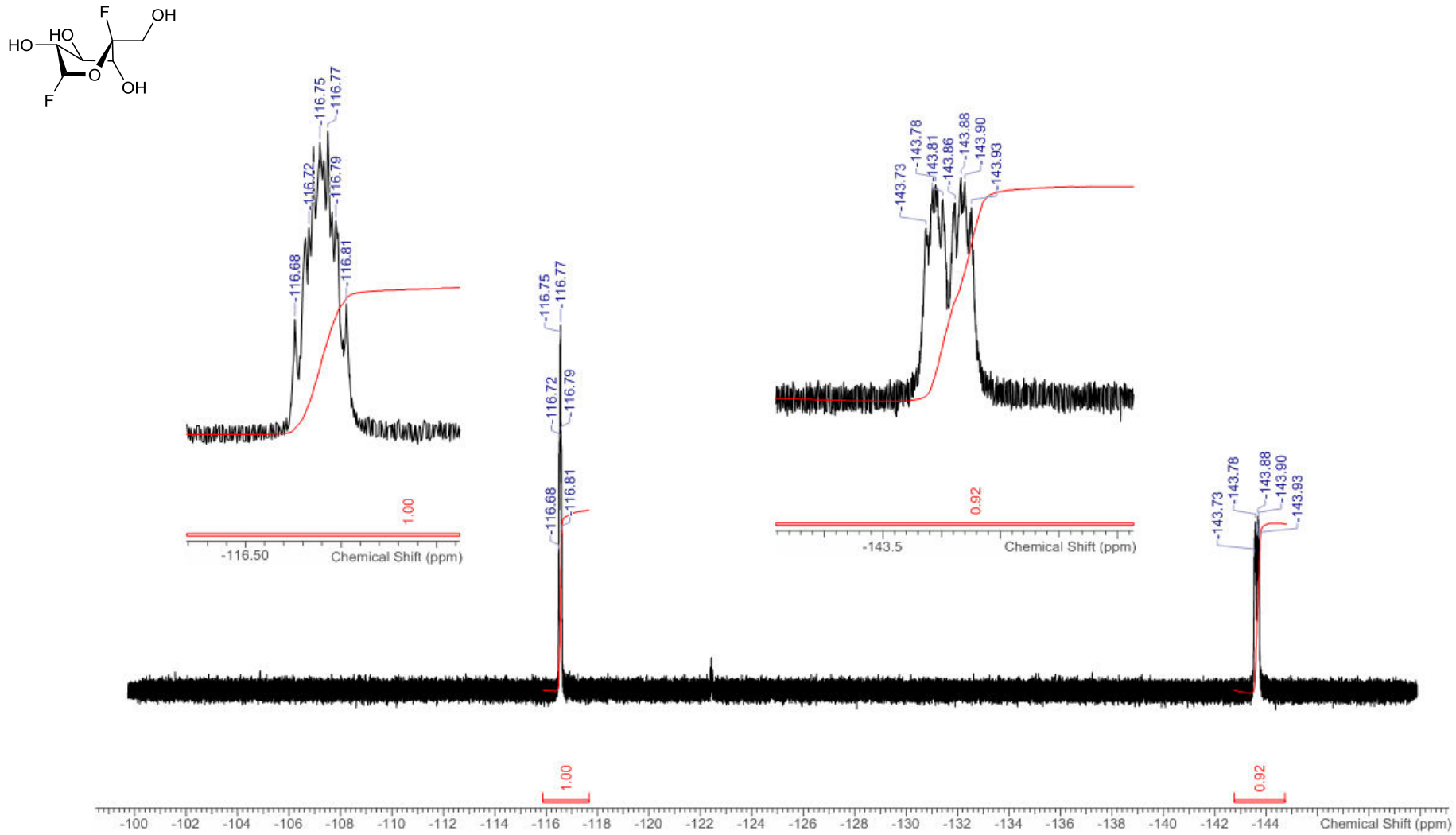
A.10 5-fluoro-β-L-idopyranosyl fluoride (2.38)

^{1}H NMR (400 MHz, CD_3OD):



Appendix A

^{19}F NMR (376 MHz, CD_3OD):



Appendix A

Appendix B: Additional data for Chapter 3

B.1 Sequencing results of pET28a(+)::Rv3031 amplified using plasmid mini prep technique

Forward sequencing results (beginning of T7 promotor region):

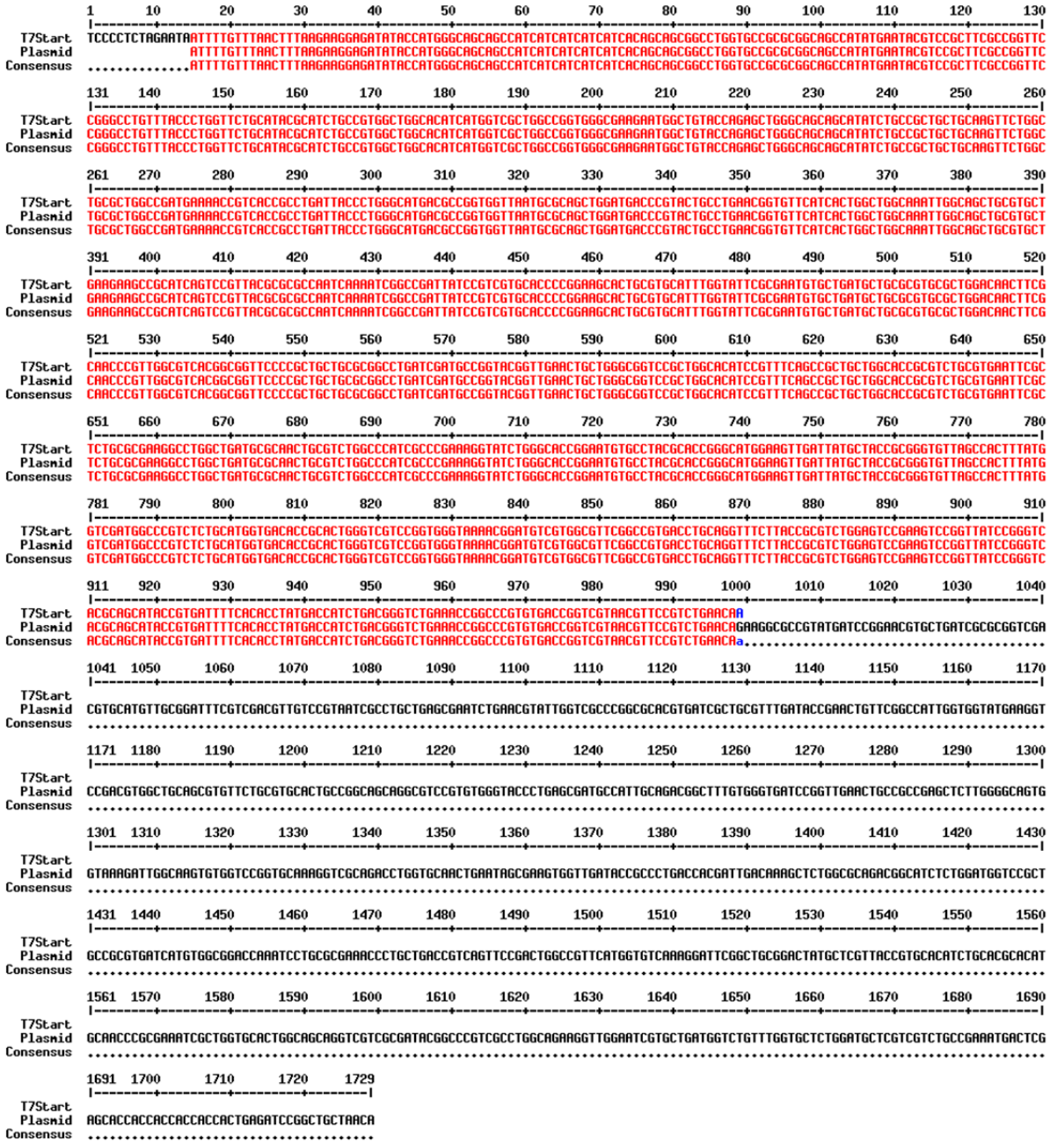


Figure B. 1 Forward sequencing results (T7Start) aligned with original plasmid order sequence (plasmid).

Reverse sequencing results (T7 terminal region):

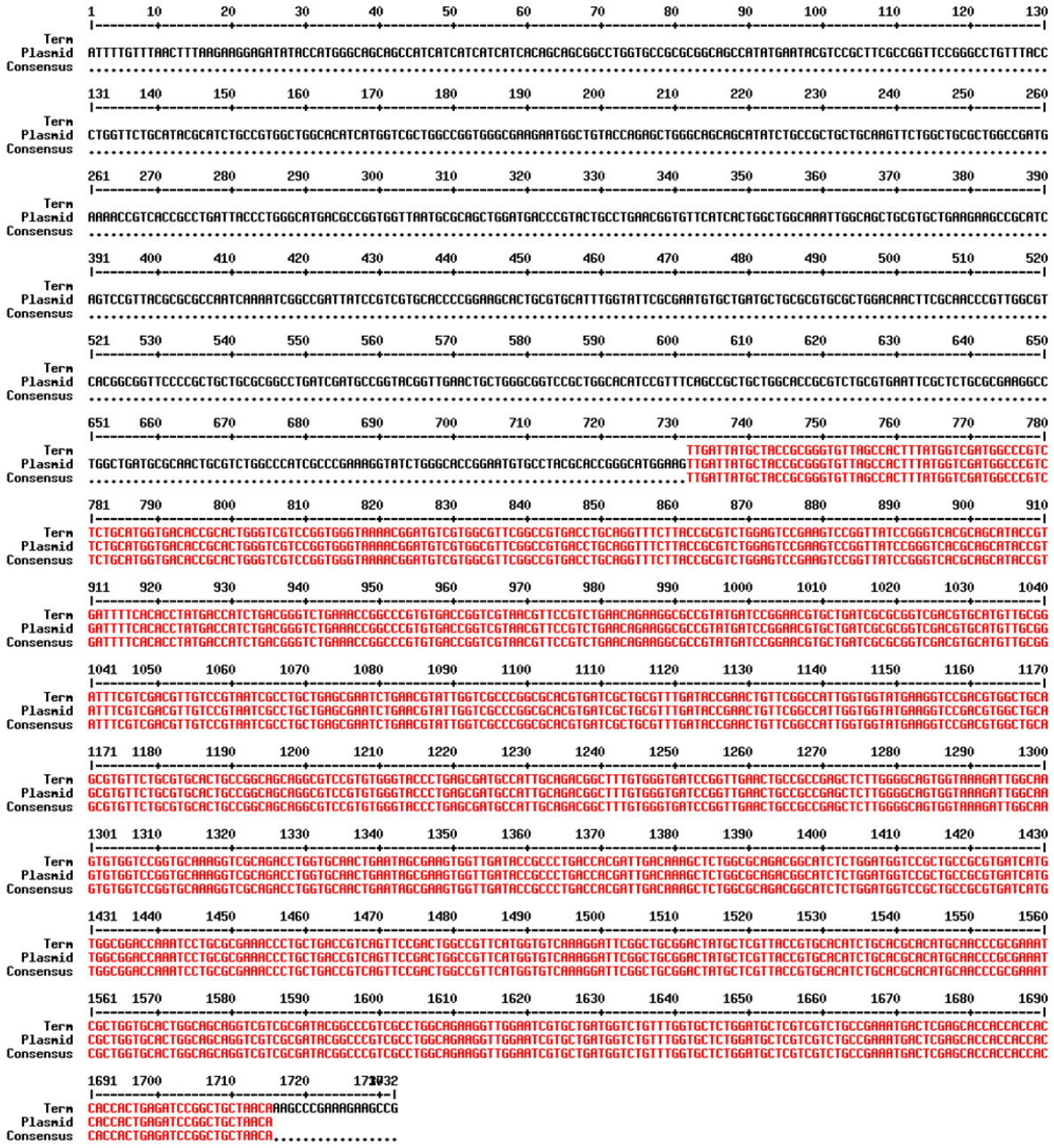


Figure B. 2 Reverse sequencing results (Term) aligned with original plasmid order sequence (plasmid).

Appendix C: Additional data for Chapter 4

C.1 Absorbance assays of MtGBE performed using a UV-vis spectrophotometer in the presence of glucosyl glycerate (see chapter 4, section 4.2.9)

Assay performed as per chapter 8, section 8.2.3.4 with 2 minutes incubation before readings taken. Reaction mixture contained 1.0 mM pNPGlc, 100 mM KCl and 1.0 mM GG in 50 mM HEPES pH 7. Results shown in figure C.1.

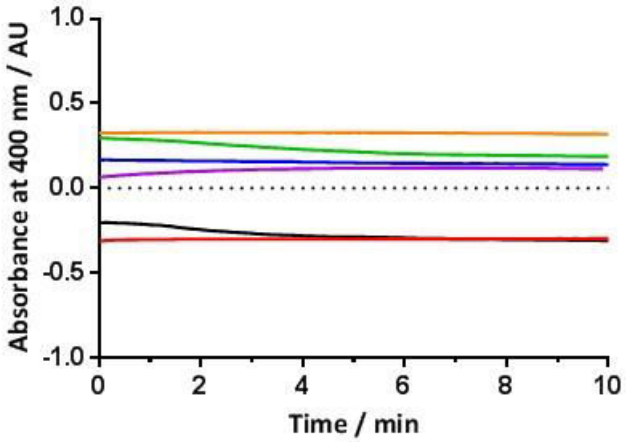


Figure C. 1 Absorbance assays of MtGBE and pNPGlc performed with fresh enzyme samples sent from collaborators. Reactions run at 37 °C for 10 minutes with 1.0 mM pNPGlc and 1.0 mM GG. Enzyme concentrations varied: 1.0 (blue), 0.5 (green), 0.25 (purple), 0.1 (orange) mg/mL. Control reactions performed with no enzyme present (black) or no GG present (red).

C.2 Absorbance scans of glycogen and maltodextrin mixed with iodine staining solution

Absorbance scan performed by mixing substrate stock solution (2.5% in DMSO) with iodine staining solution. The ratio of sample to staining solution was adjusted until a non-saturated absorbance reading was obtained. Samples were loaded into a 96 well plate (Greiner, flat-bottomed, colourless, full volume) and an absorbance scan measured from 200 – 00 nm in a Tecan Infinite M200 Pro plate reader. Results are shown in figure C.2.

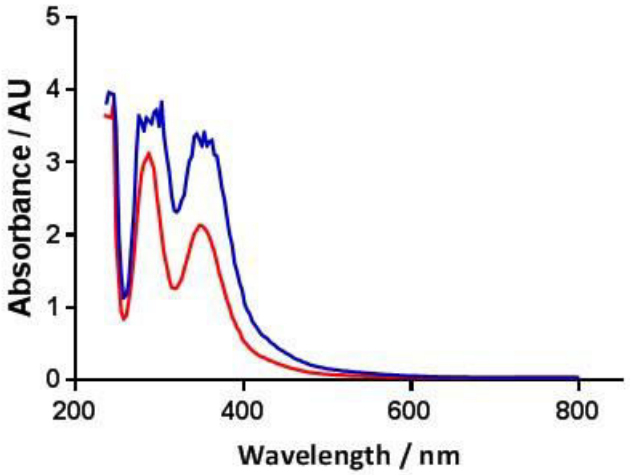


Figure C. 2 Absorbance scan of glycogen (red) and maltodextrin (blue) mixed with iodine staining solution.

C.3 Maltotriose standard curve

For sample preparation and measurement technique see 8.2.3.6 in chapter 8.

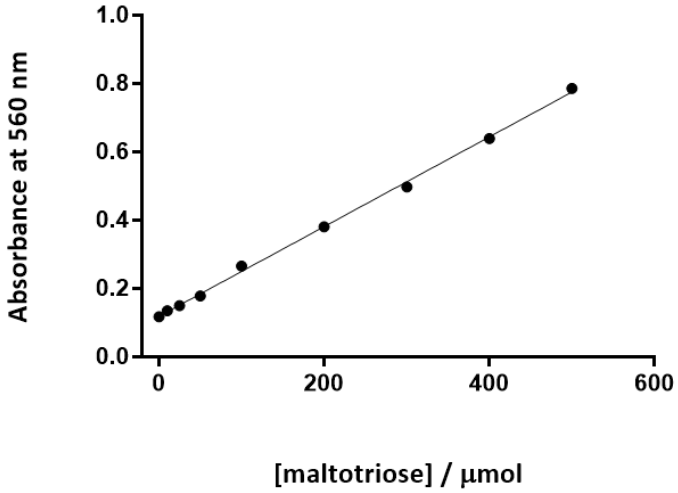


Figure C. 3 Standard curve for quantification of reducing end sugars made with maltotriose solutions and BCA reagent. Linear regression performed on GraphPad Prism 7.0.

C.4 Absorbance scans of phosphate standard curve with malachite green dye

For sample preparation and methodology, see Chapter 8, section 8.2.3.8.

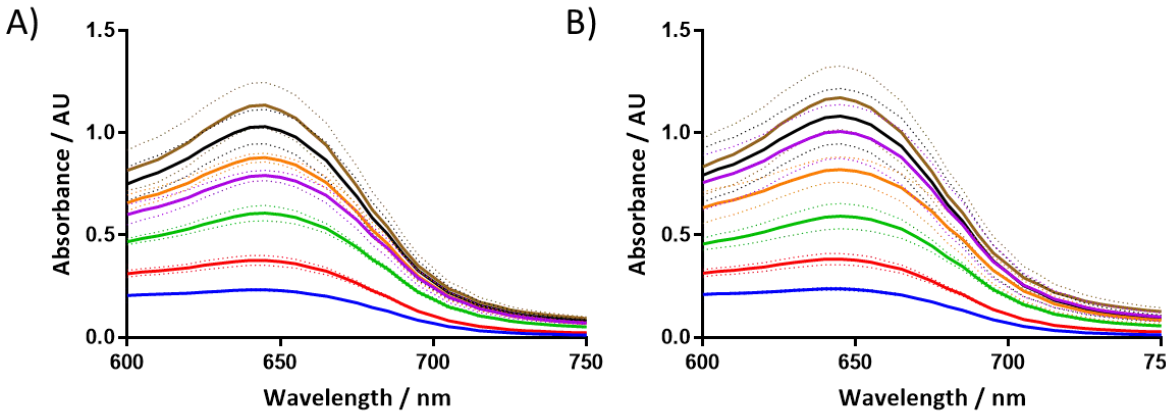


Figure C. 4 Absorbance scans of phosphate solutions of various concentrations in the presence of malachite green dye. Concentrations tested: 10 (blue), 20 (red), 40 (green), 50 (purple), 60 (orange), 80 (black), 100 (brown) µM. A: solutions in water; B: solutions in 50 mM HEPES pH 7.0.

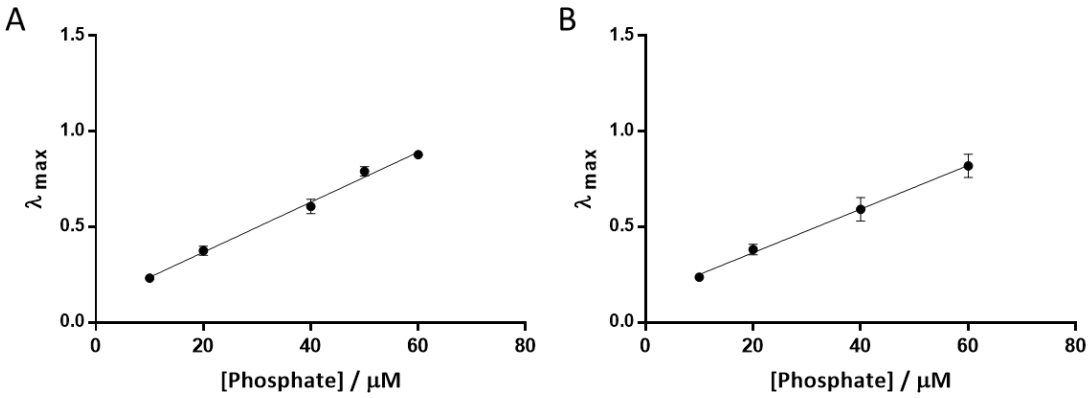


Figure C. 5 Standard curves of phosphate concentration for malachite green assay. Non-linear sections omitted for clarity. A: standard curve in water; B: standard curve in 50 mM HEPES pH 7. Linear regression performed in GraphPad Prism 7.

C.5 Absorbance scans of malachite green assays

For sample preparation and methodology, see Chapter 8, section 8.2.3.9.

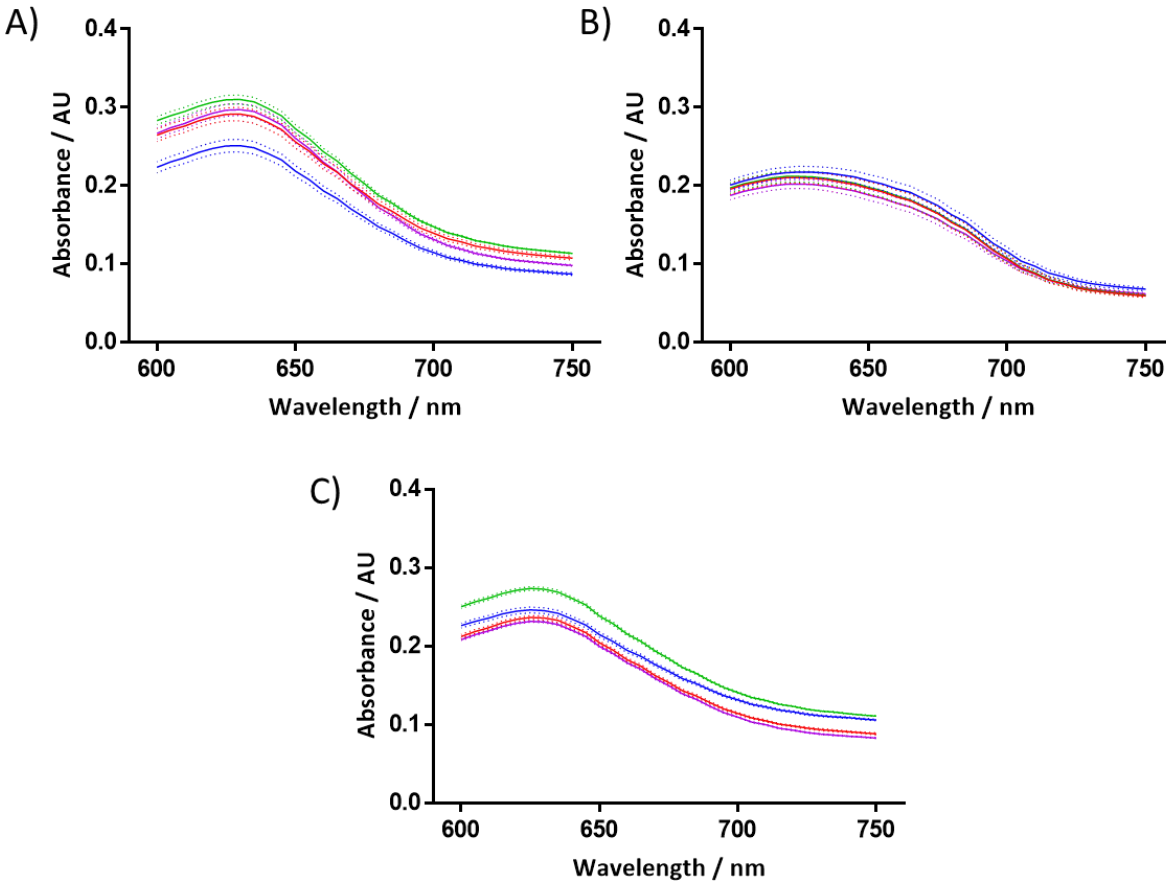


Figure C. 6 Absorbance scans of malachite green assay of MtGBE and G1P. A) MtGBE, 1.0 mM GG and 1.0 mM G1P; B) control reaction with no MtGBE present; C) control reaction with no substrate present. Aliquots taken at the following time points during the reaction: 0 (blue), 15 (red), 30 (green) and 60 (purple) minutes.

C.6 SDM primer designs

Table C.1 shows design of both forward and reverse primers used in SDM experiments. Melting temperatures and %GC content calculated using ThermoFisher Tm calculator (https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html) with settings for *Taq*-based DNA polymerase selected.

Table C.1 Primer designs for MtGBE mutants to be generated using SDM

Mutant	Forward Primer	Reverse Primer	Tm / °C	%GC
E205A	TGG GCA CCG GCG TGT GCC TAC	GTA GGC ACA CGC CGG TGC CCA	70.2	71.43
W260A	TAC CGC GTC GCG AGT CCG AAG	CTT CGG ACT CGC GAC GCG GTA	66.6	66.67
D344A	GCT GCG TTT GCG ACC GAA CTG	CAG TTC GGT CGC AAA CGC AGC	64.4	61.90
W351A	TTC GGC CAT GCG TGG TAT GAA	TTC ATA CCA CGC ATG GCC GAA	61.2	61.90
W396A	CCG AGC TCT GCG GGC AGT GGT	ACC ACT GCC CGC AGA GCT CGG	69.6	71.43
F465A	GAC TGG CCG GCG ATG GTG TCA	TGA CAC CAT CGC CGG CCA GTC	67.3	66.70
W352A	GGC CAT TGG GCG TAT GAA GGT	ACC TTC ATA CGC CCA ATG GCC	62.3	57.14
W29A	CAT GGT CGC GCG CCG GTG GGC	GCC CAC CGG CGC GCG ACC ATG	74.9	80.95
F275A	TAC CGT GAT GCG CAC ACC TAT	ATA GGT GTG CGC ATC ACG GTA	60.1	52.38
H232A	CCG TCT CTG GCG GGT GAC ACC	GGT GTC ACC CGC CAG AGA CGG	67.8	71.43
H232Y	CCG TCT CTG TAT GGT GAC ACC	GGT GTC ACC ATA CAG AGA CGG	59.0	57.14

Appendix D: Additional data for Chapter 5

D.1 Ni-NTA purification of OtsA

The purification of OtsA was carried out as per chapter 8, section 8.2.4.7. Figure D.1 shows the absorbance readings and elution gradient.

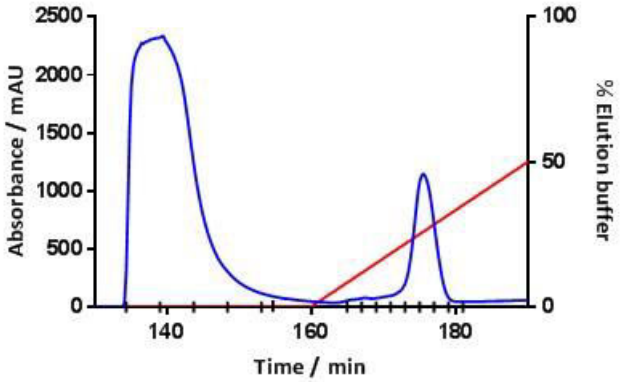


Figure D. 1 Absorbance trace (blue), % elution buffer (red) and fractions (black) from Ni-NTA purification of OtsA.

Appendix D

Bibliography

1. Daniel, T. M., *Respir. Med.* **2006,** *100* (11), 1862-1870.
2. Daniel, T. M., *Int. J. Tuberc. Lung Dis.* **2004,** *8* (5), 517-518.
3. Daniel, T. M., *Int. J. Tuberc. Lung Dis.* **2005,** *9* (11), 1181-1182.
4. Blanchard, J. S., *Annu. Rev. Biochem.* **1996,** *65*, 215-239.
5. World Health Organisation: *Global Tuberculosis Report 2016*; **2016**.
6. World Health Organisation: *Global Tuberculosis Report 2015*; **2015**.
7. World Health Organisation: *Global Tuberculosis Report 2014*; **2014**.
8. World Health Organisation: *The End TB Strategy*; **2014**.
9. Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S. V.; Eiglmeier, K.; Gas, S.; Barry Iii, C. E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, K.; Osborne, J.; Quail, M. A.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, R.; Squares, S.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G., *Nature* **1998,** *393*, 537.
10. Brennan, P. J., *Tuberculosis* **2003,** *83* (1), 91-97.
11. Mann, F. M.; Xu, M.; Chen, X.; Fulton, D. B.; Russell, D. G.; Peters, R. J., *J. Am. Chem. Soc.* **2009,** *131* (48), 17526-17527.
12. Evangelopoulos, D.; McHugh, T. D., *Chem. Biol. Drug. Des.* **2015,** *86* (5), 951-60.
13. Park, S.-H.; Bendelac, A., *Nature* **2000,** *406*, 788.
14. Janin, Y. L., *Biorg. Med. Chem.* **2007,** *15* (7), 2479-2513.
15. Migliori, G. B.; Dheda, K.; Centis, R.; Mwaba, P.; Bates, M.; O'Grady, J.; Hoelscher, M.; Zumla, A., *Trop. Med. Int. Health.* **2010,** *15* (9), 1052-1066.
16. Kolattukudy, P.; Fernandes, N. D.; Azad, A.; Fitzmaurice, A. M.; Sirakova, T. D., *Mol. Microbiol.* **1997,** *24* (2), 263-270.
17. Torrelles, J. B.; Schlesinger, L. S., *Tuberculosis* **2010,** *90* (2), 84-93.
18. Merrill, M. H., *J. Bacteriol.* **1930,** *20* (4), 235.
19. Barnes, D. D.; Lundahl, M. L. E.; Lavelle, E. C.; Scanlan, E. M., *Acs. Chem. Biol.* **2017,** *12* (8), 1969-1979.
20. Campbell, E. A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S. A., *Cell* **2001,** *104* (6), 901-912.
21. Gopal, P.; Yee, M.; Sarathy, J.; Low, J. L.; Sarathy, J. P.; Kaya, F.; Dartois, V.; Gengenbacher, M.; Dick, T., *ACS Infect. Dis.* **2016,** *2* (9), 616-626.

22. Mikusova, K.; Slayden, R. A.; Besra, G. S.; Brennan, P. J., *Antimicrob. Agents Chemother.* **1995,** *39* (11), 2484-2489.

23. Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V., *Nature* **2000,** *407* (6802), 340-348.

24. Aubry, A.; Pan, X. S.; Fisher, L. M.; Jarlier, V.; Cambau, E., *Antimicrob. Agents Chemother.* **2004,** *48* (4), 1281-1288.

25. Johnsson, K.; King, D. S.; Schultz, P. G., *J. Am. Chem. Soc.* **1995,** *117* (17), 5009-10.

26. Sacchettini, J. C.; Blanchard, J. S., *Res. Microbiol.* **1996,** *147* (1-2), 36-43.

27. Timmins, G. S.; Deretic, V., *Mol. Microbiol.* **2006,** *62* (5), 1220-1227.

28. Zhang, Y.; Heym, B.; Allen, B.; Young, D.; Cole, S., *Nature* **1992,** *358* (6387), 591-593.

29. Hazbon, M. H.; Brimacombe, M.; del Valle, M. B.; Cavatore, M.; Guerrero, M. I.; Varma-Basil, M.; Billman-Jacobe, H.; Lavender, C.; Fyfe, J.; Garcia-Garcia, L.; Leon, C. I.; Bose, M.; Chaves, F.; Murray, M.; Eisenach, K. D.; Sifuentes-Osornio, J.; Cave, M. D.; de Leon, A. P.; Alland, D., *Antimicrob. Agents Chemother.* **2006,** *50* (8), 2640-2649.

30. Unissa, A. N.; Hanna, L. E., *Tuberculosis* **2017,** *105*, 96-107.

31. Somoskovi, A.; Parsons, L. M.; Salfinger, M., *Respir. Res.* **2001,** *2* (3), 164-168.

32. Scorpio, A.; Zhang, Y., *Nat. Med.* **1996,** *2* (6), 662-667.

33. Zhang, S.; Chen, J. Z.; Shi, W. L.; Liu, W.; Zhang, W. H.; Zhang, Y., *Emerg. Microbes Infect.* **2013,** *2*, 5.

34. Shi, W. L.; Zhang, X. L.; Jiang, X.; Yuan, H. M.; Lee, J. S.; Barry, C. E.; Wang, H. H.; Zhang, W. H.; Zhang, Y., *Science* **2011,** *333* (6049), 1630-1632.

35. Dong, X.; Bhamidi, S.; Scherman, M.; Xin, Y.; McNeil, M. R., *Appl. Environ. Microbiol.* **2006,** *72* (4), 2601-2605.

36. Telenti, A.; Philipp, W. J.; Sreevatsan, S.; Bernasconi, C.; Stockbauer, K. E.; Wieles, B.; Musser, J. M.; Jacobs, W. R., *Nat. Med.* **1997,** *3* (5), 567-570.

37. Alcaide, F.; Pfyffer, G. E.; Telenti, A., *Antimicrob. Agents Chemother.* **1997,** *41* (10), 2270-2273.

38. Johansen, S. K.; Maus, C. E.; Plikaytis, B. B.; Douthwaite, S., *Mol. Cell.* **2006,** *23* (2), 173-182.

39. Finken, M.; Kirschner, P.; Meier, A.; Wrede, A.; Bottger, E. C., *Mol. Microbiol.* **1993,** *9* (6), 1239-1246.

40. Telenti, A.; Imboden, P.; Marchesi, F.; Lowrie, D.; Cole, S.; Colston, M. J.; Matter, L.; Schopfer, K.; Bodmer, T., *Lancet.* **1993,** *341* (8846), 647-650.

41. Maruri, F.; Sterling, T. R.; Kaiga, A. W.; Blackman, A.; van der Heijden, Y. F.; Mayer, C.; Cambau, E.; Aubry, A., *J. Antimicrob. Chemother.* **2012,** *67* (4), 819-831.

42. Farhat, M. R.; Jacobson, K. R.; Franke, M. F.; Kaur, D.; Sloutsky, A.; Mitnick, C. D.; Murray, M., *J. Clin. Microbiol.* **2016,** *54* (3), 727-733.

43. Sinden, R. R., B. The Structure of Nucleic Acids. In *DNA Structure and Function*, Academic Press: San Diego, **1994**; pp 3-11.

44. Zachara, N. E.; Hart, G. W., *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2006,** *1761* (5-6), 599-617.

45. Lis, H.; Sharon, N., *Chem. Rev.* **1998,** *98* (2), 637-674.

46. Mayer, S.; Raulf, M.-K.; Lepenies, B., *Histochem. Cell Biol.* **2017,** *147* (2), 223-237.

47. Blaak, E. E.; Saris, W. H. M., *Nutr Res.* **1995,** *15* (10), 1547-1573.

48. Latge, J. P., *Mol. Microbiol.* **2007,** *66* (2), 279-290.

49. McFarlane, H. E.; Doring, A.; Persson, S., *Annu. Rev. Plant Biol.* **2014***, 65*, 69-94.

50. Sutton-McDowall, M. L.; Gilchrist, R. B.; Thompson, J. G., *Reproduction* **2010,** *139* (4), 685-695.

51. Sun, L. N.; Middleton, D. R.; Wantuch, P. L.; Ozdilek, A.; Avci, F. Y., *Glycobiology* **2016,** *26* (10), 1029-1040.

52. Lasky, L. A., *Annu. Rev. Biochem* **1995,** *64*, 113-139.

53. Varki, A., *Glycobiology* **2017,** *27* (1), 3-49.

54. Davis, B. G.; Fairbanks, A. J., Chemical Glycobiology. In *Carbohydrate Chemistry*, Davies, S. G., Ed. Oxford University Press: Oxford, **2002**; pp 78-89.

55. Hakomori, S., *Annu. Rev. Immunol.* **1984,** *2*, 103-126.

56. Cazet, A.; Julien, S.; Bobowski, M.; Burchell, J.; Delannoy, P., *Breast Cancer Res.* **2010,** *12* (3).

57. Rambaruth, N. D. S.; Dwek, M. V., *Acta Histochem.* **2011,** *113* (6), 591-600.

58. Kato, K.; Ishiwa, A., *Tropical Medicine and Health* **2015,** *43* (1), 41-52.

59. Keusch, G. T., *Monogr. pathol.* **1982,** (23), 94-113.

60. Fraenkel, D. G.; Vinopal, R. T., *Annu. Rev. Microbiol.* **1973,** *27*, 69-100.

61. Schleifer, K. H.; Kandler, O., *Bacteriol. Rev.* **1972,** *36* (4), 407-477.

62. Egan, A. J. F.; Cleverley, R. M.; Peters, K.; Lewis, R. J.; Vollmer, W., *FEBS J.* **2017,** *284* (6), 851-867.

63. Postma, P. W.; Lengeler, J. W.; Jacobson, G. R., *Microbiol. Rev.* **1993,** *57* (3), 543-594.

64. Sharon, N., *Biochim. Biophys. Acta Gen. Subj.* **2006,** *1760* (4), 527-537.

65. Pieters, R. J., *Med. Res. Rev.* **2007,** *27* (6), 796-816.

66. Dorokhov, Y. L.; Sheshukova, E. V.; Kosobokova, E. N.; Shindyapina, A. V.; Kosorukov, V. S.; Komarova, T. V., *Biochemistry-Moscow+* **2016,** *81* (8), 835-857.

67. WeymouthWilson, A. C., *Nat. Prod. Rep.* **1997,** *14* (2), 99-110.

68. Kren, V.; Rezanka, T., *FEMS Microbiol. Rev.* **2008,** *32* (5), 858-889.

69. Feng, D. Y.; Shaikh, A. S.; Wang, F. S., *Acs Chem Biol* **2016,** *11* (4), 850-863.

70. Danishefsky, S. J.; Allen, J. R., *Angew. Chem. Int. Ed.* **2000,** *39* (5), 836-863.

71. Astronomo, R. D.; Burton, D. R., *Nat. Rev. Drug Discovery* **2010,** *9* (4), 308-324.

72. Nishat, S.; Andreana, P. R., *Vaccines* **2016,** *4* (2).

73. Zhang, Y.; Wang, F., *Drug discov. Ther.* **2015,** *9* (2), 79-87.

74. Davis, B. G.; Fairbanks, A. J., Introduction. In *Carbohydr. Chem.*, Davies, S. G., Ed. Oxford University Press: Oxford, **2002**; pp 1-2.

75. Davis, B. G.; Fairbanks, A. J., Open chain and ring structure of monosaccharides. In *Carbohydr. Chem.*, Davies, G., Ed. Oxford University Press: Oxford, **2002**; pp 3-11.

76. Osborn, H. M. I., *Carbohydrates*. Academic Press: Oxford, **2003**.

77. Finch, P., *Carbohydrates: Structures, Syntheses and Dynamics*. Kluwer Academic Publishers: Dordrecht, **1999**.

78. Marchand, A. P., *J. Am. Chem. Soc.* **1984,** *106* (6), 1895-1895.

79. Cao, X. F.; Liu, S. Q.; Rong, C. Y.; Lu, T.; Liu, S. B., *Chem. Phys. Lett.* **2017,** *687*, 131-137.

80. Martins, F. A.; Silla, J. M.; Freitas, M. P., *Carbohydr. Res.* **2017,** *451*, 29-35.

81. Zhou, X. Y.; Yu, D. H.; Rong, C. Y.; Lu, T.; Liu, S. B., *Chem. Phys. Lett.* **2017,** *684*, 97-102.

82. Davis, B. G.; Fairbanks, A. J., Reactions of the anomeric centre Part I. In *Carbohydr. Chem.*, Davies, G., Ed. Oxford University Press: Oxford, **2002**; pp 12-18.

83. Juaristi, E.; Cuevas, G., *Tetrahedron* **1992,** *48* (24), 5019-5087.

84. Jeffrey, P. L.; Brown, D. H.; Brown, B. I., *Biochemistry* **1970,** *9* (6), 1403-&.

85. Nigudkar, S. S.; Demchenko, A. V., *Chem. Sci.* **2015,** *6* (5), 2687-2704.

86. Davis, B. G.; Fairbanks, A. J., Chemical Disaccharide Formation. In *Carbohydr. Chem.*, Davies, S. G., Ed. Oxford University Press: Oxford, 2002; pp 53-64.

87. Staneloni, R. J.; Leloir, L. F., *Trends Biochem. Sci* **1979,** *4* (3), 65-67.

88. Villarpa.C; Larner, J., *Annu. Rev. Biochem* **1970,** *39*, 639-679.

89. Kamerling, J. P.; Boons, G.; Lee, Y. C.; Suzuki, A.; Taniguchi, N.; Voragen, A. G. J., *Comprehensive glycoscience : from chemistry to systems biology*. Elsevier Ltd.: Oxford, **2007**; Vol. 2.

90. Stadthagen, G.; Sambou, T.; Guerin, M.; Barilone, N.; Boudou, F.; Kordulakova, J.; Charles, P.; Alzari, P. M.; Lemassu, A.; Daffe, M.; Puzo, G.; Gicquel, B.; Riviere, M.; Jackson, M., *J. Biol. Chem.* **2007,** *282* (37), 27270-27276.

91. Lee, Y. C.; Ballou, C. E., *J. Biol. Chem.* **1964,** *239* (10), PC3602-PC3603.

92. Lee, Y. C., *J. Biol. Chem.* **1966,** *241* (8), 1899-1908.

93. Jankute, M.; Cox, J. A. G.; Harrison, J.; Besra, G. S., *Annu. Rev. Microbiol.* **2015***, 69*, 405-423.

94. Heijenoort, J. v., *Glycobiology* **2001,** *11* (3), 25R-36R.

95. Raymond, J. B.; Mahapatra, S.; Crick, D. C.; Pavelka, M. S., *J. Biol. Chem.* **2005,** *280* (1), 326-333.

96. Warth, A.; Strominger, J., *P. Natl. Acad. Sci.* **1969,** *64* (2), 528-535.

97. Kim, D. H.; Lees, W. J.; Kempsell, K. E.; Lane, W. S.; Duncan, K.; Walsh, C. T., *Biochemistry* **1996,** *35* (15), 4923-4928.

98. Benson, T. E.; Walsh, C. T.; Hogle, J. M., *Structure* **1996,** *4* (1), 47-54.

99. Munshi, T.; Gupta, A.; Evangelopoulos, D.; Guzman, J. D.; Gibbons, S.; Keep, N. H.; Bhakta, S., *PLoS One* **2013,** *8* (3), e60143.

100. Kurosu, M.; Mahapatra, S.; Narayanasamy, P.; Crick, D. C., *Tetrahedron Lett.* **2007,** *48* (5), 799-803.

101. Bouhss, A.; Mengin-Lecreulx, D.; Le Beller, D.; Van Heijenoort, J., *Mol. Microbiol.* **1999,** *34* (3), 576-585.

102. Ruiz, N., *Curr. Opin. Microbiol.* **2016,** *34*, 1-6.

103. Hett, E. C.; Chao, M. C.; Rubin, E. J., *PLoS Path.* **2010,** *6* (7), e1001020.

104. McNeil, M.; Daffe, M.; Brennan, P. J., *J. Biol. Chem.* **1990,** *265* (30), 18200-18206.

105. Daffe, M.; Brennan, P. J.; McNeil, M., *J. Biol. Chem.* **1990,** *265* (12), 6734-6743.

106. Besra, G. S.; Khoo, K.-H.; McNeil, M. R.; Dell, A.; Morris, H. R.; Brennan, P. J., *Biochemistry* **1995,** *34* (13), 4257-4266.

107. Bhamidi, S.; Scherman, M. S.; Rithner, C. D.; Prenni, J. E.; Chatterjee, D.; Khoo, K.-H.; McNeil, M. R., *J. Biol. Chem.* **2008,** *283* (19), 12992-13000.

108. Peng, W.; Zou, L.; Bhamidi, S.; McNeil, M. R.; Lowary, T. L., *J. Org. Chem.* **2012,** *77* (21), 9826-9832.

109. Draper, P.; Kay-Hooi, K.; Chatterjee, D.; Anne, D.; Morris, R. H., *Biochem. J* **1997,** *327* (2), 519-525.

110. Jin, Y.; Xin, Y.; Zhang, W.; Ma, Y., *FEMS Microbiol. Lett.* **2010,** *310* (1), 54-61.

111. Mikušová, K.; Beláňová, M.; Korduláková, J.; Honda, K.; McNeil, M. R.; Mahapatra, S.; Crick, D. C.; Brennan, P. J., *J. Bacteriol.* **2006,** *188* (18), 6592-6598.

112. Rose, N. L.; Completo, G. C.; Lin, S.-J.; McNeil, M.; Palcic, M. M.; Lowary, T. L., *J. Am. Chem. Soc.* **2006,** *128* (20), 6721-6729.

113. Wheatley, R. W.; Zheng, R. B.; Richards, M. R.; Lowary, T. L.; Ng, K. K., *J. Biol. Chem.* **2012,** *287* (33), 28132-28143.

114. Kremer, L.; Dover, L. G.; Morehouse, C.; Hitchin, P.; Everett, M.; Morris, H. R.; Dell, A.; Brennan, P. J.; McNeil, M. R.; Flaherty, C., *J. Biol. Chem.* **2001,** *276* (28), 26430-26440.

115. Wolucka, B. A.; McNeil, M. R.; de Hoffmann, E.; Chojnacki, T.; Brennan, P. J., *J. Biol. Chem.* **1994,** *269* (37), 23328-23335.

116. Birch, H. L.; Alderwick, L. J.; Bhatt, A.; Rittmann, D.; Krumbach, K.; Singh, A.; Bai, Y.; Lowary, T. L.; Eggeling, L.; Besra, G. S., *Mol. Microbiol.* **2008,** *69* (5), 1191-1206.

117. Pitarque, S.; Larrouy-Maumus, G.; Payré, B.; Jackson, M.; Puzo, G.; Nigou, J., *Tuberculosis* **2008,** *88* (6), 560-565.

118. Chatterjee, D.; Hunter, S.; McNeil, M.; Brennan, P., *J. Biol. Chem.* **1992,** *267* (9), 6228-6233.

119. Kaur, D.; Obregón-Henao, A.; Pham, H.; Chatterjee, D.; Brennan, P. J.; Jackson, M., *P. Natl. Acad. Sci.* **2008,** *105* (46), 17973-17977.

120. Birch, H. L.; Alderwick, L. J.; Appelmelk, B. J.; Maaskant, J.; Bhatt, A.; Singh, A.; Nigou, J.; Eggeling, L.; Geurtsen, J.; Besra, G. S., *P. Natl. Acad. Sci.* **2010,** *107* (6), 2634-2639.

121. Chatterjee, D.; Bozic, C.; McNeil, M.; Brennan, P., *J. Biol. Chem.* **1991,** *266* (15), 9652-9660.

122. Chatterjee, D.; Khoo, K.-H.; McNeil, M. R.; Dell, A.; Morris, H. R.; Brennan, P. J., *Glycobiology* **1993,** *3* (5), 497-506.

123. McNeil, M. R.; Robuck, K. G.; Harter, M.; Brennan, P. J., *Glycobiology* **1994,** *4* (2), 165-174.

124. Chatterjee, D.; Lowell, K.; Rivoire, B.; McNeil, M. R.; Brennan, P., *J. Biol. Chem.* **1992,** *267* (9), 6234-6239.

125. Joe, M.; Sun, D.; Taha, H.; Completo, G. C.; Croudace, J. E.; Lammas, D. A.; Besra, G. S.; Lowary, T. L., *J. Am. Chem. Soc.* **2006,** *128* (15), 5059-5072.

126. Turnbull, W. B.; Shimizu, K. H.; Chatterjee, D.; Homans, S. W.; Treumann, A., *Angew. Chem. Int. Ed.* **2004,** *43* (30), 3918-3922.

127. Maeda, N.; Nigou, J.; Herrmann, J.-L.; Jackson, M.; Amara, A.; Lagrange, P. H.; Puzo, G.; Gicquel, B.; Neyrolles, O., *J. Biol. Chem.* **2003,** *278* (8), 5513-5516.

128. Mishra, A. K.; Driessen, N. N.; Appelmelk, B. J.; Besra, G. S., *FEMS Microbiol. Rev.* **2011,** *35* (6), 1126-1157.

129. Nigou, J.; Gilleron, M.; Rojas, M.; García, L. F.; Thurnher, M.; Puzo, G., *Microb. Infect.* **2002,** *4* (9), 945-953.

130. Schlesinger, L. S.; Hull, S. R.; Kaufman, T. M., *J. Immunol.* **1994,** *152* (8), 4070-4079.

131. Turnbull, W. B.; Stalford, S. A., *Org. Biomol. Chem.* **2012,** *10* (30), 5698-5706.

132. Sambou, T.; Dinadayala, P.; Stadthagen, G.; Barilone, N.; Bordat, Y.; Constant, P.; Levillain, F.; Neyrolles, O.; Gicquel, B.; Lemassu, A., *Mol. Microbiol.* **2008,** *70* (3), 762-774.

133. Chapman, G. B.; Hanks, J. H.; Wallace, J. H., *J. Bacteriol.* **1959,** *77* (2), 205.

134. Rashid, A. M.; Batey, S. F. D.; Syson, K.; Koliwer-Brandl, H.; Miah, F.; Barclay, J. E.; Findlay, K. C.; Nartowski, K. P.; Khimyak, Y. Z.; Kalscheuer, R.; Bornemann, S., *Biochemistry* **2016,** *55* (23), 3270 - 3284.

135. Cywes, C.; Hoppe, H. C.; Daffé, M.; Ehlers, M., *Infect. Immun.* **1997,** *65* (10), 4258-4266.

136. Stokes, R. W.; Norris-Jones, R.; Brooks, D. E.; Beveridge, T. J.; Doxsee, D.; Thorson, L. M., *Infect. Immun.* **2004,** *72* (10), 5676-5686.

137. Gagliardi, M. C.; Lemassu, A.; Teloni, R.; Mariotti, S.; Sargentini, V.; Pardini, M.; Daffé, M.; Nisini, R., *Cell. Microbiol.* **2007,** *9* (8), 2081-2092.

138. Chandra, G.; Chater, K. F.; Bornemann, S., *Microbiol. Sgm* **2011,** *157*, 1565-1572.

139. Kalscheuer, R.; Syson, K.; Veeraraghavan, U.; Weinrick, B.; Biermann, K. E.; Liu, Z.; Sacchettini, J. C.; Besra, G.; Bornemann, S.; Jacobs, W. R., Jr., *Nat. Chem. Biol.* **2010,** *6* (5), 376-384.

140. Elbein, A. D.; Pastuszak, I.; Tackett, A. J.; Wilson, T.; Pan, Y. T., *J. Biol. Chem.* **2010,** *285* (13), 9803-9812.

141. Edson, N., *Bacteriol. Rev.* **1951,** *15* (3), 147.

142. Izumori, K.; Yamanaka, K.; Elbein, D., *J. Bacteriol.* **1976,** *128* (2), 587-591.

143. Niederweis, M., *Microbiology-Sgm* **2008,** *154*, 679-692.

144. Brzostek, A.; Śliwiński, T.; Rumijowska-Galewicz, A.; Korycka-Machała, M.; Dziadek, J., *Microbiology+* **2005,** *151* (7), 2393-2402.

145. Noy, T.; Vergnolle, O.; Hartman, T. E.; Rhee, K. Y.; Jacobs, W. R.; Berney, M.; Blanchard, J. S., *J. Biol. Chem.* **2016,** *291* (13), 7060-7069.

146. Marrero, J.; Trujillo, C.; Rhee, K. Y.; Ehrt, S., *PLoS Path.* **2013,** *9* (1), e1003116.

147. de Carvalho, L. P. S.; Fischer, S. M.; Marrero, J.; Nathan, C.; Ehrt, S.; Rhee, K. Y., *Chem. Biol.* **2010,** *17* (10), 1122-1131.

148. Brennan, P. J.; Nikaido, H., *Annu. Rev. Biochem* **1995,** *64* (1), 29-63.

149. Marrakchi, H.; Lanéelle, M.-A.; Daffé, M., *Chem. Biol.* **2014,** *21* (1), 67-85.

150. Li, W.; Upadhyay, A.; Fontes, F. L.; North, E. J.; Wang, Y.; Crans, D. C.; Grzegorzewicz, A. E.; Jones, V.; Franzblau, S. G.; Lee, R. E., *Antimicrob. Agents Chemother.* **2014,** *58* (11), 6413-6423.

151. Gavalda, S.; Bardou, F.; Laval, F.; Bon, C.; Malaga, W.; Chalut, C.; Guilhot, C.; Mourey, L.; Daffé, M.; Quémard, A., *Chem. Biol.* **2014,** *21* (12), 1660-1669.

152. Thanna, S.; Sucheck, S. J., *Medchemcomm* **2016,** *7* (1), 69-85.

153. Nguyen, L.; Chinnapapagari, S.; Thompson, C. J., *J. Bacteriol.* **2005,** *187* (19), 6603-6611.

154. Jackson, M.; Raynaud, C.; Lanéelle, M. A.; Guilhot, C.; Laurent-Winter, C.; Ensergueix, D.; Gicquel, B.; Daffé, M., *Mol. Microbiol.* **1999,** *31* (5), 1573-1587.

155. Armitige, L. Y.; Jagannath, C.; Wanger, A. R.; Norris, S. J., *Infect. Immun.* **2000,** *68* (2), 767-778.

156. Spargo, B.; Crowe, L.; Ioneda, T.; Beaman, B.; Crowe, J., *P. Natl. Acad. Sci.* **1991,** *88* (3), 737-740.

157. Glickman, M. S.; Cox, J. S.; Jacobs Jr, W. R., *Mol. Cell* **2000,** *5* (4), 717-727.

158. Shi, L.; Zhang, H.; Qiu, Y.; Wang, Q.; Wu, X.; Wang, H.; Zhang, X.; Lin, D., *Acta Biochim Biophys Sin* **2013,** *45* (10), 837-844.

159. Cambier, C.; Takaki, K. K.; Larson, R. P.; Hernandez, R. E.; Tobin, D. M.; Urdahl, K. B.; Cosma, C. L.; Ramakrishnan, L., *Nature* **2014,** *505* (7482), 218.

160. Stadthagen, G.; Jackson, M.; Charles, P.; Boudou, F.; Barilone, N.; Huerre, M.; Constant, P.; Liav, A.; Bottova, I.; Nigou, J., *Microb. Infect.* **2006,** *8* (8), 2245-2253.

161. Huet, G.; Constant, P.; Malaga, W.; Lanéelle, M.-A.; Kremer, K.; Van Soolingen, D.; Daffé, M.; Guilhot, C., *J. Biol. Chem.* **2009,** *284* (40), 27101-27113.

162. Malaga, W.; Constant, P.; Euphrasie, D.; Cataldi, A.; Daffé, M.; Reyrat, J.-M.; Guilhot, C., *J. Biol. Chem.* **2008,** *283* (22), 15177-15184.

163. Reed, M. B.; Domenech, P.; Manca, C.; Su, H.; Barczak, A. K.; Kreiswirth, B. N.; Kaplan, G.; Barry III, C. E., *Nature* **2004,** *431* (7004), 84.

164. Tsenova, L.; Ellison, E.; Harbacheuski, R.; Moreira, A. L.; Kurepina, N.; Reed, M. B.; Mathema, B.; Barry III, C. E.; Kaplan, G., *J. Infect. Dis.* **2005,** *192* (1), 98-106.

165. Reed, M. B.; Gagneux, S.; DeRiemer, K.; Small, P. M.; Barry, C. E., *J. Bacteriol.* **2007,** *189* (7), 2583-2589.

166. Machida, Y.; Bloch, K., *P. Natl. Acad. Sci. USA* **1973,** *70* (4), 1146-1148.

167. Mendes, V.; Maranha, A.; Alarico, S.; Empadinhas, N., *Nat. Prod. Rep.* **2012,** *29* (8), 834-844.

168. Maranha, A.; Moynihan, P. J.; Miranda, V.; Lourenco, E. C.; Nunes-Costa, D.; Fraga, J. S.; Barbosa Pereira, P. J.; Macedo-Ribeiro, S.; Ventura, M. R.; Clarke, A. J.; Empadinhas, N., *Sci. Rep-UK* **2015,** *5*, 1-18.

169. Sassetti, C. M.; Boyd, D. H.; Rubin, E. J., *Mol. Microbiol.* **2003,** *48* (1), 77-84.

170. Kamisango, K.; Dell, A.; Ballou, C. E., *J. Biol. Chem.* **1987,** *262* (10), 4580-4586.

171. Jackson, M.; Brennan, P. J., *J. Biol. Chem.* **2009,** *284* (4), 1949-1953.

172. Empadinhas, N.; Albuquerque, L.; Mendes, V.; Macedo-Ribeiro, S.; da Costa, M. S., *FEMS Microbiol. Lett.* **2008,** *280* (2), 195-202.

173. Gest, P.; Kaur, D.; Pham, H. T.; van der Woerd, M.; Hansen, E.; Brennan, P. J.; Jackson, M.; Guerin, M. E., *Acta Crystallogr. F* **2008,** *64*, 1121-1124.

174. Pereira, P. J. B.; Empadinhas, N.; Albuquerque, L.; Sa-Moura, B.; da Costa, M. S.; Macedo-Ribeiro, S., *Plos One* **2008,** *3* (11).

175. Urresti, S.; Albesa-Jove, D.; Schaeffer, F.; Pham, H. T.; Kaur, D.; Gest, P.; van der Woerd, M. J.; Carreras-Gonzalez, A.; Lopez-Fernandez, S.; Alzari, P. M.; Brennan, P. J.; Jackson, M.; Guerin, M. E., *J. Biol. Chem.* **2012,** *287* (29), 24649-24661.

176. Kumar, G.; Guan, S. Q.; Frantom, P. A., *Arch. Biochem. Biophys.* **2014,** *564*, 120-127.

177. Kaur, D.; Pham, H.; Larrouy-Maumus, G.; Riviere, M.; Vissa, V.; Guerin, M. E.; Puzo, G.; Brennan, P. J.; Jackson, M., *Plos One* **2009,** *4* (5).

178. Mendes, V.; Maranha, A.; Alarico, S.; da Costa, M. S.; Empadinhas, N., *Sci. Rep-UK* **2011,** *1*.

179. Zheng, Q. Q.; Jiang, D. Q.; Zhang, W.; Zhang, Q. Q.; Zhao, Q.; Jin, J.; Li, X.; Yang, H. T.; Bartlam, M.; Shaw, N.; Zhou, W. H.; Rao, Z. H., *J. Biol. Chem.* **2014,** *289* (31), 21242-21251.

180. Murakami, T.; Kanai, T.; Takata, H.; Kuriki, T.; Imanaka, T., *J. Bacteriol.* **2006,** *188* (16), 5915-5924.

181. Palomo, M.; Pijning, T.; Booiman, T.; Dobruchowska, J. M.; van der Vlist, J.; Kralj, S.; Planas, A.; Loos, K.; Kamerling, J. P.; Dijkstra, B. W.; van der Maarel, M.; Dijkhuizen, L.; Leemhuis, H., *J. Biol. Chem.* **2011,** *286* (5), 3520-3530.

182. Na, S.; Park, M.; Jo, I.; Cha, J.; Ha, N. C., *Biochem. Biophys. Res. Commun.* **2017,** *484* (4), 850-856.

183. Uitdehaag, J. C. M.; Mosi, R.; Kalk, K. H.; van der Veen, B. A.; Dijkhuizen, L.; Withers, S. G.; Dijkstra, B. W., *Nat. Struct. Biol.* **1999,** *6* (5), 432-436.

184. Suzuki, E.; Suzuki, R., *Cell. Mol. Life Sci.* **2016,** *73* (14), 2643-2660.

185. Tetlow, I. J.; Emes, M. J., *Iubmb Life* **2014,** *66* (8), 546-558.

186. Lourenco, E. C.; Maycock, C. D.; Ventura, M. R., *Carbohydr. Res.* **2009,** *344* (15), 2073-2078.

187. Crich, D.; Cai, W. L., *J. Org. Chem.* **1999,** *64* (13), 4926-4930.

188. Crich, D.; Smith, M., *J. Am. Chem. Soc.* **2001,** *123* (37), 9015-9020.

189. Crich, D.; Sun, S., *J. Am. Chem. Soc.* **1998,** *120* (2), 435-436.

190. Crich, D.; Sun, S., *Tetrahedron* **1998,** *54* (29), 8321-8348.

191. Crich, D.; Sun, S., *J. Org. Chem.* **1997,** *62* (5), 1198-1199.

192. Crich, D.; Sun, S., *J. Org. Chem.* **1996,** *61* (14), 4506-4507.

193. Lemieux, R. U., *Can. J. Chem.* **1951,** *29* (12), 1079-1091.

194. Lourenco, E. C.; Rita Ventura, M., *Carbohydr. Res.* **2011,** *346* (2), 163-168.

195. Yokoyama, H.; Otaya, K.; Kobayashi, H.; Miyazawa, M.; Yamaguchi, S.; Hirai, Y., *Org. Lett.* **2000,** *2* (16), 2427-2429.

196. Sugisaki, C. H.; Ruland, Y.; Baltas, M., *Eur. J. Org. Chem.* **2003,** (4), 672-688.

197. Ladame, S.; Bardet, M.; Perie, J.; Willson, M., *Biorg. Med. Chem.* **2001,** *9* (3), 773-783.

198. Qu, F. C.; Hong, J. H.; Du, J. F.; Newton, M. G.; Chu, C. K., *Tetrahedron* **1999,** *55* (30), 9073-9088.

199. Tanaka, A.; Yamashita, K., *Agric. Biol. Chem.* **1980,** *44* (1), 199-202.

200. Williams, R.; Galan, M. C., *Eur. J. Org. Chem.* **2017,** (42), 6247-6264.

201. Yang, L.; Qin, Q.; Ye, X. S., *Asian J. Org. Chem.* **2013,** *2* (1), 30-49.

202. Goodman, L., *Adv. Carbohyd. Chem.* **1967**, *22*, 109-175.

203. Nukada, T.; Berces, A.; Zgierski, M. Z.; Whitfield, D. M., *J. Am. Chem. Soc.* **1998,** *120* (51), 13291-13295.

204. Teruaki, M.; Masahiro, S.; Hiroyuki, C.; Hideki, J., *Chem. Lett.* **2002,** *31* (1), 56-57.

205. Eby, R.; Schuerch, C., *Carbohydr. Res.* **1974,** *34* (1), 79-90.

206. Baek, J. Y.; Lee, B.-Y.; Jo, M. G.; Kim, K. S., *J. Am. Chem. Soc.* **2009,** *131* (48), 17705-17713.

207. Lourenço, E. C.; Ventura, M. R., *Tetrahedron* **2013,** *69* (34), 7090-7097.

208. Ishiwata, A.; Munemura, Y.; Ito, Y., *Tetrahedron* **2008,** *64* (1), 92-102.

209. Andersson, F.; Fúgedi, P.; Garegg, P. J.; Nashed, M., *Tetrahedron Lett.* **1986,** *27* (33), 3919-3922.

210. Schmidt, R. R.; Rücker, E., *Tetrahedron Lett.* **1980,** *21* (15), 1421-1424.

211. L. Douglas, N.; V. Ley, S.; Lucking, U.; L. Warriner, S., *J. Chem. Soc., Perkin Trans. 1* **1998,** (1), 51-66.

212. Zhang, Z.; Ollmann, I. R.; Ye, X.-S.; Wischnat, R.; Baasov, T.; Wong, C.-H., *J. Am. Chem. Soc.* **1999,** *121* (4), 734-753.

213. Ye, X.-S.; Wong, C.-H., *J. Org. Chem.* **2000,** *65* (8), 2410-2431.

214. Bieg, T.; Szeja, W., *Carbohydr. Res.* **1990,** *205*, C10-C11.

215. Kartha, K. P. R.; Field, R. A., *Tetrahedron* **1997,** *53* (34), 11753-11766.

216. Guan, Y. Y.; Song, C.; Lei, P. S., *J. Asian Nat. Prod. Res.* **2014,** *16* (1), 43-52.

217. Lu, S. R.; Lai, Y. H.; Chen, J. H.; Liu, C. Y.; Mong, K. K. T., *Angew. Chem. Int. Ed.* **2011,** *50* (32), 7315-7320.

218. Li, Y.; Manickam, G.; Ghoshal, A.; Subramaniam, P., *Synth. Commun.* **2006,** *36* (7), 925-928.

219. Niemietz, M.; Perkams, L.; Hoffman, J.; Eller, S.; Unverzagt, C., *Chem. Commun.* **2011,** *47* (37), 10485-10487.

220. Hofstein, B. v., *Biochim. Biophys. Acta* **1961,** *48* (1), 159-163.

221. Williams, S. J.; Withers, S. G., *Carbohydr. Res.* **2000,** *327* (1-2), 27-46.

222. Brayer, G. D.; Sidhu, G.; Maurus, R.; Rydberg, E. H.; Braun, C.; Wang, Y.; Nguyen, N. T.; Overall, C. M.; Withers, S. G., *Biochemistry* **2000,** *39* (16), 4778-4791.

223. Rydberg, E. H.; Li, C.; Maurus, R.; Overall, C. M.; Brayer, G. D.; Withers, S. G., *Biochemistry* **2002,** *41* (13), 4492-4502.

224. Ly, H. D.; Withers, S. G., *Annu. Rev. Biochem* **1999,** *68*, 487-522.

225. White, A.; Tull, D.; Johns, K.; Withers, S. G.; Rose, D. R., *Nat. Struct. Biol.* **1996,** *3* (2), 149-154.

226. Notenboom, V.; Birsan, C.; Nitz, M.; Rose, D. R.; Warren, R. A. J.; Withers, S. G., *Nat. Struct. Biol.* **1998,** *5* (9), 812-818.

227. Braun, C.; Brayer, G. D.; Withers, S. G., *J. Biol. Chem.* **1995,** *270* (45), 26778-26781.

228. McCarter, J. D.; Withers, S. G., *J. Am. Chem. Soc.* **1996,** *118* (1), 241-242.

229. Kajimoto, T.; Liu, K. K. C.; Pederson, R. L.; Zhong, Z. Y.; Ichikawa, Y.; Porco, J. A.; Wong, C. H., *J. Am. Chem. Soc.* **1991,** *113* (16), 6187-6196.

230. Zechel, D. L.; Withers, S. G., *Acc. Chem. Res.* **2000,** *33* (1), 11-18.

231. Card, P. J.; Reddy, G. S., *J. Org. Chem.* **1983,** *48* (24), 4734-4743.

232. Somawardhana, C. W.; Brunngraber, E. G., *Carbohydr. Res.* **1981,** *94* (2), C14-C15.

233. Card, P. J., *J. Org. Chem.* **1983,** *48* (3), 393-395.

234. Thanna, S.; Lindenberger, J. J.; Gaitonde, V. V.; Ronning, D. R.; Sucheck, S. J., *Org. Biomol. Chem.* **2015,** *13* (29), 8080-8080.

235. McCarter, J. D. Mechanism-based inhibitors as in vitro and in vivo probes of glycosidase structure and mechanism. University of British Columbia, **1995**.

236. Ferrier, R. J.; Furneaux, R. H., *J. Chem. Soc., Perkin Trans. 1* **1977,** (18), 1996-2000.

237. Blattner, R.; Ferrier, R. J., *J. Chem. Soc., Perkin Trans. 1* **1980,** (0), 1523-1527.

238. Nelson, C. R., *Carbohydr. Res.* **1979,** *68* (1), 55-60.

239. Kocienski, P. J., *Protecting Groups*. 3rd ed.; Georg Thieme Verlag: Stuttgart, **2005**.

240. Baneyx, F., *Curr. Opin. Biotechnol.* **1999,** *10* (5), 411-421.

241. Carrio, M. M.; Cubarsi, R.; Villaverde, A., *FEBS Lett.* **2000,** *471* (1), 7-11.

242. Kopito, R. R., *Trends Cell Biol.* **2000,** *10* (12), 524-530.

243. Singh, S. M.; Panda, A. K., *J. Biosci. Bioeng.* **2005,** *99* (4), 303-310.

244. Upadhyay, A. K.; Singh, A.; Mukherjee, K. J.; Panda, A. K., *Front. Microbiol.* **2014,** *5*.

245. Ni, H.; Guo, P. C.; Jiang, W. L.; Fan, X. M.; Luo, X. Y.; Li, H. H., *J. Biotechnol.* **2016,** *231*, 65-71.

246. Pavan, M. E.; Pavan, E. E.; Cairo, F. M.; Pettinari, M. J., *Rev. Argent. Microbiol.* **2016,** *48* (1), 5-14.

247. Tripathi, N. K., *Chembioeng Reviews* **2016,** *3* (3), 116-133.

248. Upadhyay, V.; Singh, A.; Jha, D.; Singh, A.; Panda, A. K., *Microb. Cell Fact.* **2016,** *15*.

249. Upadhyay, V.; Singh, A.; Panda, A. K., *Protein Expression Purif.* **2016,** *117*, 52-58.

250. Klutts, S.; Pastuszak, I.; Edavana, V. K.; Thampi, P.; Pan, Y. T.; Abraham, E. C.; Carroll, J. D.; Elbein, A. D., *J. Biol. Chem.* **2003,** *278* (4), 2093-2100.

251. Hartl, F. U.; Bracher, A.; Hayer-Hartl, M., *Nature* **2011,** *475* (7356), 324-332.

252. Thomas, J. G.; Ayling, A.; Baneyx, F., *Appl. Biochem. Biotechnol.* **1997,** *66* (3), 197-238.

253. Kim, Y. E.; Hipp, M. S.; Bracher, A.; Hayer-Hartl, M.; Hartl, F. U., *Annu. Rev. Biochem.,* **2013**, *82*, 323-355.

254. Altamirano, M. M.; Golbik, R.; Zahn, R.; Buckle, A. M.; Fersht, A. R., *P. Natl. Acad. Sci. USA* **1997,** *94* (8), 3576-3578.

255. Johnson, W. C., *Annu. Rev. Biophys. Bio.* **1988,** *17*, 145-166.

256. Sreerama, N.; Woody, R. W., *Anal. Biochem.* **2000,** *287* (2), 252-260.

257. Farid, D. R. S., Examples of different pure secondary structures. In *Circular Dichroism (CD) Spectroscopy*, proteinchemist.com.

258. Sreerama, N.; Venyaminov, S. Y.; Woody, R. W., *Anal. Biochem.* **2000,** *287* (2), 243-251.

259. Corpet, F., *Nucleic Acids Res.* **1988,** *16* (22), 10881-10890.

260. Blesak, K.; Janecek, S., *Extremophiles* **2012,** *16* (3), 497-506.

261. Henrissat, B.; Bairoch, A., *Biochem. J* **1996,** *316* (Pt 2), 695.

262. Zona, R.; Chang-Pi-Hin, F.; O'Donohue, M. J.; Janecek, S., *Eur. J. Biochem.* **2004,** *271* (14), 2863-2872.

263. Janeček, Š., *Biologia* **2005,** *60* (Suppl 16), 177-184.

264. Erra-Pujada, M.; Chang-Pi-Hin, F.; Debeire, P.; Duchiron, F.; O'Donohue, M. J., *Biotechnol. Lett* **2001,** *23* (16), 1273-1277.

265. Imamura, H.; Fushinobu, S.; Yamamoto, M.; Kumasaka, T.; Jeon, B.-S.; Wakagi, T.; Matsuzawa, H., *J. Biol. Chem.* **2003,** *278* (21), 19378-19386.

266. Santos, C. R.; Tonoli, C. C. C.; Trindade, D. M.; Betzel, C.; Takata, H.; Kuriki, T.; Kanai, T.; Imanaka, T.; Arni, R. K.; Murakami, M. T., *Proteins* **2011,** *79* (2), 547-557.

267. Hoefsloot, L. H.; Hoogeveen-Westerveld, M.; Kroos, M.; Beeumen, J. v.; Reuser, A. J.; Oostra, B., *EMBO J* **1988,** *7* (6), 1697-1704.

268. Takata, H.; Ohdan, K.; Takaha, T.; Kuriki, T.; Okada, S., *J., Appl. Glycosci.* **2003**, *50*, 15-20.

269. Palomo, M.; Kralj, S.; van der Maarel, M. J. E. C.; Dijkhuizen, L., *Appl. Environ. Microbiol.* **2009,** *75* (5), 1355-1362.

270. Fan, Q.; Xie, Z. J.; Zhan, J. L.; Chen, H. L.; Tian, Y. Q., *Starch-Starke* **2016,** *68* (3-4), 355-364.

271. Mohtar, N. S.; Rahman, M. B. A.; Abd Rahman, R. N. Z. R.; Leow, T. C.; Salleh, A.; Isa, M. N. M., *Peerj* **2016,** *4*.

272. Moulay, S., *J. Polym. Eng.* **2013,** *33* (5), 389-443.

273. Fersht, A., The basic equations of enzyme kinetics. In *Enzyme Structure and Mechansim*, second ed.; W. H. Freeman and Company: New York, **1984**; pp 98-120.

274. Katsuya, Y.; Mezaki, Y.; Kubota, M.; Matsuura, Y., *J. Mol. Biol.* **1998,** *281* (5), 885-897.

275. Anthon, G. E.; Barrett, D. M., *Anal. Biochem.* **2002,** *305* (2), 287-289.

276. Bainor, A.; Chang, L.; McQuade, T. J.; Webb, B.; Gestwicki, J. E., *Anal. Biochem.* **2011,** *410* (2), 310-312.

277. Espada, J., *J. Biol. Chem.* **1962,** *237* (12), 3577-3581.

278. Blankenfeldt, W.; Asuncion, M.; Lam, J. S.; Naismith, J. H., *EMBO J* **2000,** *19* (24), 6652-6663.

279. Cori, C. F.; Schmidt, G.; Cori, G. T., *Science* **1939,** *89* (2316), 464-465.

280. Cori, G. T.; Cori, C. F., *J. Biol. Chem.* **1940,** *135* (2), 733-756.

281. Hanes, C. S., *P. R. Soc. London. B, Biol. Sci.* **1940,** *129* (855), 174-208.

282. Zhou, X.; Arthur, G., *J. Lipid Res.* **1992,** *33* (8), 1233-6.

283. Cogan, E. B.; Birrell, G. B.; Griffith, O. H., *Anal. Biochem.* **1999,** *271* (1), 29-35.

284. Feng, J.; Chen, Y.; Pu, J.; Yang, X.; Zhang, C.; Zhu, S.; Zhao, Y.; Yuan, Y.; Yuan, H.; Liao, F., *Anal. Biochem.* **2011,** *409* (1), 144-149.

285. Hermans, M.; Kroos, M.; Van Beeumen, J.; Oostra, B.; Reuser, A., *J. Biol. Chem.* **1991,** *266* (21), 13507-13512.

286. Frandsen, T. P.; Svensson, B., *Plant Mol. Biol.* **1998,** *37* (1), 1-13.

287. Okuyama, M.; Okuno, A.; Shimizu, N.; Mori, H.; Kimura, A.; Chiba, S., *FEBS J* **2001,** *268* (8), 2270-2280.

288. Frackman, S.; Kobs, G.; Simpson, D.; Storts, D., *Promega notes* **1998,** *65* (27-29), 27-29.

289. Henrissat, B., *Biochem. J* **1991,** *280*, 309-316.

290. Sambou, T.; Dinadayala, P.; Stadthagen, G.; Barilone, N.; Bordat, Y.; Constant, P.; Levillain, F.; Neyrolles, O.; Gicquel, B.; Lemassu, A.; Daffe, M.; Jackson, M., *Mol. Microbiol.* **2008,** *70* (3), 762-774.

291. Hoffmann, A.; Bukau, B.; Kramer, G., *Biochim. Biophys. Acta* **2010,** *1803* (6), 650-661.

292. Kramer, G.; Rauch, T.; Rist, W.; Vorderwülbecke, S.; Patzelt, H.; Schulze-Specking, A.; Ban, N.; Deuerling, E.; Bukau, B., *Nature* **2002,** *419*, 171.

293. Valent, Q. A.; Kendall, D. A.; High, S.; Kusters, R.; Oudega, B.; Luirink, J., *EMBO J* **1995,** *14* (22), 5494-5505.

294. Schaffitzel, E.; Rüdiger, S.; Bukau, B.; Deuerling, E., *Biol. Chem.*, **2001**, 382, 1235-1243.

295. Ferbitz, L.; Maier, T.; Patzelt, H.; Bukau, B.; Deuerling, E.; Ban, N., *Nature* **2004,** *431*, 590.

296. Baram, D.; Pyetan, E.; Sittner, A.; Auerbach-Nevo, T.; Bashan, A.; Yonath, A., *P. Natl. Acad. Sci. USA* **2005,** *102* (34), 12017-12022.

297. Singhal, K.; Vreede, J.; Mashaghi, A.; Tans, S. J.; Bolhuis, P. G., *PLoS Comp. Biol.* **2015,** *11* (10), e1004444.

298. Liu, K.; Maciuba, K.; Kaiser, C. M., *Biophys. J.* **2018**, *114* (3), 552a.

299. Martinez-Hackert, E.; Hendrickson, W. A., *Cell* **2009,** *138* (5), 923-934.

300. Wagner, G. K.; Pesnot, T., *ChemBioChem* **2010,** *11* (14), 1939-1949.

301. Palcic, M. M.; Sujino, K., *Trends Glycosci. Glyc.* **2001,** *13* (72), 361-370.

302. Srivastava, G.; Hindsgaul, O.; Palcic, M. M., *Carbohydr. Res.* **1993,** *245* (1), 137-144.

303. Gupta, V.; Bamezai, R. N. K., *Protein Sci.* **2010,** *19* (11), 2031-2044.

304. Holmes, R. S.; Goldberg, E., *Comput. Biol. Chem.* **2009,** *33* (5), 379-385.

305. Karlsson, O. P.; Dahlqvist, A.; Vikström, S.; Wieslander, Å., *J. Biol. Chem.* **1997,** *272* (2), 929-936.

306. Parsonage, D.; Newton, G. L.; Holder, R. C.; Wallace, B. D.; Paige, C.; Hamilton, C. J.; Dos Santos, P. C.; Redinbo, M. R.; Reid, S. D.; Claiborne, A., *Biochemistry* **2010,** *49* (38), 8398-8414.

307. Glover, K. J.; Weerapana, E.; Imperiali, B., *P. Natl. Acad. Sci. USA* **2005,** *102* (40), 14255-14259.

308. Gibson, R. P.; Lloyd, R. M.; Charnock, S. J.; Davies, G. J., *Acta Crystallogr D* **2002,** *58*, 349-351.

309. Elbein, A. D.; Pan, Y. T.; Pastuszak, I.; Carroll, D., *Glycobiology* **2003,** *13* (4), 17R-27R.

310. Penna, S., *Trends Plant Sci.* **2003,** *8* (8), 355-357.

311. De Smet, K. A. L.; Weston, A.; Brown, I. N.; Young, D. B.; Robertson, B. D., *Microbiology* **2000,** *146* (1), 199-208.

312. Avonce, N.; Mendoza-Vargas, A.; Morett, E.; Iturriaga, G., *BMC Evol. Biol.* **2006,** *6* (1), 109.

313. Errey, J. C.; Lee, S. S.; Gibson, R. P.; Fleites, C. M.; Barry, C. S.; Jung, P. M. J.; O'Sullivan, A. C.; Davis, B. G.; Davies, G. J., *Angew. Chem. Int. Ed.* **2010,** *49* (7), 1234-1237.

314. Nunes-Costa, D.; Maranha, A.; Costa, M.; Alarico, S.; Empadinhas, N., *Glycobiology* **2017,** *27* (3), 213-227.

315. Empadinhas, N.; da Costa, M. S., *Environ. Microbiol.* **2011,** *13* (8), 2056-2077.

316. Costa, J.; Empadinhas, N.; Goncalves, L.; Lamosa, P.; Santos, H.; da Costa, M. S., *J. Bacteriol.* **2006,** *188* (3), 1022-1030.

317. Costa, J.; Empadinhas, N.; da Costa, M. S., *J. Bacteriol.* **2007,** *189* (5), 1648-1654.

318. Fernandes, C.; Empadinhas, N.; da Costa, M. S., *J. Bacteriol.* **2007,** *189* (11), 4014-4019.

319. Nishihara, K.; Kanemori, M.; Kitagawa, M.; Yanagi, H.; Yura, T., *Appl. Environ. Microbiol.* **1998,** *64* (5), 1694-1699.

320. Yamada, K.; Fujita, H.; Kunishima, M., *Org. Lett.* **2012,** *14* (19), 5026-5029.

321. Schmidt, A.-K. C.; Stark, C. B. W., *Org. Lett.* **2011,** *13* (21), 5788-5791.

322. Matsuo, K.; Shindo, M., *Org. Lett.* **2011,** *13* (16), 4406-4409.

323. Goubert, M.; Toupet, L.; Sinibaldi, M.-E.; Canet, I., *Tetrahedron* **2007,** *63* (34), 8255-8266.

324. Hirooka, M.; Koto, S., *Bull. Chem. Soc. Jpn.* **1998,** *71* (12), 2893-2902.

325. Koto, S.; Morishima, N.; Shichi, S.; Haigoh, H.; Hirooka, M.; Okamoto, M.; Higuchi, T.; Shimizu, K.; Hashimoto, Y.; Irisawa, T.; Kawasaki, H.; Takahashi, Y.; Yamazaki, M.; Mori, Y.; Kudo, K.; Ikegaki, T.; Suzuki, S.; Zen, S., *Bull. Chem. Soc. Jpn.* **1992,** *65* (12), 3257-3274.

326. Tsuzuki, M.; Tsuchiya, T., *Carbohydr. Res.* **1998,** *311* (1), 11-24.

327. Chatterjee, D.; Paul, A.; Rajkamal; Yadav, S., *RSC Adv.* **2015,** *5* (38), 29669-29674.