Drug targets in *Mycobacterium tuberculosis* α-glucan synthesis

Sibyl Florence Doto Batey

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

Department of Biological Chemistry
John Innes Centre
Norwich
September 2017

Abstract

α-Glucans are important energy storage polysaccharides in bacteria, plants and animals. In *Mycobacterium tuberculosis*, α-glucan also functions as a virulence factor that is exported to the mycobacterial capsule and interacts with human immune receptors. In *M. tuberculosis* and other actinomycetes, α-glucan is synthesised from maltose-1-phosphate by the maltosyl transferase GlgE and the α-1,6-branching enzyme GlgB. These enzymes have been genetically validated as tuberculosis drug targets. The loss of the α-glucan virulence factor is exacerbated by the toxic accumulation of maltose-1-phosphate, which results in a pleiotropic stress response and cell death.

α-Glucan produced by this pathway has shorter branches than classical bacterial α-glucan, but the mechanistic basis of this was not fully understood. To address this, I produced α-glucan *in vitro*, demonstrating GlgE and GlgB alone were sufficient to generate the distinct architecture. To investigate the determinants of α-glucan branch lengths, I solved high resolution crystal structures of *Mycobacterium smegmatis* GlgB bound to different oligosaccharides. These enabled the identification of several substrate binding sites, including the donor and acceptor sites of a newly forming branchpoint, giving novel insights into the action of glucan branching enzymes.

To address *M. tuberculosis* GlgB as a drug target, I developed a high-throughput screening assay and used this to identify a potent, small-molecule inhibitor. I synthesised a selection of analogues to investigate structure-activity relationships. I also assessed the action of the compounds *in vivo*, demonstrating growth inhibition at high concentrations, but not via the GlgB target. Finally, I explored trehalose analogues as potential precursors for GlgE inhibitors, focussing on the synthesis, uptake and processing of 4-deoxy trehalose analogues.

This multidisciplinary work gives a deeper understanding of mycobacterial α-glucan synthesis and how this can be targeted to develop new tuberculosis therapeutics.

In loving memory of Doreen Owen

12th March 1926 – 1st October 2016

Acknowledgements

Contents

List of Figures

List of Tables

Abbreviations

λ	Wavelength
Å	Ångström
ACR	Acarbose
ADP	Adenosine diphosphate
AG	Arabinogalactan
AMP	Adenosine monophosphate
aq	Aqueous
ATP	Adenosine triphosphate
BCG	Bacillus Calmette-Guérin
BSA	Bovine serum albumin
CBM	Carbohydrate binding module
CceBE1	Cyancobacterial branching enzyme 1
CE	Capillary electrophoresis
CFU	Colony forming unit
COSY	Correlation spectroscopy
DAST	Diethylaminosulphur trifluoride
dH_2O	Deionised water
DMAP	4-Dimethylaminopyridine
DMSO	Dimethyl sulphoxide
DP	Degree of polymerisation
DSF	Differential scanning fluorimetry
ESI	Electrospray ionisation
GDP	Guanosine diphosphate
GH	Glycoside hydrolase
GT	Glycosyl transferase
HPAEC	High performance anion exchange chromatography
IPTG	Isopropyl β-D-thiogalactopyranoside
HRMS	High-resolution mass spectra
HTS	High-throughput screening
INH	Isonazid
ISA	Isoamylase
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
M1P	α-Maltose-1-phosphate
MALDI	Matrix assisted laser desorption / ionisation
MDR-TB	Multi-drug resistant TB
MES	2-(*N*-morpholino)ethanesulphonic acid
MGLP	Methylglucose lipopolysaccharide
MIC	Minimum inhibitory concentration
MS	Mass spectrometry
MsGlgB	*M. smegmatis* GlgB
MtGlgB	*M. tuberculosis* GlgB
MW	Molecular Weight
NMR	Nuclear magnetic resonance
OD	Optical density

PAD	Pulsed amperometric detection
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PG	Peptidoglycan
P_i	Phosphate
RIF	Rifampicin
RMSD	Root mean square deviation
SAR	Structure activity relationships
SDS	Sodium dodecyl sulphate
SPR	Surface plasma resonance
STD	Saturation transfer difference
TB	Tuberculosis
TDM	Trehalose-6,6'-dimycolate
TEM	Transmission electron microscopy
TLC	Thin layer chromatography
TLS	Translation libration screw
TOF	Time of flight
Tre	Trehalose
U	Unit
UDP	Uridine diphosphate
UV	Ultraviolet
v/v	Volume/volume
w/v	Weight/volume
WT	Wild type
XDR-TB	Extensively-drug resistant TB

1. Introduction

This thesis describes the investigation of potential drug targets in *Mycobacterium tuberculosis* α-glucan biosynthesis. Firstly, I briefly review the remarkable structural and functional diversity of carbohydrates, along with the multitude of enzymes that control carbohydrate metabolic processes. The structure, function and biosynthesis of α-glucan polysaccharides across different organisms are then discussed, before focussing on the *M. tuberculosis* pathways. I go on to discuss the bacterium in the context of the devastating infectious disease it causes and the current treatments available. Then, I detail the efforts so far to target the α-glucan biosynthetic pathway in *M. tuberculosis*. Finally, I outline the questions that this work will address in order to develop the understanding of mycobacterial α-glucan synthesis and target the enzymes involved in this process with potential anti-tuberculosis therapeutics.

1.1 Carbohydrates

1.1.1 Carbohydrate diversity

Carbohydrates are ubiquitous organic molecules that play an integral role in nearly all living processes.[1] Carbohydrates are fundamental to biological energy storage and transfer, being the main product of photosynthesis and the substrate of respiration. They are the major component of cell walls, from cellulose in plants to peptidoglycan in bacteria; and as chitin, help form the exoskeleton in crustaceans and insects. In addition to structural roles, carbohydrates are essential to cellular recognition and adhesion, for example in reproduction and immunity. They are also important signalling molecules and ligands, often when conjugated to other macromolecules, such as proteins or lipids.[2]

The name derives from hydrated carbon, and indeed, carbohydrates were historically considered to be compounds with the empirical formula $C_n(H_2O)_n$, where $n \geq 3$. Generally, these are 5- or 6- carbon polyhydroxylated aldoses, with ketoses being less common.[3] The archetypal carbohydrate is the aldohexose glucose, which is the most abundant organic molecule on earth. However, the term carbohydrate also includes many other molecules that do not conform to the empirical formula, such as the backbone of DNA, deoxyribose, and those with additional heteroatoms, such as iminosugars and glycosyl thiols. Carbohydrates exist almost exclusively as cyclic hemiacetals, with the 6-membered pyranose ring usually favoured over the 5-membered furanose ring. The anomeric carbon is that which is derived from the carbonyl carbon in the open-chain form. A single carbohydrate ring is termed a monosaccharide, and rings are linked by glycosidic bonds to

form di-, oligo- and poly- saccharides. In a chain of sugar molecules, the end that retains a free anomeric carbon is termed the reducing end, with the non-reducing end at the other extremity.[4]

The wide-ranging biological roles of carbohydrates are matched by an extensive structural diversity. Typically, they contain numerous stereogenic centres, and subtle stereochemical variations have dramatic effects on carbohydrate properties, and hence their biological functions.[5] For example, the α and β isomers of glucose play very different biological roles. Similarly, glucose and galactose differ only in the stereochemistry of one hydroxyl group, but again have markedly different properties. Another source of variation lies in the potential for linkages at different positions in the same monosaccharide ring. In turn, this creates polysaccharides with radically different secondary structures, such as the linear architecture of the β-1,4-glucose polymer cellulose,[6] compared to the triple helices that form in the β-1,3-glucose polymer paramylon,[7] **Figure 1.1**. The combination of mixed linkages in the same polysaccharide molecule creates an even greater number of structural possibilities.

Overall, this results in a vast structural complexity that far surpasses that of proteins, lipids or nucleic acids. Furthermore, carbohydrate structure is not encoded in genetic information, making it much more challenging to manipulate. This is compounded by the difficulties in distinguishing between different isomers and linkages. Carbohydrates are therefore the most challenging class of biomolecule to investigate, resulting in glycobiology historically lagging behind other fields.[8] The advent of automated carbohydrate synthesis[9] and significant advances in mass spectrometry methods[10-11] go some way to addressing these challenges, but there is clearly much more to be done in glycobiology to match the progress made in other disciplines.

A) Fischer projection of D-glucose; **B)** D-glucopyranose and D-glucofuranose; Pyranose; Furanose

C) α; β

D) Galactose

Triplex; β-1,3-glucan

Sheet; β-1,4-glucan
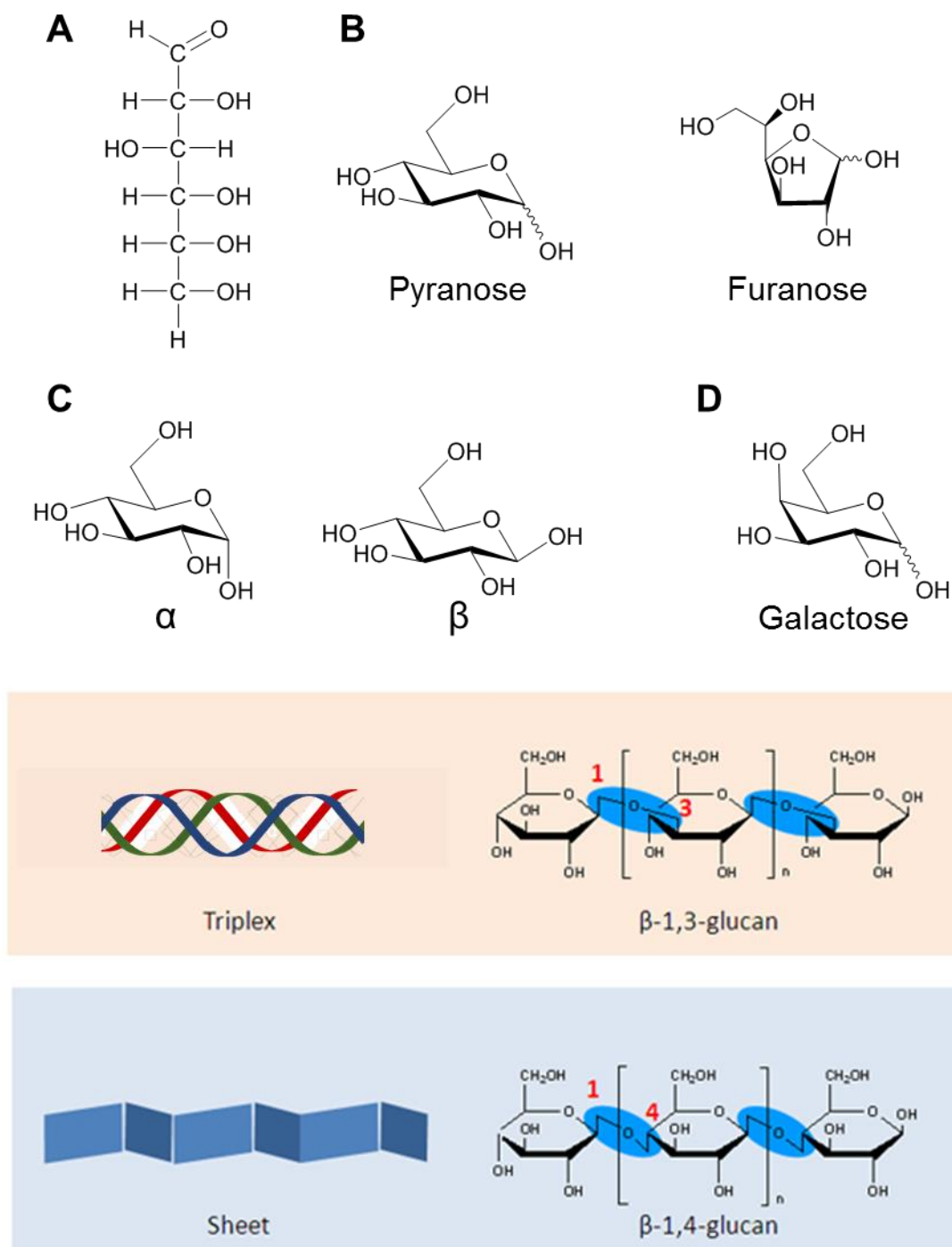
Figure 1.1 Carbohydrate structures. A) Fischer projection of D-glucose; B) D-glucopyranose and D-glucofuranose are the two cyclic hemiacetal forms of D-glucose; C) α-D-glucopyranose and β-D-glucopyranose differ in the stereochemistry at the 1-OH; D) D-galactopyranose differs from D-glucopyranose in the stereochemistry of the 4-OH; E) Structures of paramylon and cellulose, forming a triple helix or a sheet respectively, adapted from 'Development of *Euglena*-based bioplastics'.[12]

1.1.2 Carbohydrate active enzymes

A plethora of carbohydrate metabolic enzymes is required to orchestrate the immense structural variety of carbohydrates. These are usually highly specific for a particular stereochemistry or glycosidic linkage, although some enzymes can accommodate a range of substrates. Over the last two decades, the carbohydrate active enzymes database[13] (CAZy; http://www.cazy.org/) has been developed, in which enzymes are classified according to their amino acid sequence. The major CAZy classes are glycoside hydrolases[14] (GHs) and glycosyltransferases[15] (GTs), with the remaining classes comprising polysaccharide lyases, carbohydrate esterases[16] and auxiliary activities.[17] GHs catalyse the cleavage and formation of glycosidic bonds, whilst GTs catalyse the transfer of saccharides from activated phosphoester donors to a specific acceptor. There are two types of mechanisms for GHs and GTs: retaining and inverting, referring to retention or inversion of the anomeric configuration. Each CAZy class is divided into a numerical family according to the protein sequence. Since this usually correlates with the three-dimensional structure, characteristic folds can be identified in each family, such as a $(\beta/\alpha)_8$ barrel. CAZymes often contain discrete units that bind specific saccharides, termed carbohydrate binding modules (CBMs), which are themselves classified into families based on sequence.[18]

This thesis focuses on enzymes belonging to the GH13 family, which act on substrates containing α-glycosidic linkages and includes diverse functionalities, such as isomerases, polymerases, branching enzymes and debranching enzymes. Reactions of GH13 enzymes retain the α-configuration, following a classical Koshland double displacement mechanism.[19-20] The aspartate nucleophile attacks the face opposite to the anomeric hydroxyl to form a covalent intermediate. The glutamate acid/base residue then aids attack by water or a sugar hydroxyl to release the product with a retained configuration. The reaction proceeds via oxocarbenium ion-like transition states, **Figure 1.2**. The subsites that bind each monosaccharide moiety are labelled -*n* to +*n*, where -*n* is the non-reducing end of the sugar and +*n* is the reducing end. Cleavage takes places between the -1 and +1 subsites,[21] **Figure 1.3**. When the reaction results in the formation of a new glycosidic bond, the moiety that contains the resultant anomeric carbon is termed the donor, whereas the moiety that contains the hydroxyl nucleophile is called the acceptor.[4]
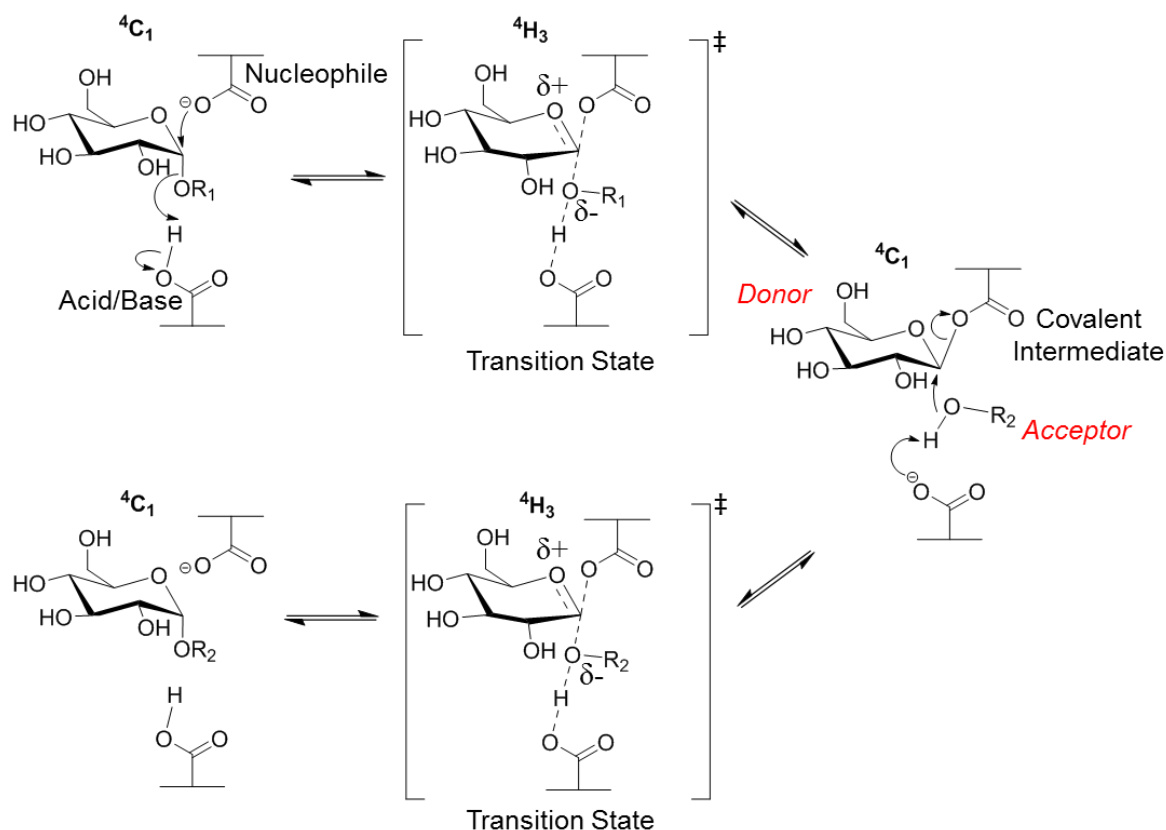
Figure 1.2 Mechanism of an α-retaining GH enzyme. The classical Koshland double displacement mechanism gives overall retention of the α-configuration. R_1 = a glycosyl moiety leaving group; R_2 = either a glycosyl moiety in the case of bond formation or H in the case of hydrolysis. The mechanism proceeds via an oxocarbenium ion-like transition state in which the pyranose most likely adopts a $^4H_3$ conformation, in contrast to the ground-state $^4C_1$ conformation.
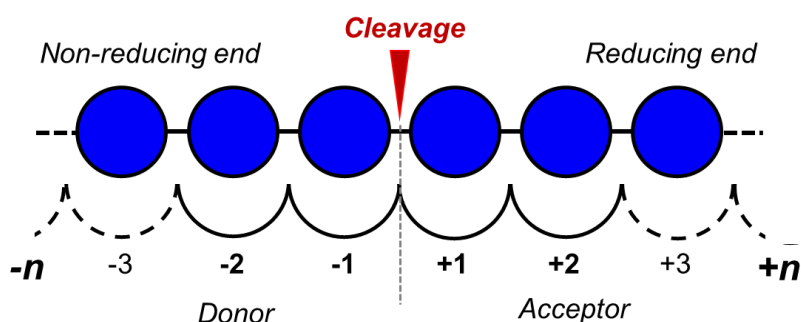
Figure 1.3 Subsite nomenclature in GH enzymes. Protein subsites are labelled from -n to +n, with the oligosaccharide orientated such that the non-reducing end is located at the -n subsite and the reducing end subsite is located in the +n subsite. Cleavage takes place between the -1 and +1 subsites. In the case of bond formation, the -n subsites comprise the donor site and the +n subsites comprise the acceptor site.

1.2   α-Glucans

α-Glucans are a subset of carbohydrate polymers composed of α-D-glucose units, which can be linked by 1→2, 1→3, 1→4 or 1→6 α-glycosidic bonds.[1] In this thesis I exclusively refer to α-glucans as those consisting of 1→4 linked chains, which are connected by 1→6 α-glycosidic bonds, known as branchpoints. α-Glucans have a number of biological roles, including being involved in the formation of dental caries by bacteria.[22] Most commonly they function as energy storage molecules and include the ubiquitous polysaccharides starch and glycogen.[23]

1.2.1   Structure of α-glucans

The structural features of α-glucan polymers are illustrated in **Figure 1.4**. One feature of the polysaccharide architecture is the average length of the α-1,4 chains, which is measured in units of degree of polymerisation (DP), pertaining to the number of glucose residues in the chain. This has a reciprocal relationship with the percentage of branching, which can be independently determined by quantifying the number of α-1,6 linkages versus α-1,4. The internal chain length can also be measured, giving the average distance between one branchpoint and the furthest branch it bears,[24] **Figure 1.5**.
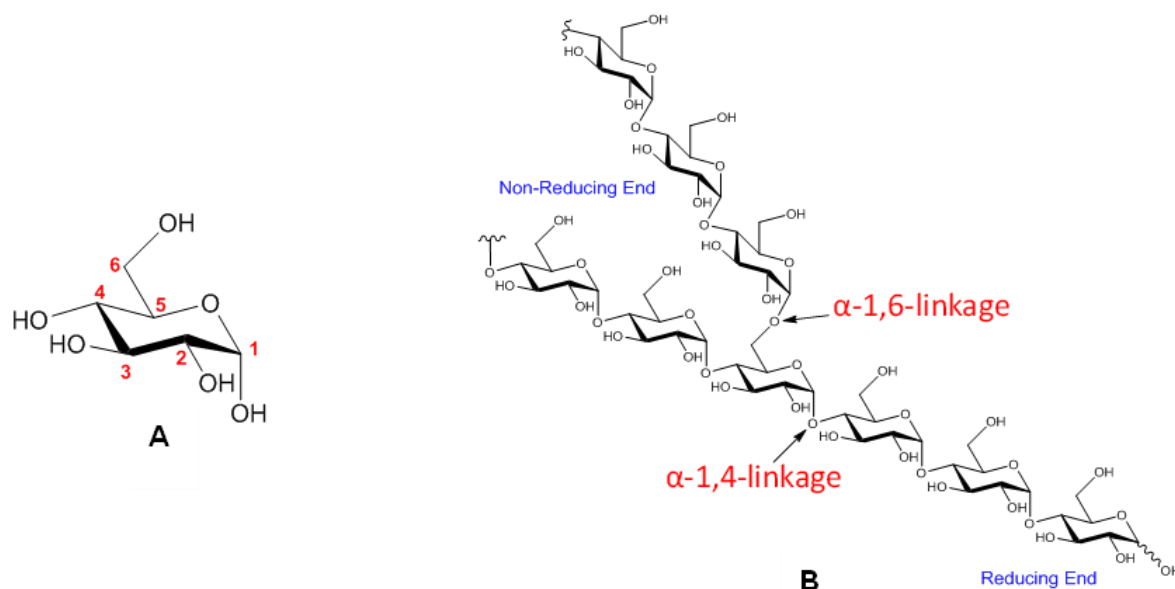
Figure 1.4 Structure of α-glucans. A) α-D-glucopyranose is linked via 1,4 or 1,6 α-glycosidic bonds. B) Basic structure of α-glucan, consisting of two α-1,4 chains linked by an α-1,6 bond. The free anomeric carbon is the reducing end and the free 4-hydroxyls are the non-reducing ends.

1.2.2 α-Glucan synthesis

Synthesis of α-glucans is carried out by the iterative action of an α-1,4-glucan polymerase and an α-1,6-branching enzyme. The polymerase extends an α-glucan chain and, when a threshold length is surpassed, the branching enzyme cleaves an α-1,4 linkage within the chain and transfers this to a 6-OH of another glucose unit in the molecule, thus creating an α-1,6 branchpoint. This process results in three types of chains within an α-glucan molecule: a progenitor C chain, which retains a free reducing end; a B chain that originates from a branchpoint and itself bares one or more branches; and an A chain that originates from a branchpoint but itself bares no branches, **Figure 1.5**. Another measure of α-glucan structure is the A:BC chain ratio, which defines the distribution of branching in the molecule, and is directly related to the number of branches emanating from each B/C chain.[24]
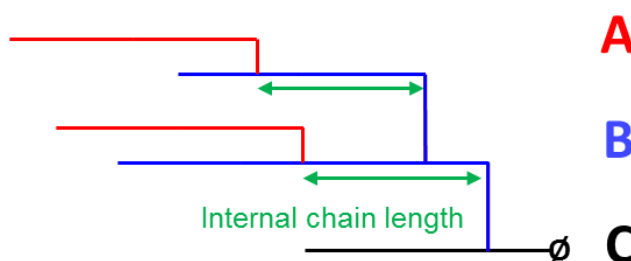
Figure 1.5 Types of chains in α-glucan. The external A chains bare no branches, the internal B chains bares one or more branches and the C chain is the only one that does not originate from a branch and hence retains a free reducing end, represented with ø. The internal chain length is the distance from one branchpoint to the next.

1.2.3 Starch

Starch is the major energy reserve of photosynthetic eukaryotes and the most common carbohydrate in the human diet.[25] It is formed as insoluble, semi-crystalline granules which consist of two polysaccharide fractions: amylose and amylopectin.[26] Amylose contains very few α-1,6 branches and is typically folded as single, left-handed helix, denoted V-amylose, although A and B double helical structures are also known.[27] Amylopectin usually accounts for at least 70% of the granule and contains numerous α-1,6 branches, which are arranged in clusters at regular intervals along the main chain. Double helices form between cluster strands and these pack together to form lamellae within the granule, **Figure 1.6**.[25] Amyloses are typically reported to have average chain lengths of DP 200-700, whereas amylopectin has average chain lengths of DP 20-25, corresponding to 4-5% α-1,6 linkages in the molecule.[28]
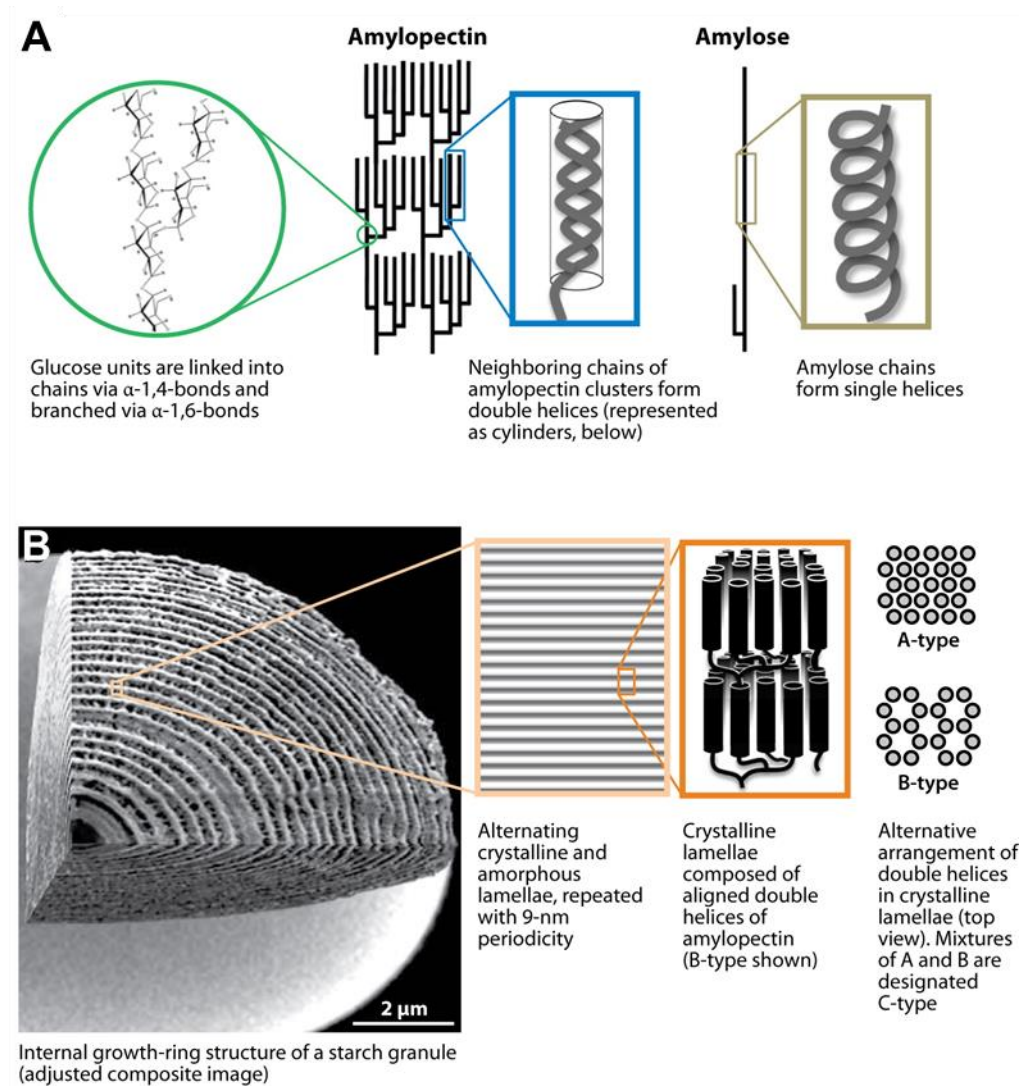
Figure 1.6 Starch structure. A) The arrangement of amylose and amylopectin; B) A cross-section of a starch granule from potato. Adapted from Zeeman SC, *et al*. 2010,[25] with permission.

1.2.4 Glycogen

Glycogen is the analogous carbohydrate energy store in animals, fungi and bacteria.[23] Glycogen has an arboreal structure, with branches dispersed more evenly through the molecule. Average branch lengths are typically DP 10-12, giving a higher overall percentage of α-1,6 linkages (8-10%).[29] It forms dendrimeric β-particles with a diameter of 10-100 nm, and these can aggregate to give larger α-particles, 100s of nm in diameter, **Figure 1.7**. Unlike its botanic counterpart, glycogen is largely soluble and can be dissolved in water at ambient temperature.[30] In mammals, a priming protein, glycogenin, initiates glycogen synthesis through covalent attachment of glucose molecules to its Tyr195 residue. Until very recently this was assumed to be essential, however, glycogenin-deficient mice were recently shown to accumulate high levels of glycogen.[31]
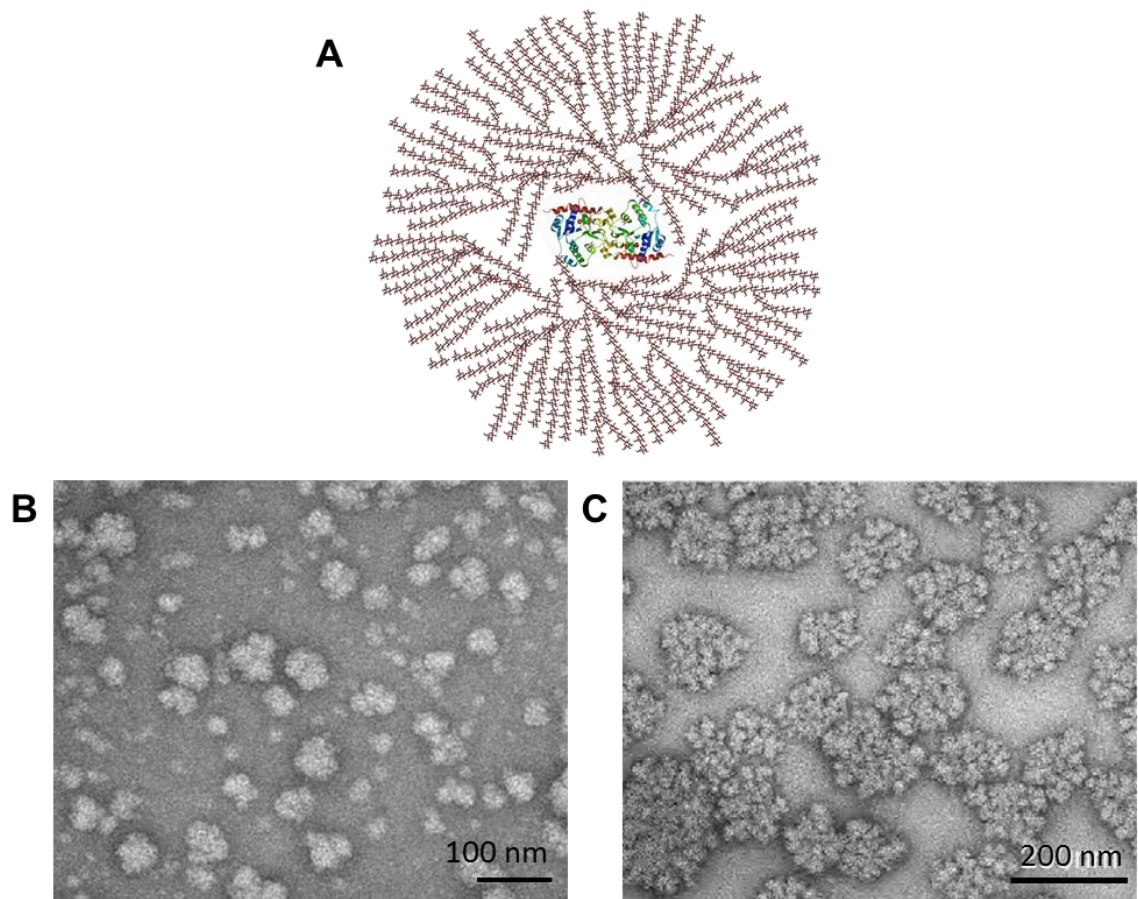
Figure 1.7 Glycogen structure. A) Representation of the molecular structure of mammalian glycogen originating from a glycogenin dimer. Image by Mikael Häggström. B) β-particles of *E. coli* glycogen.[32] C) α-particles of rat liver glycogen.[30] All images used with permission.

1.2.5 Bacterial glycogen synthesis

Classical glycogen biosynthesis in bacteria does not require glycogenin and takes place via the GlgABC pathway, which was first described in *Escherichia coli*. In this pathway, a glucose unit is activated by the addition of ADP via the action of GlgC on α-glucose-1-phosphate, using an AMP unit from ATP. ADP-glucose is then polymerised via α-1,4 glucosidic linkages by glycogen synthase, GlgA, to generate a linear α-glucan. Finally, α-1,6 branches are introduced by the branching enzyme, GlgB, which cleaves oligosaccharides at the non-reducing end and transfers them to a C6-hydroxyl group within or between chains,[29] **Figure 1.8**. Glycogen can be degraded to α-glucose-1-phosphate by the action of glycogen phosphorylase, GlgP,[33] and the debranching enzyme, GlgX,[34] thus completing the metabolic cycle. These enzymes have been biochemically characterised from a number of different bacteria, but their structure-activity relationships remain poorly understood.
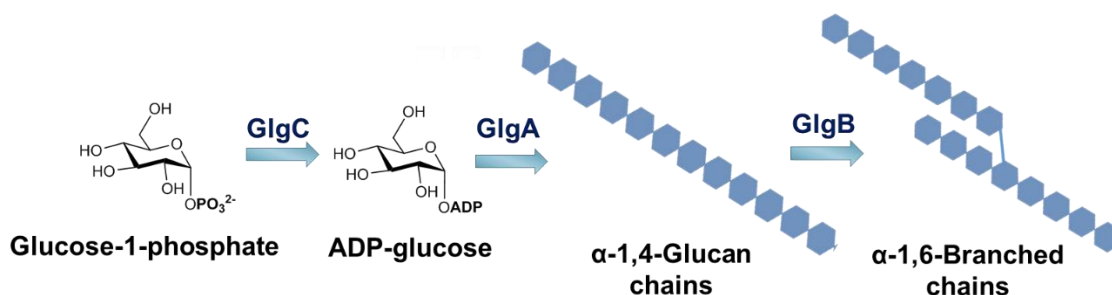
Figure 1.8 Classical glycogen synthesis pathway. Glucose-1-phosphate is converted into ADP-glucose by GlgC. This activated sugar is a substrate for the α-1,4 polymerase GlgA and GlgB introduces α-1,6 branches. Pyranose units are represented as blue hexagons.

Glycogen synthesis by this route is controlled by allosteric regulation of GlgC whereby glycolytic intermediates activate GlgC activity and where ADP, AMP and P_i are common inhibitors. For *M. smegmatis* GlgC, fructose-1,6-bisphosphate and fructose-6-phosphate were found to be activators, and AMP and ADP found to be inhibitors.[35]

1.3 α-Glucan metabolism in *M. tuberculosis*

1.3.1 α-Glucans in *M. tuberculosis*

It had long been established that, in common with many other bacteria,[22] mycobacteria accumulated a glycogen-like α-glucan polysaccharide in nitrogen-limiting growth conditions.[36] More recently, it was found that α-glucan is the main component of the loosely associated capsule, the outermost layer of the mycobacterial cell envelope.[37] Here, it

functions as a virulence factor and mutant *M. tuberculosis* strains lacking α-glucan had significantly attenuated growth in the both the acute and chronic phases of infection in a murine model.[38] *M. tuberculosis* α-glucan binds to the human immune receptors DC-SIGN and CR3 *in vitro*, and the interaction which each is mediated by internal glucosyl residues.[39] The capsular α-glucan has also been implicated in the persistence of *M. tuberculosis* in mice[40] and can induce monocytes to differentiate into altered dendritic cells, allowing mycobacteria to evade the immune host response.[41]

A further type of α-glucan found in mycobacteria and some *Nocardia* species are the 6-*O*-methylglucose lipopolysaccharides (MGLPs), whose metabolism was thought to be linked to capsular α-glucan metabolism. MGLPs consist of a predominantly 6-*O*-methylated α-1,4 backbone, 15-20 carbons in length, with an α-1,6 linked di-glucosylglycerate moeity at the reducing end.[42] Two β-1,3-glucose branches are found near the reducing end and there is extensive acylation across the molecule, **Figure 1.9**. MGLPs are thought to function as regulators/chaperones of fatty acid biosynthesis, with the nature of acylation determining fatty acid binding specificity.[43]
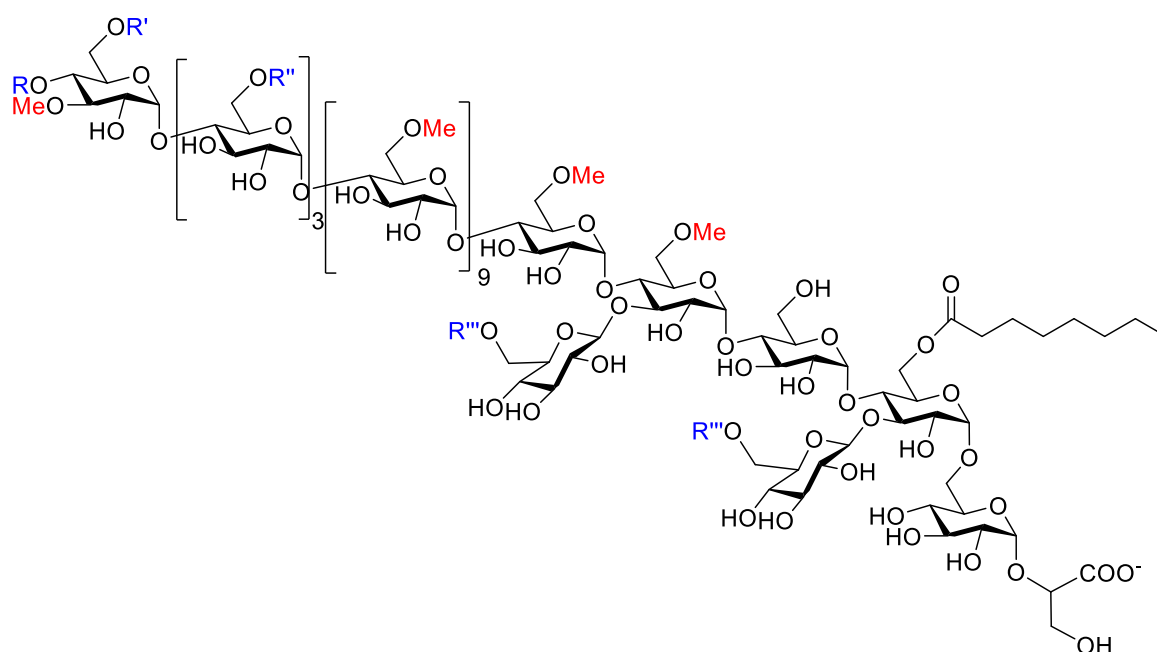
Figure 1.9 The structure of MGLP The molecule consists of an α-1,4 chain linked by an α-1,6 bond to glycosylglycerate at the reducing end. Two β-D-glucose branches emanate from the second and fourth residues in the chain. R = isobutyrate; R' = propionate, R' = acetate, R''' = succinate.[43-44]

1.3.2 The GlgE pathway

It had been presumed that *M. tuberculosis* α-glucan synthesis was analogous to that in other bacteria, because homologues of the *E. coli* genes *glgA, glgB* and *glgC* had been

identified.[40, 45] However, unlike in *E. coli*, *glgB* was found on a separate operon with other genes implicated in α-glucan metabolism, including *glgE*.[40] The homologue of *glgE* in *M. smegmatis* was thought to encode for a glucanase, due to a perceived increase in glycogen accumulation when the activity of GlgE was impaired by a temperature sensitive mutation.[46] The *glgB* and *glgE* genes also clustered with *treS* and *pep2*, mirroring a similar arrangement of genes previously seen in duplicated operons in *Streptomyces coelicolor*.[47] TreS was known to interconvert trehalose and maltose[48] and had previously been proposed to function as a link between trehalose and glycogen metabolism.[49-50] This was finally confirmed by the discovery of the novel GlgE-mediated α-glucan synthesis pathway,[51] **Figure 1.10**.
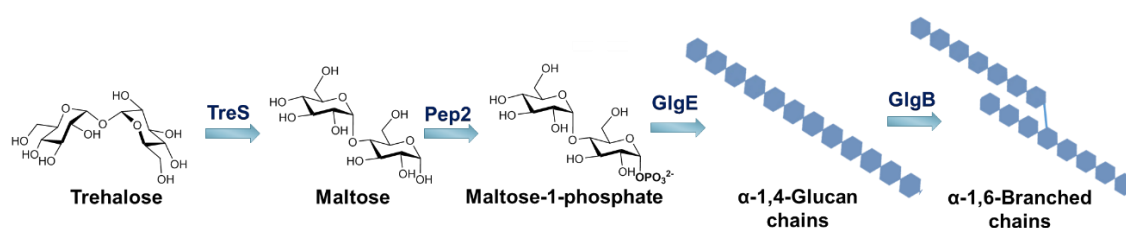
Figure 1.10 The GlgE pathway. Trehalose is isomerised to maltose by TreS and this is then phosphorylated by Pep2 to make maltose-1-phosphate. This is the substrate for the α-1,4 polymerase GlgE and α-1,6 branches are subsequently introduced by the branching enzyme GlgB. Pyranose units are represented as blue hexagons.

In this pathway trehalose is firstly isomerised to α-maltose by TreS. Interestingly, the equilibrium position of TreS in isolation lies towards trehalose, however the flux *in vivo* is in the α-maltose direction.[52] α-Maltose is then phosphorylated by the maltokinase Pep2, to generate α-maltose-1-phosphate (M1P). TreS and Pep2 act together in a hetero-octameric complex.[53] M1P is then the substrate for GlgE, which is in fact a maltosyltransferase that uses M1P as a donor to extend α-1,4-glucan chains, releasing phosphate in the process.[51] Once the maltooligosaccharides reach a DP ≥ ~16, GlgB can introduce an α-1,6 branch. The enzymes GlgE and GlgB work in tandem to create the α-glucan polysaccharide, analogous to the GlgA and GlgB pair in *E. coli*.[32] Genetic knockouts of *glgE* or *glgB* are lethal in *M. tuberculosis in vitro* and in mouse models, due to the accumulation of the M1P substrate, which leads to a pleiotropic stress response, including inhibition of respiration, down-regulation of ATP synthase genes and the induction of DNA damage. Moreover, this effect is exacerbated by an upregulated feedback loop, whereby trehalose production is increased in an apparent attempt to compensate for the attenuated α-glucan synthesis,

leading to self-poisoning. GlgE and GlgB were therefore validated as *M. tuberculosis* drug targets and this is discussed in more detail below.[51]

The discovery of the GlgE pathway presented a number of questions. The first was the contrary function of GlgE compared to its original annotation as a glucanase in *M. smegmatis*. This can be attributed to the assay that the earlier study used to measure the accumulation of glycogen, as this simply measured the glucose released after amyloglucosidase treatment. It would therefore not discriminate between M1P and glycogen, resulting in the erroneous assignment of GlgE function. More recently, work by myself, K. Syson and colleagues confirmed this was the case, by demonstrating a build-up of M1P in a *M. smegmatis* strain with the temperature sensitive *glgE* mutation which was investigated in the previous study, **Figure 1.11**.
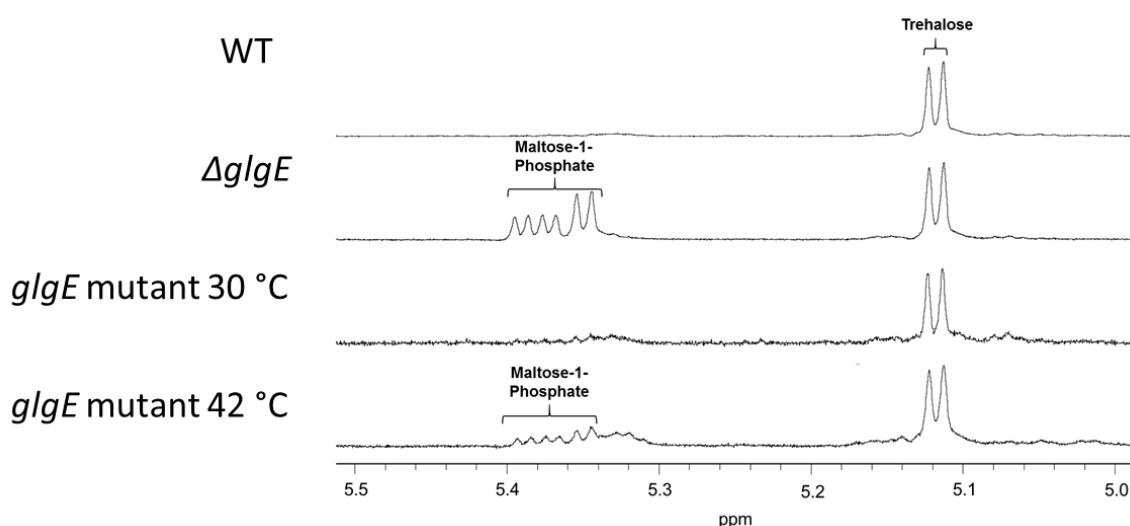
Figure 1.11 Accumulation of M1P in *M. smegmatis* with temperature sensitive *glgE* mutation. NMR spectra in the region of 5.0 – 5.5 ppm. Peaks corresponding to trehalose anomeric protons seen at 5.15 ppm for the WT (first panel). When GlgE function is abolished M1P accumulates with peaks corresponding to its anomeric protons being observed between 5.3 and 5.4 ppm (second panel). A strain with the temperature sensitive *glgE* mutation accumulates M1P when grown at a non-permissive temperature of 42 °C (fourth panel) but not at 30 °C (third panel). This confirms that the suggested glycogen accumulation in the original study was in fact a build-up of M1P. I generated these data and figure using strains provided by R. Kalscheuer.

The presence of the *glgA* and *glgC* genes in *M. tuberculosis* raised a second question: do multiple α-glucan synthesis pathways co-exist in this bacterium? This was answered during the investigation of the unexplained presence of M1P in a Δ*glgE* Δ*pep2* mutant strain. Characterisation of *M. tuberculosis* GlgA revealed that, contrary to its annotated function as an α-1,4 glucan synthase, it in fact produced M1P from the condensation of α-glucose-1-phosphate and ADP-glucose, and accordingly the enzyme was renamed as GlgM. The

function of GlgC is as predicted and, as in *E. coli*, this enzyme generates ADP-glucose from glucose-1-phosphate. However, this is instead used to make M1P and the formation of M1P by GlgA was confirmed to be dependent on GlgC. This represents a significant reconfiguration of α-glucan synthesis in *M. tuberculosis* relative to *E.coli*, **Figure 1.12**.

A further potential complication was the proposed interplay between the MGLP and α-glucan synthetic pathways. In the synthesis of MGLP, di-glucosylglycerate is thought to be extended by Rv3032, using a UDP-glucose donor.[54] Rv3032 was initially implicated in α-glucan synthesis because its deletion appeared to reduce intracellular α-glucan content and the combined Δ*glgA* Δ*Rv3032* was said to be not viable.[40] However, both of these findings were recently refuted, suggesting that Rv3032 has no role in α-glucan synthesis and that MGLP synthesis is an independent process.[38]
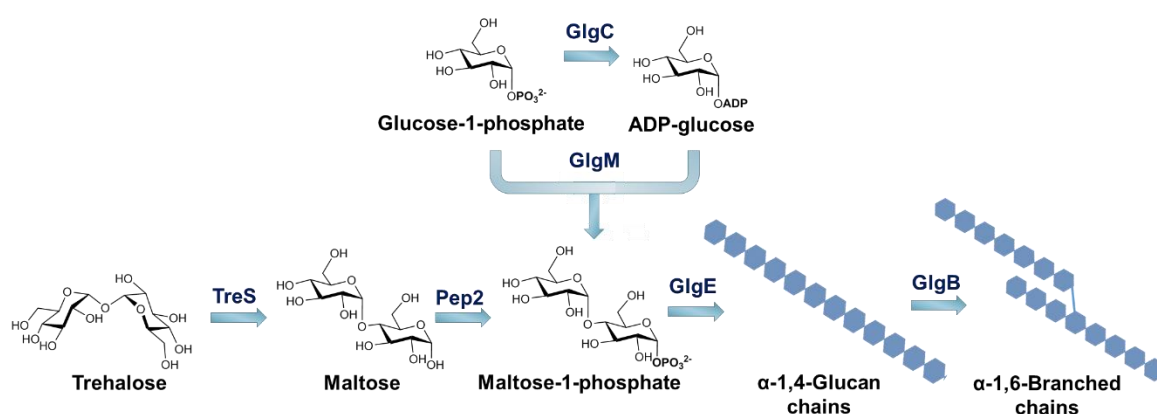
Figure 1.12 Synthesis of α-glucan in *M. tuberculosis*. M1P can be generated by either TreS/Pep2 or GlgM, converging to one unified route to α-glucan in mycobacteria. Pyranose units are represented as blue hexagons.

The complete GlgE pathway has been identified in 14% of sequenced bacterial genomes and is especially common in *Streptomyces*, *Mycobacterium* and *Corynebacterium* species.[42] Some organisms contain only this set of α-glucan synthesis genes, for example, *Streptomyces venezuelae*, and the GlgE pathway has been confirmed as the sole α-glucan synthetic pathway in this species.[43]

1.3.3 Mycobacterial α-glucan structure

Since α-glucan is made by a different biosynthetic pathway in actinomycetes, it follows that it might have altered structural properties. Recently, an in-depth analysis of *M. tuberculosis* α-glucan structure was carried out and it was compared to α-glucan originating from the classical glycogen synthesis pathways, such as in *E. coli*.[32] Interestingly, *M. tuberculosis* had an average chain length of DP 7.1 ± 2.0, markedly shorter than classical bacterial and

mammalian glycogen, which had average chain lengths of 11.0 ± 2.9 and 10.8 ± 4.6, respectively. The internal chain length of *M. tuberculosis* α-glucan was 4.7 ± 1.6, with an A:BC chain ratio of 0.6, compared to 8.3 ± 3.3 and 0.8 for *E. coli* glycogen. *M. tuberculosis* α-glucans extracted from the cytosol or capsule were essentially identical, consistent with their singular biosynthetic origin. Furthermore, α-glucan extracted from *S. venezuelae* had very similar properties to that from mycobacteria, consistent with the hypothesis that α-glucan made via the GlgE pathway has notably shorter average branch lengths, internal chain lengths and a lower A:BC chain ratio, compared to classical glycogen.[32]

Comparison of the known types of α-glucan polysaccharide demonstrates that mycobacterial α-glucan has the shortest branches, with the corresponding highest percentage of branching, **Table 1.1**.

Table 1.1 Features of different types of α-glucan. Data from Tester *et al.*[28] and Rashid *et al.*[32] Decreasing chain length correlates with increased aqueous solubility and an increased percentage branching.

α-Glucan type	Aqueous solubility	Mean chain length (DP)	% Branching	Biological role
Amylose	X	50 – 1,000	<1	Energy storage
Amylopectin	X	20 – 25	~4	Energy storage
Classical glycogen	✓	10 – 12	~9	Energy storage
M. tuberculosis α-glucan	✓	7 – 8	~14	Energy storage Virulence

However, this work does not provide direct evidence for the formation of mycobacterial α-glucan by GlgE and GlgB alone, and without this, the distinct structure of mycobacterial α-glucan cannot be directly linked to these enzymes.

1.3.4 Determinants of mycobacterial α-glucan branch length specificity

The shorter average chain length of α-glucan derived from the GlgE pathway implies that the actinomycete GlgB enzymes generate shorter branches than those from the classical glycogen synthesis pathways, such as *E. coli* GlgB. It would therefore be interesting to understand the underlying basis of these different specificities. However, whilst a wide range of substrate preferences, branch length specificities and branchpoint positions have

been reported for different α-1,6 branching enzymes,[55] the factors controlling these have not yet been identified.

Most branching enzymes belong to the GH13 CAZy family, and are further classified into the GH13_8 or GH13_9 subfamilies, which usually correlates with a eukaryotic or prokaryotic origin, respectively, however there are some exceptions.[55] More recently, branching enzymes have been discovered that are classified into the GH57 CAZy family.[56-57] All GH13 branching enzymes share the GH13 family (β/α)$_8$ barrel, which contains the catalytic aspartate and glutamate, along with a further aspartate and arginine that typically coordinate the pyranose in the -1 subsite.[58] The catalytic domain is flanked by the CBM48 and C-terminal β-sandwich domains, along with an N-terminal β-sandwich domain in many species.[55] This domain has been linked to branch length specificity, with some N-terminal truncations increasing branch length in *E.coli* GlgB, although overall there was not a clear relationship between truncation size and glucan properties.[59] The lack of a clear correlation between amino acid sequence and branch length specificity implies that the structural features responsible are not directly distinguishable from protein sequence. It is therefore necessary to consider the three-dimensional structures of branching enzymes, some of which have been determined with X-ray crystallography.
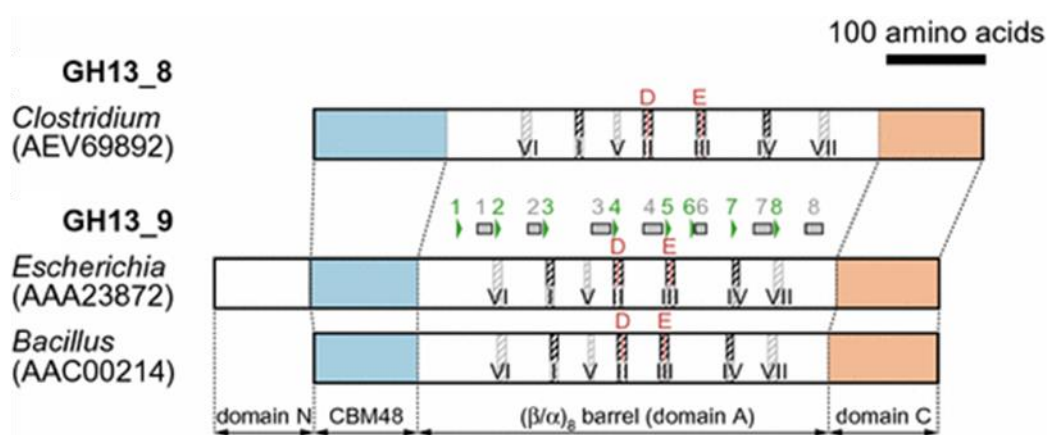
Figure 1.13 Domains and sequence motifs of GH13 family branching enzymes. GH13 branching enzymes consist of domain N, which is absent in some enzymes, CBM48, the catalytic domain A, which contains the (β/α)$_8$ barrel, and domain C. Conserved sequence regions are numbered I-VII, with the catalytic Asp and Glu residues located in II and III, respectively. Adapted from Suzuki and Suzuki, 2010,[55] with permission.

The crystal structure of full length *M. tuberculosis* GlgB (MtGlgB) has been solved,[60] along with that of *E. coli*,[61] human[62] and rice,[63] all lacking domain N. Recently, the full length structure of a cyanobacterial branching enzyme was also solved by Hayashi *et al.*, which had domain N located at the back of the protein,[64] in contrast to the linear arrangement of domains along the longitudinal axis in the *M. tuberculosis* enzyme. Structures of the *E. coli*,

human and rice enzymes in complex with maltooligosaccharides or cyclodextrins have been solved, but these ligands are all located at a significant distance from the active site, [62-63, 65-66] therefore their influence on the branch length specificity is unclear.The recently characterised cyanobacterial branching enzyme is the only branching enzyme structure with a maltooligosaccharide bound at the active site cleft, identifying the donor subsites.[64] Unusually, this GH13_9 enzyme is involved in the synthesis of starch, and generates branch lengths of DP 6-7.[67] Hayashi *et al.* proposed an acceptor site based on other maltooligosaccharide binding sites in the C-terminal domain, both at a distance of ~30 Å from the catalytic nucleophile. Site-directed mutagenesis of some of the residues at these distal sites resulted in decreased activity of the enzyme or slight shifts in the distribution of branch lengths in the products. However, it was not determined whether these directly result from the disruption of maltooligosaccharide binding at these sites or whether mutagenesis results in conformational changes in the protein. The suggested model does not fully explain the reported branch length specificity of DP 6-7, since the B chains created by this extended acceptor site are DP10, **Figure 1.14**.
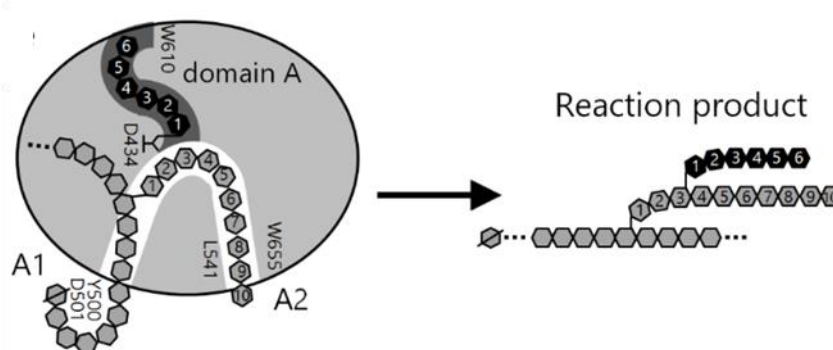
Figure 1.14 Proposed substrate binding model in cyanobacterial branching enzyme. The donor site is coloured dark grey and bound pyranose units are represented by black hexagons and numbered according to the donor (-) subsites. The proposed acceptor site is coloured white and is based on binding at the A1 and A2 sites in domain C, with interacting protein residues labelled. The predicted acceptor maltooligosaccharide is represented by grey hexagons, with the reducing end marked by a slash. The expected reaction product is shown on the right, which has an A chain of DP6 and B chain of DP10. Adapted from Hayashi *et al.* with permission.[64]

For the *M. tuberculosis* enzyme, only the apo crystal structure exists[60] and therefore the donor and acceptor subsites have not yet been identified. It is therefore not yet possible to compare the structural features involved in substrate binding that might underlie the branch length specificity in mycobacteria.

1.4   Trehalose metabolism in *M. tuberculosis*

Trehalose is a non-reducing disaccharide consisting of two α-D-glucose units linked by a 1,1-glycosidic bond, **Figure 1.15**. It is biosynthesised by bacteria, fungi, plants and invertebrates, but not mammals, and is known to play an important role in energy storage in insects, as well as protection from desiccation, cold and oxidation stress in bacteria.[68] Trehalose metabolism in mycobacteria has previously attracted considerable interest due to the incorporation of trehalose into cell wall glycolipids, such as trehalose-6,6'-dimycolate (TDM), which plays an important role in the virulence and persistence of *M. tuberculosis* within the host, as well as the structural integrity of the *M. tuberculosis* cell wall.[69-72]
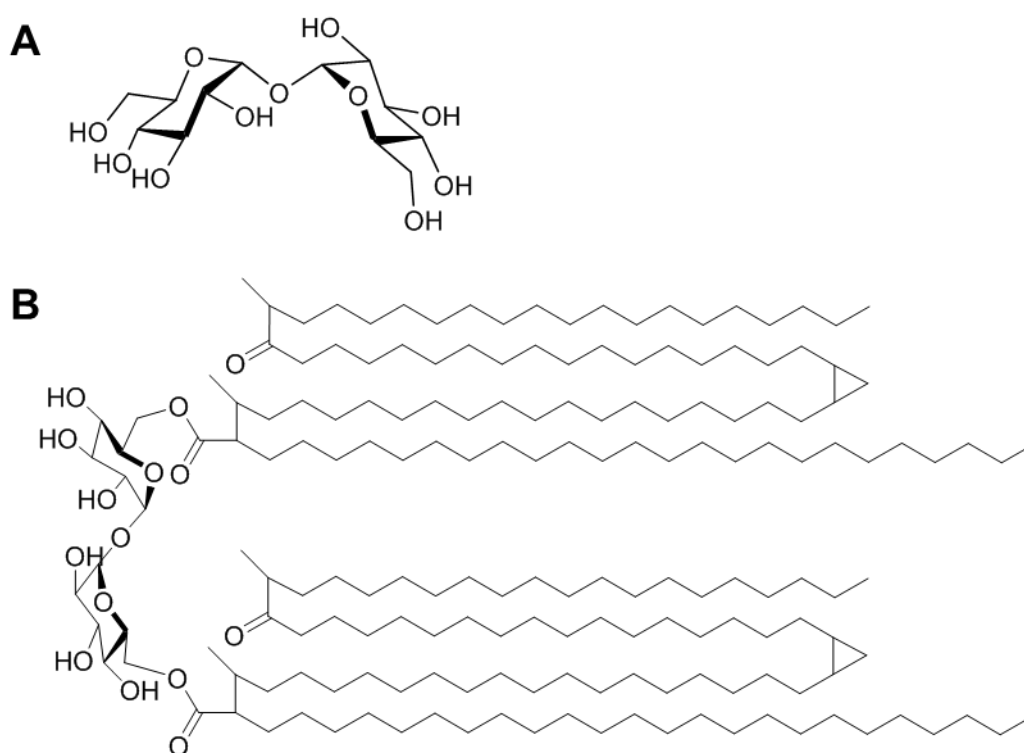
Figure 1.15 Trehalose Structures. A) Trehalose consists of two α-D-glucopyranose units linked by an α-1,1 glycosidic bond B) TDM consists of two mycolate groups attached at the 6 and 6' positions of trehalose.

Classical trehalose biosynthesis occurs via the OtsAB pathway whereby a nucleotide diphosphate-activated glucose unit is transferred onto glucose-6-phosphate by trehalose phosphate synthase, OtsA. The resultant trehalose-6-phosphate is then converted into free trehalose by trehalose phosphate phosphatase, OtsB. OtsA enzymes identified in yeast, insects and *Streptomyces* are specific for either UDP-glucose or GDP-glucose, whereas those in mycobacteria have much broader substrate specificities and can utilise all of the naturally occurring glucose nucleotides.[68] This pathway is the predominant trehalose synthesis route in *M. tuberculosis*.[73]

Trehalose can also be generated by the degradation of cytosolic α-glucan via the TreXYZ pathway, which was first identified in *Sulfolobus acidocaldarius*.[74] In this route, a glycosyl residue at the reducing end of an α-1,4-glucan is inverted by TreY to form a trehalose moiety at the end of the chain which is cleaved by TreZ to release free trehalose.[75] This process would have to be aided by the glycogen debranching enzyme GlgX (also called TreX) to get past branchpoints.[76] The discovery that trehalose is also the starting point for the GlgE pathway underlines the importance of trehalose as a metabolic hub in mycobacteria.

1.5 Tuberculosis

Tuberculosis (TB) is an infectious disease that has plagued humanity for millennia.[77] It is still the second most common cause of death worldwide due to a single infectious agent after HIV/AIDS. It is estimated that one third of the world's population is infected by *M. tuberculosis* and that 5 – 10% of these cases will go on to develop the active form of TB, with 1.8 million dying each year as a result. It is an airborne infection, which is typically spread through the transmission of respiratory fluids via coughs and sneezes. Upon inhalation, *M. tuberculosis* infects alveolar macrophages. In an immunocompetent host these are physically contained in granulomas consisting of tiny collections of modified macrophages that surround the TB bacilli. These structures present one of the major barriers for complete TB elimination by antibiotics, leading to progressive chronic disease.[78] Immunocompromised patients by contrast are also particularly vulnerable to acute TB infection, and TB is therefore the leading cause of death for HIV patients. Poverty is the other key risk factor associated with developing the disease.[79]

1.5.1 The tubercle bacillus

M. tuberculosis was first identified as the aetiological agent of TB in 1882 by Robert Koch, for which he received the 1905 Nobel Prize in Physiology or Medicine.[80] Mycobacteria are a type of actinobacteria, a group of Gram-positive bacteria, which contain a high level of guanine and cytosine in their DNA. As well as *M. tuberculosis,* the genus also includes *Mycobacterium leprae*, which is the causative agent of leprosy. The *M. tuberculosis* complex comprises at least five distinct members: *M. tuberculosis*, *Mycobacterium bovis*, *Mycobacterium africanum*, *Mycobacterium canettii* and *Mycobacterium microti*, all of which are pathogenic, slow growing species. The rapidly growing, non-pathogenic species, *Mycobacterium smegmatis,* is commonly used as a model organism for the study of *M.*

tuberculosis.[81] The complete genome sequence of the H37Rv strain of *M. tuberculosis* was published in 1998,[82] and represented a significant milestone in over a century of TB study.

Although mycobacteria are classified among the Gram-positive bacteria, they possess a cell envelope structurally related to, but arguably more complex than that of, Gram-negative bacteria.[83] The mycobacterial envelope consists of an inner plasma membrane and a thick cell wall comprised of peptidoglycan (PG) covalently linked to arabinogalactan (AG), which is esterified by mycolic acids. The outermost compartment consists of the loosely bound capsule structure and this surface forms the interface with the host.[84] This cell envelope structure confers impermeability to small molecules and presents a further obstacle to TB therapeutics,[85] **Figure 1.16**.
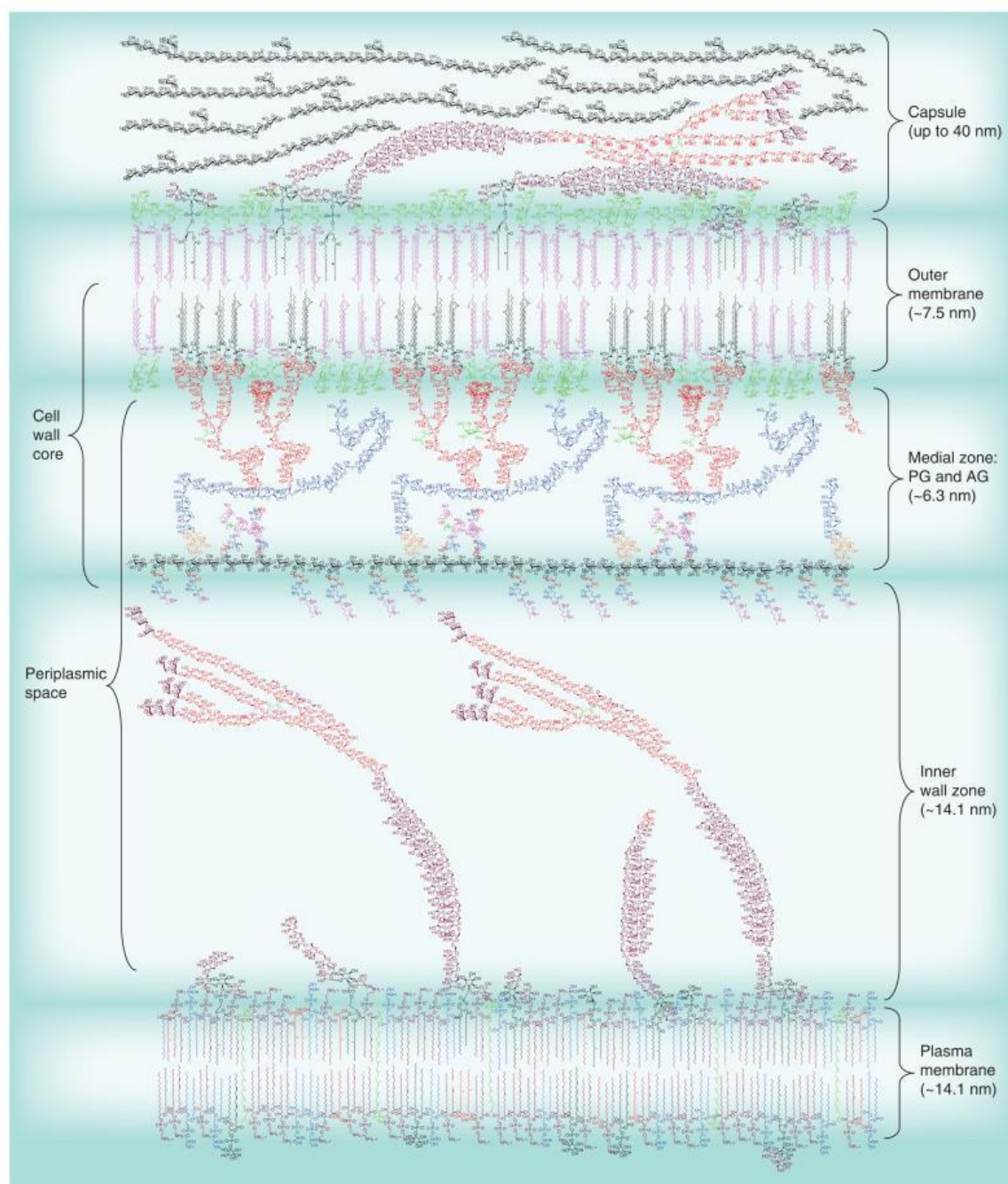
Figure 1.16 The *M. tuberculosis* cell envelope. The innermost layer is the phospholipid plasma membrane, separated from the outer membrane (phospholipids not shown) by the periplasmic space. This space is divided into the inner wall and medial zones, which are separated by a PG layer. The final, outermost layer attached to the outer membrane is the capsule. The cell wall core consists of the PG layer attached to the mycolate layer of the outer membrane via AG. Used with permission from Jackson *et al.* 2013.[84]

Key: Medial Zone: PG, black with coloured peptides; galactofuranosyl residues, blue; arabinofuranosyl residues in AG, red. Outer Membrane: mycolic acids attached to arabinan, black; mycolic acids attached to trehalose, pink. Capsule: α-glucan (not structurally representative), black; mannan, purple; arabinomannan, purple and red.

1.5.2 TB treatments

The attenuated *M. bovis* bacillus Calmette-Guérin (BCG) strain has been used as an anti-TB vaccine for the last century, however its efficacy is highly variable.[86] Since the 1970s, TB has been routinely treated with the first line drugs isoniazid (INH) and rifampicin (RIF) which target fatty acid metabolism and RNA synthesis respectively. Second line treatments include the injectable antibiotics kanamycin, capreomycin and amikacin, which target protein synthesis, as well as the fluoroquinolones, which target DNA gyrase.[78] Lengthy and expensive treatment regimens have been streamlined into the 6-month directly observed treatment short-course therapy, which has been successful in improving patient compliance and completion of the full course of drugs.[87]

Multi-drug resistant TB (MDR-TB) is defined as strains that are resistant to INH and RIF and accounted for 3.6% of the 8.6 million new cases of TB in 2012, and 20% of previously treated cases. Worst affected countries include South Africa and India, along with most of Eastern Europe. Extensively-drug resistant TB (XDR-TB) is defined as strains resistant to at least one of the three injectable antibiotics and a member of the quinolone family, in addition to INH and RIF. There was at least one case of XDR-TB in all the countries surveyed by WHO in 2013 and it makes up about 10% of all MDR-TB cases.[79] Despite receiving considerable research investment and attention, only one new anti-TB drug has been approved in the last 40 years: bedaquiline, which has been shown to be active against MDR-TB.[88] There is clearly an urgent need to develop novel anti-TB treatments with new targets.

1.6   Targeting α-glucan metabolism in *M. tuberculosis*

The α-glucan metabolic pathways in *M. tuberculosis* present a number of potential routes for the development of novel therapeutic agents. GlgE and GlgB are attractive drug targets, with a dual mechanism of action.[89] Inhibition leads to the lethal accumulation of M1P that is escalated in an upregulated feedback loop,[51] as well as the loss of the capsular α-glucan virulence factor.[38] As there are two biosynthetic routes to M1P, mutations in both branches of the pathway would be required to prevent M1P build-up, and even if both routes are blocked, the reduced virulence resulting from the loss of capsular α-glucan remains. Trehalose metabolism in *M. tuberculosis* has already been targeted for inhibition or the incorporation of diagnostic probes, and its link with α-glucan metabolism may provide an indirect route to target the enzymes of the GlgE pathway.

1.6.1 Targeting GlgE
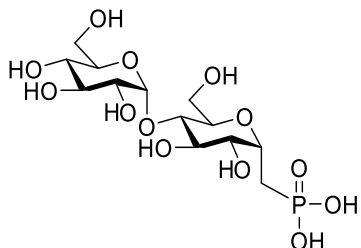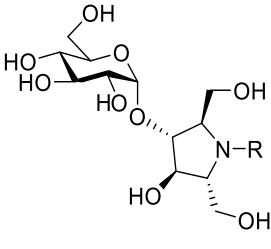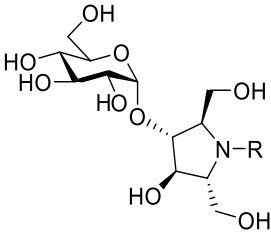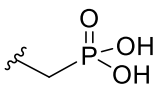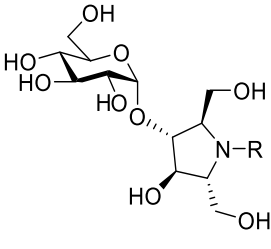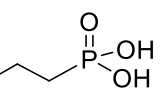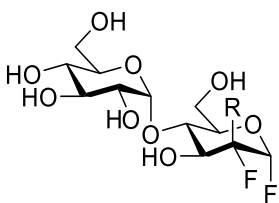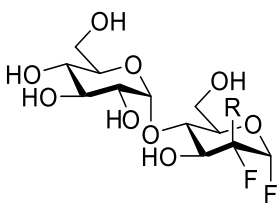
The biochemistry, structure and mechanism of GlgE have been extensively characterised by K. Syson *et al.*. As high-resolution structures of *M. tuberculosis* GlgE have been difficult to obtain, *S. coelicolor* GlgE isoform 1 has been used as a structural model.[90] These two enzymes have highly similar active site architectures and catalytic properties, with *S. coelicolor* GlgEI having $k_{\mathrm{cat}}^{\mathrm{app}}$ values an order of magnitude greater than *M. tuberculosis* GlgE, and the latter enzyme having a slightly lower optimum acceptor chain length of DP5 vs. DP6.[91] GlgE is the defining member of the GH13_3 subfamily and exists as a homodimer in solution,[91-92] with a disulphide bridge between the two monomers in the *M. tuberculosis* structure.[90] The predicted α-retaining double displacement mechanism is supported by a covalent catalytic intermediate, which could be trapped with a 2-deoxy-2-fluoro-α-maltosyl fluoride substrate analogue, aided by the removal of the catalytic acid/base residue.[92] The carbohydrate binding sites -2 to +6 have been defined, along with secondary binding sites that can accommodate branched α-glucans.[93] The structure of *Mycobacterium thermoresistible* GlgE revealed that the enzyme can adopt different conformational states, which are predicted to be important for the enzyme's activity.[94] GlgE conforms to a ping-pong enzyme mechanism, whereby the donor and acceptor binding is sequential and not simultaneous.[91] The enzyme is negatively regulated by PknB via Ser/Thr phosphorylation.[95]

However, despite the detailed knowledge of the enzyme's structure and activity, only a few inhibitors of GlgE have been reported to date. These are substrate and transition state analogues, which have been shown to bind to GlgE *in vitro*. Maltose *C*-phosphonate is a non-hydrolysable and isosteric analogue of M1P, which bound to GlgE and had an IC$_{50}$ value of 230 ± 24 µM. The transition state analogue 2,5-dideoxy-3-O-α-D-glucopyranosyl-2,5-imino-D-mannitol had a K_i value of 237 ± 27 µM for *M. tuberculosis* GlgE, and had an approximately two-fold higher affinity for the *S. coelicolor* enzyme. The addition of a phosphonate moiety improved the K_i to 45 ± 4 µM and, whilst this is relatively weak, it is the most effective GlgE inhibitor reported to date. Maltosyl fluorides also bind to GlgE, but are slow substrates rather than inhibitors,[92, 96] **Table 1.2**. Other putative GlgE inhibitors have been predicted by virtual docking, however these have not been verified experimentally.[97] Screening of small-molecule compound libraries was carried out by K. Syson in collaboration with MRCT, but also yielded no hit compounds (K. Syson, personal communication with permission).

Whilst substrate analogues have been the most promising avenue for the development of GlgE inhibitors *in vitro*, carbohydrate uptake is highly limited in *M. tuberculosis* and it lacks the capacity to import maltose-based compounds. Therefore, the compounds shown in

Table 1.2 will be ineffective *in vivo* as they cannot reach the cytosolic GlgE target. The development of a *M. tuberculosis* carbohydrate-based inhibitor is limited to the known specificities of carbohydrate importers, namely for glycerophosphocholine, amino sugars and trehalose.[98-100]

Table 1.2 Analogues that bind to and/or inhibit GlgE. [a] = IC_{50} value, [b] = K_i value, [*] = WT *S. coelicolor* GlgEI; nd = not determined.

Compound		*S. coelicolor* GlgEI V279S		*M. tuberculosis* GlgE
		K_i (µM)	Crystal Structure (Å)	IC_{50}/ K_i (µM)
Maltose *C*-phosphonate		nd	1.9 [90]	230 ± 24[a] [101]
2,5-dideoxy-3-O-α-D-glucopyranosyl-2,5-imino-D-mannitol compounds	R = H	102 ± 8 [102]	2.5 [90]	237 ± 27[b] [102]
	R =	45 ± 4 [103]	3.2 [103]	nd
	R =	95 ± 16 [103]	2.5 [103]	nd
Maltosyl fluoride compounds	R = H	Slow substrate[*] [92]	2.4[*] [92]	nd
	R = F	No inhibition [96]	2.3 [96]	nd

1.6.2 Targeting trehalose metabolism

Trehalose metabolism has attracted attention as a target for anti-TB agents due to its absence in mammals and the integral role trehalose plays in cell wall glycolipids. The specificity of trehalose for mycobacteria has also been utilised to target therapeutics to the bacteria *in vivo*, using functionalised nanoparticles.[104] Unlike almost all other types of carbohydrates, trehalose is readily taken up into the bacterium via the LpqY-SugABC transporter. This is highly specific for the α-1,1 linkage, but can accommodate a wide range of modified trehaloses.[98] This has been exploited to incorporate fluorescent or radioactive labels for diagnostic imaging.[105-108] Inhibiting the synthesis of trehalose is more challenging due to the multiple routes for trehalose biosynthesis. Efforts have instead focused on the inhibition of trehalose mycolate processing enzymes, such as Pks13, MmpL3 and Ag85, with a number of highly potent inhibitors identified.[89] One approach that has not yet been considered is the indirect inhibition of GlgE via trehalose analogues, which could be taken into the cell and converted to GlgE inhibitors by the enzymes of the GlgE pathway.

1.6.3 Targeting GlgB

MtGlgB has been biochemically characterised and is a monomeric protein with an optimal activity at 30 °C, which is surprisingly low for a human pathogen.[45] Due to exposed cysteine residues, different redox forms of the protein were proposed, however the redox environment does not affect enzymatic activity.[109] One challenge for the development of GlgB inhibitors is the inherent difficulty in determining kinetic parameters, since branching reactions are difficult to monitor directly. Indirect kinetic parameters were previously determined by monitoring the amylase side reaction of the enzyme,[60] which may not directly relate to the branching activity. Furthermore, this study used long-chain, linear amyloses that are not representative of GlgB substrates *in vivo* and therefore may not reflect the activity of the enzyme *in vivo*.

MtGlgB has received less attention as a drug target due to the existence of a human homologue.[62] However, these enzymes only share 27% sequence identity, therefore the development of a selective inhibitor is feasible. Other known inhibitors of GH13 enzymes such as ADP, ADP-glucose, tunicamycin, casternospermine, nojirimycin and acarbose were ineffective against GlgB.[60] So far, high-throughput screening to identify inhibitors of GlgB has not been carried out. Some potential inhibitors of GlgB were claimed to have been identified through a computational docking approach and two of these also showed inhibition in whole-cell assays.[110] However, the hit compounds were not validated using

counter-screens, whereby the effect of the compound on the amylose-iodine interaction is assessed to rule out false positives. The binding of compounds to the enzyme was not confirmed with any other method, therefore it cannot be ascertained whether the reported *in vivo* activity is specific for MtGlgB.

As discussed above, the apo *M. tuberculosis* MtGlgB crystal structure exists, but it is low quality, which may preclude the determination of detailed SAR for future GlgB inhibitors. An improved structural model of mycobacterial GlgB would aid the rational design of inhibitors with increased potency or specificity. Additionally, substrate-bound structures would identify specific sites in the protein that can be targeted, such as the substrate donor or acceptor sites.

1.7   Project aims

The main goals of this work were: to firstly improve the understanding of mycobacterial α-glucan synthesis through structural studies, secondly to identify a small-molecule inhibitor of GlgB and thirdly to investigate targeting GlgE using trehalose analogues. These were achieved by addressing the following questions:

Mycobacterial α-glucan synthesis (Chapter 2)

- Are GlgE and GlgB sufficient to generate α-glucan *in vitro*?
- Can the structural model of mycobacterial GlgB be improved?
- Can mycobacterial GlgB substrate binding sites be identified?

Targeting GlgB (Chapters 3 and 4)

- Can an inhibitor of GlgB be identified by high-throughput screening?
- Can SAR be established with a selection of modified compounds?
- How do GlgB inhibitors behave *in vivo*?

Targeting GlgE (Chapter 5)

- Can a synthetic route to 4-deoxy trehalose analogues be established?
- Are 4-deoxy trehalose compounds taken up in mycobacterial cells?
- Can 4-deoxy trehalose analogues be processed in the GlgE pathway?

Finally, the broader conclusions and implications of these results are discussed, along with future work.

1.8   References

1. Varki, A., *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press: New York, 2017.

2. Varki, A., Biological roles of oligosaccharides: All of the theories are correct. *Glycobiology* **1993,** *3* (2), 97-130.

3. Davis, B. G.; Fairbanks, A. J., *Carbohydrate chemistry*. Oxford University Press: Oxford, 2002.

4. Lindhorst, T. K., *Essentials of carbohydrate chemistry and biochemistry*. Wiley-VCH: Weinheim, 2007.

5. Stoddart, J. F., *Stereochemistry of carbohydrates*. Wiley-Interscience: Hoboken, 1971.

6. Klemm, D.; Heublein, B.; Fink, H.-P.; Bohn, A., Cellulose: Fascinating biopolymer and sustainable raw material. *Angew. Chem. Int. Ed.* **2005,** *44* (22), 3358-3393.

7. Deslandes, Y.; Marchessault, R. H.; Sarko, A., Triple-helical structure of (1- 3)-β-D-glucan. *Macromolecules* **1980,** *13* (6), 1466-1471.

8. Solís, D.; Bovin, N. V.; Davis, A. P.; Jiménez-Barbero, J.; Romero, A.; Roy, R.; Smetana, K.; Gabius, H.-J., A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code. *Biochim. Biophys. Acta* **2015,** *1850* (1), 186-235.

9. Seeberger, P. H., Automated oligosaccharide synthesis. *Chem. Soc. Rev.* **2008,** *37* (1), 19-28.

10. Damerell, D.; Ceroni, A.; Maass, K.; Ranzinger, R.; Dell, A.; Haslam, S. M., Annotation of glycomics MS and MS/MS spectra using the GlycoWorkbench software tool. In *Glycoinformatics*, Lütteke, T.; Frank, M., Eds. Springer New York: New York, NY, 2015; pp 3-15.

11. Aoki-Kinoshita, K.; Agravat, S.; Aoki, N. P.; Arpinar, S.; Cummings, R. D.; Fujita, A.; Fujita, N.; Hart, G. M.; Haslam, S. M.; Kawasaki, T.; Matsubara, M.; Moreman, K. W.; Okuda, S.; Pierce, M.; Ranzinger, R.; Shikanai, T.; Shinmachi, D.; Solovieva, E.; Suzuki, Y.; Tsuchiya, S.; Yamada, I.; York, W. S.; Zaia, J.; Narimatsu, H., GlyTouCan 1.0 – The international glycan structure repository. *Nucleic Acids Res.* **2016,** *44* (D1), D1237-D1242.

12. Science X. Development of *Euglena*-based bioplastics. https://phys.org/news/2013-02-euglena-based-bioplastics.html (accessed Sep 1, 2017).

13. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014,** *42* (Database issue), D490-5.

14. Henrissat, B., A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **1991,** *280* (2), 309.

15. Campbell, J. A.; Davies, G. J.; Bulone, V.; Henrissat, B., A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **1997,** *326 (Pt 3),* 929-39.

16. Lombard, V.; Bernard, T.; Rancurel, C.; Brumer, H.; Coutinho, P. M.; Henrissat, B., A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J.* **2010,** *432* (3), 437-44.

17. Levasseur, A.; Drula, E.; Lombard, V.; Coutinho, P. M.; Henrissat, B., Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for biofuels* **2013,** *6* (1), 41.

18. Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J., Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **2004,** *382* (Pt 3), 769-81.

19. Koshland, D. E., Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews* **1953,** *28* (4), 416-436.

20. Davies, G.; Henrissat, B., Structures and mechanisms of glycosyl hydrolases. *Structure* **1995,** *3* (9), 853-859.

21. Davies, G. j.; Wilson, K. s.; Henrissat, B., Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **1997,** *321* (2), 557-559.

22. Preiss, J.; Romeo, T., Physiology, biochemistry and genetics of bacterial glycogen-synthesis. *Adv. Microb. Physiol.* **1989,** *30*, 183-238.

23. Dumitriu, S., *Polysaccharides: Structural diversity and functional versatility, Second edition.* CRC Press: Boca Raton, 2004.

24. Manners, D. J., Recent developments in our understanding of glycogen structure. *Carbohydr. Polym.* **1991,** *16* (1), 37-82.

25. Zeeman, S. C.; Kossmann, J.; Smith, A. M., Starch: Its metabolism, evolution, and biotechnological modification in plants. *Annu. Rev. Plant Biol.* **2010,** *61* (1), 209-234.

26. Ball, S. G.; Morell, M. K., From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu. Rev. Plant Biol.* **2003,** *54*, 207-33.

27. Gessler, K.; Usón, I.; Takaha, T.; Krauss, N.; Smith, S. M.; Okada, S.; Sheldrick, G. M.; Saenger, W., V-Amylose at atomic resolution: X-ray structure of a cycloamylose with 26 glucose residues (cyclomaltohexaicosaose). *Proc. Natl. Acad. Sci. USA* **1999,** *96* (8), 4246-4251.

28. Tester, R. F.; Karkalas, J.; Qi, X., Starch - composition, fine structure and architecture. *J. Cereal Sci.* **2004,** *39* (2), 151-165.

29. Preiss, J., *Bacterial glycogen inclusions: Enzymology and regulation of synthesis.* Springer-Verlag: Berlin, 2006; Vol. 1, p 71-108.

30. Ryu, J. H.; Drain, J.; Kim, J. H.; McGee, S.; Gray-Weale, A.; Waddington, L.; Parker, G. J.; Hargreaves, M.; Yoo, S. H.; Stapleton, D., Comparative structural analyses of purified glycogen particles from rat liver, human skeletal muscle and commercial preparations. *Int. J. Biol. Macromol.* **2009,** *45* (5), 478-482.

31. Testoni, G.; Duran, J.; García-Rocha, M.; Vilaplana, F.; Serrano, A. L.; Sebastián, D.; López-Soldado, I.; Sullivan, M. A.; Slebe, F.; Vilaseca, M.; Muñoz-Cánoves, P.; Guinovart, J. J., Lack of glycogenin causes glycogen accumulation and muscle function impairment. *Cell Metabolism* **2017,** *26* (1), 256-266.e4.

32. 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., Assembly of α-glucan by GlgE and GlgB in mycobacteria and streptomycetes. *Biochemistry* **2016,** *55* (23), 3270-3284.

33. Alonso-Casajus, N.; Dauvillee, D.; Viale, A. M.; Munoz, F. J.; Baroja-Fernandez, E.; Moran-Zorzano, M. T.; Eydallin, G.; Ball, S.; Pozueta-Romero, J., Glycogen phosphorylase, the product of the glgP gene, catalyzes glycogen breakdown by removing glucose units from the nonreducing ends in *Escherichia coli. J. Bacteriol.* **2006,** *188* (14), 5266-5272.

34. Dauvillee, D.; Kinderf, I. S.; Li, Z. Y.; Kosar-Hashemi, B.; Samuel, M. S.; Rampling, L.; Ball, S.; Morell, M. K., Role of the *Escherichia coli* glgX gene in glycogen metabolism. *J. Bacteriol.* **2005,** *187* (4), 1465-1473.

35. Ballicora, M. A.; Iglesias, A. A.; Preiss, J., ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. *Microbiol. Mol. Biol. Rev.* **2003,** *67* (2), 213-225.

36. Antoine, A. D.; Tepper, B. S., Environmental control of glycogen and lipid content of *Mycobacterium tuberculosis. J. Bacteriol.* **1969,** *100* (1), 538-539.

37. Lemassu, A.; Daffé, M., Structural features of the exocellular polysaccharides of *Mycobacterium tuberculosis. Biochem. J.* **1994,** *297* (Pt 2), 351-357.

38. Koliwer-Brandl, H.; Syson, K.; van de Weerd, R.; Chandra, G.; Appelmelk, B.; Alber, M.; Ioerger, T. R.; Jacobs, W. R., Jr.; Geurtsen, J.; Bornemann, S.; Kalscheuer, R., Metabolic network for the biosynthesis of intra- and extracellular α-glucans required for virulence of *Mycobacterium tuberculosis. PLoS Path.* **2016,** *12* (8), e1005768.

39. Geurtsen, J.; Chedammi, S.; Mesters, J.; Cot, M.; Driessen, N. N.; Sambou, T.; Kakutani, R.; Ummels, R.; Maaskant, J.; Takata, H.; Baba, O.; Terashima, T.; Bovin, N.; Vandenbroucke-Grauls, C.; Nigou, J.; Puzo, G.; Lemassu, A.; Daffe, M.; Appelmelk, B. J., Identification of mycobacterial α-glucan as a novel ligand for DC-SIGN: Involvement of mycobacterial capsular polysaccharides in host immune modulation. *J. Immunol.* **2009,** *183* (8), 5221-5231.

40. Sambou, T.; Dinadayala, P.; Stadthagen, G.; Barilone, N.; Bordat, Y.; Constant, P.; Levillain, F.; Neyrolles, O.; Gicquel, B.; Lemassu, A.; Daffe, M.; Jackson, M., Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: Biosynthesis and impact on the persistence in mice. *Mol. Microbiol.* **2008,** *70* (3), 762-774.

41. Gagliardi, M. C.; Lemassu, A.; Teloni, R.; Mariotti, S.; Sargentini, V.; Pardini, M.; Daffe, M.; Nisini, R., Cell wall-associated alpha-glucan is instrumental for *Mycobacterium tuberculosis* to block CD1 molecule expression and disable the function of dendritic cell derived from infected monocyte. *Cell. Microbiol.* **2007,** *9* (8), 2081-2092.

42. Jackson, M.; Brennan, P. J., Polymethylated polysaccharides from *Mycobacterium* species revisited. *J. Biol. Chem.* **2009,** *284* (4), 1949-1953.

43. Nunes-Costa, D.; Maranha, A.; Costa, M.; Alarico, S.; Empadinhas, N., Glucosylglycerate metabolism, bioversatility and mycobacterial survival. *Glycobiology* **2017,** *27* (3), 213-227.

44. Kaur, D.; Pham, H.; Larrouy-Maumus, G.; Riviere, M.; Vissa, V.; Guerin, M. E.; Puzo, G.; Brennan, P. J.; Jackson, M., Initiation of methylglucose lipopolysaccharide biosynthesis in mycobacteria. *PLoS One* **2009,** *4* (5).

45. Garg, S. K.; Alam, M. S.; Kishan, K. V. R.; Agrawal, P., Expression and characterization of α-(1,4)-glucan branching enzyme Rv1326c of *Mycobacterium tuberculosis* H37Rv. *Protein Expression Purif.* **2007,** *51* (2), 198-208.

46. Belanger, A. E.; Hatfull, G. F., Exponential-phase glycogen recycling is essential for growth of *Mycobacterium smegmatis. J. Bacteriol.* **1999,** *181* (21), 6670-6678.

47. Schneider, D.; Bruton, C. J.; Chater, K. F., Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor* A3(2). *Mol. Gen. Genet.* **2000,** *263* (3), 543-553.

48. Pan, Y. T.; Edavana, V. K.; Jourdian, W. J.; Edmondson, R.; Carroll, J. D.; Pastuszak, I.; Elbein, A. D., Trehalose synthase of *Mycobacterium smegmatis* - Purification, cloning, expression, and properties of the enzyme. *Eur. J. Biochem.* **2004,** *271* (21), 4259-4269.

49. Pan, Y.-T.; Carroll, J. D.; Asano, N.; Pastuszak, I.; Edavana, V. K.; Elbein, A. D., Trehalose synthase converts glycogen to trehalose. *FEBS J.* **2008,** *275* (13), 3408-3420.

50. Elbein, A. D.; Pastuszak, I.; Tackett, A. J.; Wilson, T.; Pan, Y. T., Last step in the conversion of trehalose to glycogen: A mycobacterial enzyme that transfers maltose from maltose-1-phosphate to glycogen *J. Biol. Chem.* **2010,** *285* (13), 9803-9812.

51. Kalscheuer, R.; Syson, K.; Veeraraghavan, U.; Weinrick, B.; Biermann, K. E.; Liu, Z.; Sacchettini, J. C.; Besra, G.; Bornemann, S.; Jacobs, W. R., Self-poisoning of *Mycobacterium tuberculosis* by targeting GlgE in an α-glucan pathway. *Nat. Chem. Biol.* **2010,** *6* (5), 376-384.

52. Miah, F.; Koliwer-Brandl, H.; Rejzek, M.; Field, R. A.; Kalscheuer, R.; Bornemann, S., Flux through trehalose synthase flows from trehalose to the α-anomer of maltose in mycobacteria. *Chem. Biol.* **2013,** *20* (4), 487-493.

53. Roy, R.; Usha, V.; Kermani, A.; Scott, D. J.; Hyde, E. I.; Besra, G. S.; Alderwick, L. J.; Futterer, K., Synthesis of α-glucan in mycobacteria involves a hetero-octameric complex of trehalose synthase TreS and maltokinase Pep2. *ACS Chem. Biol.* **2013,** *8* (10), 2245-55.

54. 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., Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in *Mycobacterium tuberculosis. J. Biol. Chem.* **2007,** *282* (37), 27270-27276.

55. Suzuki, E.; Suzuki, R., Distribution of glucan-branching enzymes among prokaryotes. *Cell. Mol. Life Sci.* **2016**, 1-18.

56. Murakami, T.; Kanai, T.; Takata, H.; Kuriki, T.; Imanaka, T., A novel branching enzyme of the GH57 family in the hyperthermophilic archaeon *thermococcus kodakaraensis* KOD1. *J. Bacteriol.* **2006,** *188* (16), 5915-5924.

57. 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. J. E. C.; Dijkhuizen, L.; Leemhuis, H., *Thermus thermophilus* Glycoside Hydrolase family 57 branching enzyme: crystal structure, mechanism of action, and products formed. *J. Biol. Chem.* **2011,** *286* (5), 3520-3530.

58. MacGregor, E. A.; Janecek, S.; Svensson, B., Relationship of sequence and structure to specificity in the α-amylase family of enzymes. *Biochim. Biophys. Acta* **2001,** *1546* (1), 1-20.

59. Wang, L.; Regina, A.; Butardo, V. M.; Kosar-Hashemi, B.; Larroque, O.; Kahler, C. M.; Wise, M. J., Influence of *in situ* progressive N-terminal is still controversial truncation of glycogen branching enzyme in *Escherichia coli* DH5α on glycogen structure, accumulation, and bacterial viability. *BMC Microbiol.* **2015,** *15* (1), 1-14.

60. Pal, K.; Kumar, S.; Sharma, S.; Garg, S. K.; Alam, M. S.; Xu, H. E.; Agrawal, P.; Swaminathan, K., Crystal structure of full-length *Mycobacterium tuberculosis* H37Rv glycogen branching enzyme: Insights of N-terminal β-sandwich in subtrate specificity and enzymatic activity *J. Biol. Chem.* **2010,** *285* (27), 20897-20903.

61. Abad, M. C.; Binderup, K.; Rios-Steiner, J.; Arni, R. K.; Preiss, J.; Geiger, J. H., The X-ray crystallographic structure of *Escherichia coli* branching enzyme. *J. Biol. Chem.* **2002,** *277* (44), 42164-42170.

62. Froese, D. S.; Michaeli, A.; McCorvie, T. J.; Krojer, T.; Sasi, M.; Melaev, E.; Goldblum, A.; Zatsepin, M.; Lossos, A.; Álvarez, R.; Escribá, P. V.; Minassian, B. A.; von Delft, F.; Kakhlon, O.; Yue, W. W., Structural basis of glycogen branching enzyme deficiency and pharmacologic rescue by rational peptide design. *Hum. Mol. Genet.* **2015,** *24* (20), 5667-5676.

63. Chaen, K.; Noguchi, J.; Omori, T.; Kakuta, Y.; Kimura, M., Crystal structure of the rice branching enzyme I (BEI) in complex with maltopentaose. *Biochem. Biophys. Res. Commun.* **2012,** *424* (3), 508-511.

64. Hayashi, M.; Suzuki, R.; Colleoni, C.; Ball, S. G.; Fujita, N.; Suzuki, E., Bound substrate in the structure of cyanobacterial branching enzyme supports a new mechanistic model. *J. Biol. Chem.* **2017**.

65. Feng, L.; Fawaz, R.; Hovde, S.; Gilbert, L.; Chiou, J.; Geiger, J. H., Crystal structures of *Escherichia coli* branching enzyme in complex with linear oligosaccharides. *Biochemistry* **2015,** *54* (40), 6207-6218.

66. Feng, L.; Fawaz, R.; Hovde, S.; Sheng, F.; Nosrati, M.; Geiger, J. H., Crystal structures of *Escherichia coli* branching enzyme in complex with cyclodextrins. *Acta crystallographica. Section D, Structural biology* **2016,** *72* (Pt 5), 641-7.

67. Suzuki, R.; Koide, K.; Hayashi, M.; Suzuki, T.; Sawada, T.; Ohdan, T.; Takahashi, H.; Nakamura, Y.; Fujita, N.; Suzuki, E., Functional characterization of three (GH13) branching enzymes involved in cyanobacterial starch biosynthesis from *Cyanobacterium* sp. NBRC 102756. *Biochim. Biophys. Acta* **2015,** *1854* (5), 476-484.

68. Elbein, A. D.; Pan, Y. T.; Pastuszak, I.; Carroll, D., New insights on trehalose: a multifunctional molecule. *Glycobiology* **2003,** *13* (4), 17R-27R.

69. Takayama, K.; Wang, C.; Besra, G. S., Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.* **2005,** *18* (1), 81-101.

70. Rose, J. D.; Maddry, J. A.; Comber, R. N.; Suling, W. J.; Wilson, L. N.; Reynolds, R. C., Synthesis and biological evaluation of trehalose analogs as potential inhibitors of mycobacterial cell wall biosynthesis. *Carbohydr. Res.* **2002,** *337* (2), 105-120.

71. Wang, J.; Elchert, B.; Hui, Y.; Takemoto, J. Y.; Bensaci, M.; Wennergren, J.; Chang, H.; Rai, R.; Chang, C. W. T., Synthesis of trehalose-based compounds and their inhibitory activities against *Mycobacterium smegmatis*. *Biorg. Med. Chem.* **2004,** *12* (24), 6397-6413.

72. Chang, C. W. T.; Elchert, B.; Hui, Y.; Wang, J. H.; Wennergren, J.; Rai, R.; Takemoto, J.; Bensaci, M., Novel trehalose-based compounds against *Mycobacterium*, synthesis and antibacterial studies. *Abstr. Am. Chem. Soc.* **2003,** *226*, 212.

73. Murphy, H. N.; Stewart, G. R.; Mischenko, V. V.; Apt, A. S.; Harris, R.; McAlister, M. S. B.; Driscoll, P. C.; Young, D. B.; Robertson, B. D., The OtsAB pathway is essential for trehalose biosynthesis in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **2005,** *280* (15), 14524-14529.

74. Maruta, K.; Mitsuzumi, H.; Nakada, T.; Kubota, M.; Chaen, H.; Fukuda, S.; Sugimoto, T.; Kurimoto, M., Cloning and sequencing of a cluster of genes encoding novel enzymes of trehalose biosynthesis from thermophilic archaebacterium *Sulfolobus acidocaldarius*. *Biochim. Biophys. Acta* **1996,** *1291* (3), 177-181.

75. Ruhal, R.; Kataria, R.; Choudhury, B., Trends in bacterial trehalose metabolism and significant nodes of metabolic pathway in the direction of trehalose accumulation. *Microb. Biotechnol.* **2013,** *6* (5), 493-502.

76. Song, H. N.; Jung, T. Y.; Park, J. T.; Park, B. C.; Myung, P. K.; Boos, W.; Woo, E. J.; Park, K. H., Structural rationale for the short branched substrate specificity of the glycogen debranching enzyme GlgX. *Proteins: Struct. Funct. Bioinform.* **2010,** *78* (8), 1847-1855.

77. Hershkovitz, I.; Donoghue, H. D.; Minnikin, D. E.; Besra, G. S.; Lee, O. Y. C.; Gernaey, A. M.; Galili, E.; Eshed, V.; Greenblatt, C. L.; Lemma, E.; Bar-Gal, G. K.; Spigelman, M., Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a neolithic settlement in the eastern mediterranean. *PLoS One* **2008,** *3* (10).

78. Zumla, A.; Nahid, P.; Cole, S. T., Advances in the development of new tuberculosis drugs and treatment regimens. *Nat. Rev. Drug Discov.* **2013,** *12* (5), 388-404.

79. WHO. WHO Global tuberculosis report 2016. http://who.int/tb/publications/global_report/en/ (accessed Sep 1, 2017).

80. Sakula, A., Robert Koch: Centenary of the discovery of the tubercle bacillus, 1882. *Thorax* **1982,** *37* (4), 246-251.

81. Woodruff, P. J.; Carlson, B. L.; Siridechadilok, B.; Pratt, M. R.; Senaratne, R. H.; Mougous, J. D.; Riley, L. W.; Williams, S. J.; Bertozzi, C. R., Trehalose is required for growth of *Mycobacterium smegmatis*. *J. Biol. Chem.* **2004,** *279* (28), 28835-28843.

82. Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S. V.; Eiglmeier, K.; Gas, S.; Barry, 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.; Hornby, 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., Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **1998,** *393* (6685), 537-544.

83. Daffe, M.; Draper, P., *The envelope layers of mycobacteria with reference to their pathogenicity*. Academic Press Ltd: London, 1998; Vol. 39, p 131-203.

84. Jackson, M.; McNeil, M. R.; Brennan, P. J., Progress in targeting cell envelope biogenesis in *Mycobacterium tuberculosis*. *Future Microbiol.* **2013,** *8* (7), 855-875.

85. Brennan, P. J.; Nikaido, H., The envelope of mycobacteria. *Annu. Rev. Biochem.* **1995,** *64*, 29-63.

86. Colditz, G. A.; Brewer, T. F.; Berkey, C. S.; Wilson, M. E.; Burdick, E.; Fineberg, H. V.; Mosteller, F., Efficacy of BCG vaccine in the prevention of tuberculosis - metaanalysis of the published literature. *J. Am. Med. Assoc.* **1994,** *271* (9), 698-702.

87. WHO. The End TB Strategy. http://www.who.int/tb/post2015_strategy/en/ (accessed Sep 1, 2017).

88. Diacon, A. H.; Pym, A.; Grobusch, M.; Patientia, R.; Rustomjee, R.; Page-Shipp, L.; Pistorius, C.; Krause, R.; Bogoshi, M.; Churchyard, G.; Venter, A.; Allen, J.; Palomino, J. C.; De Marez, T.; van Heeswijk, R. P. G.; Lounis, N.; Meyvisch, P.; Verbeeck, J.; Parys, W.; de Beule, K.; Andries, K.; Mc Neeley, D. F., The diarylquinoline TMC207 for multidrug-resistant tuberculosis. *New Engl. J. Med.* **2009,** *360* (23), 2397-2405.

89. Thanna, S.; Sucheck, S. J., Targeting the trehalose utilization pathways of *Mycobacterium tuberculosis. MedChemComm* **2016,** *7* (1), 69-85.

90. Lindenberger, J. J.; Kumar Veleti, S.; Wilson, B. N.; Sucheck, S. J.; Ronning, D. R., Crystal structures of *Mycobacterium tuberculosis* GlgE and complexes with non-covalent inhibitors. *Sci. Rep.* **2015,** *5*, 12830.

91. Syson, K.; Stevenson, C. E. M.; Rejzek, M.; Fairhurst, S. A.; Nair, A.; Bruton, C. J.; Field, R. A.; Chater, K. F.; Lawson, D. M.; Bornemann, S., Structure of *Streptomyces* maltosyltransferase GlgE, a homologue of a genetically validated anti-tuberculosis target. *J. Biol. Chem.* **2011,** *286* (44), 38298-38310.

92. Syson, K.; Stevenson, C. E. M.; Rashid, A. M.; Saalbach, G.; Tang, M.; Tuukkanen, A.; Svergun, D. I.; Withers, S. G.; Lawson, D. M.; Bornemann, S., Structural insight into how *Streptomyces coelicolor* maltosyl transferase GlgE binds α-maltose 1-phosphate and forms a maltosyl-enzyme intermediate. *Biochemistry* **2014,** *53* (15), 2494-2504.

93. Syson, K.; Stevenson, C. E. M.; Miah, F.; Barclay, J. E.; Tang, M.; Gorelik, A.; Rashid, A. M.; Lawson, D. M.; Bornemann, S., Ligand-bound structures and site-directed mutagenesis identify the acceptor and secondary binding sites of *Streptomyces coelicolor* maltosyltransferase GlgE. *J. Biol. Chem.* **2016,** *291* (41), 21531-21540.

94. Mendes, V.; Blaszczyk, M.; Maranha, A.; Empadinhas, N.; Blundell, T. L., Structure of *Mycobacterium thermoresistibile* GlgE defines novel conformational states that contribute to the catalytic mechanism. *Sci. Rep.* **2015,** *5*, 17144.

95. Leiba, J.; Syson, K.; Baronian, G.; Zanella-Cleon, I.; Kalscheuer, R.; Kremer, L.; Bornemann, S.; Molle, V., *Mycobacterium tuberculosis* maltosyltransferase GlgE, a genetically validated antituberculosis target, is negatively regulated by Ser/Thr phosphorylation. *J. Biol. Chem.* **2013,** *288* (23), 16546-16556.

96. Thanna, S.; Lindenberger, J. J.; Gaitonde, V. V.; Ronning, D. R.; Sucheck, S. J., Synthesis of 2-deoxy-2,2-difluoro-α-maltosyl fluoride and its X-ray structure in complex with *Streptomyces coelicolor* GlgEI-V279S. *Org. Biomol. Chem.* **2015,** *13* (27), 7542-7550.

97. Billones J. B, V. A. M. F., Structure-based design of inhibitors against maltosyltransferase GlgE. *Orient. J. Chem.* **2014,** *30* (3).

98. Kalscheuer, R.; Weinrick, B.; Veeraraghavan, U.; Besra, G. S.; Jacobs, W. R., Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **2010,** *107* (50), 21761-21766.

99. Jiang, D.; Zhang, Q.; Zheng, Q.; Zhou, H.; Jin, J.; Zhou, W.; Bartlam, M.; Rao, Z., Structural analysis of *Mycobacterium tuberculosis* ATP-binding cassette transporter subunit UgpB reveals specificity for glycerophosphocholine. *FEBS J.* **2014,** *281* (1), 331-341.

100. Fullam, E.; Prokes, I.; Fütterer, K.; Besra, G. S., Structural and functional analysis of the solute-binding protein UspC from *Mycobacterium tuberculosis* that is specific for amino sugars. *Open Biol.* **2016,** *6* (6).

101. Veleti, S. K.; Lindenberger, J. J.; Ronning, D. R.; Sucheck, S. J., Synthesis of a *C*-phosphonate mimic of maltose-1-phosphate and inhibition studies on *Mycobacterium tuberculosis* GlgE. *Biorg. Med. Chem.* **2014,** *22* (4), 1404-1411.

102. Veleti, S. K.; Lindenberger, J. J.; Thanna, S.; Ronning, D. R.; Sucheck, S. J., Synthesis of a poly-hydroxypyrolidine-based inhibitor of *Mycobacterium tuberculosis* GlgE. *J. Org. Chem.* **2014,** *79* (20), 9444-9450.

103. Veleti, S. K.; Petit, C.; Ronning, D. R.; Sucheck, S. J., Zwitterionic pyrrolidene-phosphonates: inhibitors of the glycoside hydrolase-like phosphorylase *Streptomyces coelicolor* GlgEI-V279S. *Org. Biomol. Chem.* **2017,** *15* (18), 3884-3891.

104. Jayawardana, K. W.; Jayawardena, H. S. N.; Wijesundera, S. A.; De Zoysa, T.; Sundhoro, M.; Yan, M., Selective targeting of *Mycobacterium smegmatis* with trehalose-functionalized nanoparticles. *Chem. Commun.* **2015,** *51* (60), 12028-12031.

105. Swarts, B. M.; Holsclaw, C. M.; Jewett, J. C.; Alber, M.; Fox, D. M.; Siegrist, M. S.; Leary, J. A.; Kalscheuer, R.; Bertozzi, C. R., Probing the mycobacterial trehalome with bioorthogonal chemistry. *J. Am. Chem. Soc.* **2012,** *134* (39), 16123-16126.

106. Foley, H. N.; Stewart, J. A.; Kavunja, H. W.; Rundell, S. R.; Swarts, B. M., Bioorthogonal chemical reporters for selective in situ probing of mycomembrane components in mycobacteria. *Angew. Chem. Int. Ed.* **2016,** *55* (6), 2053-2057.

107. Rundell, S. R.; Wagar, Z. L.; Meints, L. M.; Olson, C. D.; O'Neill, M. K.; Piligian, B. F.; Poston, A. W.; Hood, R. J.; Woodruff, P. J.; Swarts, B. M., Deoxyfluoro-D-trehalose (FDTre) analogues as potential PET probes for imaging mycobacterial infection. *Org. Biomol. Chem.* **2016,** *14* (36), 8598-8609.

108. Backus, K. M.; Boshoff, H. L.; Barry, C. S.; Boutureira, O.; Patel, M. K.; D'Hooge, F.; Lee, S. S.; Via, L. E.; Tahlan, K.; Barry, C. E.; Davis, B. G., Uptake of unnatural trehalose analogs as a reporter for *Mycobacterium tuberculosis*. *Nat. Chem. Biol.* **2011,** *7* (4), 228-235.

109. Garg, S.; Alam, M. S.; Bajpai, R.; Kishan, K. V. R.; Agrawal, P., Redox biology of *Mycobacterium tuberculosis* H37Rv: protein-protein interaction between GlgB and WhiB1 involves exchange of thiol-disulfide. *BMC Biochem.* **2009,** *10*.

110. Dkhar, H. K.; Gopalsamy, A.; Loharch, S.; Kaur, A.; Bhutani, I.; Saminathan, K.; Bhagyaraj, E.; Chandra, V.; Swaminathan, K.; Agrawal, P.; Parkesh, R.; Gupta, P., Discovery of *Mycobacterium tuberculosis* α-1,4-glucan branching enzyme (GlgB) inhibitors by structure- and ligand-based virtual screening. *J. Biol. Chem.* **2015,** *290* (1), 76-89.

2. Structural insights into the control of branch lengths in mycobacterial α-glucan

2.1 Introduction

α-Glucans are important energy storage polysaccharides in bacteria, plants and animals. In *M. tuberculosis* α-glucan also functions as a virulence factor that is exported to the capsular layer of the mycobacterial cell envelope, where it can interact with human immune receptors.[1-2] In *M. tuberculosis* and other actinomycetes, α-glucan is synthesised from M1P, by the maltosyl transferase GlgE and the α-1,6-branching enzyme GlgB.[3] Prior to the work described in this chapter, α-glucans from different actinomycete sources were extracted and analysed, demonstrating that actinomycete α-glucan is structurally distinct from classical *E. coli* or mammalian glycogen. It had shorter average branch lengths of DP ~7, compared to DP ~11 for classical glycogen. Mycobacterial α-glucan also had shorter internal chain lengths of DP ~5.0 and a smaller A:BC chain ratios of 0.2-0.6, compared to DP ~8 and 0.8 for classical glycogen.[4] However, it had not been shown that GlgE and GlgB were directly responsible for the distinct α-glucan architecture. Therefore, I aimed to answer the question: Are GlgE and GlgB sufficient to generate α-glucan *in vitro*?

Following this, I wanted to investigate the structural features of mycobacterial GlgB, to improve the understanding of how this enzyme generates the distinct architecture of the α-glucan product. This would also be useful in order to identify target sites for MtGlgB inhibitors and help with the determination of structure-activity relationships of potential inhibitors. The structural determinants that control the product architecture would also be relevant to the general understanding of branching enzymes, as these had not been clearly established for any enzyme of this type. Whilst all GH13 branching enzymes share the same domains, motifs and GH α-retaining mechanism, **Figure 2.1**, there is no obvious correlation with amino acid sequence and branch length specificity.[5] Crystal structures of branching enzymes from *E. coli*, rice, human, *M. tuberculosis* and cyanobacteria have been solved,[6-11] however only the latter has a maltooligosaccharide bound at the donor site, which allowed the identification of the donor subsites. Several maltooligosaccharide binding sites were also identified in the *E. coli*, rice and human enzymes,[7-8, 11-12] but these were far removed from the catalytic residues. An acceptor site was proposed for the cyanobacterial enzyme, based on similar distal maltooligosaccharide binding sites, however no maltooligosaccharides bound closer than ~30 Å to the catalytic residues were seen in this region.[10]

For the *M. tuberculosis* homologue only a low quality, apo structure had been reported,[9] so it was not yet possible to relate substrate binding to branch length control. Therefore, I firstly aimed to answer the question: Can the structural model of mycobacterial GlgB be improved? Secondly, I aimed to answer the question: can mycobacterial GlgB substrate binding sites be identified? This would not only serve as a tool to investigate the structural determinants of branch length specificity, but would also reveal new target sites for small-molecule GlgB inhibitors.
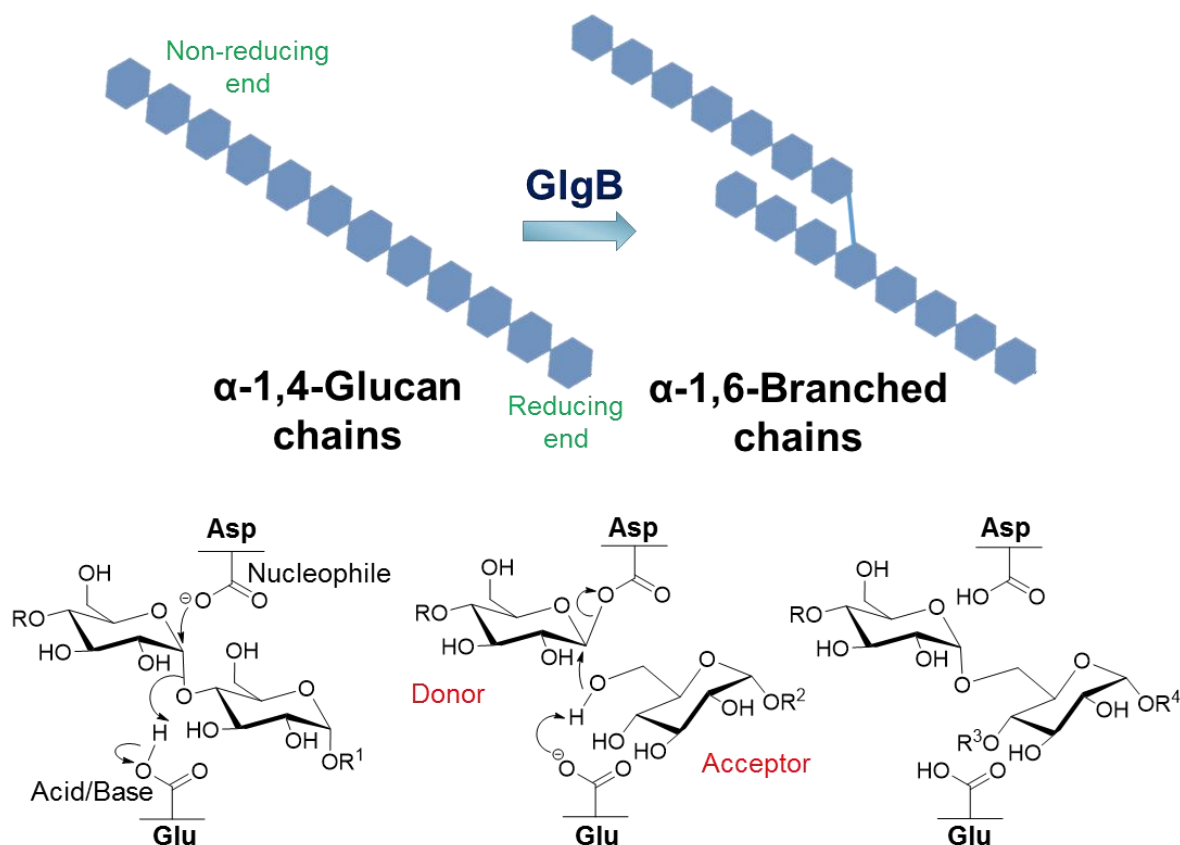
Figure 2.1 Branching enzyme mechanism. The branching enzyme GlgB cleaves an α-1,4 glucan chain and transfers the non-reducing end on to a 6-OH of another glucose residue, creating a branchpoint. GH13 branching enzymes utilise an α-retaining mechanism, with the catalytic residues of an Asp nucleophile and a Glu acid/base. The reaction proceeds via a covalent intermediate (middle panel).

2.2 GlgB enzymes from actinomycetes generate α-glucan branch lengths of DP 7-8 *in vitro*

Some of the work described in this section has been published as part of Rashid *et al*. 2016[4]. **Appendix 1**, and appears with permission.

2.2.1 Production of recombinant GlgB enzymes

To investigate the activity of actinomycete branching enzymes *in vitro*, I firstly produced recombinant GlgB proteins. I optimised the expression and purification protocols to produce MtGlgB and MsGlgB each with an N-terminal His$_6$ tag in high yield, **Figure 2.2**. Other GlgB enzymes from *S. coelicolor*, *S. venezuelae* and *M. tuberculosis* CCDC5180 had been previously produced by A. Rashid.[4]
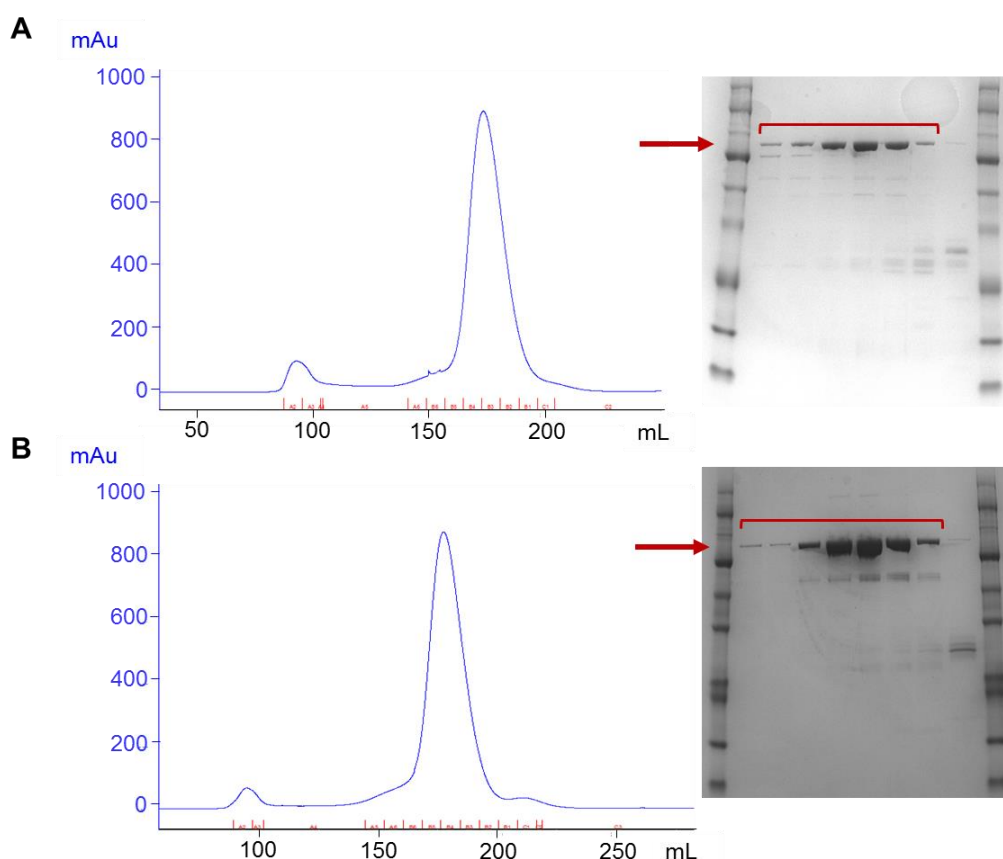
Figure 2.2 Production of recombinant GlgB proteins.A) *M. tuberculosis* **H37Rv GlgB; B)** *M. smegmatis* **mc² 155 GlgB.** Proteins were purified by Ni⁺ affinity and size exclusion chromatographies, which showed one major peak. The protein purity was confirmed by analysing fractions with SDS-PAGE and Coomassie staining. The expected mass of the recombinant GlgB protein is shown with a red arrow adjacent to the major component in both cases. For each protein, the fractions indicated with a red bracket were combined and concentrated for use in further experiments.

2.2.2 Synthesis of α-glucan *in vitro*

The GlgB enzyme was hypothesised to work in tandem with GlgE, which extends α-1,4 oligosaccharides using M1P to generate α-1,4 glucan substrates for branching.[3] It was previously established that *M. tuberculosis* and *S. coelicolor* GlgE enzymes have similar kinetic properties, apart from an order of magnitude lower k_{cat}^{app} for the *M. tuberculosis* enzyme.[13] Therefore, to generate α-glucan *in vitro*, I incubated M1P and a maltohexaose acceptor with either *M. tuberculosis* or *S. coelicolor* GlgE along with MtGlgB. To ensure completion of the reactions, two further aliquots of M1P were added over the course of the 90 h incubation period. Residual low molecular weight oligosaccharides of mass ≤ 5 kDa were removed by filtration and the resultant α-glucan particles were imaged with transmission electron microscopy (TEM) by E. Barclay, **Figure 2.3**. Similar spherical particles were observed, with average sizes of 32 ± 14 nm and 30 ± 15 nm for *M. tuberculosis* GlgE and *S. coelicolor* GlgE respectively, indicating that the source of GlgE does not affect α-glucan particle size in these conditions. Particles were slightly smaller than in α-glucan extracted from *M. tuberculosis*, which had an average size of 58 ± 2 nm. I went on to generate *in vitro* α-glucan with the other actinomycete GlgB enzymes detailed above, in combination with *S. coelicolor* GlgE due to its higher catalytic activity.
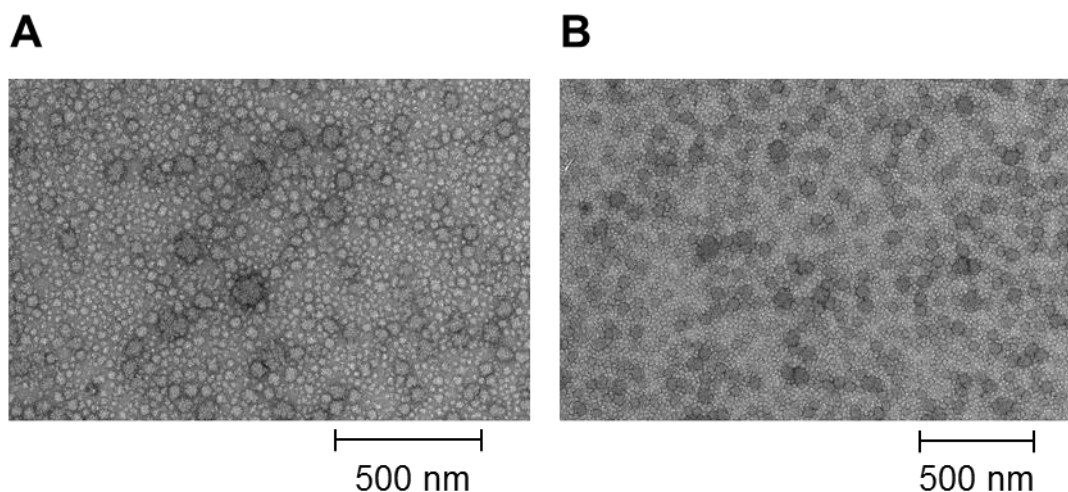
Figure 2.3 Synthetic α-Glucan particles. A) α-Glucan particles generated by *M. tuberculosis* GlgE and MtGlgB. Mean particle size ± standard deviation: 32 ± 14 from 48 measurements. **B)** α-Glucan particles generated by *S. coelicolor* GlgE and MtGlgB. Mean particle size 30 ± 15 nm from 40 measurements. Branch lengths DP 8.1 ± 2.7 (MALDI mass spectrometry) or 7.5 ± 3.5 (CE).

2.2.3 Synthetic α-glucan has branch lengths of DP 7-8

α-Glucan branch lengths were then determined by debranching with isoamylase and analysis by matrix assisted laser desorption / ionisation (MALDI) mass spectrometry. Results were verified by capillary electrophoresis (CE), which gave very similar branch lengths, showing that MALDI mass spectrometry was reliable for these analyses.

Comparison of reactions of *M. tuberculosis* GlgE or *S. coelicolor* GlgE in combination with MtGlgB showed similar branch lengths of DP 8.6 ± 2.5 and 8.1 ± 2.7 respectively, demonstrating the source of GlgE also has no discernible effect on α-glucan branch lengths. The branch length specificities of the other actinomycete GlgB enzymes were then determined, **Table 2.1**. Branch lengths ranged from DP 7-8, which correlated with those of α-glucan extracted from actinomycete cells, therefore, GlgE and GlgB are sufficient for synthesis of an α-glucan which is representative of the polysaccharide *in vivo*. This work confirms that these enzymes are responsible for the notably shorter branches in actinomycete α-glucan compared to those in classical glycogen, which has branch lengths of 10.8 ± 4.6 (rabbit liver) and 11.0 ± 2.9 (*E. coli*).[4]

Table 2.1 α-Glucan branch lengths generated by actinomycete GlgB enzymes. α-Glucan was generated as described above, debranched with isoamylase and branch lengths were analysed by MALDI mass spectrometry and CE. Length distributions are shown as mean ± standard deviation of products DP ≥ 4. Each result is representative of at least two independent experiments.

Source of GlgB	Source of GlgE	Branch Length DP	
		MALDI Mass Spectrometry	CE
M. tuberculosis	*M. tuberculosis*	8.6 ± 2.5	8.0 ± 3.1
	S. coelicolor	8.1 ± 2.7	7.5 ± 3.5
M. tuberculosis CCDC5180	*S. coelicolor*	8.4 ± 2.5	-
M. smegmatis		8.5 ± 2.5	-
S. venezuelae		7.9 ± 2.4	-
S. coelicolor		7.2 ± 2.2	7.1 ± 3.1

2.3   A high-resolution structural model of MtGlgB

Having established that MtGlgB and other actinomycete glucan branching enzymes generate markedly shorter branches than those of classical glycogen, my next goal was to investigate the structural basis for this differing branching activity. This firstly required the generation of high-resolution GlgB crystal structures, with which substrate binding could be investigated.

2.3.1 Crystallisation of MtGlgB

The crystal structure of apo MtGlgB at 2.33 Å was previously solved by Pal and co-workers (PDB code 3K1D),[9] however no substrate-bound structures had been reported for this enzyme. I therefore initially set out to replicate the crystallisation of MtGlgB, according to the reported conditions, comprising a precipitant solution of 30% polyethylene glycol (PEG) 4,000 and 0.1 M MES at pH 6.5 with 12 mg/mL protein. The MtGlgB protein with an N-terminal His$_6$ tag was produced as described above, and crystallisation trials were set up with these conditions, but no protein crystals were observed. Systematic variation of the components of the precipitant solution, pH and protein concentration, along with the addition of reducing agents, metal ions or maltooligosaccharides did not improve protein crystallisation. The MtGlgB protein was then subjected to crystallisation trials with a range of different commercial screens, which also yielded no protein crystals. Cleavage of the N-terminal His$_6$ tag did not improve protein crystallisation under any of the investigated conditions.

I then explored the possibility that the tags used for protein production affected the protein crystallisation. The MtGlgB protein produced by Pal and co-workers included an N-terminal S-tag and a C-terminal His$_6$ tag, compared with a single N-terminal His$_6$ tag in my construct. Modifying the expression construct to locate the His$_6$ tag at the C terminus resulted in much a lower yield of protein that was mostly insoluble. Investigation of the plasmid stability indicated a higher proportion than expected of mutants that had lost the ability to produce MtGlgB-CHis$_6$, **Figure 2.4A**. This normally suggests that the target protein is toxic to the cell, although it is difficult to rationalise how this would result from such a small modification to the protein. Both the solubility and overall yield were not improved with different *E. coli* expression strains.

Subsequently, I investigated truncating the protein sequence as an approach to enable protein crystallisation. Firstly, I tested the removal of amino acid residues 1-108, to remove domain N, because this is the most disordered region in the protein according to predictions

generated with XtalPred-RF,[14] and this approach has been successful for the crystallisation of other branching enzymes.[6] In contrast to the previous expression construct, plasmid stability tests indicated that the resultant recombinant protein had no detrimental effect on *E. coli* growth, **Figure 2.4A.** Whilst it was produced in relatively high yield, it was insoluble in all expression conditions tested. I then tested the removal of just the first nine amino acids, which were predicted to be the most disordered region in domain N, by both XtalPred-RF[14] and their absence in the MtGlgB electron density map. The resulting protein was both soluble and could be produced in high yield. Gratifyingly, crystallisation trials with various commercial screens resulted in protein crystals with two different precipitant solutions, namely 12.5% polyacrylate 2100 with 0.5 M $NH_4H_2PO_4$ and 30% Sokalan® CP42 with 0.1 M Tris-HCl at pH 8.5. Sokalan® is also an acrylic polymer, suggesting that this reagent may play an essential role the crystallisation of ΔN9-MtGlgB. Protein crystallisation strategies are summarised in **Table 2.2** and crystal images are shown in **Figure 2.4**.

Table 2.2 Summary of approaches for MtGlgB crystallisation. Full length MtGlgB with either an N or C-terminal His_6 tag did not crystallise. Removal of the N domain resulted in insoluble protein, whereas removal of only the first nine amino acids gave protein crystals, however these diffracted very weakly.

Protein ID	Vector	Tag	Soluble Protein	Crystallisation	Diffraction
MtGlgB	pET21a	N-His_6	✓	✗	-
MtGlgB-CHis_6	pET101a	C-His_6	✓ (low)	✗	-
ΔN9-MtGlgB	pET21a	N-His_6	✓	✓	✗
ΔN108-MtGlgB	pOPINA	C-His_6	✗	-	-
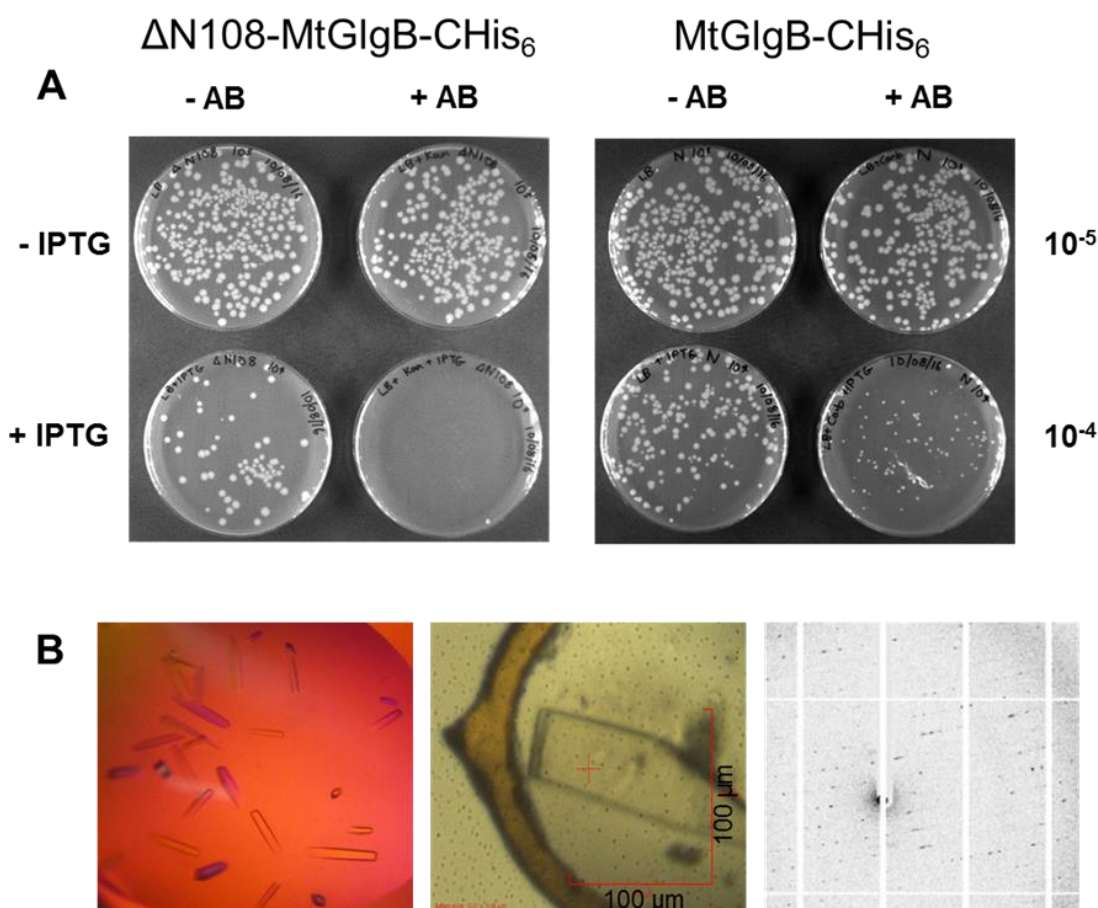
Figure 2.4 Crystallisation of MtGlgB. A) Plasmid stability tests for MtGlgB-C-His$_6$ ΔN108-MtGlgB-C-His$_6$. Colonies on IPTG + antibiotic (AB) plate for MtGlgB-C-His$_6$ indicates that production of this protein has an unexplained detrimental effect on cell function, whereas the absence of colonies on the equivalent plate for ΔN108-MtGlgB-C-His$_6$ demonstrates that this protein is non-toxic. **B)** Images of (i) ΔN9-MtGlgB crystals generated with polyacrylate precipitant imaged with a polarising optic; (ii) rod-shaped crystal in cryoloop; (iii) part of the diffraction pattern, showing very weak diffraction.

I attempted to optimise the production of protein crystals by varying the components of the precipitant solutions and micro-seeding with crystal fragments. Any modifications to the polyacrylate precipitant solution decreased the quality of the crystals. For the other crystallisation condition, decreasing the proportion of Sokalan to ~10% increased the size of the crystals. A protein concentration of ~5 mg/mL was found to be optimal as higher or lower concentrations resulted in precipitation or no crystals, respectively. In all cases, cleavage of the N-terminal His$_6$ tag, addition of maltooligosaccharides and varying the ratio of protein and precipitant solution had no discernible effect on the crystals generated. Likewise, a commercial additive screen did not lead to any improvements or variations in the type or quality of crystals produced. Finally, crystallisation trials with commercial screens that did not previously yield crystals were repeated using seeding with crystal fragments. This resulted in protein crystallisation in two further conditions, however the crystals were

too small to be used in X-ray diffraction experiments and were not improved by optimisations.

2.3.2 The structure of MtGlgB could not be solved

Crystals of MtGlgB were cryoprotected, frozen and X-ray diffraction data were collected. In general, the diffraction was very poor, with crystals from the polyacrylate condition diffracting slightly more strongly. Data processing was attempted for the best data set, for which the solvent content was estimated at 65%. Attempts to process the data with the given cell parameters resulted in two overlapped molecules with large gaps in the P2 crystal lattice. Doubling the b axis resulted in a reasonable solution with no overlaps, which contained four molecules per asymmetric unit. However, attempts to refine this model were unsuccessful and an R_{free} value of less than 45% could not be obtained. Crystals from the attempted optimisations described above gave similar or worse data. I concluded that it was not feasible to investigate the structural determinants of branching activity with MtGlgB.

2.3.3 Crystallisation of MsGlgB

Since it was not possible to obtain structural data for MtGlgB, I went on to investigate the MsGlgB homologue. These two proteins share 82% amino acid sequence identity and earlier in this chapter I showed they have very similar branch length specificities. Therefore, it was reasonable to expect that the *M. smegmatis* homologue would provide a good structural model for MtGlgB. The MsGlgB protein was produced as described in section **2.2.1** and subjected to crystallisation trials with a range of commercial screens. After one week, crystals were observed with precipitant solution: 0.06 M MgCl$_2$, 0.06 M CaCl$_2$, 20% glycerol, 10% PEG 4,000, 0.1 M imidazole, 0.1 M MES at pH 6.5. Variation of the components of the precipitant solution did not improve the quality of crystals, but micro-seeding did lead to improvements in the crystal size and morphology, **Figure 2.5**.
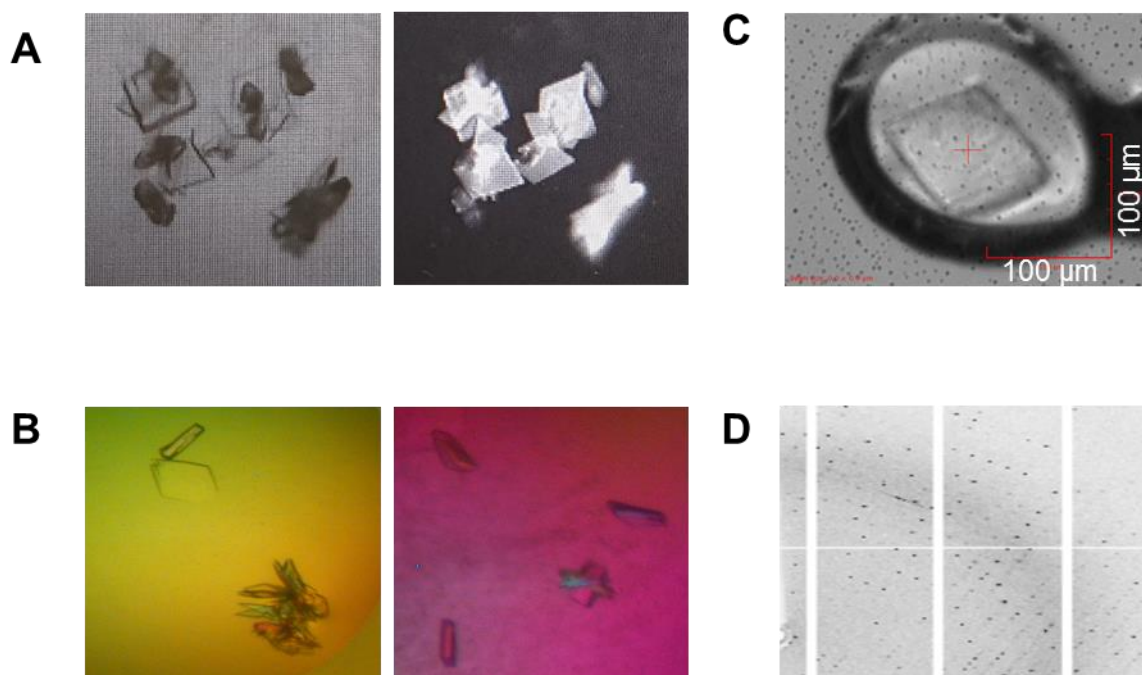
Figure 2.5 Crystallisation of MsGlgB. A) Initial crystals observed for MsGlgB with and without UV imaging. **B)** Optimised crystals of MsGlgB imaged with a polarising optic. **C)** Diamond-shaped crystal in cryoloop. **D)** Section of the diffraction pattern showing strong diffraction.

2.3.4 The structure of MsGlgB was solved at 1.72 Å resolution

Crystals of MsGlgB were flash-frozen without additional cryoprotectant and X-ray diffraction data were recorded. The data sets were processed in the $P2_12_12$ space group, with cell parameters of approximately a = 98, b = 164, c = 54 Å. The initial structure was solved by molecular replacement in PHASER,[15] using MtGlgB (PDB code 3K1D) as a template. One molecule was placed in the asymmetric unit, corresponding to the monomeric nature of the protein in solution. The structure was rebuilt and completed through several iterations of refinement with REFMAC5,[16] and manual model building with COOT.[17] This model was used as a starting point to refine further MsGlgB structures, and the best data set for the apo wildtype (WT) protein, **Apo-WT-MsGlgB**, was solved at 1.72 Å resolution, **Figure 2.6** and **Table 2.4.**
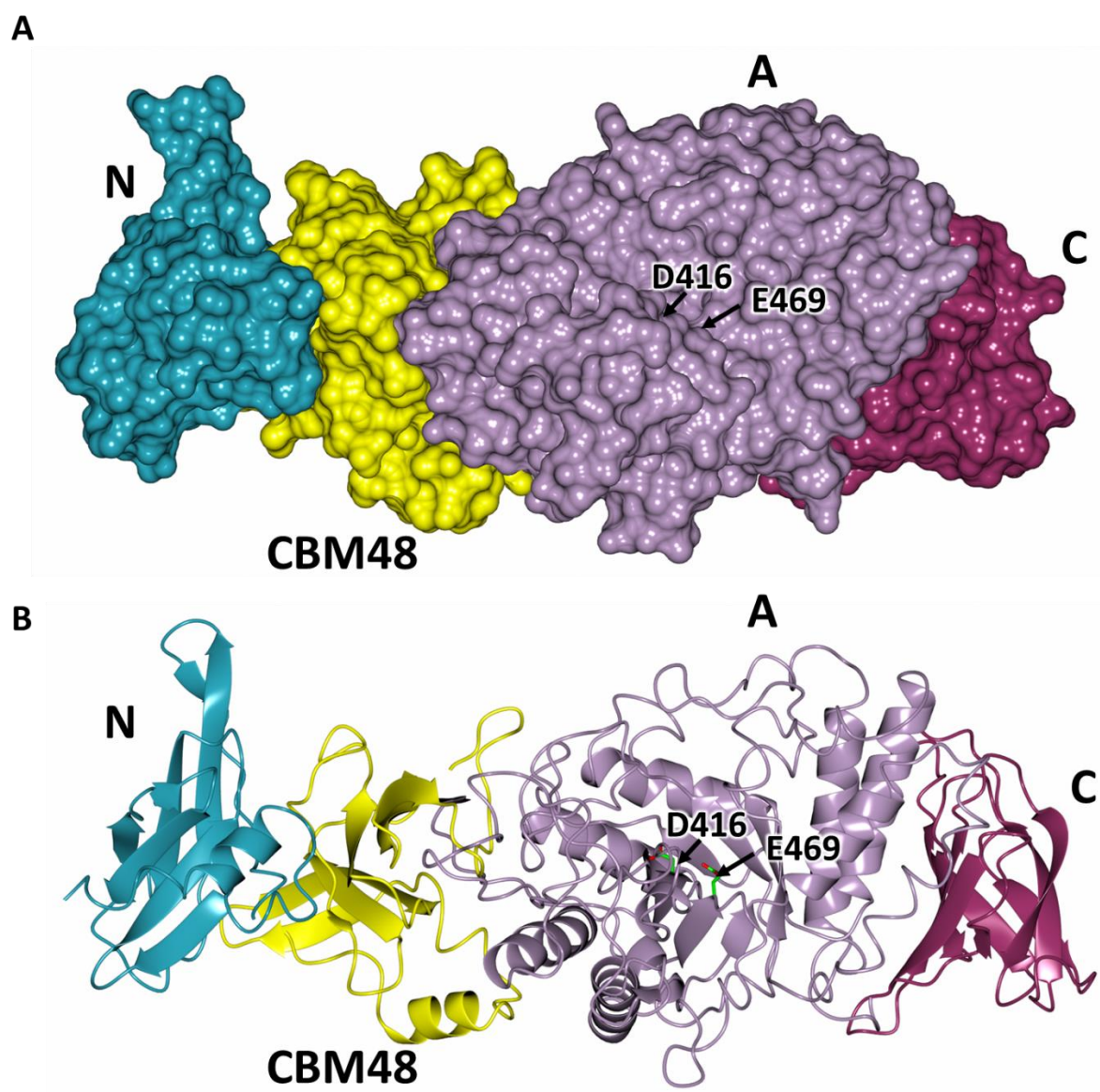
Figure 2.6 Structure of MsGlgB coloured by domain. Apo-WT-MsGlgB structure shown as **A)** surface model and **B)** ribbon model. Cyan = domain N, yellow = CBM48, lilac= domain A, dark purple = domain C. Catalytic residues D416 and E469 are indicated with arrows.

2.3.5 Comparison of MsGlgB and MtGlgB apo structures

In common with MtGlgB and other branching enzymes, [6-8, 10] MsGlgB comprised domain N (residues 9-110), CBM48 (111-231), the catalytic domain A (232-635) and domain C (637-736), arranged sequentially along the longitudinal axis of the protein, **Figure 2.6**. Reflecting their conserved amino acid sequences and catalytic specificities, the 3D structures of the enzymes were highly similar, with a C-α root mean square deviation (RMSD) of 0.49 Å. Differences were mainly concentrated in domain N and some surface loops, **Figure 2.7**. Overall, this work validated MsGlgB as a high-resolution structural model for MtGlgB. This model represents a significant improvement on the previously published structure, and can

be used for both the rational design of GlgB inhibitors and to investigate the structural determinants of the branch length specificity.
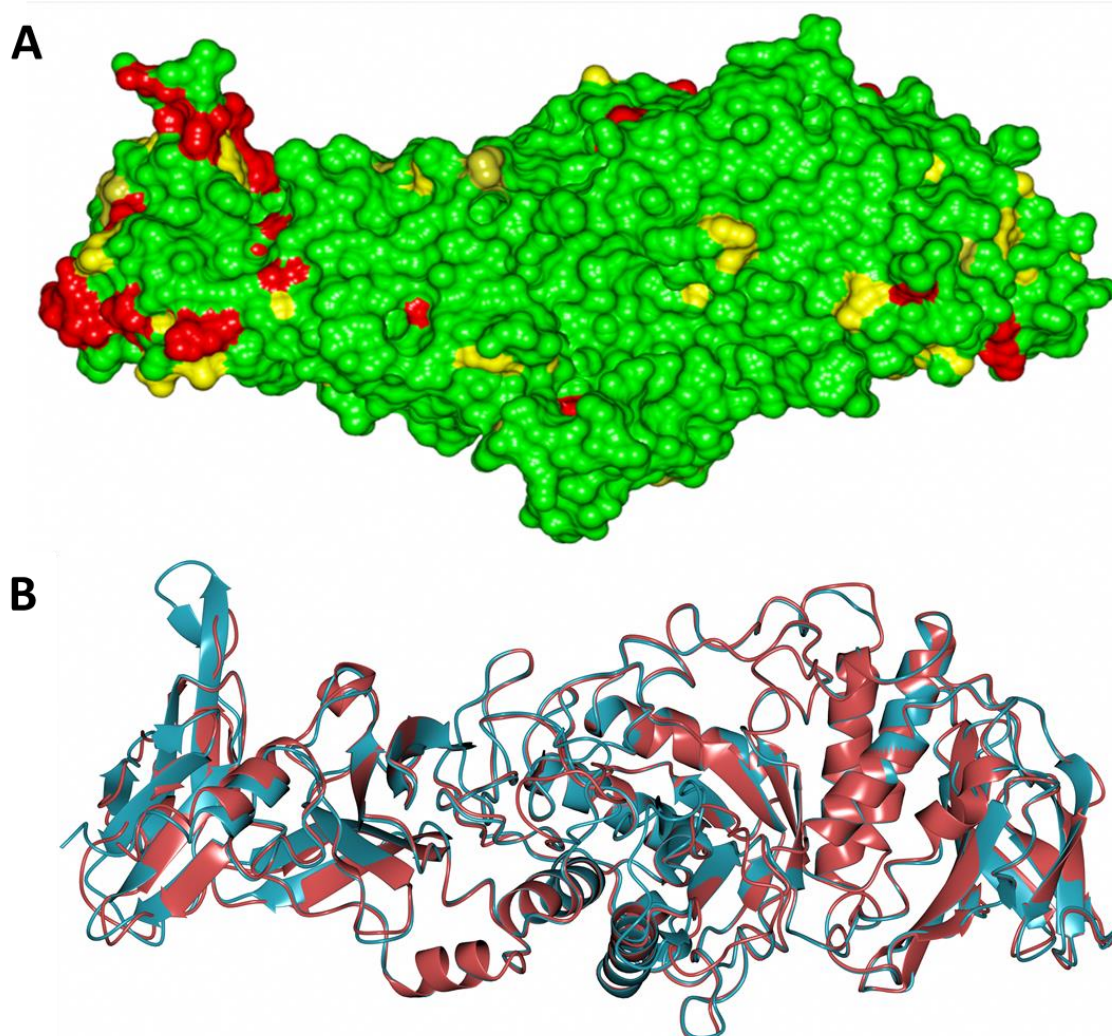
Figure 2.7 MsGlgB is structurally equivalent to MtGlgB. A) Surface model of MsGlgB coloured by amino acid sequence conservation with MtGlgB. Green = identical, yellow = similar, red = different. The proteins share 82% sequence identity, with most differences found in their N-termini. **B)** Superposition of MsGlgB (cyan) and MtGlgB (coral) structures, shown as ribbon models. As would be expected from their high sequence conservation, the structures are very similar, with a C_α RMSD of 0.49 Å.

2.4   Identification of oligosaccharide binding sites in MsGlgB

Having obtained a high-resolution model for mycobacterial GlgB, my next objective was to identify oligosaccharide binding sites to understand how this enzyme binds to its substrate and creates branches of a specific length. In total, 10 distinct binding sites could be identified from five oligosaccharide-bound crystal structures of wild type and modified GlgB proteins. An overview of oligosaccharide binding is shown in **Figure 2.8** and the identification of these sites is detailed in the following sections. For clarity, the structures referred to in this section are summarised in **Table 2.3**.

Table 2.3 Summary of structures discussed in Section 2.4. Ligand binding sites are defined and colour-coded according to **Figure 2.7** and the identity of the maltooligosaccharide ligands is given in parenthesis.
* DP3-ACR binds across the N3 and A1 sites in **DP3-ACR-MsGlgB**.

Structure ID	X-ray data collection statistics	Ligand binding sites
Apo-WT-MsGlgB	**Table 2.4**	None
DP4-WT-MsGlgB	**Table 2.4**	C4 (DP4)
Apo-D416A-MsGlgB	**Appendix 2**	None
Apo-E469A-MsGlgB	**Appendix 2**	None
Apo-D416A-E469A-MsGlgB	**Appendix 2**	None
DP8-D416A-MsGlgB	**Table 2.5**	N1 (DP1), N2 (DP2), N3 (DP2), Donor (DP8), C1 (DP6), C2 (DP6), C3 (DP3), C4 (DP7)
DP8-D416A-E469A-MsGlgB	**Appendix 2**	N1 (DP1), N2 (DP2), N3 (DP2), Donor (DP8), C1 (DP6), C2 (DP6), C3 (DP3), C4 (CP7)
ACR-E469A-MsGlgB	**Table 2.5**	N2 (DP3), A2 (ACR), C3 (ACR)
DP3-ACR-MsGlgB	**Table 2.5**	N2 (DP2), N3 (DP3*), A1 (ACR*), C2 (ACR)
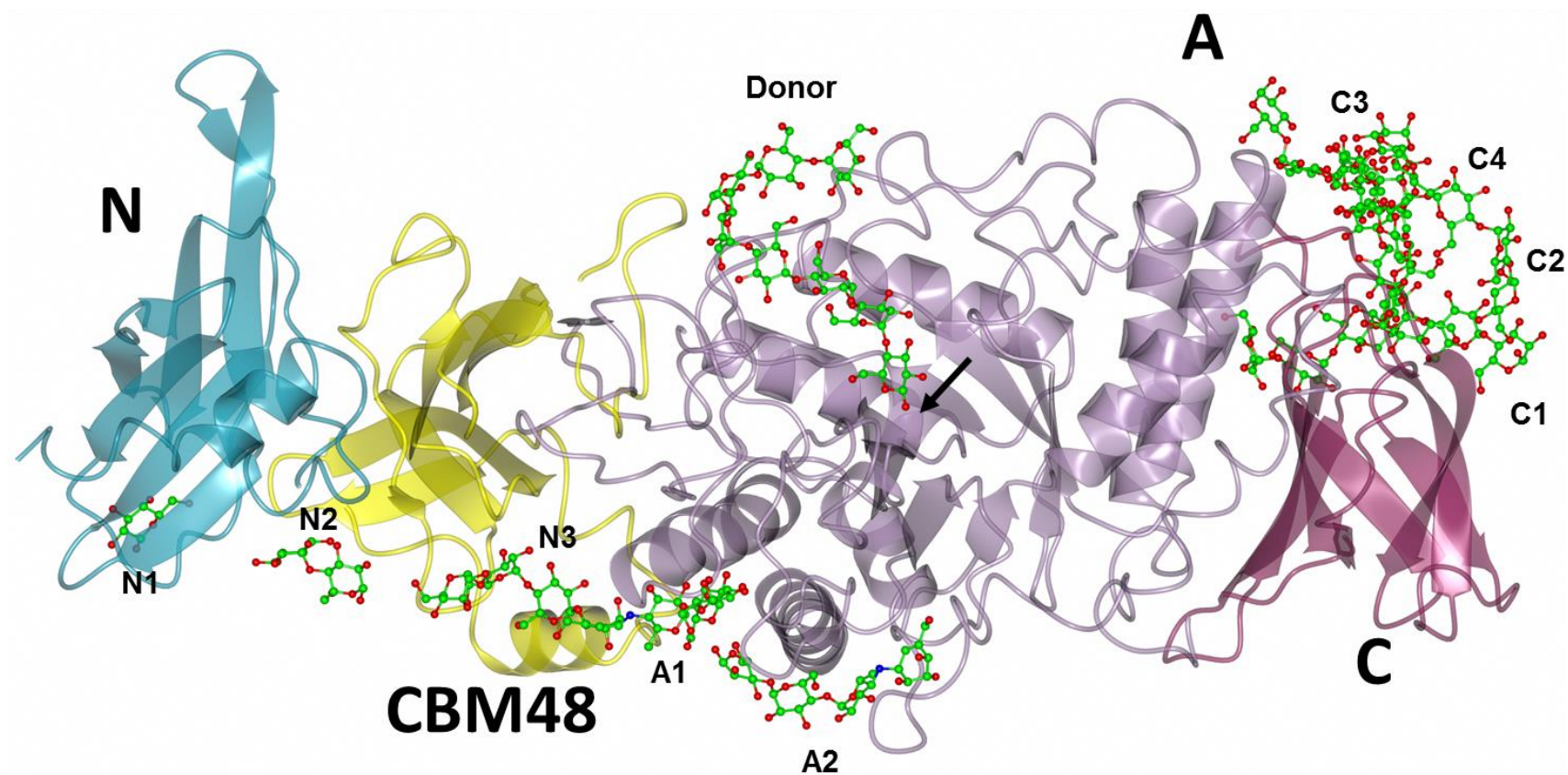
Figure 2.8 Overview of oligosaccharide binding in MsGlgB. Ribbon model of **Apo-WT-MsGlgB** superposed with all oligosaccharide binding sites observed in the crystal structures **DP8-D416A-MsGlgB**, **ACR-E469A-MsGlgB** and **DP3-ACR-E469A-MsGlgB** described in **Table 2.5** and the text below. Bindings sites comprise: N1 (1), N2 (2), N3 (3), A1 (4), A2 (4), Donor (8), C1 (6), C2 (6), C3 (3) and C4 (7), with the maximum number of glucose units or pseudoglucose units accommodated specified in parentheses. Colour scheme according to **Figure 2.6**.

2.4.1 Identification of surface binding sites in WT MsGlgB

MsGlgB WT protein crystals were subjected to soaks and/or co-crystallisations with various commercially available α-glucans, including maltose, maltooligosaccharides, maltodextrin, amyloses from different plant sources and cyclodextrins at concentrations of up to ~10 mM. In most cases these did not yield oligosaccharide-bound structures, with the exception of weak density C4 or N2 sites. Only by soaking crystals with much greater concentrations of maltooligosaccharides or maltodextrin (≥ 100 mM) could binding sites N1-N3 and C1-C4 be distinguished. The binding of oligosaccharides tended to improve the resolution of the structures and a structure with a tetrasaccharide bound at the C4 site was determined at 1.46 Å resolution, **Table 2.4,** the highest reported for any branching enzyme. However, no binding of oligosaccharides at or near the active site was observed in the WT protein and therefore I investigated whether this could be achieved using mutagenesis.

Table 2.4 X-ray data collection and refinement statistics for WT MsGlgB structures.
Values in parentheses are for the outer resolution shell; r.m.s., root mean square.

Dataset ID	Apo-WT-MsGlgB	DP4-WT-MsGlgB
Data Collection		
Beamline	I04-1	I03
Wavelength (Å)	0.9282	0.9763
Detector	Pilatus 6M	Pilatus3 6 M
Resolution range (Å)	84.49-1.72 (1.75-1.72)	54.10-1.46 (1.48-1.46)
Space group	P21212	P21212
a, b, c (Å)	98.49, 164.41, 54.52	97.36, 162.31, 53.75
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Total observations	912,423 (46,131)	1,949,958 (86,737)
Unique reflections	94,955 (4,676)	148,349 (7219)
Multiplicity	9.6 (9.9)	13.1 (12.0)
Mean $I/\sigma(I)$	9.6 (1.2)	16.3 (1.2)
Completeness (%)	100.0 (100.00)	100.0 (99.9)
Rmerge[a]	0.152 (1.846)	0.073 (2.015)
Rmeas[b]	0.161 (1.949)	0.076 (2.105)
$CC_{1/2}$[c]	0.998 (0.606)	1.000 (0.738)
Wilson *B* value (Å^2)	20.7	21.0
Refinement		
Resolution range (Å)	84.49-1.72 (1.75-1.72)	54.10-1.46 (1.48-1.46)
Reflections: working/free[d]	90,195/4,684	140,775/7,483
Final *Rwork*	0.164	0.125
Final *Rfree*[e]	0.186	0.160
Estimated coordinate error (Å)[f]	0.089	0.052
r.m.s bond distance deviation (Å)	0.0152	0.0103
r.m.s. bond angle deviation (°)	1.511	1.411
No. of protein residues per chain (ranges)	723 (9-375; 380-736)	723 (9-375; 380-736)
No. of heterogen residues: sugar/water/other[g]	0/759/9	4/737/9
Mean B factors: protein/ligands/water/overall (Å^2)	25/34/37/25	27/36/41/27
Ramachandran: favoured/allowed/disallowed[h] (%)	97.6/2.3/0.1	98.2/1.7/0.1
Protein Data Bank accession code	N/A	N/A

[a] $Rmerge = \sum_{hkl} \sum_i |Ii(hkl) - \langle I(hkl)\rangle|/ \sum_{hkl} \sum_i Ii(hkl)$.

[b] $Rmeas = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |Ii(hkl) - \langle I(hkl)\rangle|/ \sum_{hkl} \sum_i Ii(hkl)$, where Ii(hkl) is the ith observation of reflection hkl, $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl.

[c] $CC_{1/2}$ is the correlation coefficient between symmetry equivalent intensities from random halves of the data set.

[d] The data set was split into "working" and "free" sets consisting of 95 and 5% of the data, respectively. The free set was not used for refinement.

[e] The *R* factors *Rwork* and *Rfree* were calculated as follows: $R = \sum(| Fobs - Fcalc |)/\sum| Fobs |$ where *Fobs* and *Fcalc* are the observed and calculated structure factor amplitudes, respectively.

[f] Based on *Rfree* as calculated by REFMAC5.1

[g] "Other" includes glycerol and sodium ions

[h] As calculated using MolProbity.[2]

2.4.2 Crystallisation of MsGlgB proteins with mutated active site residues

One approach that has been successfully used to generate substrate-bound crystal structures of CAZymes is the mutation of active site residues. I investigated this by generating modified GlgB proteins in which either or both of the catalytic residues D416 and E469 were mutated to alanine. MsGlgB expression plasmids were modified using a QuikChange Lightning site-directed mutagenesis kit and proteins D416A-MsGlgB, E469A-MsGlgB and D416A-E469A-MsGlgB were produced as described previously, **Figure 2.9**. As would be expected, no enzymatic activity was detected for any of the modified proteins. In addition, no negative effects on protein production were observed as a result of these changes. In fact, modification of D416 increased protein thermostability, as determined by nano differential scanning fluorimetry (DSF), for which differences in melting temperatures (T_m) $\geq$ 0.5 °C are considered significant. T_ms were 65.0 °C for D416A-MsGlgB and 64.4 °C for D416A-E469A-MsGlgB, compared with 63.9 °C for the WT protein and 62.9 °C for E469A-MsGlgB. The proteins were crystallised, X-ray diffraction data were recorded and models were refined as previously, **Figure 2.9** and **Appendix 2.** The electron density maps were consistent with the expected amino acid substitutions. Overall the structures were very similar to that of the WT protein, apart from a slight rearrangement of residues adjacent to E469A in E469A-MsGlgB and D416A-E469A-MsGlgB**.**
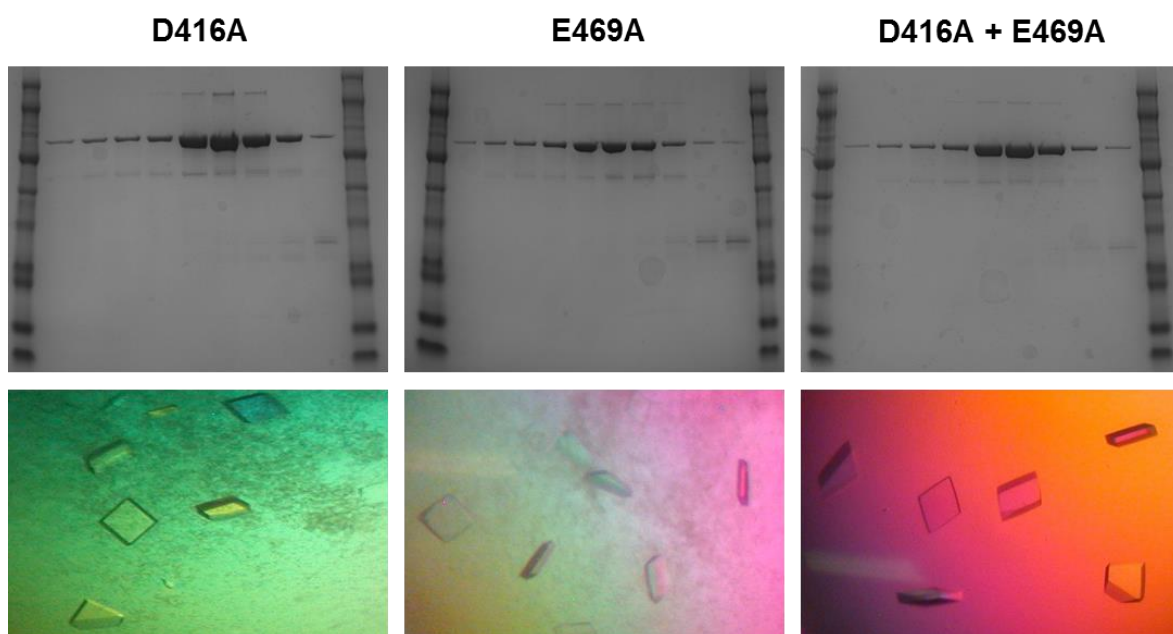
Figure 2.9 Production and crystallisation of MsGlgB proteins with mutated active site residues. Proteins were produced and crystallised as described above. Analysis of fractions by SDS-PAGE and the corresponding protein crystals imaged with polarised light are shown for each MsGlgB protein.

2.4.3 Attempts to trap a covalent intermediate with 2-deoxy-2-fluoro maltosyl fluoride oligosaccharides

Having solved the structures of GlgB proteins with the activity of either or both of the catalytic residues abolished, I firstly attempted to use E469A-MsGlgB and a fluorinated substrate mimic to obtain structures with oligosaccharides covalently bound at the active site. This was achieved previously for *S. coelicolor* GlgE, whereby a covalent intermediate was trapped using 2-deoxy-2-fluoro-α-maltosyl fluoride and a modified GlgE protein with the acid/base residue substituted for Ala, equivalent to E469A-MsGlgB.[18] With this substrate, the electron withdrawing character of the 2-fluoro substituent destabilises the oxocarbenium transition states in the α-retaining double displacement mechanism, **Figure 1.2**. In the formation of the covalent bond with the nucleophilic residue, this is partly compensated for by the presence of the anomeric fluoride, which is a better leaving group relative to hydroxyl in the normal glucan substrate. However, formation of the α-1,4-glycosidic bond in the second step is much slower, impeding the overall reaction mechanism. With WT GlgE 2-deoxy-2-fluoro-α-maltosyl fluoride served as a slow substrate that could be used to extend maltotetraose.[13] The additional loss of the acid/base catalysis in the modified GlgE protein further slows the formation of the α-1,4-glycosidic bond, effectively trapping the covalent intermediate, which was then detected by X-ray crystallography, **Figure 2.10A**.[18]

Initially, I soaked crystals of E469A-MsGlgB with high concentrations of 2-deoxy-2-fluoro-α-maltosyl fluoride, but this did not result in any detectable ligand binding in the active site or anywhere else in the protein. Previous studies have suggested that maltooligosaccharides ≥ DP4 are required to detect any binding to branching enzymes.[8, 11] I hypothesised that extended 2-deoxy-2-fluoro-α-maltosyl fluorides could provide the additional binding interactions that may be needed to trap a covalent enzyme intermediate in E469A-MsGlgB. I then tested whether 2-deoxy-2-fluoro-α-maltosyl fluoride could be extended by *S. coelicolor* GlgE in the absence of a DP4 acceptor. MALDI spectra of the reaction products demonstrated that this was the case and peaks had an m/z value of DPn + (n + 2), where n is an even integer up to DP24, **Figure 2.10B**. This is consistent with a 2-deoxy-2-fluoro substitution on alternate glucose residues and the retention of the reducing end fluoride, which might have otherwise been hydrolysed. After removal of GlgE by precipitation and filtration, the product mixture was used for E469A-MsGlgB crystal soaks. However, as before, no additional density was observed in the crystal structures, and therefore this approach to generate active site bound GlgB structures was not pursued further.

A

B
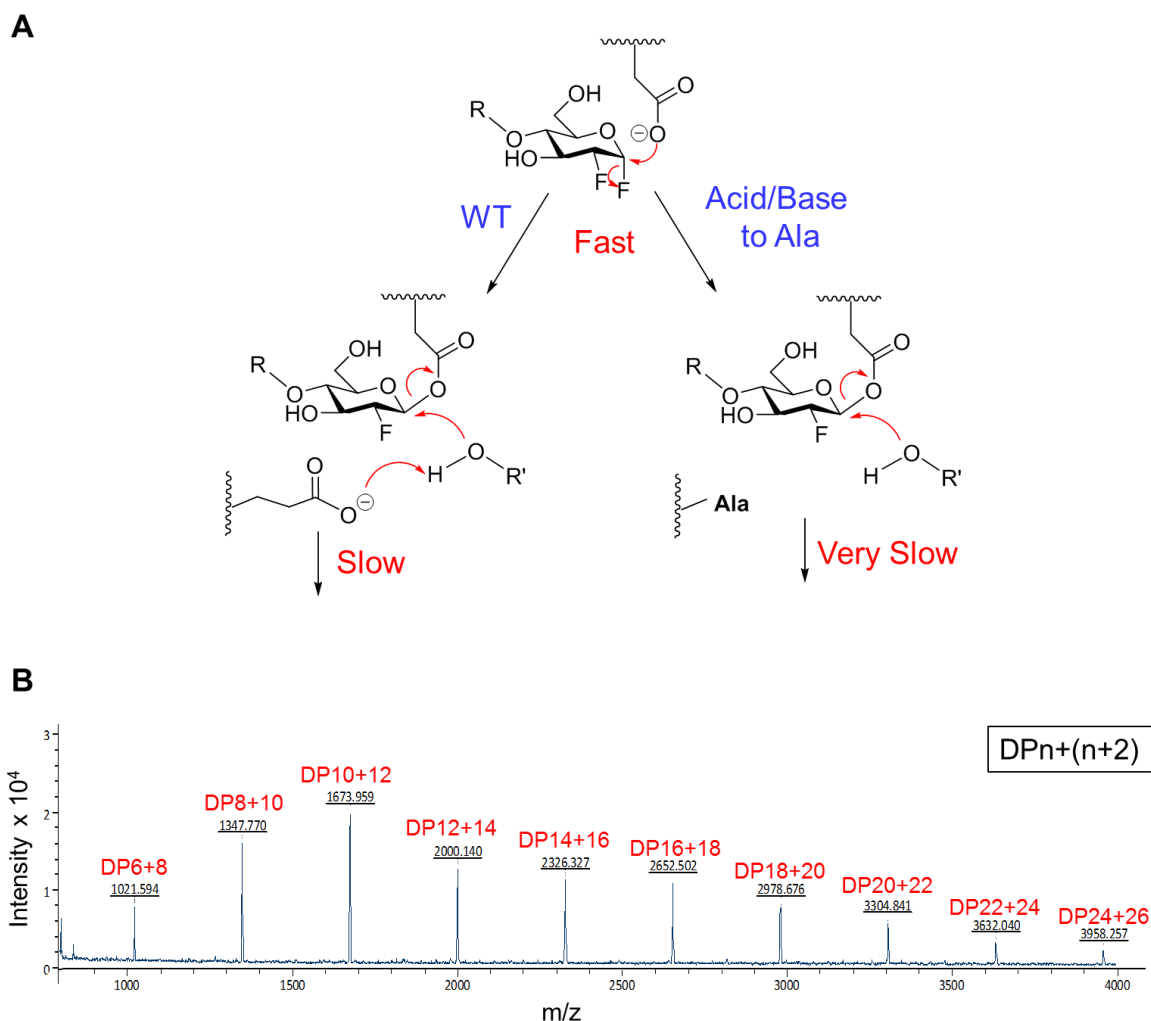
Figure 2.10 Using 2-deoxy-2-fluoro maltosyl fluoride oligosaccharides to trap covalent intermediates. A) Reaction of 2-deoxy-2-fluoro maltosyl fluoride moiety in the active site of an GH13 α-retaining enzymes with and without mutation of acid/base residue to alanine. **B**) MALDI mass spectrum of products from the incubation of 2-deoxy-2-fluoro-α-maltosyl fluoride with WT *S. coelicolor* GlgE. The m/z values are increased by two for each fluorine atom, with the general formulae of m/z = DPn + (n + 2), indicating that the reducing end fluorine atom was retained. Crystal soaks with this mixture did not yield substrate-bound structures.

2.4.4 Identification of donor site with maltooctaose-bound D416A-MsGlgB

I subsequently carried out crystal soaks of D416A-MsGlgB, E469A-MsGlgB and D416A-E469A-MsGlgB with high concentrations of maltooctaose and maltodextrin. For D416A-MsGlgB and D416A-E469A-MsGlgB maltooligosaccharides were observed bound in the active cleft, but not in E469A-MsGlgB. Maltooligosaccharide binding at the N1-N3 and C1-C4 sites was observed in all of the modified proteins, indicating that, under these conditions, the presence of the nucleophilic residue D416 was the main barrier to binding at the active site. The best data sets were processed to give **DP8-D416A-MsGlgB** and **DP8-D416A-**

E469A-MsGlgB, both with maltooctaose-bound at the active site, at resolutions of 1.88 Å and 1.42 Å, respectively, **Table 2.5**. Comparison with other GH13 enzyme structures clearly indicated that maltooctaose occupied the donor binding site.

Interestingly, the reducing end pyranose occupying donor subsite -1 had a different configuration in each structure. In **DP8-D416A-MsGlgB** this pyranose has a β-configuration, with O1 positioned towards the D416A residue, consistent with the orientation in the covalent intermediate present in the α-retaining mechanism. The O1 forms a hydrogen bonding interaction with the acid/base catalytic residue E469 as would be expected from the branching enzyme catalytic mechanism. Conversely, in the **DP8-D416A-E469A-MsGlgB** structure the reducing end pyranose has an α-configuration. The rearrangement around E469A, described above for the apo-E469A-MsGlgB structure, allows E470 to coordinate to the α-O1, **Figure 2.11**. Since this is unlikely to be relevant to the catalytic mechanism, only **DP8-D416A-MsGlgB** was used in further analyses of ligand binding.
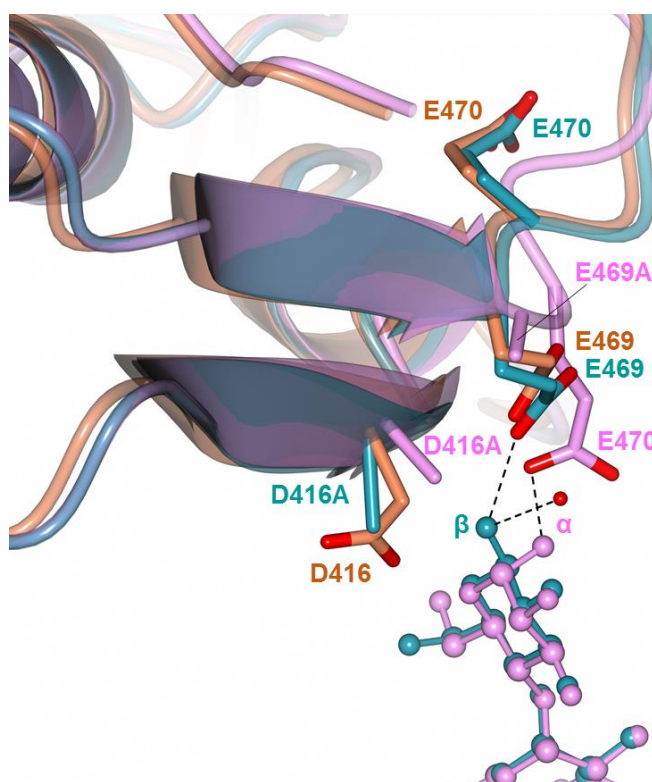
Figure 2.11 Alternative configurations of the reducing end pyranose observed in DP8-D416A-MsGlgB and DP8-D416A-E469A-MsGlgB. Superposition of **Apo-WT-MsGlgB**, (coral, no ligand-bound) **DP8-D416A-MsGlgB** (cyan, β-maltooctaose-bound) and **DP8-D416A-E469A-MsGlgB** (pink, α-maltooctaose-bound) structures, showing the reducing end pyranose in ball and stick form and the main chain of residues 400-420 and 460-490 as a ribbon model. Rearrangement of residues E469A - A471 in **DP8-D416A-E469A-MsGlgB** allows E470 to coordinate to α-O1, whereas in **DP8-D416A-MsGlgB** E469 coordinates to β-O1.

The electron density of the oligosaccharide ligand was clearly resolved, so I could unambiguously assign donor subsites -1 to -8, **Figure 2.12**. In total, 17 amino acid residues coordinate to maltooctaose in **DP8-D416A-MsGlgB** via hydrogen bonding or van der Waals interactions, **Figure 2.13**. The bound maltooctaose adopts a twisted "S" conformation, as seen previously in *Chlamydomonas reinhardtii* isoamyase (ISA) (CrISA; pdb code 4OKD)[19] and *Cyanothece* sp. ATCC 51142 BE1 (CceBE1; pdb code 5GQX), the only other known branching enzyme structure with a maltooligosaccharide bound at the active site.[10] In the latter structure, an unusual 91 ° rotation of the reducing end pyranose was observed. On the other hand, I observed an orientation of this ring that corresponds to the covalently bound glycosyl enzyme intermediate of the *Cr*ISA, **Figure 2.12**, suggesting that this structure reflects the orientation of the sugar in the covalent intermediate. Accordingly, the distance between the C_α of D416A and the C1 of pyranose is 4.42 Å, compared to 4.59 Å in the covalently bound *Cr*ISA structure.

In the -1 subsite, there are hydrogen bonding interactions with His351 (O6), His536 (O3), Asp337 (O2 & O3) and Arg414 coordinates O1 & O2 via a water molecule. These residues are highly conserved across GH13 family enzymes.[20] A further CH/π stacking interaction with Tyr311 coordinates this reducing-end pyranose. Subsites -2 to -6 are mainly conserved between **DP8-D416A-MsGlgB** and CceBE1. However, in CceBE1 the pyranose in subsite -7 is rotated at a slightly different angle. Whilst both pyranoses in the -7 subsites are coordinated to a C-α NH via O2, that of CceBE1 is additionally coordinated by Arg216 side chain (equivalent to Pro227 in **DP8-D416A-MsGlgB**), Asp324 (Gly306 in **DP8-D416A-MsGlgB**) and Val282 C=O, in an extra loop not found in **DP8-D416A-MsGlgB**. The pyranose in the MsGlgB -8 subsite is coordinated by Glu591 and Glu592, residues that are not conserved in CceBE1. The O4 of pyranose in the -8 subsite points into solvent, meaning that it is possible that oligosaccharides longer than DP8 can extend from the MsGlgB donor site.
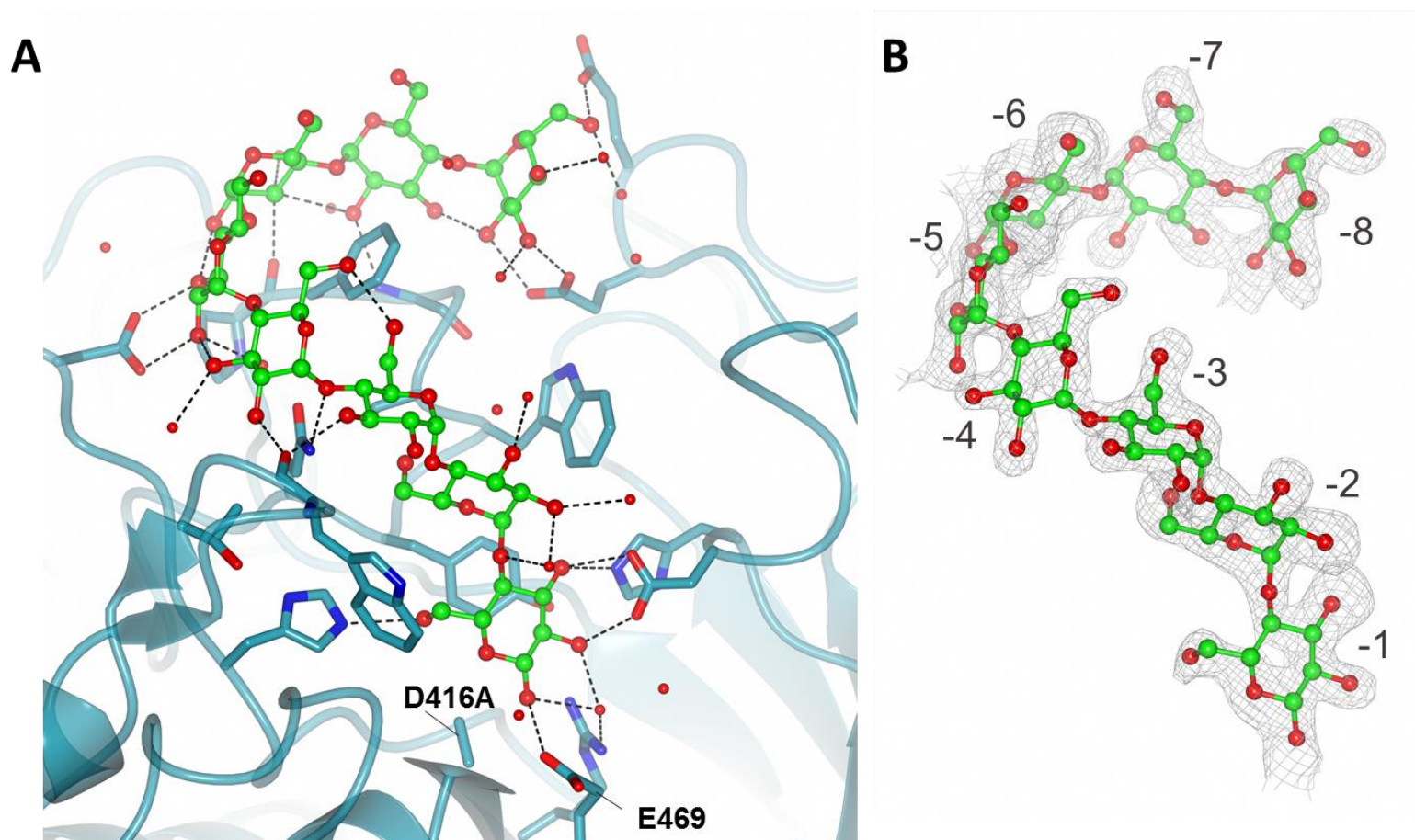
Figure 2.12 Maltooctaose-bound in the donor site of D416A-MsGlgB. A) Maltooctaose-bound in the donor site of MsGlgB showing hydrogen bonding and coordinating residues, also shown in **Figure 2.13**. Protein shown in cyan, maltooligosaccharide carbons in green, oxygen atoms in red and nitrogen atoms in blue. Catalytic residues D416A and E469A are indicated. **B)** The omit electron density map (1.88-Å resolution; contoured at ~3.0σ) indicating the binding of β-maltooctaose. Donor subsites are labelled -1 to -8.
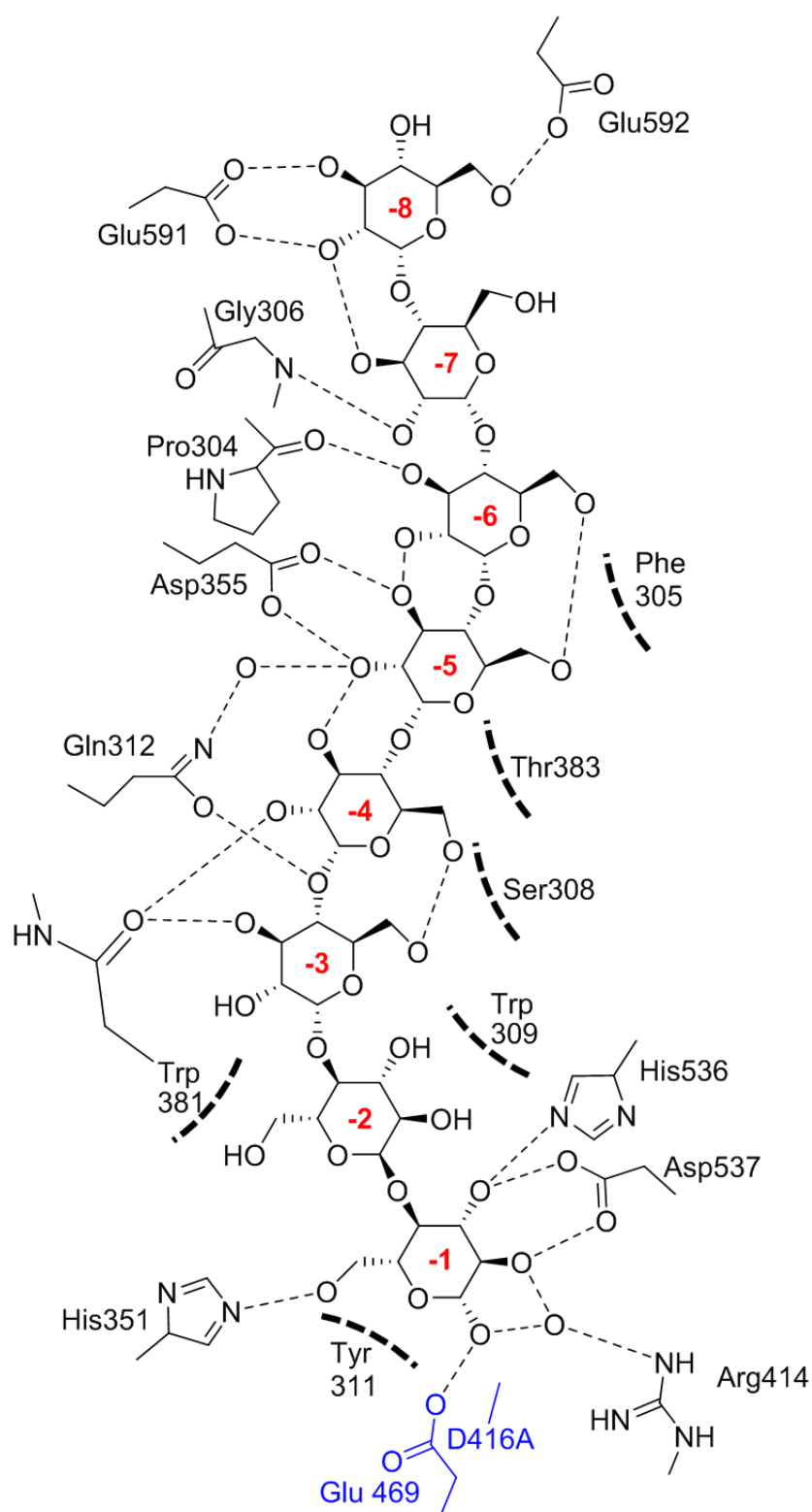
Figure 2.13 Interactions between maltooctaose and D416A-MsGlgB. 2D plot of maltooctaose-bound in donor site in a "twisted-S" conformation and protein residue side chains. Hydrogen bonds are shown by dashed lines and dashed arc indicate hydrophobic interactions. Catalytic residues are shown in blue and donor subsites labelled in red.
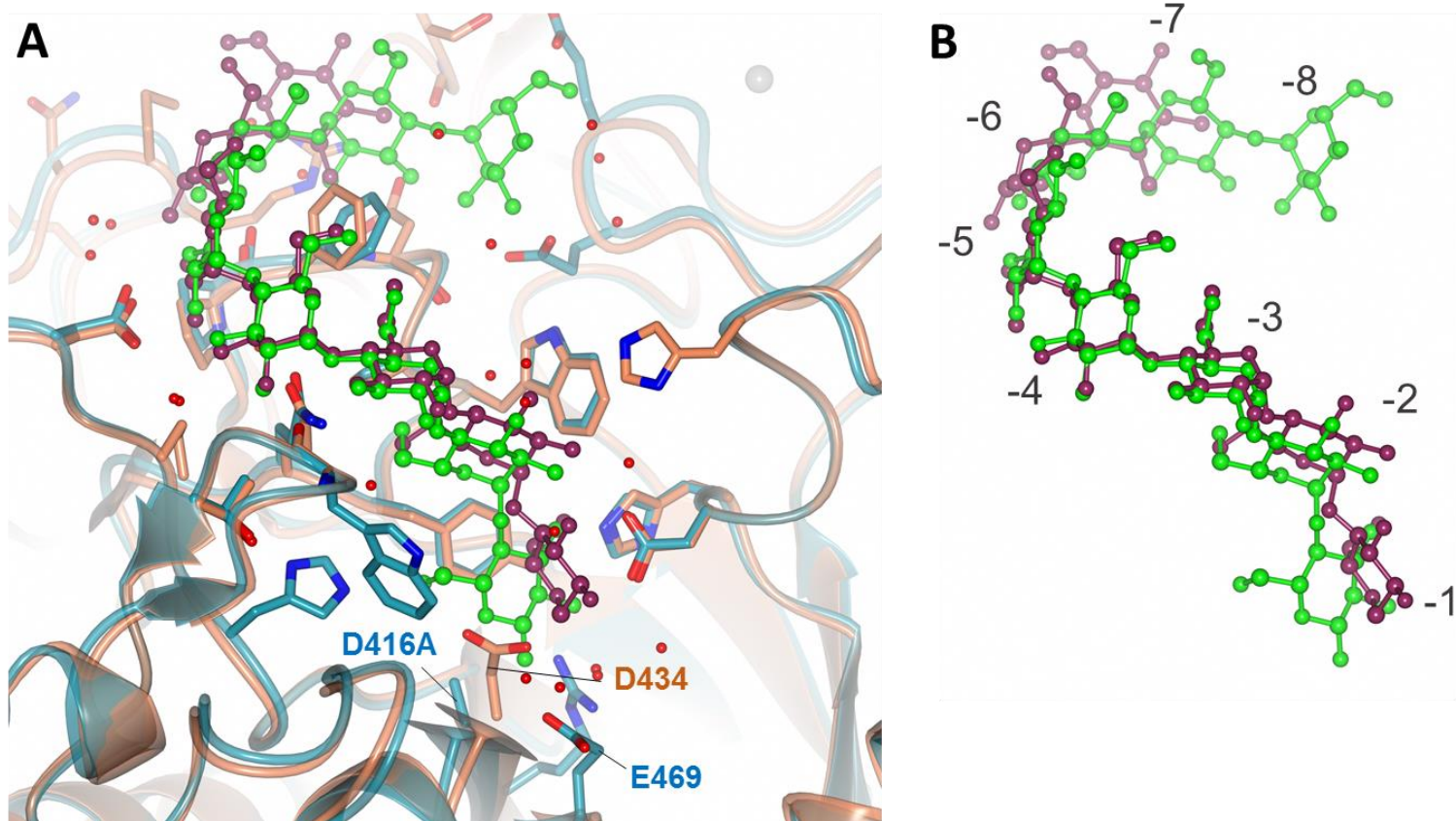
Figure 2.14 MsGlgB and Cce BE1 donor sites are similar. A) Superposition of **DP8-D416A-MsGlgB** and DP7-CceBE1 donor binding sites, showing interacting protein side chains. For **DP8-D416A-MsGlgB** protein is shown in cyan and the maltooligosaccharide in green. For DP7-CceBE1 the protein is shown in coral and the maltooligosaccharide is shown in purple. Catalytic residues are labelled in blue for **DP8-D416A-MsGlgB** and orange for DP7-CceBE1. **B)** Superposition of maltooligosaccharides only, with MsGlgB donor subsites labelled.

The structures of **Apo-D416A-MsGlgB** and **DP8-D416A-MsGlgB** were then directly compared, to ascertain whether maltooligosaccharide binding at the donor site is facilitated by conformational changes in the protein. In **DP8-D416A-MsGlgB** there was a slight shift in the main chain between residues 350-355, with His351 NH moving 2.2 Å to accommodate the reducing end pyranose and allow H-bonding to O6, which is precluded by proximity in the apo structure. Rotation of Gln312 was also observed, with Oδ moving 3.13 Å to allow a H-bonding interaction between it and the O4 of the glycosidic bond between pyranoses in the -3 and -4 subsites. In **Apo-D416A-MsGlgB** Phe305 has an alternate conformation that overlaps with the pyranose in the -6 subsite; this is not seen in the ligand-bound structure. Additionally, Glu591 shifts in **DP8-D416A-MsGlgB** to accommodate the pyranose in the -8 subsite and allow H-bonding to O2 and O3, with Oδ moving 2.85 Å.

2.4.5 Acarbose soaks reveal additional binding sites in the catalytic domain

Having assigned the donor site of MsGlgB, my next objective was to identify the acceptor site. To this end, crystals of D416A-MsGlgB, E469A-MsGlgB and D416A-E469A-MsGlgB were soaked with α-glucans that contained α-1,6 bonds, including pullulan, β-limit dextrin, rabbit liver glycogen, amylopectin and maltotriosyl maltotriose. Mostly, very weak or no additional density was detected, though for the latter soak, maltotriose was observed in the donor subsites -1 to -3 in D416A-MsGlgB and D416A-E469A-MsGlgB. I subsequently investigated the binding of the pseudotetrasaccharide acarbose, a widely-used inhibitor of α-amylase enzymes,[21] even though this had previously been shown not to inhibit MtGlgB.[9] Surprisingly, this revealed a new binding site in the catalytic domain of E469A-MsGlgB, site A2. In D416A-MsGlgB and D416A-E469A-MsGlgB, acarbose instead bound at the donor site, suggesting that when binding is not obstructed by the presence of nucleophile D416, this is the preferred binding site. In **ACR-E469A-MsGlgB** O4 at the non-reducing end of acarbose was located 19.5 Å from the D416 Oδ, in an orientation that suggested the A2 site could be a distal part of the main acceptor site, **Figure 2.15**.

In an attempt to extend binding towards the catalytic residues I soaked E469A-MsGlgB crystals with acarbose that had been extended by six glucose units at the non-reducing end (DP6-ACR). However, this molecule was instead orientated such that the acarbose moiety was rotated ~180° and located in the adjacent A1 site, with three of the glucose units bound at the N3 site. Superposition of **ACR-E469A-MsGlgB** and **DP3-ACR-E469A-MsGlgB** suggested that these binding sites could represent an α-1,6 branchpoint between the O1 of acarbose in the A2 site and an O6 of acarbose in the A1 site, **Figure 2.15**. The O1 and the carbon of the O6 were 3.2 Å apart.

Table 2.5 X-ray data collection and refinement statistics for MsGlgB complexes
Values in parentheses are for the outer resolution shell, r.m.s., root mean square

Dataset ID	DP8-D416A-MsGlgB	ACR-E469A-MsGlgB	DP3-ACR-MsGlgB
Data Collection			
Beamline	I04-1	I03	I04
Wavelength (Å)	0.9282	0.9762	0.9795
Detector	Pilatus 6M	Pilatus3 6 M	Pilatus 6M
Resolution range (Å)	84.7-1.88 (1.92-1.88)	48.50-1.60 (1.63-1.60)	62.28-1.45 (1.47-1.45)
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
a, b, c (Å)	98.79, 164.57, 54.42	97.00, 162.04, 53.64	97.67, 161.69, 53.79
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Total observations	972,120 (60,979)	729,137 (37,025)	2,005,551 (97,080)
Unique reflections	73,128 (4,408)	111,249 (5,452)	151,467 (7,404)
Multiplicity	13.3 (13.8)	6.6 (6.8)	13.2 (13.1)
Mean $I/\sigma(I)$	11.3 (1.5)	9.7 (1.3)	16.8 (1.4)
Completeness (%)	100.0 (100.0)	99.2 (98.8)	100.0 (99.9)
R_{merge}[a]	0.184 (1.851)	0.095 (1.688)	0.082 (1.779)
R_{meas}[b]	0.192 (1.922)	0.104 (1.828)	0.086 (1.874)
$CC_{1/2}$[c]	0.998 (0.753)	0.998 (0.463)	1.000 (0.685)
Wilson B value (Å^2)	25.0	22.0	25.0
Refinement			
Resolution range (Å)	84.7-1.88 (1.92-1.88)	48.50-1.60 (1.63-1.60)	62.28-1.45 (1.47-1.45)
Reflections: working/free[d]	69,343/3,715	105,686/5501	143,858/7,518
Final R_{work}	0.173	0.176	0.124
Final R_{free}[e]	0.206	0.192	0.161
Estimated coordinate error (Å)[f]	0.125	0.078	0.054
r.m.s bond distance deviation (Å)	0.0102	0.0102	0.0099
r.m.s. bond angle deviation (°)	1.393	1.394	1.484
No. of protein residues per chain (ranges)	722 (9-94; 99-736)	722 (9-94; 101-736)	722 (9-94; 100-736)
No. of heterogen residues: sugar/water/other[g]	35/636/6	13/411/15	17/901/18
Mean B factors: protein/ligands/water/overall (Å^2)	13/42/40/30	32/41/40/32	23/35/39/23
Ramachandran: favoured/allowed/disallowed[h] (%)	98.5/1.4/0.1	97.8/2.1/0.1	98.0/1.9/0.1
Protein Data Bank accession code	N/A	N/A	N/A

[a] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

[b] $R_{meas} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl, $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl.

[c] $CC_{1/2}$ is the correlation coefficient between symmetry equivalent intensities from random halves of the data set.

[d] The data set was split into "working" and "free" sets consisting of 95 and 5% of the data, respectively. The free set was not used for refinement.

[e] The R factors R_{work} and R_{free} were calculated as follows: $R = \sum(|F_{obs} - F_{calc}|)/\sum|F_{obs}|$ where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[f] Based on R_{free} as calculated by REFMAC5.[16]

[g] Sugars includes acarbose as four residues. "Other" includes glycerol, imidazole, sodium, calcium, nickel and chloride ions.

[h] As calculated using MolProbity.[22]
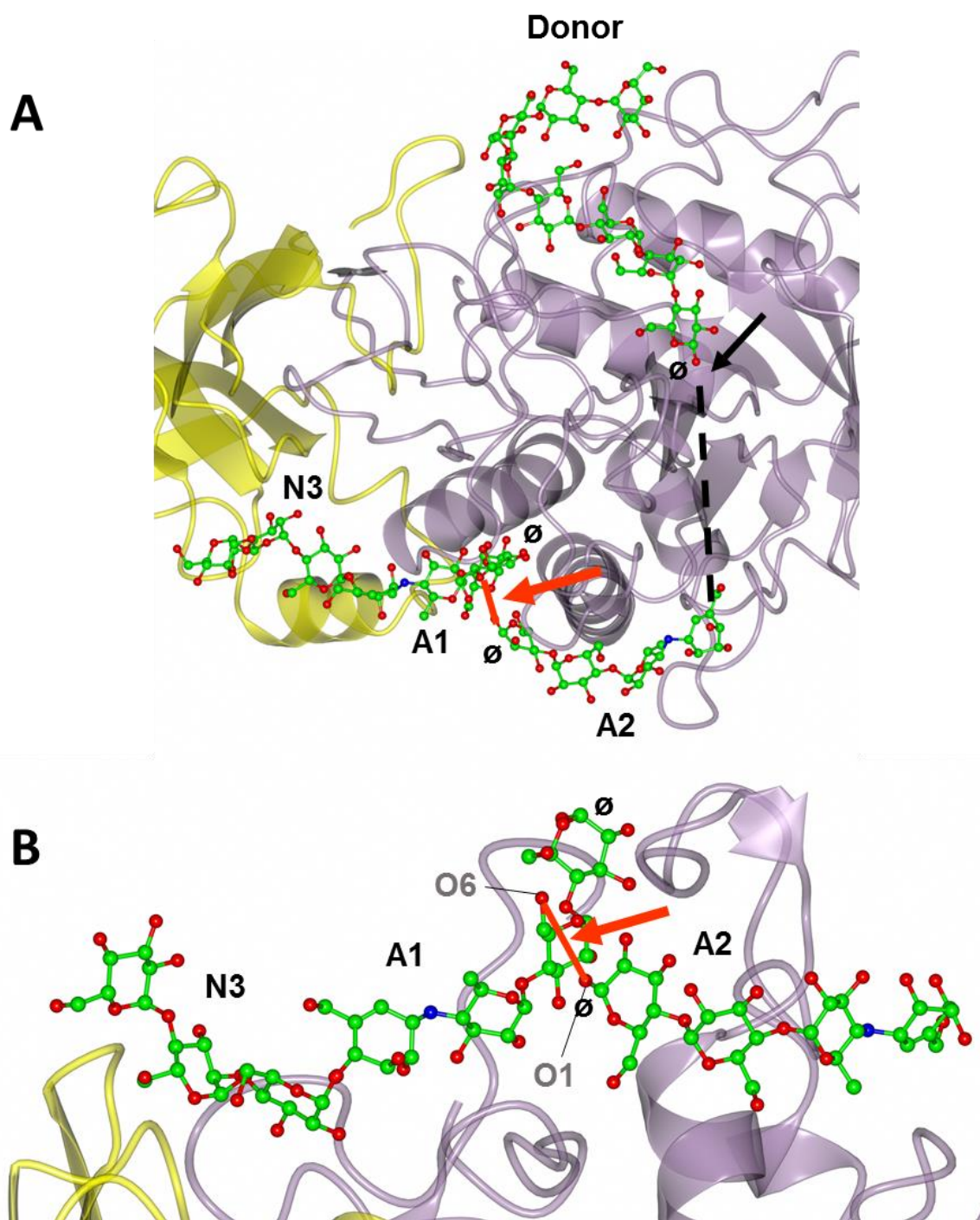
Figure 2.15 Binding at the A1 and A2 sites suggests coordination of an α-1,6-branchpoint. A) Superposition of **DP8-6416A-MsGlgB**, **ACR-E469A-MsGlgB**, **DP3-E469A-MsGlgB** and structures showing binding of maltooctaose at the donor site, acarbose at the A2 site and DP3-acarbose at the N3 and A1 sites. Reducing ends of the oligosaccharaides are indicated by ø. A black dashed line indicates the distance of 19.5 Å between the non-reducing end of acarbose in A2 and the D416 catalytic residue, which is marked by a black arrow. An orange solid line shows the putative α-1,6 branchpoint. **B)** A closer view of the N3, A1 and A2 sites rotated ~180° relative to **A**. The O1 and the carbon of the O6 were 3.2 Å apart. Protein and oligosaccharides are coloured according to **Figure 2.8**.

2.4.6 Residues 370-383 form a flexible loop that is involved in substrate binding

Comparison of crystal structures with oligosaccharides bound at different sites revealed that residues 370-383 formed a loop that could adopt different conformations, **Figure 2.16**. In apo structures, this loop is disordered, with no or very weak electron density detected. When an oligosaccharide is bound in the donor site, such as in **DP8-D416A-MsGlgB**, the loop adopts a fixed conformation that can be clearly resolved for most residues. This allows coordination of Trp381 to the oligosaccharide, both via CH/π stacking between the tryptophan side chain and pyranose in the -2 subsite, as well as hydrogen bonding of the Trp381 backbone carbonyl to the -3 pyranose O3 and the -4 pyranose O2. Conversely, when acarbose is bound at the A2 site in **ACR-E469A-MsGlgB**, I observed an alternative conformation of this loop, which instead allows coordination of His370 and Arg375 to acarbose. Surprisingly, in this latter conformation, the loop occupies the same space as the maltooligosaccharide units bound at the -3 and -4 subsites in **DP8-D416A-MsGlgB**, indicating that this would obstruct binding at the donor site.
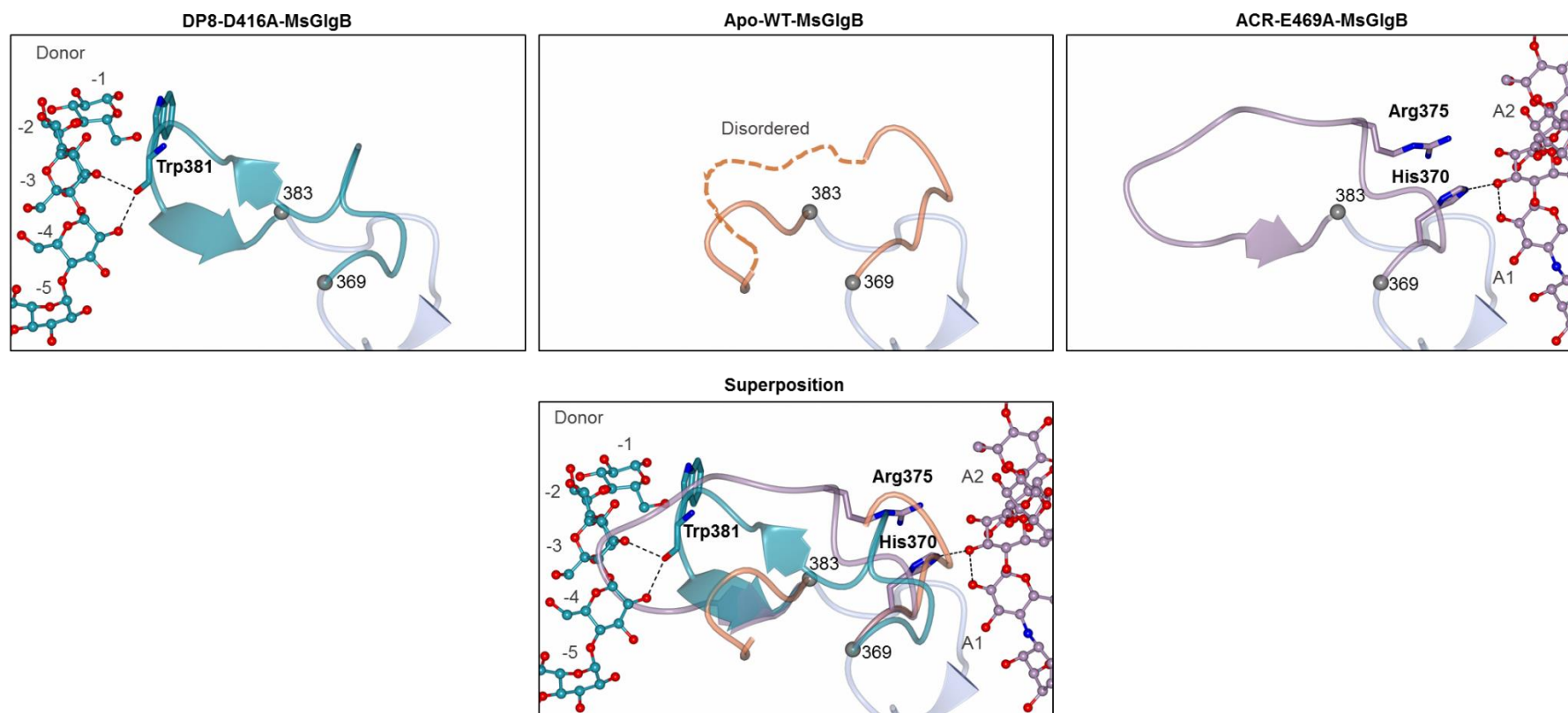
Figure 2.16 Residues 370-383 make up a flexible loop region that adopts distinct conformations in oligosaccharide-bound structures. The top three panels show the same view of residues 364-392 and bound oligosaccharides in **DP8-D416A-MsGlgB** (cyan), **Apo-WT-MsGlgB** (orange) and **ACR-E469A-MsGlgB** (lilac). The hinge points of the loop at residues 369 and 383 indicated by grey spheres and residues 364-369 and 383-392 are shown in light blue. In **Apo-WT-MsGlgB** residues 374-381 are disordered and represented by a dashed orange line. In **DP8-D416A-MsGlgB** the loop adopts the conformation shown in cyan that allows coordination of the oligosaccharide in the donor site and Trp381. In **ACR-E469A-MsGlgB** the loop adopts the conformation shown in lilac that allows coordination of His370 and Arg375 to acarbose. The lower panel shows a superposition of the top three panels.

2.4.7 There is extensive maltooligosaccharide binding in domain C of MsGlgB

Examination of domain C across all crystal structures demonstrated that of the four distinct binding sites, C1 can accommodate an oligosaccharide up to DP6, C2 DP7, C3 DP3 and C4 DP6. Although these sites are adjacent to each other, the orientation of the sugars and the positions of the atoms do not allow for the possibility of α-1,6 linkages between the oligosaccharides in any pairs of sites, **Figure 2.17**. This pattern of binding sites may only make sense in the context of a much larger α-glucan polysaccharide. Under the crystallography conditions, oligosaccharides were most commonly bound at the C4 site and oligosaccharides here were most consistently and clearly resolved. However, this may simply be the most accessible site in the crystal structure and this does not necessarily reflect the affinity of this site for maltooligosaccharide binding *in vivo*.
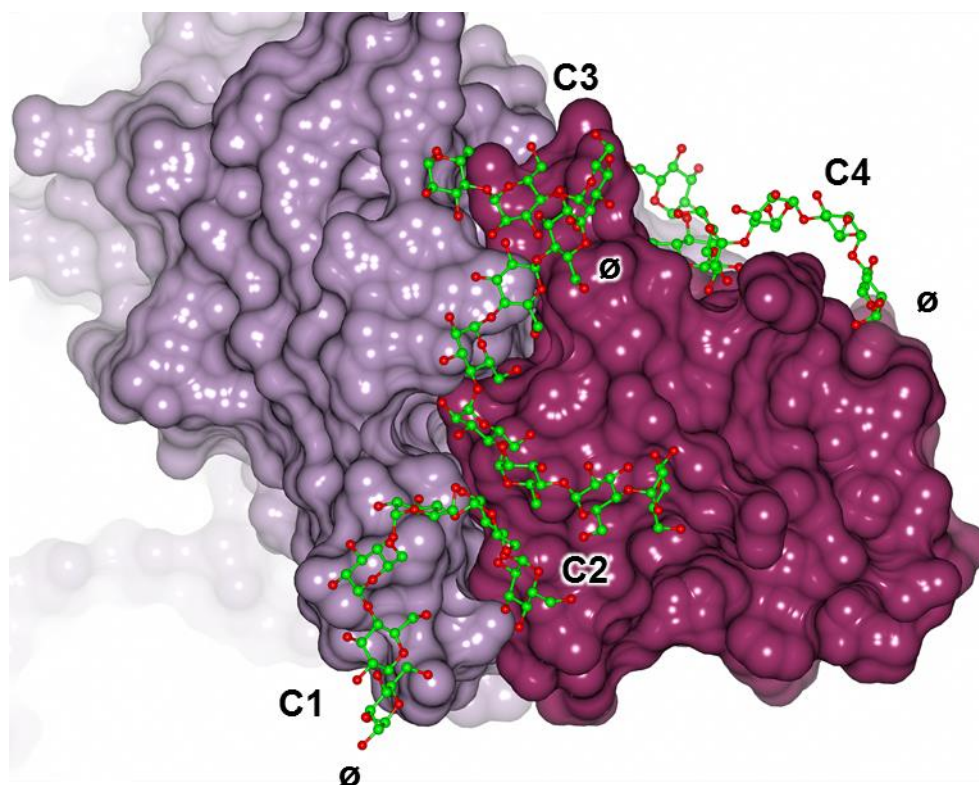
Figure 2.17 C-terminal surface binding sites. Maltooligosaccharide binding at the interface of domain A and C reveals four distinct sites, C1- C4, accommodating maltooligosaccharides of DP 6, 7, 3 and 6 respectively. Domains are coloured according to **Figure 2.6**, with the view along the longitudinal axis of the protein. The reducing ends of the oligosaccharides are indicated by **Ø**.

2.4.8 Amino acid residues that coordinate oligosaccharides are largely conserved between MsGlgB and MtGlgB

Having identified numerous amino acid residues that interact with oligosaccharides in MsGlgB, I investigated the conservation of these residues in MtGlgB, **Figure 2.18**. As would be expected from the 82% sequence identity, residues that interacted with carbohydrates were generally highly conserved. Interacting amino acid residues in donor subsites -1 to -6 were completely conserved. In the -7 subsite, glycine (Gly306 MsGlgB) was changed to alanine (Ala301 MtGlgB), however as this interaction with the O2 of this pyranose is via the backbone NH, this is likely to be unchanged. Likewise, there is a glutamate (Glu592 MsGlgB) to glutamine (Gln587 MtGlgB) change in the -8 subsite, but this is again unlikely to change the interaction with the side chain carbonyl and O6 of the final pyranose. The flexible loop region was conserved between the two sequences, apart from a change of arginine (Arg374 MsGlgB) to lysine (Lys369 MtGlgB), which is unlikely to have much effect as these residues are very similar. Overall, this analysis suggests that, with the exception of some minor modifications, all of the identified binding sites are conserved between the *M. smegmatis* and *M. tuberculosis* branching enzymes.

This work represents a significant step forward in the understanding of how a mycobacterial branching enzyme binds to its α-glucan substrate. The wider implications of these results are discussed below.
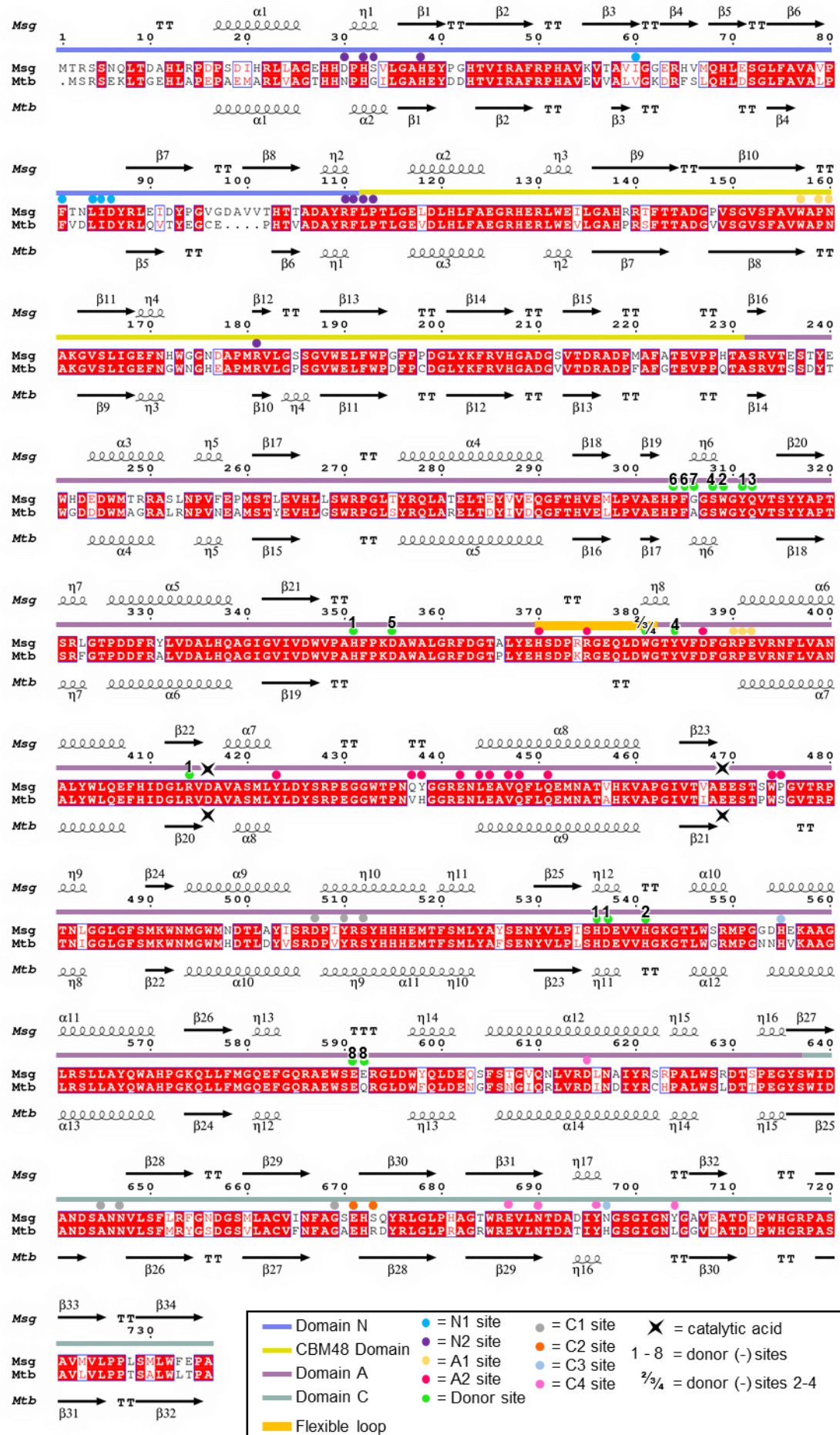
Figure 2.18 Alignment of MtGlgB and MsGlgB amino acid sequences. Numbering refers to the MsGlgB sequence. Secondary structural features are indicated with consecutive numbering. Domains are indicated with coloured bars and residues interacting with oligosaccharides are marked with coloured circles. Conserved residues are boxed with identity shown with a red background and similarity with red lettering. The alignment was generated with CLUSTALW2[23] and displayed with ESPript.[24]

2.5   Discussion

The overarching aim of this chapter was to develop the understanding of the action of the branching enzyme GlgB in the synthesis of mycobacterial α-glucan. This was achieved by both the investigation of GlgB activity *in vitro* and the generation of oligosaccharide-bound crystal structures.

Firstly, I investigated the *in vitro* activity of GlgB enzymes from mycobacteria and other actinomycetes by generating synthetic α-glucan with GlgE and different GlgB enzymes, then determining the average branch lengths, which ranged from DP 7.2 ± 2.2 to 8.6 ± 2.5. This could be directly compared to α-glucan generated *in vivo*, that had previously been extracted from *M. tuberculosis* and *S. venezuelae* cells, which had average branch lengths of DP 7.0 ± 2.0 – 7.3 ± 1.8. Therefore, whilst the synthetic α-glucan tended to have slightly higher average branch lengths, they were overall very similar and clearly distinct from the longer branch lengths seen in *E. coli* and mammalian glycogen, of ~DP11. My results confirmed that GlgE and GlgB are necessary and sufficient to generate the α-glucan polysaccharide. This work is part of a suite of experimental analyses that determine the fine detail of the actinomycete α-glucan structure, **Appendix 1**.[4]

It is noteworthy that I determined very similar average branch lengths by MALDI mass spectrometry and CE. The latter method is considered the gold standard for quantitative analysis of oligosaccharides,[25] but requires lengthy derivatisation steps with expensive reagents and typically 30-60 min to analyse each sample. In contrast, MALDI mass spectrometry is a rapid method, that does not require derivatisation, but has been reported to be not suitable for quantitative analyses without internal standards that have been properly calibrated.[26] Furthermore, MALDI mass spectrometry allows the direct identification of the DP, whereas with CE, the identity of peaks is inferred by comparison to standards. This work supports the use of MALDI mass spectrometry for the analysis of the proportions of linear α-1,4 oligosaccharides and suggests that, at least in this DP range, the method is more quantifiable than previously thought.

The *in vitro* characterisation of GlgB enzymes led directly on to the structural investigation of the mycobacterial GlgB enzyme, to understand the structural basis for the shorter branch lengths created by actinomycete GlgB enzymes relative to other bacterial and eukaryotic homologues. The crystallisation of MtGlgB proved challenging, despite the existence of a previously solved structure.[9] Crystallisation could only be achieved by the removal of the first nine amino acid residues, however these crystals diffracted poorly and the structure could not be solved. Fortuitously, MsGlgB crystallised without the need for modifications of the recombinant protein and the structure was solved at high-resolution. Importantly, the 3D structures of MsGlgB and MtGlgB were highly similar, with an RMSD of just 0.49 Å, validating MsGlgB as an equivalent structural model. This also provides the means to generate inhibitor-bound GlgB structures, which will inform structure-activity relationships for the development of lead compounds active against GlgB.

Other researchers have also found it difficult to replicate the published MtGlgB structure (Geiger and co-workers, personal communication). Therefore, it appears that MsGlgB provides a more tractable model for crystallographic studies. This difference is reflected in the higher thermostability of MsGlgB, with a T_m of 64.0 °C compared to 57.7 °C for MtGlgB, as determined by nano DSF. The predicted disulphide bonds in MtGlgB,[27] which are absent in MsGlgB, may lead to different populations of MtGlgB molecules, creating disorder that impedes crystallisation. It is interesting to note that PEG 4,000 and 0.1 M MES at pH 6.5 are present in both final crystallisation conditions, and these components may play a role in formation of the crystal lattice, but neither were detected in electron density maps. Although the crystal structure of MsGlgB was obtained much more readily relative to MtGlgB, protein crystals were observed in only one condition out of the 768 tested and required a number of optimisation steps before high-quality X-ray diffraction data could be reliably obtained. This is consistent with branching enzymes being challenging to crystallise, and there are a low number of reported structures, even though they are widely studied.

Using the MsGlgB model I went on to solve the first reported ligand-bound structures for an actinomycete branching enzyme. Carbohydrate-bound structures could only be determined using very high concentrations of oligosaccharides, ≥ 0.1 M, as was seen previously for the cyanobacterial BE1.[10] Generally, soaks with polysaccharides did not result in any ligand binding, most likely because the shorter fragments that could be detected by crystallographic studies were not present at high enough concentrations. The approach of using 2-deoxy-2-fluoro-maltosyl fluoride oligosaccharides in combination with the MsGlgB-E469A protein also did not yield donor-bound structures. This could be attributed to the alternate 2-deoxy-2-fluoro substitutions in the pseudooligosaccharide interfering with

binding or, more likely, that maltooligosaccharides ≤ DP8, which can be accommodated in the donor site, were present at too low concentrations in the reaction mixture for this to be achieved. Further optimisation of reaction and purification conditions to increase the proportion of products of appropriate DP might result in a fluorinated oligosaccharide product mixture suitable for structural studies of other GH enzymes.

Oligosaccharides were only bound in the donor site of MsGlgB when the aspartate nucleophile was mutated to alanine. This suggests that the presence of the aspartate residue precluded ligand binding, most likely by blocking the binding at the -1 subsite, which appeared to be necessary for binding at any of the other donor subsites. This contrasts with CceBE1, in which oligosaccharides bound at only subsites -2 to -6, and, when they did also bind at the -1 subsite, the reducing-end pyranose was rotated by an unusual angle. This necessitated a W610N modification, whereas this was not the case in MsGlgB, and the equivalent residue Glu592 coordinated the pyranose in the -8 subsite, suggesting that longer oligosaccharides are more easily accommodated in MsGlgB. Donor subsites -1 to -6 are almost completely conserved between MsGlgB and CceBE1, and sequence alignments suggest these may be conserved more widely across the bacterial species, **Figure 2.19**.

Figure 2.19 Alignment of bacterial branching enzyme sequences. Protein sequence alignment of MsGlgB, *S. coelicolor* GlgB, *E. coli* GlgB and CceBE1, with residues in the MsGlgB donor site highlighted in green. With the exception of a couple of residues these are conserved between these species, suggesting that this site may be conserved across different bacterial species.

The relative orientations of the A1 and A2 sites suggested that MsGlgB can coordinate an α-1,6 branched structure at a distance from the active site. Previously, analysis of mycobacterial α-glucan structure showed that branches in mycobacterial α-glucan are accommodated on the 4th or 5th glucose units from an existing branch point and that most B chains carry one or two branches.[4] Therefore, if O1 of acarbose at the A2 site was linked in a branch then the subsequent branchpoint would emanate from the O6 of the non-reducing end pyranose. The next branchpoint would be expected 4-5 residues from here, at approximately the site of catalytic residues, where a new branchpoint would be formed, **Figure 2.20**. In this model, the branch emanating from the non-reducing end of acarbose in A2 is therefore a proximal branchpoint, and the A1/A2 1,6 linkage therefore represents a

distal branchpoint. Whilst this remains speculative without more substrate-bound structures, the coordination of a distal branchpoint could have interesting implications for the activity of MsGlgB. Coordination to a distal branchpoint might be suggestive of a mechanism whereby a substrate with ≥ two existing branches is more likely to be branched again. It would make sense for MsGlgB to prefer branched substrates as this would mean it would tend to act to create larger α-glucan molecules, rather than generate many smaller branched oligosaccharides.
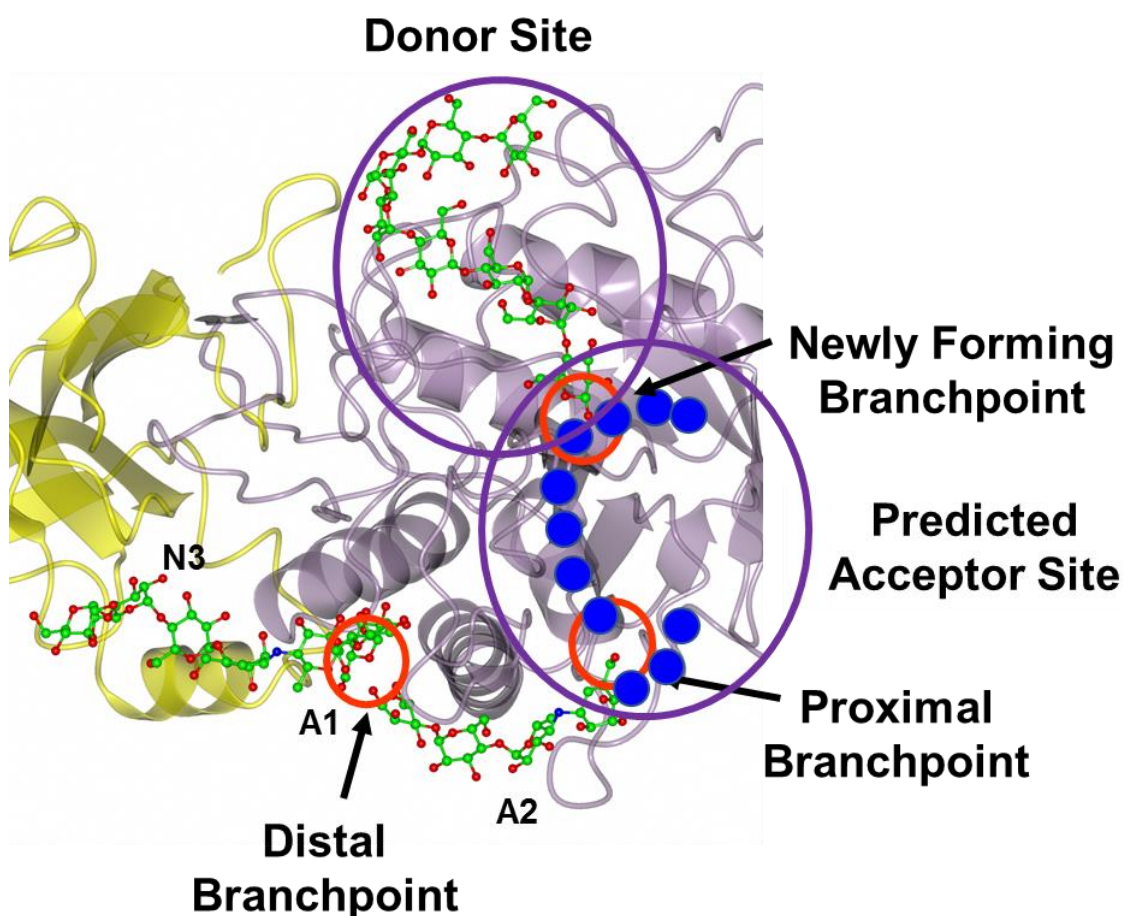
Figure 2.20 Proposed model of substrate binding for MsGlgB. Protein and bound oligosaccharirdes shown as in **Figure 2.15A**. Maltooligosaccharide binding in the proposed acceptor site is symbolised by blue circles, with each representing an α-glucopyranose unit. Suggested substrate branchpoints are labelled and indicated by red circles.

The acceptor subsites were not identified by this work, but the A2 site is a promising candidate for a distal acceptor site. Acarbose was bound at the A2 site, with a distance of 19.5 Å between the non-reducing end of the pseudotetrasaccharide and the nucleophilic residue D416. The acceptor site could be modelled by a 4-5 glucose unit extension from the non-reducing end of the A2 acarbose, **Figure 2.20**. In this model, MsGlgB residues 470-475 would be implicated in binding the acceptor oligosaccharide. This contrasts with the

acceptor cleft proposed for cyanobacterial BE1 based on oligosaccharide binding site near domain C, **Figure 2.21** and **Figure 2.22**. Although CceBE1 and MsGlgB generate similar average branch lengths of DP 6-7 and DP 7-8 respectively, they are involved in the synthesis of different α-glucan products, namely starch and glycogen and this could explain the different acceptor sites that have been proposed.

One other aspect of the branching enzyme mechanism is whether transfer is intra or inter chain. For MtGlgB chain transfer was shown to be strictly intrachain with a short oligosaccharide substrate *in vitro*.[4] Based on the average branch lengths of DP7-8 and internal chain lengths of DP4.6-5.0, intrachain transfer necessitates an average movement of a newly cleaved chain by ~3 glucose residues. This would require a significant conformational change in the acceptor binding cleft and/or strained binding of the substrate to create the optimal orientation for chain transfer. A cartoon of a possible mechanism for the strained orientation of an acceptor substrate facilitating intrachain branch formation is shown in **Figure 2.22**. This may mean that identifying maltooligosaccharides bound at the MsGlgB proximal acceptor site will be challenging due to the inherent disorder these mechanisms entail.
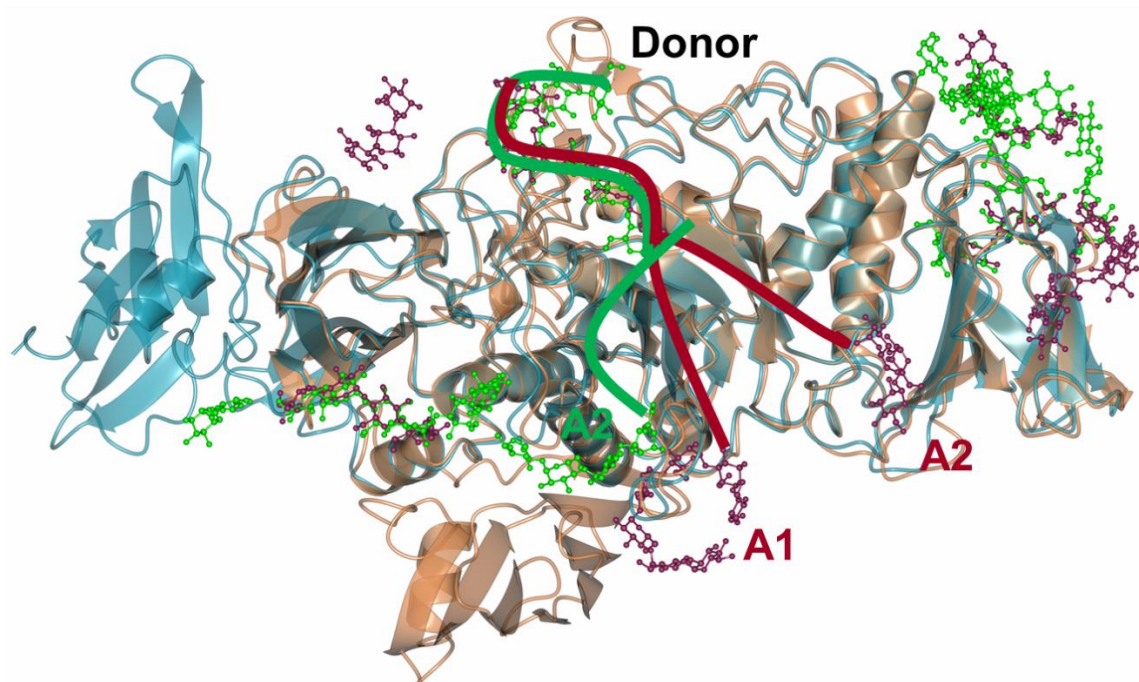
Figure 2.21 Comparison of maltooligosaccharide binding in MsGlgB and CceBE1. Superposition of MsGlgB and CceBE1 structures showing all maltooligosaccharide binding sites. Proteins are shown as ribbon models, oligosaccharides as ball and stick models. MsGlgB: protein = cyan, oligosaccharide = green; CceBE1: protein = orange, oligosaccharide = dark purple. The paths of the donor site and proposed acceptor sites for each enzyme are indicated by coloured lines.
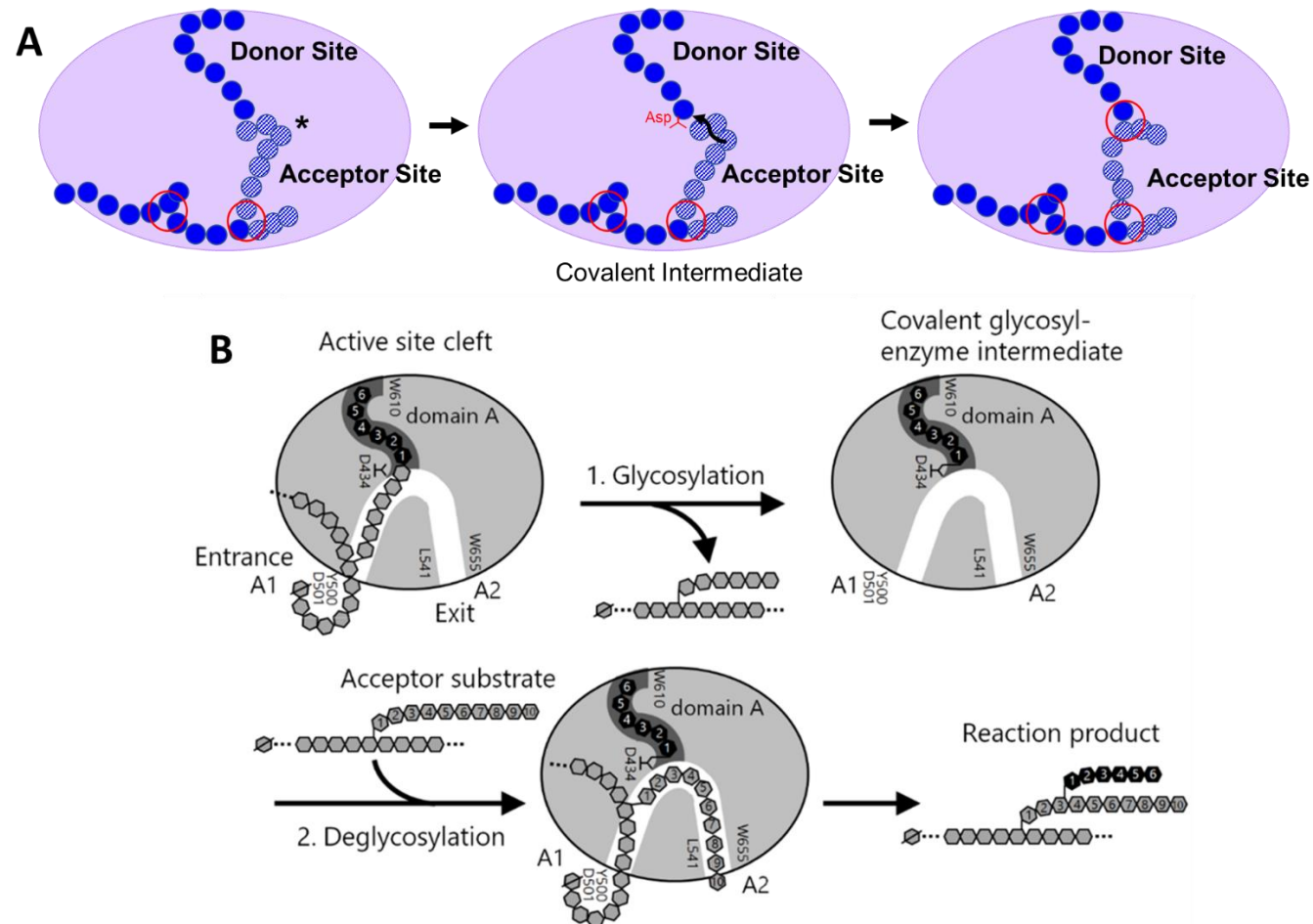
Figure 2.22 Comparison of the branching mechanisms of MsGlgB and CceBE1. A) Branching mechanism proposed for MsGlgB based on intrachain transfer facilitated by a strained oligosaccharide conformation in the proximal acceptor site, marked by *. Oligosaccharides are represented by blue circles for each glucose unit, with filled circles representing those observed in crystal structures and striped circles for predicted binding. **B)** Proposed mechanism of CceBE1, labelled according to **Figure 1.14**. Adapted from Hayashi *et al.*[10] with permission.

2.6   Summary

This chapter describes the biochemical and structural characterisation of mycobacterial GlgB. Biochemical analyses demonstrated that these enzymes generate notably shorter branches than their *E. coli* or mammalian counterparts, consistent with what is seen in the *in vivo* α-glucan product. The existing MtGlgB structural model was greatly improved with high-resolution, substrate-bound MsGlgB structures. These enabled the identification of a number of structural features, including the donor site, a flexible loop region and the putative distal acceptor site.

2.7   Materials and methods

Unless otherwise specified, all chemicals were purchased from Sigma Aldrich Company Ltd.

2.7.1   Growth media

All bacterial growth was carried out using lysogeny broth (LB) consisting of tryptone (10 g/L), yeast extract (5 g/L), NaCl (10 g/L) in dH_2O, with the addition of agar (10 g/L) for solid media. The antibiotics carbenicillin (100 μg/mL) or kanamycin (50 μg/mL) (ForMedium) were added as appropriate to select for bacterial strains.

2.7.2 Primers

Oligonucleotide sequences used in molecular biology experiments are listed in **Table 2.6**.

Table 2.6 Oligonucleotide sequences.

Oligonucleotide ID	Sequence (5' - 3')	Notes
T7	TAATACGACTCACTATAGGG	For plasmid sequencing
T7 terminal	TAGTTATTGCTCAGCGGTGG	
MtGlgB-Internal1	GTGACCGATCGTGCAGATCC	For plasmid sequencing
MtGlgB-Internal2	GCTACGATACACCGGATCACG	
MtGlgB-Internal3	CATCATGAAATGACCTTTAG	
MtGlgB-Nterm-DT	CACCATGAGCCGTAGCGAAAAACTGACCG	To generate PCR fragments for Directional TOPO® cloning to make MtGlgB-CHis$_6$ expression construct
MtGlgB-CHis$_6$-DT	TTAGTGGTGGTGGTGGTGGTGCGCCGGGGTCAGCCACAG	
MtGlgB-ΔN9-QC-F	GAAAACCTGTATTTTCAGGGTGAACATCTGGCCCCG	To remove the region encoding for amino acids 2-8 using the QuikChange Lightning kit to make ΔN9-MtGlgB expression construct
MtGlgB-ΔN9-QC-R	CGGGGCCAGATGTTCACCCTGAAAATACAGGTTTTC	
MtGlgB-ΔN108-IF	AGGAGATATACCATGACCCTGGGCGAAGTGGATCT	To generate PCR fragments for In-Fusion® HD cloning to make ΔN108-MtGlgB expression construct
MtGlgB-C-His$_6$-IF	GTGGTGGTGGTGTTTCGCCGGGGTCAGCCACAGCG	
MsGlgB-Internal1	GTGGGAACTGTTTTGGCCG	For plasmid sequencing
MsGlgB-Internal2	GTGACTAATCGGCAGCACG	
D416A-MsGlgB-F	CAGGACGTACTTTCTGCGGCAACGGTCACGA	To change codon for Asp416 to Ala in MsGlgB using the QuikChange Lightning kit
D416A-MsGlgB-R	GTCTGCGCGTTGCCGCGGTCGCCAG	
E469A-MsGlgB-F	TCGTGACCGTTGCCGCAGAAAGTACGTCCTG	To change codon for Glu469 to Ala in MsGlgB using the QuikChange Lightning kit
E469A-MsGlgB-R	CTGGCGACCGCGGCAACGCGCAGAC	

2.7.3 Cloning methods

DNA fragments were amplified by PCR thermal cycling using Phusion® High-fidelity DNA polymerase (NEB). DNA was analysed by DNA gel electrophoresis using a 1% (w/v) TAE agarose gel with 1 µg/mL ethidium bromide with UV visualisation. DNA fragments were purified using a QIAquick PCR purification kit (Qiagen) and concentrations were determined using the Nanodrop spectrophotometer (Nanodrop®). DNA fragments were incorporated into vectors using Directional TOPO® cloning with transformation into One Shot® TOP10 competent cells (Invitrogen) or In-Fusion® HD cloning with transformation into Stellar Competent cells (Takara) according to the protocols set out in the relevant user manuals. Site-directed mutagenesis was carried out using a QuikChange Lightning kit (Agilent) following the protocols for thermal cycling, *Dpn*I digestion and transformation of XL10-Gold ultracompetent cells as set out in the user manual.

To verify successful transformants, a single colony from the resultant transformation plate was used to inoculate LB media supplemented with the appropriate antibiotic and incubated at 37 °C with shaking for 18 h. Plasmids were purified using a QIAprep Spin Miniprep kit (Qiagen) and quantified using the Nanodrop spectrophotometer as above. All modifications were confirmed by DNA sequencing (Eurofins), using the T7 and T7 terminal primers, as well as internal *glgB* primers, **Table 2.6**. Sequencing data were analysed using Chromas Lite (Technelysium).

2.7.4 Plasmids for protein expression

Expression plasmids for producing GlgB and GlgE proteins with an N-terminal, TEV cleavable His_6 tag had previously been prepared by K. Syson as follows. Codon-optimised DNA sequences of *glgB* and *glgE* genes from *M. tuberculosis* H37Rv and *S. coelicolor* A3(2) (isoform 1 of *glgB* and *glgE*), and the *glgB* gene from *M. smegmatis* mc^2 155 were ligated into a pET-21a(+) vector (Novagen). The *glgB* gene from *S. venezuelae* ATCC 10712 was ligated into a pBAD43 vector (Novagen). The *M. tuberculosis* H37Rv *glgB* gene sequence was mutated into that of *M. tuberculosis* CCDC5180 by changing the codon for Ser470 into one for Pro using a QuikChange Lightning kit (Agilent).

All modifications to the *M. tuberculosis* and *M. smegmatis glgB* expression plasmids were carried out by myself as follows. The construct for the expression of ΔN9-MtGlgB was generated using QuikChange Lightning as above to remove the region encoding the first nine amino acids from the MtGlgB expression construct described above. The construct for the expression of MtGlgB-C-His_6 was made by incorporating the *M. tuberculosis glgB* gene with an extension for a C-terminal His_6 tag into a pET101 vector (Invitrogen) using Directional TOPO® cloning. To make the construct for the expression of ΔN108-MtGlgB with a C-His_6 tag, the *M. tuberculosis glgB* gene sequence was incorporated into a pOPINA vector (OPPF-UK) using In-Fusion™ cloning. For the modified MsGlgB expression constructs the *M. smegmatis glgB* gene sequence was mutated with QuikChange Lightning to either change the codon for Asp416 into Ala or Glu469 into Ala or both.

2.7.5 Production of recombinant proteins

Plasmids for protein expression were transformed into either One Shot™ BL21 (DE3), One Shot™ BL21 Star™ (DE3) (Invitrogen) or soluBL21™ (DE3) (Genlantis) chemically competent *E. coli*. Cells containing the expression plasmid were cultured in LB containing appropriate antibiotic at 37 °C. Expression was induced when the cells reached an OD_{600nm} of 0.6, by the addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG). After

incubation at 30 °C for 2 h and then 16 °C for 18 h, the cells were harvested by centrifugation at 5,000 × g for 10 min at 4 °C and resuspended in 50 mM Tris-HCl, pH 8.0, containing 15 mM imidazole, 300 mM NaCl, DNase 1, and a complete protease inhibitor cocktail tablet (Roche). The cells were disrupted with a TS Series Benchtop 1.1 kW cell disruptor (Constant Systems Ltd.) at 25 kPSI. The resulting cell lysate was separated from the cell debris by centrifugation at 18,000 × g for 25 min at 4 °C.

Proteins were purified from each cell lysate by application onto a 1 mL FF HisTrap column (GE Healthcare). Protein was eluted with a 15–300 mM imidazole gradient and further purified using a HiLoad 26/60 Superdex 200 gel filtration column (GE healthcare) equilibrated with 20 mM sodium phosphate buffer, pH 7.0, 50 mM NaCl. The eluate was collected in fractions, which were analysed by SDS-polyacrylamide gel electrophoresis (PAGE) using pre-cast 12% (w/v) SDS-PAGE gels and staining with InstantBlue (Expedeon Ltd). Fractions containing the desired recombinant protein were combined and concentrated using an Amicon Ultra-15 centrifugal filter unit with a 10 kDa cut-off (MerckMillipore) and centrifugation at 5,000 × g. Aliquots of protein were flash-frozen in liquid N_2 and stored at -80 °C. Protein concentrations were determined using a Direct Detect® infrared spectrometer (Merck).

In some cases, the His_6 tag was cleaved by incubating the recombinant protein with TEV protease in a buffer containing 50 mM Tris, pH 8 and 150 mM NaCl. This was followed by the addition of imidazole at a final concentration of 30 mM and the solution was applied onto a FF HisTrap column to remove TEV protease. The flow-through was collected, concentrated and then solvent exchanged by dialysis.

2.7.6   Plasmid stability test

Plasmid stability tests were carried out according to the pET system manual (Novagen). Bacterial strains transformed with an expression plasmid were cultured in LB and grown at 37 °C until the cells reached an OD_{600nm} of 0.6. Serial dilutions of the cell suspension were then prepared up to 10^{-6} dilution. The 10^{-6} dilution was plated on LB-agar (allowing growth of all viable cells) and LB-agar with antibiotic (allowing growth of cells retaining the plasmid). The 10^{-6} dilution was plated on LB-agar with 1 mM IPTG (allowing growth of cells that have lost the plasmid or the ability to express the gene) and LB-agar with 1 mM IPTG and antibiotic (allowing growth of cells that retain the plasmid but cannot express the gene). After incubation at 37 °C the number of colonies on each plate was counted. It was expected <2% of cells will form a colony on the plate containing IPTG and < 0.01% on the plate

containing IPTG and antibiotic. Higher values were indicative of plasmid instability and hence a toxic gene product.

2.7.7   Nano differential scanning fluorimetry

Nano DSF experiments were performed on the Prometheus NT.Plex (NanoTemper Technologies Ltd.). For each protein sample 10 µL was loaded into a glass capillary and subjected to a temperature ramp of 2 °C/min from 20 °C to 95 °C. Fluorescence at 350 nm and 330 nm was constantly monitored and data were analysed with PR.ThermControl software (NanoTemper Technologies Ltd.). A difference ≥ 0.5 °C is considered significant.

2.7.8   *In vitro* glucan synthesis

To generate synthetic α-glucan 1 mM maltohexaose and 10.1 mM M1P (synthesised by A. Rashid as previously described[13]) were incubated with 7 µg each of recombinant GlgE and GlgB (giving < 2 µM final concentrations) in 100 mM bis-Tris propane, pH 7.0, containing 50 mM sodium chloride in a total volume of 50 µL. After 18 h at 22 °C, a second aliquot of M1P was added. After a further 90 h, the sample was heated at 95 °C for 5 min, cooled, and washed five times using a 10 kDa cut-off membrane with 100 mM sodium acetate buffer, pH 4.0, to remove small maltooligosaccharides. The sample was analysed by TEM, MALDI mass spectrometry, or CE with or without debranching as appropriate.

2.7.9   Transmission electron microscopy

TEM imaging was carried out by E. Barclay as follows. Aqueous polymer samples (typically 0.1 mg/mL) were allowed to settle briefly onto a 400 mesh copper grid with a carbon-coated pyroxylin support film before being negatively stained. α-Glucan was stained with 2% (w/v) aqueous uranyl acetate, pH 4.5. The grids were viewed in a FEI Tecnai 20 transmission electron microscope (Eindhoven, Netherlands) at 200 kV, and digital TIFF images were taken using an AMT XR60B digital camera (Deben).

2.7.10   Debranching α-glucans with isoamylase

Polymer samples (typically in 25 µL) were debranched using *Pseudomonas* sp. isoamylase (with typically 2 µL of a 1,000 U/mL stock solution) from Megazyme in 100 mM sodium acetate buffer, pH 4.0, at 37 °C for 4 h. Reactions were terminated by heating at 95 °C for 5 min, and the products were analysed either directly or after precipitation with ethanol (5 vol). Precipitated samples were dissolved in water (50 µL) and freeze-dried to yield an amorphous powder.

2.7.11 MALDI mass spectrometry

Matrix assisted laser desorption/ionisation mass spectra were acquired on Bruker Ultraflex or Autoflex Speed MALDI-Time of flight (TOF)/TOF mass spectrometers, equipped with a pulsed nitrogen laser emitting at 337 nm or Smartbeam-II 2 kHz laser, respectively. Samples were typically mixed 1:2 v/v with a 20 mg/mL solution of 2,5-dihydroxybenzoic acid in 50% aqueous acetonitrile and spotted on a target plate (Bruker MTP 384 Polished Steel TF Target). Data were processed using Bruker flexAnalysis software. The relative abundance of maltooligosaccharides of DP ≥ 4 was determined at least in duplicate by measuring the area under the peaks of all isotopes associated with each species of a given degree of polymerisation using the Sophisticated Numerical Annotation Procedure with sugar-sodium adduct average composition.

2.7.12 Capillary electrophoresis

Reducing saccharides were labelled with 8-aminopyrene-1,3,6-trisulphonic acid to facilitate their separation on the basis of charge and their detection by fluorescence spectrophotometry. Samples (typically 5 µL) were treated with 8-aminopyrene-1,3,6-trisuphonic acid (1 µL of 50 mM in 15% aqueous glacial acetic acid) and freshly prepared $NaCNBH_3$ in tetrahydrofuran (1 µL of 1 M). Reaction mixtures were incubated at 37 °C for 4 h and diluted 10-fold with water. Samples were subjected to capillary electrophoresis using a Beckman Coulter PA800 ProteomeLab Protein Characterisation System using laser-induced fluorescence detection with an excitation wavelength of 488 nm and an emission band-pass filter of 520 nm ±10 nm. An SCIEX N-CHO capillary (50 µm internal diameter × 50.2 cm; 40 cm effective length to detector) was used with a constant voltage of −30 kV, the anode at the detector, and a temperature of 25 °C. The capillary was rinsed with SCIEX N-CHO buffer, and the samples were introduced at a pressure of 0.5 psi for 5 s.

2.7.13 Protein crystallisation

Proteins were concentrated to 5-12 mg/mL using Amicon Ultra-0.5 centrifugal filter unit with a 10 kDa cut-off (MerckMillipore) and subsequently filtered using a 0.1 µm Ultrafree® filter (MerckMillipore). Protein crystallisation trials were carried out using the sitting drop vapour-diffusion method in a 96-well MRC plate (Molecular Dimensions) format using the following screens: KISS (John Innes Centre), JCSG-plus, PACT premier, Structure, Morpheus, Midas (Molecular Dimensions), Ammonium sulphate and PEG suite (Qiagen). A volume of 40 µL was transferred to each well using the Freedom evo liquid handling robot (Tecan). Drops

were dispensed using the OryxNano robot (Douglas Instruments Ltd.) comprising 0.3 µL of precipitant solution and 0.3 µL of protein solution. The plates were sealed and stored at 19 °C in a crystal hotel (CrystlPro, TriTek Corp.) and crystal formation was monitored by imaging plates with and without UV light. Crystals were optimised by varying the protein concentration, varying the components of the precipitant solution, varying the ratio of precipitant and protein solutions in each drop, and adding micro-seed solution, which was prepared by crushing existing crystals and preparing serial dilutions using the precipitant solution in the well reservoir. Crystals were imaged using a SMZ800 microscope (Nikon) with and without a polarising filter to observe birefringence. Photographs were taken with a Nikon Coolpix E995 digital camera (Nikon).

2.7.14 Crystal soaks

Crystal soaks were carried out using maltohexaose (Sigma Aldrich), maltoheptaose, maltooctaose (Carbosynth), β-cyclodextrin (Sigma Aldrich), γ-cyclodextrin (Wacker Chemie AG), cycloamylose (DP26; Ezaki Glico Wako), maltodextrin (Sigma Aldrich), amylose 2800 (av. molecular weight (MW) 2800; Tokyo Chemical Industry Ltd), amylose BAR-5K-1 (av. MW 4,500; PS-Biotec Inc), amylose EX-I (av. DP17; Hayashibara Co. Ltd), amylose from potato (av. MW 150,000; Sigma), glucosyl maltotriosyl maltotriose (Megazyme), rabbit liver glycogen (Megazyme), amylopectin from potato (Sigma Aldrich), pullulan (Sigma Aldrich), β-limit dextrin (Megazyme), acarbose (Toronto Research Chemicals), DP6-acarbose (prepared by M. Rejzek as described previously[28]) and 2-deoxy-2-fluoro maltosyl fluoride (prepared by A. Rashid as described previously[18]). 2-Deoxy-2-fluoro maltosyl fluoride (5 – 50 mM) was extended by incubation with *S. coelicolor* GlgEI (~1 µM) in a 100 mM bis-Tris propane, pH 7.0, 50 mM NaCl buffer. The reaction was terminated by heating at 95 °C for 5 min and GlgE was removed by filtration or amberlite resin.

Solutions for crystal soaks were prepared by dissolving oligosaccharides in MsGlgB crystallisation solution: 0.06 M $MgCl_2$, 0.06 M $CaCl_2$, 20% glycerol, 10% PEG 4,000, 0.1 M imidazole, 0.1 M MES at pH 6.5, without additional cryoprotection agents. Initially, oligosaccharide concentrations of 1 mM – 50 mM were prepared, which were then increased to 100 - 500 mM, or saturating solutions in cases where solubility was limiting. Crystals were soaked for 15 min – 48 h prior to mounting.

2.7.15 Data collection

In collaboration with C. Stevenson, crystals were collected in litho-loops (Molecular Dimensions) and, in the case of MtGlgB crystals, were soaked in cryo-protectant solution comprising the crystallisation solution with the addition of 25% ethylene glycol. All crystals were flash-cooled by plunging into liquid N_2 and stored in Unipuck casettes (MiTeGen) for transport to Diamond Light Source (Oxfordshire, UK) for data collection. Crystals were then robotically mounted onto the geniostat on stations I02, I03, I04 or I04-1 and maintained at 100 K with a Cryojet cryocooler (Oxford instruments). Diffraction data were recorded using a Pilatus 6M hybrid photon counting detector (Dectris).

2.7.16 Data processing

The resultant datasets were processed using the XIA2 expert system[29] and merged using AIMLESS.[30] For MsGlgB, the data sets were processed in the space group $P2_12_12$ with cell parameters of approximately $a = 98$, $b = 164$, $c = 54$ Å. The initial structure was solved by molecular replacement in PHASER,[14] using MtGlgB (PDB code 3K1D) as the template. One molecule was placed in the asymmetric unit, corresponding to the monomeric nature of the protein in solution. The model was rebuilt using BUCCANEER[31] and completed through several iterations of refinement with REFMAC5[15] with isotropic thermal parameters and manual adjustment with COOT.[16] In the later stages, translation libration screw (TLS) refinement was used with a total of five TLS domains, which were defined by the TLS motion determination server.[32] This model was used for the refinement of subsequent datasets. For higher resolution data, anisotropic thermal parameters were used instead of TLS parameters. Electron density maps were consistent with expected point mutations and provided evidence of ligand binding at various sites in the protein.

2.7.17 Structural validation

Model geometries were validated using MOLPROBITY[22] for protein residues and PRIVATEER[33] for sugar ligands. To generate $mF_{obs} - dF_{calc}$ difference electron density maps ligands were deleted from the coordinates of the final models and these were used as input to REFMAC5.[16] All structural figures were prepared using CCP4mg.[34]

2.8   References

1. Koliwer-Brandl, H.; Syson, K.; van de Weerd, R.; Chandra, G.; Appelmelk, B.; Alber, M.; Ioerger, T. R.; Jacobs, W. R., Jr.; Geurtsen, J.; Bornemann, S.; Kalscheuer, R., Metabolic network for the biosynthesis of intra- and extracellular α-glucans required for virulence of *Mycobacterium tuberculosis*. *PLoS Path.* **2016,** *12* (8), e1005768.

2. Geurtsen, J.; Chedammi, S.; Mesters, J.; Cot, M.; Driessen, N. N.; Sambou, T.; Kakutani, R.; Ummels, R.; Maaskant, J.; Takata, H.; Baba, O.; Terashima, T.; Bovin, N.; Vandenbroucke-Grauls, C.; Nigou, J.; Puzo, G.; Lemassu, A.; Daffe, M.; Appelmelk, B. J., Identification of mycobacterial α-glucan as a novel ligand for DC-SIGN: Involvement of mycobacterial capsular polysaccharides in host immune modulation. *J. Immunol.* **2009,** *183* (8), 5221-5231.

3. Kalscheuer, R.; Syson, K.; Veeraraghavan, U.; Weinrick, B.; Biermann, K. E.; Liu, Z.; Sacchettini, J. C.; Besra, G.; Bornemann, S.; Jacobs, W. R., Self-poisoning of *Mycobacterium tuberculosis* by targeting GlgE in an α-glucan pathway. *Nat. Chem. Biol.* **2010,** *6* (5), 376-384.

4. 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., Assembly of α-glucan by GlgE and GlgB in mycobacteria and streptomycetes. *Biochemistry* **2016,** *55* (23), 3270-3284.

5. Suzuki, E.; Suzuki, R., Distribution of glucan-branching enzymes among prokaryotes. *Cell. Mol. Life Sci.* **2016**, 1-18.

6. Abad, M. C.; Binderup, K.; Rios-Steiner, J.; Arni, R. K.; Preiss, J.; Geiger, J. H., The X-ray crystallographic structure of *Escherichia coli* branching enzyme. *J. Biol. Chem.* **2002,** *277* (44), 42164-42170.

7. Chaen, K.; Noguchi, J.; Omori, T.; Kakuta, Y.; Kimura, M., Crystal structure of the rice branching enzyme I (BEI) in complex with maltopentaose. *Biochem. Biophys. Res. Commun.* **2012,** *424* (3), 508-511.

8. Froese, D. S.; Michaeli, A.; McCorvie, T. J.; Krojer, T.; Sasi, M.; Melaev, E.; Goldblum, A.; Zatsepin, M.; Lossos, A.; Álvarez, R.; Escribá, P. V.; Minassian, B. A.; von Delft, F.; Kakhlon, O.; Yue, W. W., Structural basis of glycogen branching enzyme deficiency and pharmacologic rescue by rational peptide design. *Hum. Mol. Genet.* **2015,** *24* (20), 5667-5676.

9. Pal, K.; Kumar, S.; Sharma, S.; Garg, S. K.; Alam, M. S.; Xu, H. E.; Agrawal, P.; Swaminathan, K., Crystal structure of full-length *Mycobacterium tuberculosis* H37Rv glycogen branching enzyme: Insights of N-terminal β-sandwich in subtrate specificity and enzymatic activity *J. Biol. Chem.* **2010,** *285* (27), 20897-20903.

10. Hayashi, M.; Suzuki, R.; Colleoni, C.; Ball, S. G.; Fujita, N.; Suzuki, E., Bound substrate in the structure of cyanobacterial branching enzyme supports a new mechanistic model. *J. Biol. Chem.* **2017**.

11. Feng, L.; Fawaz, R.; Hovde, S.; Gilbert, L.; Chiou, J.; Geiger, J. H., Crystal structures of *Escherichia coli* branching enzyme in complex with linear oligosaccharides. *Biochemistry* **2015,** *54* (40), 6207-6218.

12. Feng, L.; Fawaz, R.; Hovde, S.; Sheng, F.; Nosrati, M.; Geiger, J. H., Crystal structures of *Escherichia coli* branching enzyme in complex with cyclodextrins. *Acta crystallographica. Section D, Structural biology* **2016,** *72* (Pt 5), 641-7.

13. Syson, K.; Stevenson, C. E. M.; Rejzek, M.; Fairhurst, S. A.; Nair, A.; Bruton, C. J.; Field, R. A.; Chater, K. F.; Lawson, D. M.; Bornemann, S., Structure of *Streptomyces* maltosyltransferase GlgE, a homologue of a genetically validated anti-tuberculosis target. *J. Biol. Chem.* **2011,** *286* (44), 38298-38310.

14. Slabinski, L.; Jaroszewski, L.; Rychlewski, L.; Wilson, I. A.; Lesley, S. A.; Godzik, A., XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **2007,** *23* (24), 3403-5.

15. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (4), 658-674.

16. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* **1997,** *53* (3), 240-255.

17. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **2004,** *60* (12 Part 1), 2126-2132.

18. Syson, K.; Stevenson, C. E. M.; Rashid, A. M.; Saalbach, G.; Tang, M.; Tuukkanen, A.; Svergun, D. I.; Withers, S. G.; Lawson, D. M.; Bornemann, S., Structural insight into how *Streptomyces coelicolor* maltosyl transferase GlgE binds α-maltose 1-phosphate and forms a maltosyl-enzyme intermediate. *Biochemistry* **2014,** *53* (15), 2494-2504.

19. Sim, L.; Beeren, S. R.; Findinier, J.; Dauvillée, D.; Ball, S. G.; Henriksen, A.; Palcic, M. M., Crystal Structure of the Chlamydomonas Starch Debranching Enzyme Isoamylase ISA1 Reveals Insights into the Mechanism of Branch Trimming and Complex Assembly. *J. Biol. Chem.* **2014,** *289* (33), 22991-23003.

20. Janeček, Š., How many conserved sequence regions are there in the α-amylase family.

21. Qian, M.; Haser, R.; Buisson, G.; Duee, E.; Payan, F., The Active Center of a Mammalian α-Amylase: Structure of the Complex of a Pancreatic α-Amylase with a Carbohydrate Inhibitor Refined to 2.2 Å Resolution. *Biochemistry* **1994,** *33* (20), 6284-6294.

22. Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2010,** *66* (Pt 1), 12-21.

23. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** *23* (21), 2947-2948.

24. Gouet, P.; Robert, X.; Courcelle, E., ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **2003,** *31* (13), 3320-3323.

25. Morell, M. K.; Samuel, M. S.; O'Shea, M. G., Analysis of starch structure using fluorophore-assisted carbohydrate electrophoresis. *Electrophoresis* **1998,** *19* (15), 2603-2611.

26. Szájli, E.; Fehér, T.; Medzihradszky, K. F., Investigating the Quantitative Nature of MALDI-TOF MS. *Mol. Cell. Proteomics* **2008,** *7* (12), 2410-2418.

27. Garg, S. K.; Alam, M. S.; Kishan, K. V. R.; Agrawal, P., Expression and characterization of α-(1,4)-glucan branching enzyme Rv1326c of *Mycobacterium tuberculosis* H37Rv. *Protein Expression Purif.* **2007,** *51* (2), 198-208.

28. Yoon, S. H.; Robyt, J. F., Study of the inhibition of four alpha amylases by acarbose and its 4IV-alpha-maltohexaosyl and 4IV-alpha-maltododecaosyl analogues. *Carbohydr. Res.* **2003,** *338* (19), 1969-80.

29. Winter, G., xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **2010,** *43* (1), 186-190.

30. Evans, P. R.; Murshudov, G. N., How good are my data and what is the resolution? *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2013,** *69* (Pt 7), 1204-1214.

31. Cowtan, K., The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D* **2006,** *62* (9), 1002-1011.

32. Painter, J.; Merritt, E. A., TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **2006,** *39* (1), 109-111.

33. Agirre, J.; Iglesias-Fernandez, J.; Rovira, C.; Davies, G. J.; Wilson, K. S.; Cowtan, K. D., Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 833-4.

34. McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E., Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **2011,** *67* (Pt 4), 386-94

3. Identification of a small-molecule inhibitor of MtGlgB

3.1   Introduction

MtGlgB is an α-1,6 branching enzyme that has been genetically validated as a drug target, with a novel mechanism of toxicity.[1] A number of putative MtGlgB inhibitors have previously been identified by virtual screening, including two which inhibited *M. tuberculosis* growth.[2] However, questions remain over the validity of the *in vitro* evaluation of these compounds due to the lack of counter-screens or verification that the compounds bind to the enzyme. Other known inhibitors of GH13 enzymes such as ADP, ADP-glucose, tunicamycin, casternospermine, nojirimycin and acarbose were ineffective against MtGlgB.[3] The only verified inhibitor of any α-1,6 branching enzyme reported to date is the pseudooligosaccharide BAY e4609, a fraction of glycosylated acarbose derivatives with DP 9 – 32. This showed slow-binding inhibition of *E. coli* GlgB,[4] but this heterogeneous mixture of large molecules is unsuitable for therapeutic application. Therefore, in this chapter I address the question: can an inhibitor of MtGlgB be identified by high-throughput screening of compound libraries?

This firstly necessitates the establishment of an assay suitable for HTS. The branching enzyme reaction involves the cleavage of an α-1,4-glycosidic linkage and the formation of an α-1,6-glycosidic bond.[5] Enzyme activity is typically monitored by treatment of the branched products with isoamylase and subsequent quantification of the increase in reducing ends with colorimetric methods, with one or more heating steps.[6-8] These are therefore unattractive for HTS due to the variability associated with heating microplates and the multiple steps involved.[9] A more simple method for monitoring branching activity uses iodine staining, which can be used to distinguish branched and unbranched α-glucan chains by colour.[10]

The dark blue colour resulting from the addition of iodine solution to amylose was first described in 1814.[11] The structure of the complex was established by Rundle and co-workers in the 1940s as a left-handed amylose helix with six glucose residues per turn, with the iodine component arranged axially in the inner cavity.[12-13] The nature of the polyiodide species has been investigated with numerous different methods, with the prevailing interpretation being a mixture of I_3^- and I_5^-,[14-16] although more recently an infinite iodide species, $I_\infty^{\delta-}$, has been suggested.[17] The relationship between the DP of amylose and λ_{max} has been established, with λ_{max} increasing from 500 – 650 nm with increasing DP.[18]

In branched α-glucans, such as amylopectin and glycogen, the glucan helices are disrupted, resulting in a lower λ_{max} value and a reduced peak in the absorbance spectrum.[19] The use of the amylose-iodine assay in monitoring the activity of branching enzymes is well-documented,[20-21] and therefore, I firstly aimed to optimise this assay for HTS of MtGlgB.

3.2   Development of an amylose-iodine assay for HTS of MtGlgB

The amylose-iodine assay is the most attractive for HTS because no other steps are required and the absorbance can be monitored directly in a microplate. The first step was to optimise the amylose and iodine components of the assay.

3.2.1   Development of a MtGlgB amylose-iodine benchtop assay

The first consideration was the amylose substrate. Amylose has a very low solubility in aqueous solution, with insolubility increasing with chain length.[22] Assays to monitor BE activity typically use partially solubilised long-chain amyloses that are highly viscous and likely to precipitate.[20] This would result in erroneous absorbance readings, and has the potential to block HTS equipment. Long-chain amyloses were therefore incompatible with HTS. It was previously shown that MtGlgB can act on amylose-2800, a shorter, more soluble type of amylose. This has a mean DP of 12.6 ± 5.3 and, after MtGlgB treatment and subsequent debranching, chains with DP 6-9 increased, whilst chains with DP > 15 were depleted, demonstrating the branching activity of the enzyme.[23] I determined that 2 mg/mL amylose-2800, that was solubilised by heating, would remain in solution for over 48 h, giving a suitable window for screening many compounds.

The next consideration was the iodine detection reagent. Lugol's solution is an aqueous preparation of iodine named after the French chemist Jean Lugol, and has been used as a disinfectant or for the treatment for thyroid diseases since its discovery in 1829.[24] It is a 2:1 mixture of I_2 and KI (aq), to create a solution of iodide ions that is commonly used to investigate amylose.[25] Addition of Lugol's solution to amylose-2800 gives a purple/red colour with an absorbance maximum around 530 nm, **Figure 3.1** (green line). Addition of Lugol's solution to amylose-2800 that had been incubated with MtGlgB for 1 h gave a lower absorbance signal, as expected from the branching activity **Figure 3.1** (blue line). Relatively high concentrations of enzyme were required to see a clear shift, most likely reflecting that this is not an ideal substrate for GlgB, because chain lengths are around the minimum DP required for branching, ~DP 15.[23] Aqueous dilutions of Lugol's solution of 10-25% gave absorbance readings in a suitable range for microplate measurements.
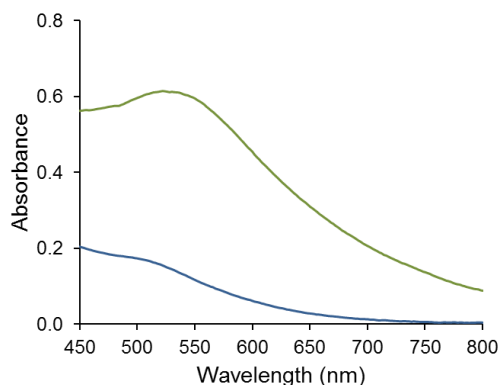
Figure 3.1 Amylose-2800 before and after incubation with MtGlgB. Absorbance spectra of amylose-2800 before (green line) and after (blue line) incubation with ~500 nM MtGlgB for 1 h. A 10% aqueous solution prepared from commercial Lugol's solution is used. Branching results in depletion of the absorbance signal and the loss of the peak at ~530 nm.

A titration of amylose-2800 with a set dilution of Lugol's solution demonstrated an approximately linear relationship between amount of amylose-2800 and absorbance, **Figure 3.2A**. Once an aqueous dilution of Lugol's solution was prepared, it gave consistent initial absorbance readings when added to the same preparation of amylose-2800 every hour, indicating that the diluted Lugol's solution is stable over this time-period, **Figure 3.2B**. However, after initial addition of Lugol's solution, I observed that the absorbance signal decreased over time. I determined that this decrease was linear, **Figure 3.2C**, and was unaffected by other factors, such as light or temperature. Therefore, this signal drift can be accounted for in the assay protocol by recording absorbance measurements at a set timepoint after the addition of Lugol's solution.

Time-course experiments demonstrated that the addition of Lugol's solution was sufficient to stop the enzymatic reaction, and at 22 °C the optimal assay window was achieved after a 90 min incubation period. However, I found that when Lugol's solution was added directly, with no incubation period, the absorbance signal decreased with increasing enzyme concentration. This suggested that the presence of enzyme was affecting the amylose-iodine interaction independently of the enzymatic activity. This could be due to amino acid residues such as lysine interacting with iodine.[26] A heat-inactivated preparation of the enzyme was found to serve as a convenient control that would account for these interactions, **Figure 3.2D.** Both the active and heat-inactivated enzyme preparations were found to give consistent absorbance reading when stored on ice over a time-period of at least four hours, **Figure 3.2E**.
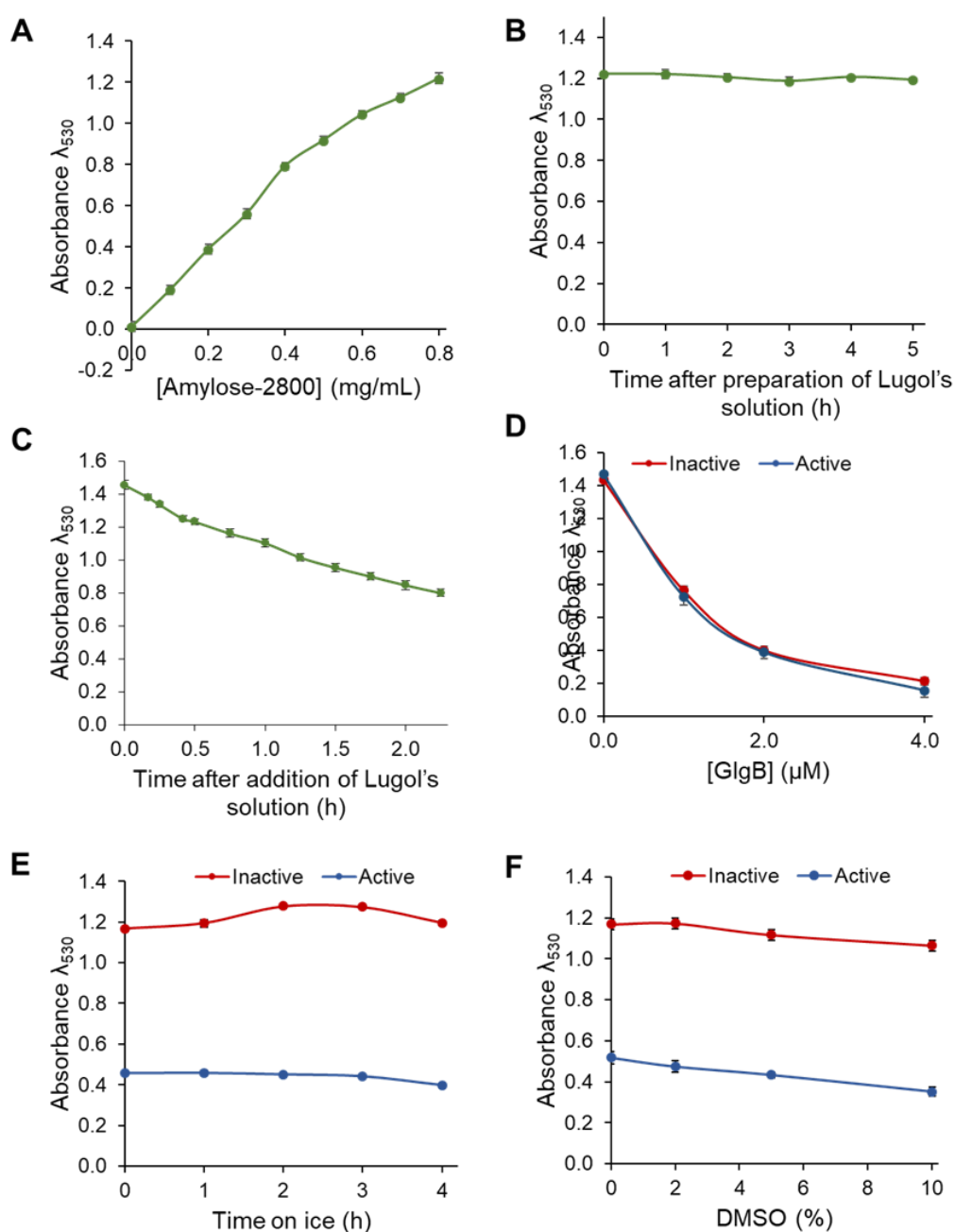
Figure 3.2 Investigation of amylose-iodine assay parameters. A) Titration of amylose-2800 with Lugol's solution**,** demonstrating a linear relationship. **B**) Initial absorbance readings from the addition of diluted Lugol's solution to amylose-2800, every hour after preparation, demonstrating stability of the diluted Lugol's solution over time. **C)** Depletion of the absorbance signal over time after the initial addition of Lugol's solution to amylose-2800. **D)** Comparison of absorbance at t(0) with either active or heat-inactivated MtGlgB for increasing concentrations of MtGlgB, suggesting that MtGlgB is reducing the absorbance signal independently of its enzymatic activity. **E)** Absorbance readings at t(90) with preparations of active and heat-inactivated MtGlgB stored on ice, demonstrating their stability. **F)** Effect of DMSO % on absorbance signal for active and heat-inactivated MtGlgB at t(90). Unless otherwise stated: 0.8 mg/mL amylose-2800, absorbance read at 530 nm, 30 min after addition of 10-25% (aq.) Lugol's solution, 22 °C. At least three repeats were carried out for all data sets.

Compound libraries for HTS are typically stored in DMSO and therefore the influence of DMSO on the assay was also tested. The presence of DMSO slightly decreased the amylose-iodine absorbance signal, **Figure 3.2F**, however up to 10% DMSO had no effect on the enzymatic activity of MtGlgB as the difference between the two signals remained the same. Therefore, a final concentration of 10% DMSO was used for testing compound inhibition with this assay.

Considering the variables discussed above, assay conditions were optimised on a 96-well plate format for 100 μL reaction volumes with the subsequent addition of 50 μL aq. Lugol's solution as the detection reagent. Final enzyme reactions comprised 0.8 mg/mL amylose-2800, 860 nM GlgB (active or heat-inactivated) and 10% DMSO, in 100 μL. After 90 min incubation at 22 °C, 50 μL of 20% aq. Lugol's solution was added with mixing, and after a further 30 min incubation at 22 °C, the absorbance was recorded at 530 nm. The approximate K_m for MtGlgB under these conditions was determined as 2.2 mg/mL, **Figure 3.3**. However, the limited solubility of amylose-2800 precluded more accurate calculations because the full extent of the hyperbolic curve could not be determined.

Finally, the overall assay quality was assessed according to the Z' values, defined as:[27]

$$Z' = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|}$$

where σ_p and σ_n = the standard deviation of the positive and negative controls, respectively, and μ_p and μ_n = the mean of the of the positive and negative controls, respectively. A value of Z' = 1 represents an ideal assay, 1 > Z' ≥ 0.5 represents an excellent assay, 0.5 > Z' > 0 represents a marginal assay and Z < 0 means that the assay is essentially unusable.

Values for the assay described above were consistently ≥ 0.5, demonstrating it was an excellent assay with a large signal to noise ratio, hence it could be taken forward for HTS.
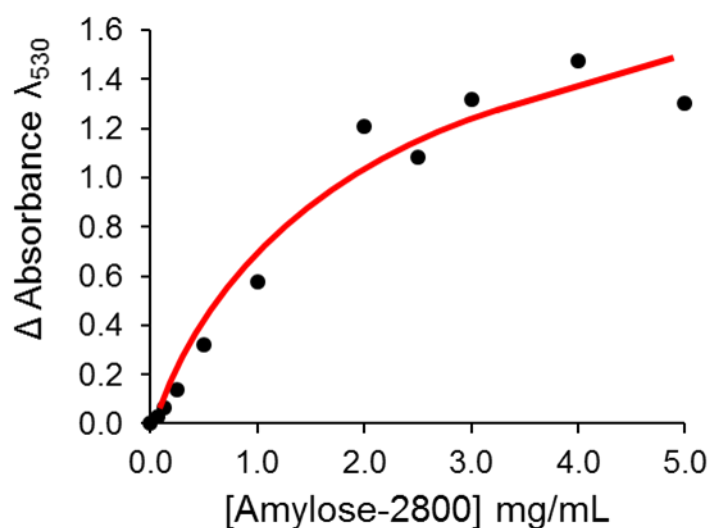
Figure 3.3 MtGlgB has a K_m of ~2.2 mg/mL with amylose-2800 under assay condtions. Determined by the difference between absorbance at 530 nm in positive (active MtGlgB) and negative (heat-inactivated MtGlgB) controls, with increasing concentrations of amylose-2800. In 100 μL, 860 nM MtGlgB was added to amylose-2800 with 10% DMSO. After 90 min incubation at 22 °C, 50 μL 20% aq. Lugol's solution was added and the absorbance was measured after a further 30 min at 22 °C. As the full extent of the hyperbolic curve could not be determined, a K_m value was estimated using Sigmaplot. Data were recorded in triplicate.

3.2.2 Comparison with published MtGlgB amylose-iodine assay

During the course of this work, Dkhar *et al.* published a study in which potential inhibitors of MtGlgB were identified by virtual screening. The inhibition of MtGlgB was tested experimentally and some compounds were also shown to inhibit the growth of *M. tuberculosis* H37Rv.[2] Due to the lack of counter-screens or other methods to verify compound binding, questions remained about the conclusions of this work. To investigate this further, I replicated the amylose-iodine assay reported in the Dkhar *et al.* study, which utilised a long-chain amylose substrate that was solubilised in 1 M NaOH and neutralised with 1 M HCl. According to the protocol, this was incubated with MtGlgB for up to 30 min, then a preparation of iodine solution which contained HCl was added. Under these conditions, I observed little difference between positive and negative controls for reaction times up to 1 h, with corresponding Z' values of ~0. Furthermore, upon the addition of the acidic iodine solution a significant amount of amylose precipitation was observed, precluding accurate absorbance readings. Different preparations of MtGlgB gave similar results, and increased enzyme concentrations and/or reaction times made no significant difference. I therefore concluded that the amylose-iodine assay reported by Dkhar *et al.* was not suitable for accurate evaluation of MtGlgB inhibition.

3.2.3   Automation of amylose-iodine MtGlgB assay

For screening the compound libraries at MRCT, it was necessary to miniaturise the assay to a 384-well plate format and develop an automated method. The assay volume was reduced to 20 µL and the assay protocol was adjusted so that enzyme and substrate solutions were mixed in approximately equal volumes to reduce errors resulting from the transfer of small volumes. Standard polystyrene 384-well plates were compared with Nonbinding Surface plates, which have been treated to prevent protein binding to the plastic. No differences between the two were found and therefore standard plates were used. Automation methods for the assay were developed for Biomek (Beckman Coulter) and Flexidrop (Perkin Elmer) instruments in collaboration with Puneet Khurana, MRCT. Several steps were required to optimise the transfer of the amylose-2800 solution, which tended to result in plate-striping due to its viscosity. The final method was again verified by calculating Z' values, which were in the range of 0.7 – 0.8, demonstrating an excellent signal:noise ratio.

For testing compounds, a 20 min pre-incubation of the enzyme and compound before substrate-addition was used, to allow for slow binders. Pilot runs were carried out with a set of about 1,000 compounds (MRCT) at 10 µM, which gave a hit rate of 0.14%, indicating the assay was unlikely to give a high number of false positives. However, because this could also indicate that the assay had a low sensitivity I increased the concentration of compounds in further screening to 30 µM. The final parameters for screening are shown in **Table 3.1**.

Table 3.1 Optimised amylose-iodine assay conditions for HTS of MtGlgB.
Concentrations are given for the enzyme reaction in 20 µL before the addition of 10 µL Lugol's solution.

Component	Condition
Substrate	0.8 mg/mL amylose-2800
Enzyme	860 nM recombinant MtGlgB
DMSO	10% (v/v)
Compound	30 µM
Pre-incubation time	20 min
Reaction time	90 min
Detection reagent	20% Lugol (aq.), 30 min incubation
Absorbance wavelength	530 nm

3.3 Screening compound libraries against MtGlgB

3.3.1 Screening MRCT compound libraries

The optimised assay protocol, **Table 3.1**, was used to screen MRCT compound libraries consisting of 13,472 compounds, at 30 µM. These were divided over 44 384-well plates, with each plate normalised to the positive and negative controls. The assay parameters were monitored over the course of the screening process and the Z' values were maintained at 0.7 – 0.8 throughout. In total, there were 17 compounds with a percentage inhibition over 25%, **Figure 3.4**, representing a hit rate of 0.13%. These compounds, alongside any library compounds with a similarity score over 0.5 were rescreened to identify consistent hitters. Two consistent hitters were identified, compounds **3a** and **3b**, **Figure 3.5**. The kinetic solubility of these were determined as 212 µM and 5 µM, respectively, by David Tickle (MRCT), indicating only **3a** was suitable for further investigation. Counter-screens in the absence of enzyme demonstrated that compound **3a** did not interfere with the amylose-iodine interaction independently of MtGlgB.
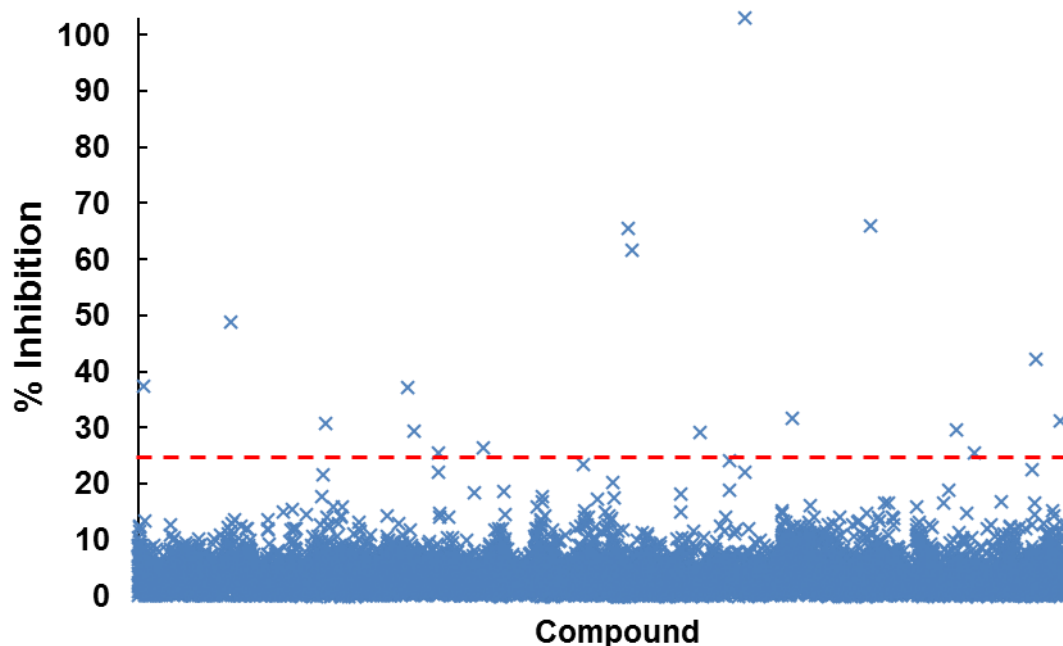
Figure 3.4 Inhibition of MtGlgB by compounds from MRCT libraries. Compounds (×) are plotted on the x-axis according to MRCT compound number (arbitrary, numbers not shown). Calculated % inhibition values lower than 0 are not shown. The cut-off for further testing at ≥25% inhibition is illustrated with a red dashed line.

3a   **3b**
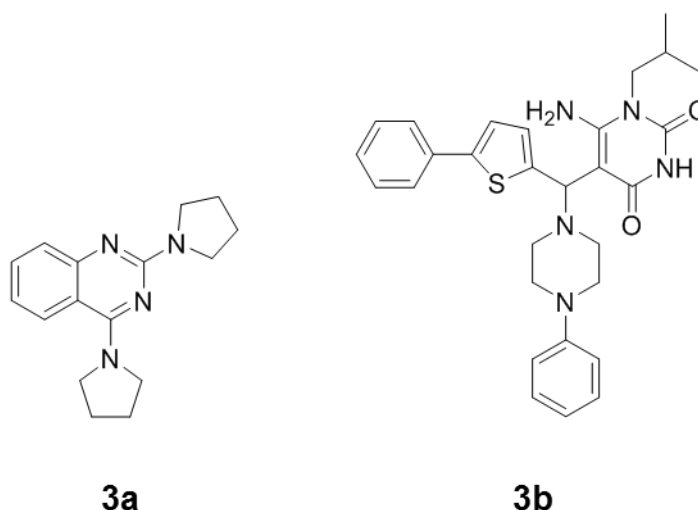
Figure 3.5 Hit Compounds identified from screening of MtGlgB. Compound **3b** was excluded due to low solubility, whereas **3a** was investigated further.

3.3.2   Screening compounds from open-source GSK libraries

Recently, Glaxo-Smith-Kline made available two sets of potent *M. tuberculosis* H37Rv inhibitors that have favourable properties but unknown targets.[28-29] These consisted of 227 compounds, which were screened at 50 µM with the amylose-iodine assay in a 96-well plate format. No inhibitors of MtGlgB were identified.

3.4   Validation of hit compound 3a

3.4.1 IC_{50} estimation of 3a

Firstly, I attempted to determine the IC_{50} of **3a**. Due to the solubility limit of the compound it was not possible to prepare a stock solution that gave 100% inhibition of MtGlgB. An estimated IC_{50} value was therefore calculated as 71 µM by assuming 100% inhibition at an arbitrary high concentration, **Figure 3.6**. This estimate is comparable with the IC_{50} and K_i values of the published GlgE inhibitors and in the expected range for a primary hit from a screen.
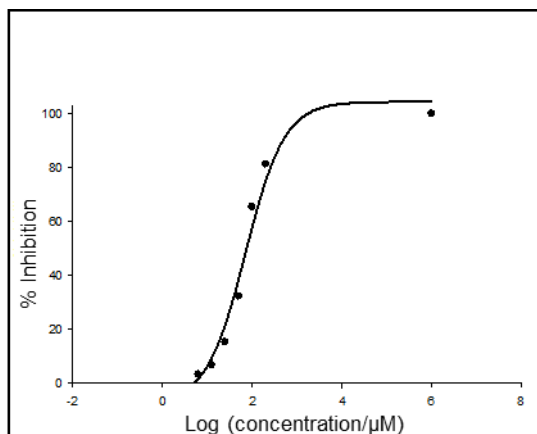
Figure 3.6 Compound 3a has an approximate IC$_{50}$ of 71 μM. The percentage inhibition was determined with the standard amylose-iodine assay conditions in a 96-well plate format, for increasing concentrations of **3a**. An approximate IC$_{50}$ value was calculated using Sigmaplot, with an artificial 100% inhibition data point added. Data recorded in triplicate.

3.4.2 Isothermal titration calorimetry verifies the binding of 3a to MtGlgB

Whilst the counter-screen data indicate that **3a** does not appear to interfere with the iodine-amylose interaction in the absence of enzyme, it cannot be ruled out that it has an indirect effect, for example by sequestering the amylose-2800 substrate such that it is less accessible to MtGlgB. To determine whether **3a** interacted directly with GlgB, several methods to investigate protein-ligand interactions were explored, including surface plasma resonance (SPR). However, this was found to be not suitable for measuring the interaction because the size difference between an immobilized protein (~83 kDa) and ligand (268 Da) is too great for detection of a μM binding interaction. On the other hand, isothermal titration calorimetry (ITC) measurements are independent of size and ITC is considered the gold standard for the measurement of protein-ligand interactions.[30]

ITC experiments showed a clear binding interaction between **3a** and MtGlgB with an average molar binding ratio of 0.89, suggestive of 1:1 binding, and an average K_d of 3.1 ± 0.7 μM (n = 3); representative data shown in **Figure 3.7**. Protein concentrations of 10 μM – 50 μM were tested, with ratio of ligand:protein 10:1 – 16:1. Control experiments with buffer titrated into MtGlgB and **3a** titrated into buffer or control proteins such as BSA, showed no heat change, confirming that this binding interaction was genuine. The shape of the curve and binding heats were consistent with a low value of the Wiseman "c" parameter,[31] which ranged from 3-10.[32] The average calculated thermodynamic energies for the binding were: ΔH: -3.30 kcal/mol, ΔS: 13.5 cal/mol/deg; -TΔS: -3.98 kcal/mol, to give ΔG: -7.28 kcal/mol, suggestive of binding dominated by hydrophobic interaction. Thus, **3a**, is confirmed as a true binder of MtGlgB, with a binding affinity significantly lower than that suggested by the

estimated IC_{50}, representing a reasonable start for compound development. The binding ratio is approximately 1:1 with the protein, suggestive of a single binding site.
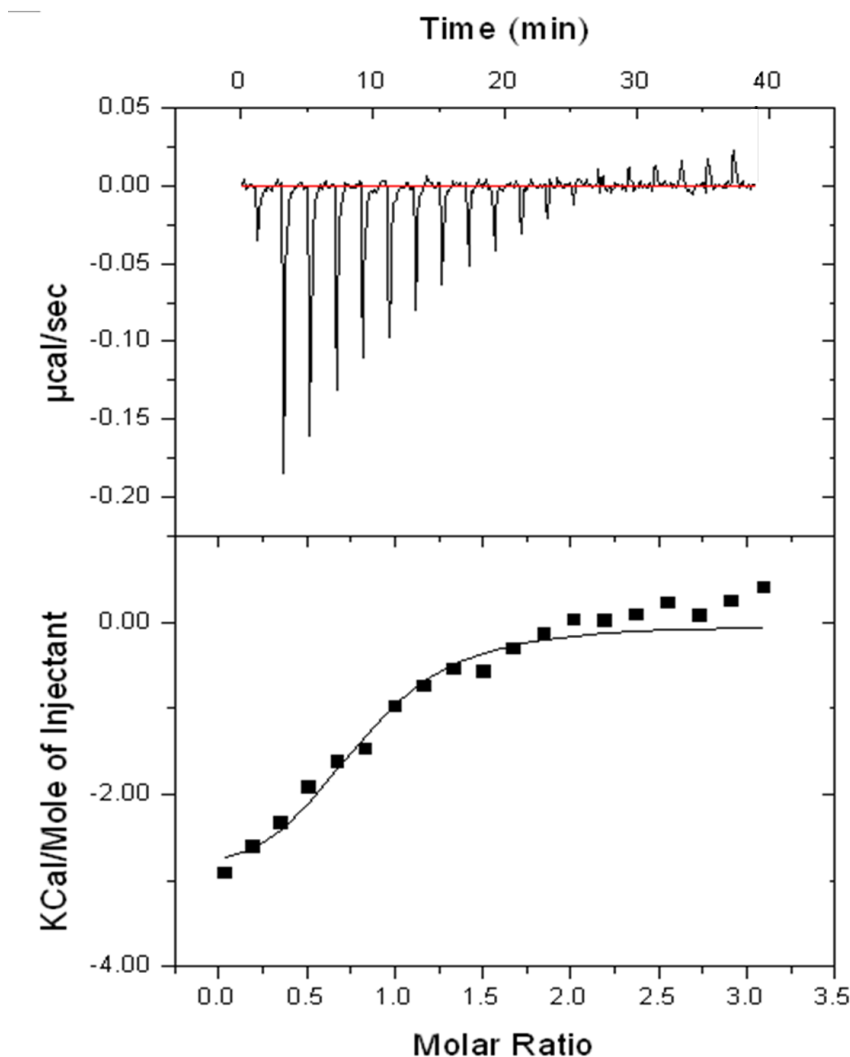
Figure 3.7 ITC confirms that 3a binds to MtGlgB. 3a binds to MtGlgB with an average K_d =3.1 ± 0.7 µM and 0.89 molar ratio, suggestive of ~1:1 binding. A representative dataset is shown here with a calculated K_d = 3.9 µM and a molar ratio of 0.80.

3.4.3 Compound 3a inhibits MsGlgB *in vitro*

To gain insight into the cross-species spectrum of **3a** inhibition, I investigated the inhibition of MsGlgB. The benchtop amylose-iodine assay detailed in section **3.2.1** was adapted slightly to account for the slightly higher activity observed for MsGlgB, compared to MtGlgB. As a modified MsGlgB protein lacking the catalytic residues was available, direct comparison could be made between this and the heat-inactivated control. Absorbance measurements were found to be very similar in both negative controls, confirming each to be valid. The percentage inhibition of *M. smegmatis* and MtGlgB with 50 µM **3a** was then

determined under comparable conditions. The average percentage inhibition was 41% for MtGlgB and 36% for MsGlgB, demonstrating that **3a** inhibits MsGlgB similarly, **Figure 3.8**.
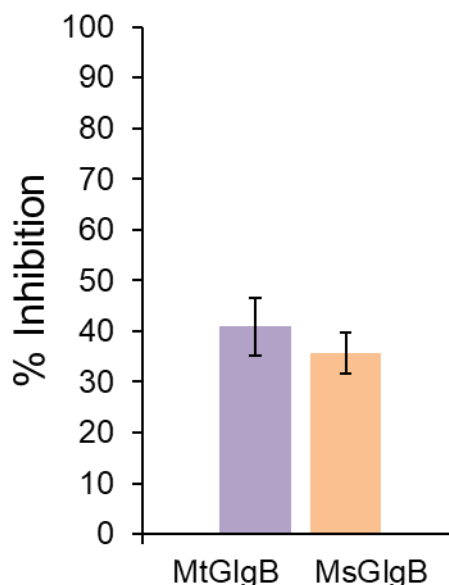
Figure 3.8 The compound 3a inhibits both MtGlgB and MsGlgB *in vitro*. Assay conditions were as described for the benchtop amylose-iodine MtGlgB assay in section **3.2.1**. At least three repeats were carried out with each enzyme.

3.5   Co-crystallisation trials of 3a and MsGlgB

Since **3a** also inhibited MsGlgB and crystals of this enzyme were readily available (as described in **Chapter 2**) I attempted to determine inhibitor-bound crystal structures. Initially co-crystallisation trials were explored, whereby MsGlgB and **3a** were combined and pre-incubated for 1 h, before being mixed with precipitant solution and dispensed in vapour-diffusion sitting drop plates. This either prevented crystal formation, or resulted in small needle-like crystals that were not suitable for diffraction experiments. To avoid this, I generated MsGlgB crystals first, and then soaked them for up to 2 h with **3a**, which was diluted in precipitant solution to give a final concentration of DMSO ≤ 5%. The crystals were collected, frozen and subject to X-ray diffraction experiments, as described previously. Crystals diffracted to resolutions of around 2 Å and the structure was solved as before. In all cases, no additional electron density that could correspond to **3a** was detected.

To increase the concentration of **3a** in the crystal soak, I tested the tolerance of the MsGlgB crystal to increased concentrations of DMSO. I found that the MsGlgB crystals could tolerate a 1:1 DMSO:precipitant solution for up to 3 h before they could be seen to degrade. I then prepared 1:1 mixtures of **3a** in DMSO and precipitant solution, to achieve final

concentrations of **3a** up to 50 mM. These were used for crystal soaks of up to 3 h and crystals were then used for diffraction experiments as above. Again, no additional electron density that could correspond to **3a** was detected and I concluded that it would not be possible to obtain a structure of **3a** bound to MsGlgB by this approach.

3.6   ITC shows that compound MB16695 does not bind to MtGlgB

In section **3.2.2** I investigated the amylose-iodine MtGlgB assay previously described by Dhkar *et al.*[2] and found that it was not possible to distinguish between positive and negative controls with their protocol. This casts doubt on the validity of the MtGlgB inhibition values reported in this study, therefore I investigated the interaction between one of the putative inhibitors and MtGlgB in more detail. Compound MB16695 was reported to inhibit MtGlgB by 36.5% at a concentration of 3 nM, with a reaction time of 10 min. Furthermore, it was stated to have a minimum inhibitory concentration (MIC) of 15.6 µg/mL in growth inhibition assays with *M. tuberculosis* H37Rv, which was assumed to be due to it targeting MtGlgB in the cell.

Firstly, I tested the inhibition of MtGlgB by MB16695 with the benchtop amylose-iodine assay described in section **3.2.1**. I did not observe any inhibition of MtGlgB up to 50 µM MB16695, and above this concentration there was significant bleaching of the purple colour. I then investigated a possible binding interaction between MB16695 and MtGlgB using ITC, which had previously been used to confirm the binding of **3a** to MtGlgB. Under similar conditions I saw no binding curve for MB16695, **Figure 3.9**. These data suggest that MB16695 does not inhibit MtGlgB and therefore the growth inhibition of *M. tuberculosis* H37Rv by this compound was not specific for the MtGlgB target. Further investigation is required to determine whether the other inhibitors reported in this study truly inhibit MtGlgB.
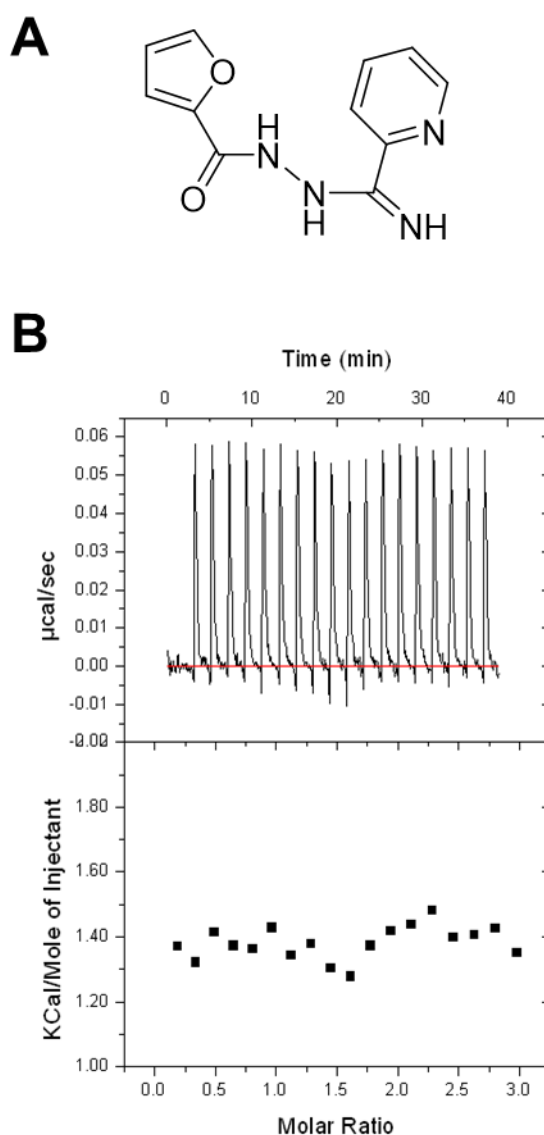
Figure 3.9 MB16695 does not bind to MtGlgB. A) Structure of MB16695. **B)** No binding response was observed for MB16695 and MtGlgB. ITC experiments were carried out as described above.

3.7 Discussion

In this chapter I aimed to identify small-molecule inhibitors of MtGlgB using HTS, which was achieved by the identification of **3a**. This firstly required the establishment of a HTS assay for MtGlgB. The branching enzyme reaction can be challenging to monitor, because substrates and products are indistinguishable by mass or functional group, and there is no release of side-products.[5] The amylose-iodine assay takes advantage of the difference in absorbance with iodine between branched products and unbranched substrates. Whilst this method has a long history of use for amylolytic enzymes and branching enzymes,[25] this is the first time the assay parameters have been systematically investigated to give an assay

with high Z' values that can be used to screen 1,000s of compounds in quick succession. Colorimetric assays are usually preferred for screening due to their inherent simplicity, however there are some drawbacks that must be considered. Most significant is the interference of compounds with absorbance measurements, independent of interactions with the enzyme;[33] a particular problem for coloured compounds. Indeed, a number of the compounds identified as false positives were coloured compounds that artificially increased absorbance measurements at 530 nm. Likewise, it cannot be ruled out that there were false negatives that artificially reduced absorbance readings, for example by interacting with the iodine reagent. This is an important consideration for all colorimetric assays and must be taken into account when taking this assay forward.

In the past, amylose-iodine assays have been perceived as non-quantitiative.[7] This work may encourage other researchers that the amylose-iodine assay has utility as a quantitative method for investigating branching enzymes, as long as the appropriate controls are implemented. For example, previous assays have not checked for the potential interaction of the enzyme with the iodine reagent, was seen here for MtGlgB with Lugol's solution. This is important to consider as would lead to enzyme activities being overestimated. One caveat for using this assay with other branching enzymes is that the reported minimum substrate specificity tends to be much longer than the DP range of amylose-2800. However, as for MtGlgB, it may be the case that whilst the shorter chain amylose is not an ideal substrate, it is sufficient in the context of the assay.

 A previously published amylose-iodine assay for MtGlgB[2] was not reproducible in my hands and the long-chain amylose led to significant precipitation. The amylose-2800 substrate was used because it is more soluble than higher molecular weight amyloses that are typically utilised, however the overall aqueous solubility is still limited. The K_m of amylose-2800 with MtGlgB under the assay conditions was estimated as 2.2 mg/mL, with more accurate determination precluded by the limited solubility of amylose-2800 at high concentrations. The substrate concentration used in the assay was 0.8 mg/mL, approximately three-fold lower than K_m, with a higher concentration again prohibited by the limited solubility. Typically, a substrate concentration at or close to the K_m is used, however, a substrate concentration below K_m still allows for the identification of competitive inhibitors, in addition to uncompetitive and mixed inhibitors.[9] The substrate concentration therefore does not limit the capacity to identify inhibitors with this assay.

Compound **3a** was identified as a MtGlgB inhibitor and the binding of the compound to the enzyme was confirmed with ITC. There was a difference between the estimated IC_{50} value

of 71 µM determined with the assay and the K_d of 3.1 ± 0.7 µM determined by ITC. As the IC_{50} value was an estimation due to the limited solubility of **3a**, the true value may well be closer to the K_d. The relationship between the IC_{50} and the K_d depends on the type of inhibition and the substrate concentration. Nevertheless, IC_{50} is always greater than or equal to K_d, as was seen here.

Since an inhibitor-bound crystal structure could not be obtained, the binding site of **3a** remains unknown. ITC data suggest binding dominated by hydrophobic interaction. Since **3a** has a planar, quinazoline scaffold, it is possible to speculate that it interacts via CH(π) stacking with an aromatic residue such as phenylalanine, tyrosine or tryptophan. A number of these residues interact with maltooligosaccharides in the ligand-bound MsGlgB structures, hence this could be a mechanism to inhibit the action of the enzyme. Knowledge of the binding site will be invaluable to the future development of **3a**, as this will allow rational design of **3a** analogues with improved potency. As efforts to identify this with crystallography have been unsuccessful thus far, other methods such as STD-NMR could be utilised.[34]

I also investigated the interaction of MB16695 with MtGlgB, since this compound had previously been reported to inhibit the enzyme at low nM concentrations.[2] Conversely, I saw no inhibition of MtGlgB by MB16695 in my amylose-iodine assay, but did observe bleaching of the colour and hence lower absorbance readings at concentrations of MB16695 > 50 µM. Since this would actually give a false-negative result, it does not explain the reported inhibition in the Dkhar *et al.* study, especially as this was for much lower concentrations of MB16695. ITC data confirmed that MB16695 does not bind to MtGlgB. This finding, in addition to the unreproducible assay, suggests that the inhibition of MtGlgB reported for other compounds may also not be valid. Whilst there is no suggestion that there are inaccuracies in the *M. tuberculosis* H37Rv growth inhibition studies, the observed inhibition must be enacted via targets other than MtGlgB.

These findings indicate that compound **3a** represents the first true small-molecule inhibitor of MtGlgB. The next steps are the investigation of structure-activity relationships and *in vivo* activity: these are explored in Chapter **4**.

3.8 Summary

I developed an amylose-iodine assay for HTS of compounds against MtGlgB. This was used to screen compound libraries of approximately 13,500 compounds. Compound **3a** was validated as an inhibitor of MtGlgB, and ITC experiments demonstrated a 1:1 molar ratio with a low µM K_d. This will be further investigated with medicinal chemistry and *in vivo* assays in the following chapter.

3.9 Materials and methods

Methods for protein production and purification, crystal soaks of MsGlgB, data collection and refinement were carried out as described in section **2.7**.

3.9.1 Amylose-iodine assay

Amylose-2800 was purchased from Tokyo Chemical Industry. Compound **3a** was purchased from Specs. Compound MB16695 was purchased from Ambinter. Lugol solution, amylose from potato (av. MW 150,000) DMSO and all other chemicals were purchased from Sigma Aldrich. 96-well, 384-well or 384-well Nonbinding Surface flat-bottom microplates were purchased from Corning. Amylose and MtGlgB stock solutions and dilutions were prepared in 20 mM sodium phosphate buffer, pH 7.0, containing 50 mM NaCl (henceforth referred to as phosphate buffer), Lugol solution was diluted in ultrapure water (Milli-Q). The heat-inactivated MtGlgB was prepared by heating at 90 °C for 5 min. Absorbance measurement were recorded using a CLARIOstar® plate reader (BMG LabTech) and data were analysed in Microsoft Excel (Microsoft Office) or SigmaPlot (Systat Software Inc.).

3.9.2 Screening conditions

Compound libraries were provided by MRCT and included a range of compounds from commercial suppliers: Maybridge (Trevillett, UK), Interbioscreen (Moscow, Russia), SPECs (Delft, Netherlands), Peakdale (Chapel-en-le-Frith, UK) and Biofocus (Chesterford Pk, UK). Library compounds were solubilised in DMSO at 1 mM in 384-well plates in preparation for screening and dilutions were prepared in DMSO. Automation methods were developed using Biomek (Beckman Coulter) and Flexidrop (Perkin Elmer) instruments in collaboration with Puneet Khurana, MRCT.

3.9.3   Kinetic solubility

The kinetic solubilities of **3a** and **3b** were determined by David Tickle (MRCT). A stock solution in DMSO was diluted with 1 × PBS (Sigma Aldrich) and the concentration at which precipitate is first observed was recorded

3.9.4   Isothermal titration calorimetry

ITC experiments were carried out using a MicroCal VP-ITC instrument (Malvern) and data were analysed using Origin® (Malvern). Aliquots of MtGlgB were thawed on ice, combined, thoroughly mixed and centrifuged, before the determination of the protein concentration using a Direct Detect® infrared spectrometer (Merck) and a Nanodrop spectrophotometer (Nanodrop®). Dilutions were made as appropriate with the same batch of phosphate buffer the protein was stored in. Protein solutions were allowed to warm to 25 °C shortly before initiating ITC experiments, which were all carried out at this temperature. Compound **3a** was solubilised in the same batch of phosphate buffer at pH 7.0, using sonication and heating. The compound solution was monitored by eye and by absorbance measurements to check for precipitation over the course of the experiments. Typically, experiments consisted of 19 injections of 2 μL every 2 min, with MtGlgB in the cell and **3a** in the syringe. Protein concentration ranged from 10 – 50 μM, with the ratio to ligand ranging from 1:10 to 1:16. After experiments were complete, the cell solution was also checked for precipitation. Control experiments consisted of injections of phosphate buffer into MtGlgB, **3a** into buffer and **3a** into control proteins such as BSA, all of which resulted in consistent heat changes that were not suggestive of binding.

3.10   References

1. Kalscheuer, R.; Syson, K.; Veeraraghavan, U.; Weinrick, B.; Biermann, K. E.; Liu, Z.; Sacchettini, J. C.; Besra, G.; Bornemann, S.; Jacobs, W. R., Self-poisoning of *Mycobacterium tuberculosis* by targeting GlgE in an α-glucan pathway. *Nat. Chem. Biol.* **2010,** *6* (5), 376-384.

2. Dkhar, H. K.; Gopalsamy, A.; Loharch, S.; Kaur, A.; Bhutani, I.; Saminathan, K.; Bhagyaraj, E.; Chandra, V.; Swaminathan, K.; Agrawal, P.; Parkesh, R.; Gupta, P., Discovery of *Mycobacterium tuberculosis* α-1,4-glucan branching enzyme (GlgB) inhibitors by structure- and ligand-based virtual screening. *J. Biol. Chem.* **2015,** *290* (1), 76-89.

3. Pal, K.; Kumar, S.; Sharma, S.; Garg, S. K.; Alam, M. S.; Xu, H. E.; Agrawal, P.; Swaminathan, K., Crystal structure of full-length *Mycobacterium tuberculosis* H37Rv glycogen branching enzyme: Insights of N-terminal β-sandwich in substrate specificity and enzymatic activity *J. Biol. Chem.* **2010,** *285* (27), 20897-20903.

4. Binderup, K.; Libessart, N.; Preiss, J., Slow-binding inhibition of branching enzyme by the pseudooligosaccharide BAY e4609. *Arch. Biochem. Biophys.* **2000,** *374* (1), 73-78.

5. Sinnott, M.; Williams, A.; Page, M., *Carbohydrate chemistry and biochemistry: Structure and mechanism*. Royal Society of Chemistry: 2013.

6. Takeda, Y.; Guan, H.-P.; Preiss, J., Branching of amylose by the branching isoenzymes of maize endosperm. *Carbohydr. Res.* **1993,** *240*, 253-263.

7. Utsumi, Y.; Yoshida, M.; Francisco, J. P. B.; Sawada, T.; Kitamura, S.; Nakamura, Y., Quantitative assay method for starch branching enzyme with bicinchoninic acid by measuring the reducing terminals of glucans. *J. Appl. Glycosci.* **2009,** *56* (3), 215-222.

8. Nakamura, Y.; Sawada, T.; Ohdan, T.; Aihara, S.; Fujita, N., New assay method for starch branching enzyme and starch synthase by the chain-length distribution analysis. *J. Appl. Glycosci.* **2011,** *58* (3), 119-123.

9. Brooks, H. B.; Geeganage, S.; Kahl, S. D.; Montrose, C.; Sittampalam, S.; Smith, M. C.; Weidner, J. R., *Basics of enzymatic assays for HTS*. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences: 2012.

10. Manners, D. J.; Stark, J. R., α-(1→4)-D-Glucans. Part XXII. The iodine staining properties of linear maltosaccharides. *Starch/Stärke* **1974,** *26* (3), 78-81.

11. Colin, J.-J.; Gauthier de Claubry, H.-F., Mémoire sure les combinaisons de l'iode avec les substances végétales et animales. *Ann. Chim Ser.* **1814,** *90*, 87-100.

12. Rundle, R. E.; Baldwin, R. R., The configuration of starch and the starch-iodine complex. I. The dichroism of flow of starch-iodine solutions. *J. Am. Chem. Soc.* **1943,** *65* (4), 554-558.

13. Rundle, R. E., The configuration of starch in the starch-iodine complex. 5. Fourier projections from X-ray diagrams. *J. Am. Chem. Soc.* **1947,** *69* (7), 1769-1772.

14. Teitelbaum, R. C.; Ruby, S. L.; Marks, T. J., A resonance Raman-iodine mossbauer investigation of the starch-iodine structure - aqueous solution and iodine vapor preparations. *J. Am. Chem. Soc.* **1980,** *102* (10), 3322-3328.

15. Yu, X.; Houtman, C.; Atalla, R. H., The complex of amylose and iodine. *Carbohydr. Res.* **1996,** *292* (0), 129-141.

16. Nimz, O.; Gessler, K.; Uson, I.; Laettig, S.; Welfle, H.; Sheldrick, G. M.; Saenger, W., X-ray structure of the cyclomaltohexaicosaose triiodide inclusion complex provides a model for amylose-iodine at atomic resolution. *Carbohydr. Res.* **2003,** *338* (9), 977-986.

17. Madhu, S.; Evans, H. A.; Doan-Nguyen, V. V. T.; Labram, J. G.; Wu, G.; Chabinyc, M. L.; Seshadri, R.; Wudl, F., Infinite polyiodide chains in the pyrroloperylene–iodine complex: Insights into the starch–iodine and perylene–iodine complexes. *Angew. Chem.* **2016,** *128* (28), 8164-8167.

18. Bailey, J. M.; Whelan, W. J., Physical properties of starch: I. Relationship between iodine stain and chain length. *J. Biol. Chem.* **1961,** *236* (4), 969-973.

19. Archibald, A. R.; Fleming, I. D.; Liddle, A. M.; Manners, D. J.; Mercer, G. A.; Wright, A., 232. α-1,4-Glucosans. Part XI. The absorption spectra of glycogen- and amylopectin-iodine complexes. *J. Chem. Soc* **1961,** (0), 1183-1190.

20. Rydberg, U.; Andersson, L.; Andersson, R.; Åman, P.; Larsson, H., Comparison of starch branching enzyme I and II from potato. *Eur. J. Biochem.* **2001,** *268* (23), 6140-6145.

21. Roussel, X.; Lancelon-Pin, C.; Vikso-Nielsen, A.; Rolland-Sabate, A.; Grimaud, F.; Potocki-Veronese, G.; Buleon, A.; Putaux, J. L.; D'Hulst, C., Characterization of substrate and product specificity of the purified recombinant glycogen branching enzyme of Rhodothermus obamensis. *Biochim. Biophys. Acta* **2013,** *1830* (1), 2167-77.

22. Green, M. M.; Blankenhorn, G.; Hart, H., Which starch fraction is water-soluble, amylose or amylopectin? *J. Chem. Educ.* **1975,** *52* (11), 729.

23. 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., Assembly of α-glucan by GlgE and GlgB in mycobacteria and streptomycetes. *Biochemistry* **2016,** *55* (23), 3270-3284.

24. Pankaj, S.; Thakur, S. D.; Sarvesh; Dogra, V.; Vasishta, N. K.; Singh, M. M.; Gandhotra, V. K., Evaluation of efficacy, intrauterine safety and storage conditions of Lugol's iodine. *Indian J. Anim. Nutr.* **2012,** *33* (1), 71-73.

25. Tomasik, P.; Schilling, C. H., Complexes of starch with inorganic guests. *Adv. Carbohydr. Chem. Biochem.* **1998,** *53*, 263-343.

26. Kortagere, S.; Ekins, S.; Welsh, W. J., Halogenated ligands and their interactions with amino acids: Implications for structure–activity and structure–toxicity relationships. *J. Mol. Graphics Model.* **2008,** *27* (2), 170-177.

27. Zhang, J.-H.; Chung, T. D. Y.; Oldenburg, K. R., A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* **1999,** *4* (2), 67-73.

28. Ballell, L.; Bates, R. H.; Young, R. J.; Alvarez-Gomez, D.; Alvarez-Ruiz, E.; Barroso, V.; Blanco, D.; Crespo, B.; Escribano, J.; Gonzalez, R.; Lozano, S.; Huss, S.; Santos-Villarejo, A.; Martin-Plaza, J. J.; Mendoza, A.; Rebollo-Lopez, M. J.; Remuinan-Blanco, M.; Lavandera, J. L.; Perez-Herran, E.; Gamo-Benito, F. J.; Garcia-Bustos, J. F.; Barros, D.;

Castro, J. P.; Cammack, N., Fueling Open-Source Drug Discovery: 177 Small-Molecule Leads against Tuberculosis. *ChemMedChem* **2013,** *8* (2), 313-321.

29. Rebollo-Lopez, M. J.; Lelièvre, J.; Alvarez-Gomez, D.; Castro-Pichel, J.; Martínez-Jiménez, F.; Papadatos, G.; Kumar, V.; Colmenarejo, G.; Mugumbate, G.; Hurle, M.; Barroso, V.; Young, R. J.; Martinez-Hoyos, M.; González del Río, R.; Bates, R. H.; Lopez-Roman, E. M.; Mendoza-Losana, A.; Brown, J. R.; Alvarez-Ruiz, E.; Marti-Renom, M. A.; Overington, J. P.; Cammack, N.; Ballell, L.; Barros-Aguire, D., Release of 50 new, drug-like compounds and their computational target predictions for open source anti-tubercular drug discovery. *PLoS One* **2015,** *10* (12), e0142293.

30. Vega, S.; Abian, O.; Velazquez-Campoy, A., A unified framework based on the binding polynomial for characterizing biological systems by isothermal titration calorimetry. *Methods* **2015,** *76*, 99-115.

31. Turnbull, W. B.; Daranas, A. H., On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **2003,** *125* (48), 14859-14866.

32. Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N., Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **1989,** *179* (1), 131-7.

33. Reymond, J. L., *Enzyme Assays*. Wiley: 2006.

34. Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J., Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011,** *88* (7), 990-994.

35. Thon, V. J.; Khalil, M.; Cannon, J. F., Isolation of human glycogen branching enzyme cDNAs by screening complementation in yeast. *J. Biol. Chem.* **1993,** *268* (10), 7509-7513.

36. Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L., Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov.* **2007,** *6* (1), 29-40.

37. Sakharkar, K. R.; Sakharkar, M. K.; Chow, V. T. K., Biocomputational strategies for microbial drug target identification. In *New Antibiotic Targets*, Champney, W. S., Ed. Humana Press: Totowa, NJ, 2008; pp 1-9.

38. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L., How many drug targets are there? *Nat. Rev. Drug Discov.* **2006,** *5* (12), 993-996.

39. Yonath, A., Antibiotics targeting ribosomes: Resistance, selectivity, synergism, and cellular regulation. *Annu. Rev. Biochem.* **2005,** *74* (1), 649-679.

40. Soares da Costa, T. P.; Tieu, W.; Yap, M. Y.; Pendini, N. R.; Polyak, S. W.; Sejer Pedersen, D.; Morona, R.; Turnidge, J. D.; Wallace, J. C.; Wilce, M. C. J.; Booker, G. W.; Abell, A. D., Selective inhibition of Biotin Protein Ligase from *Staphylococcus aureus*. *J. Biol. Chem.* **2012,** *287* (21), 17823-17832.

41. Arya, T.; Reddi, R.; Kishor, C.; Ganji, R. J.; Bhukya, S.; Gumpena, R.; McGowan, S.; Drag, M.; Addlagatta, A., Identification of the molecular basis of inhibitor selectivity between the human and streptococcal type I methionine aminopeptidases. *J. Med. Chem.* **2015,** *58* (5), 2350-2357.

42. Kasbekar, M.; Fischer, G.; Mott, B. T.; Yasgar, A.; Hyvönen, M.; Boshoff, H. I. M.; Abell, C.; Barry, C. E.; Thomas, C. J., Selective small molecule inhibitor of the *Mycobacterium tuberculosis* fumarate hydratase reveals an allosteric regulatory site. *Proc. Natl. Acad. Sci. USA* **2016,** *113* (27), 7503-7508.

4. Structure-activity studies of MtGlgB inhibitor 3a

4.1   Introduction

The previous chapter describes the identification of a small-molecule inhibitor of MtGlgB from an amylose-iodine assay screen; compound **3a**, **Figure 4.1**. This was validated by ITC, which gave a K_d of 3.1 ± 0.7 µM and showed a 0.89 molar binding ratio, suggestive of a single protein binding site. Compound **3a** has a number of favourable properties, such as conforming to Lipinski's rule of 5,[1-2] a high weighted quantitative estimation of drug likeness (QED)[3] and a ligand efficiency greater than 0.3,[2] **Figure 4.1**. The quinazoline scaffold is also generally attractive as it is found in a number of marketed drugs.[4] To investigate **3a** further I addressed the question: Can structure-activity relationships be established with a selection of modified compounds? I aimed to investigate which features of **3a** were important for inhibition, with a view to improving the efficacy and/or solubility of **3a**.

I also wanted to address the question: How do **3a** and its analogues behave *in vivo*? I aimed to gain further insight into the activity of **3a** and synthesised compounds by testing their effect on *M. smegmatis* growth. As discussed previously, this is commonly used as a faster growing and non-pathogenic model for *M. tuberculosis*.[5-6] In Chapter 2, I demonstrated the high similarity of MtGlgB and MsGlgB and in Chapter 3, I showed that **3a** inhibits MtGlgB and MsGlgB by a similar amount. *M. smegmatis* is therefore a relevant model for the GlgB target.
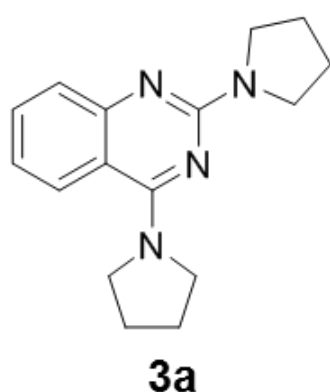
Molecular weight	268.36
Lipinski's Ro5	✓
QED (weighted)	0.781
LogP (predicted)	2.24 ± 0.76
K_d	3.1 ± 0.7 µM
Ligand efficiency	0.364

Figure 4.1 Drug-like properties of compound 3a. Structure of **3a** with properties calculated using DruLiTo.[1-3]

4.2   Chemical synthesis of 3a analogues

4.2.1   Selection of target compounds

The structure of compound **3a** comprises a quinazoline core, consisting of fused benzene and pyrimidine rings, with two pyrrolidine side groups at positions 2 and 4. The aim of chemical synthesis was therefore to create a range of compounds varied in both the aromatic core and amine side groups, **Figure 4.2**. Both the availability of commercial reagents and the predicted ease of synthesis were also considerations for the selection of target compounds. Since 2,4-dichloro pyrimidine compounds were commercially available and could theoretically be substituted at the 2 and 4 positions with amine groups via a nucleophilic addition/elimination,[7] these were chosen as the basis for the synthesised compounds.

Firstly, substitutions with a single amine group were explored, with the 4-amino 2-chloro compounds being more synthetically accessible. Then, to investigate both the effect of the ring size and cyclic nature of the amine group, pyrrolidine was substituted by either piperidine or butylamine. Further substitutions, to introduce more heteroatoms via piperazine and proline were considered, however these proved more challenging to synthesise, as discussed below. Finally, I wanted to investigate modifications to the quinazoline scaffold, by retaining the pyrimidine core without the additional benzene ring. I explored different pyrimidine substitutions with trifluromethane and thiophene. This proposed set of compounds would provide a good starting point for preliminary structure-activity studies of **3a**.
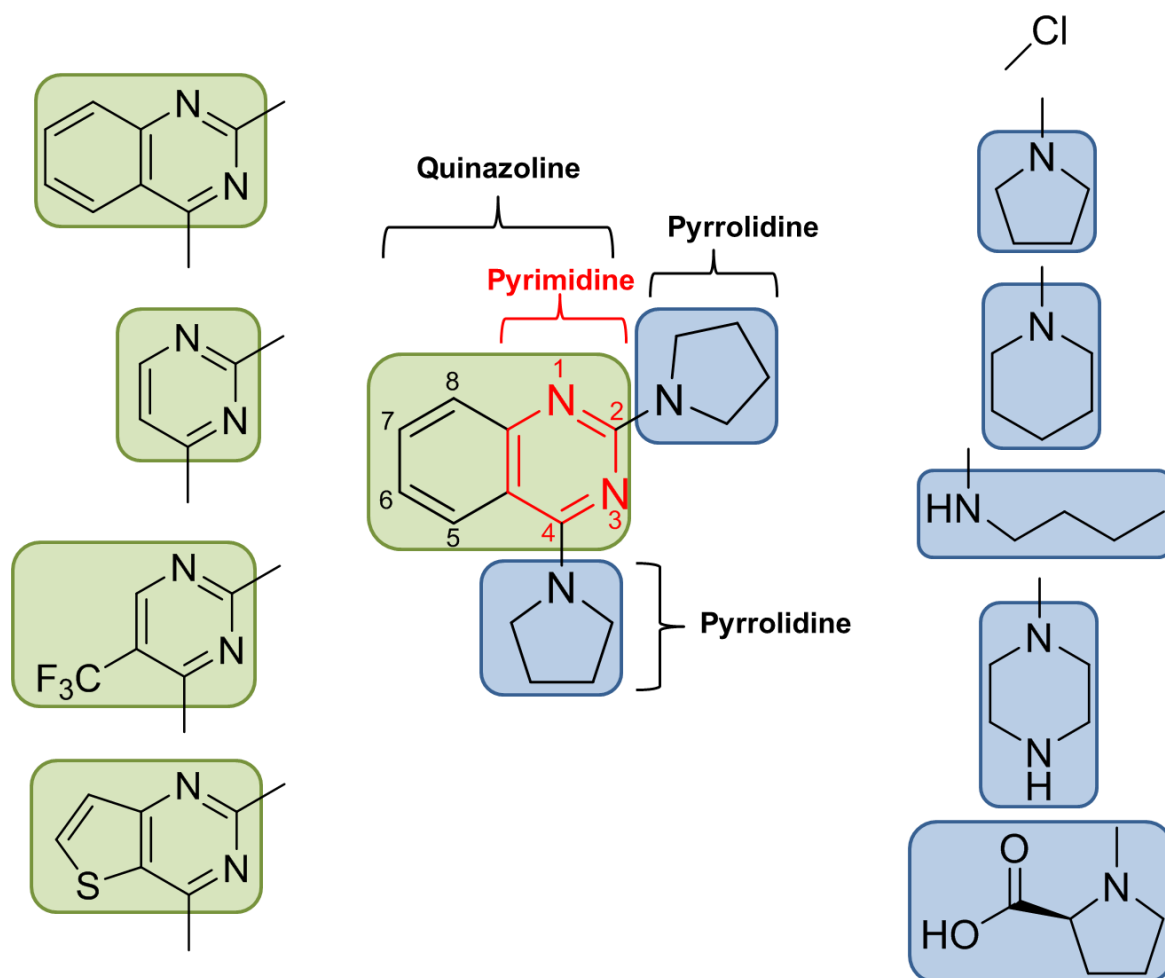
Figure 4.2 Proposed structural modifications for compound 3a. Compound **3a** comprises a quinazoline core consisting of fused pyrimidine and benzene rings (green box) with pyrrolidine side groups at the 2 and 4 positions (blue boxes). Proposed modifications of these groups are shown in green and blue boxes respectively.

4.2.2   Chemical synthesis

The general synthetic strategy was nucleophilic substitution of 2,4-dichloro pyrimidine compounds with simple aliphatic amines. The 4-position is known to be much more reactive with respect to nucleophilic substitution, compared to substitution at the 2-position.[8] This differential reactivity could be exploited in some cases, to sequentially add different amine groups to the pyrimidine scaffolds.

4.2.2.1 Synthesis of Quinazoline Compounds

I initially explored substitution of 2,4-dichloroquinazoline at the 4-position with simple amines, **Figure 4.3**. Stirring with pyrrolidine in the presence of triethylamine (TEA) gave **4a**, and none of the di-substituted product. Likewise, reaction with piperidine and TEA readily gave **4b**. Reaction with *n*-butylamine was slower, therefore heat was used to drive the reaction and form **4c.1**. Reactions with piperazine or proline were also slow and resulted in a complex mixture of products. The troubleshooting of the piperazine and proline reactions included exploration of different bases, solvents and temperatures, but no significant improvements were seen. These target molecules were therefore not pursued further and I moved on to the synthesis of di-substituted quinazolines.

I firstly investigated 2-position substitutions of **4a** with different amines, using higher temperatures and longer reaction times. Substitution with piperidine was successful, to form **4d**, however reaction with pyrrolidine to make **3a** was much slower and required microwave radiation conditions to form the desired product. Substitution of **4a** with *n*-butylamine could not be achieved despite considerably extended reaction times and temperatures. Consistent with previously attempted reactions, reaction of **4a** with piperazine or proline did not give any of the desired products, and hence these reagents were not used in further synthesis of quinazoline derivatives.

To create the di-piperidinyl quinazoline, **4b** was further reacted with piperidine, which fortuitously gave **4e** as a crystalline solid, that could be easily isolated by filtration. On the other hand, reaction of **4b** with pyrrolidine surprisingly yielded no detectable product, despite exploring different reaction times and temperatures. As with **4a**, reaction with *n*-butylamine also did not give the desired product. The mono-substituted *n*-butylamine compound **4c.1** was then reacted with piperidine, which readily gave **4f**. Unlike the previous sets of reactions, it was possible to substitute with *n*-butylamine at the 2-position to give the di-substituted *n*-butylamine compound **4g**. However, as before, reaction with pyrrolidine did not yield the desired product. Interestingly, after purification of the reaction mixture by flash column chromatography, an inseparable mixture of **4c.1** and **4c.2** was isolated, indicating a possible migration of the *n*-butylamine group in **4c.1** under these conditions.

Yields were generally modest, but sufficient for compound characterisation and further testing, and the reactions were therefore not optimised further.
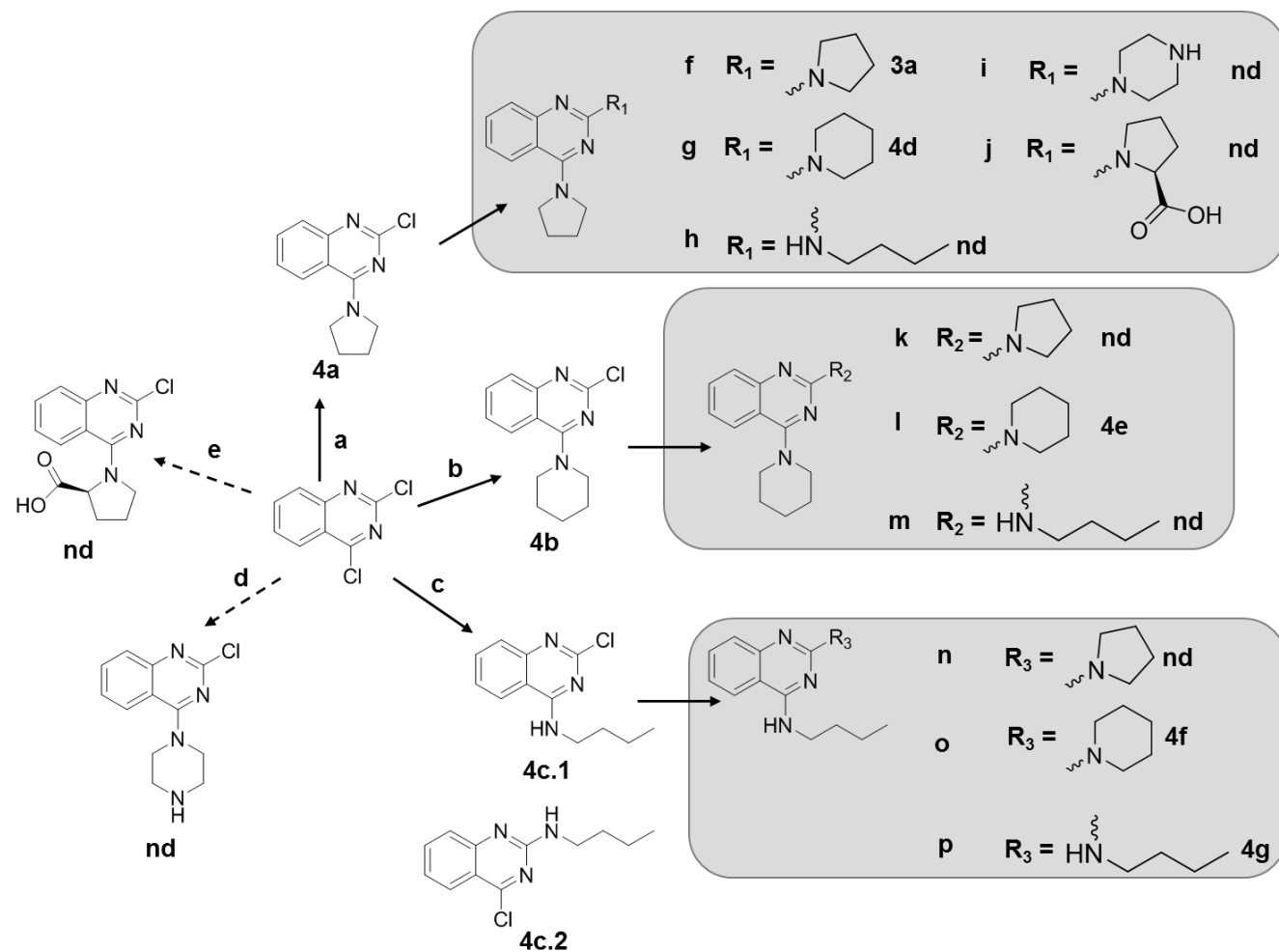
Figure 4.3 Synthetic scheme for quinazoline-based compounds. All reactions carried out in ethanol with 1.5-2 equivalence of the respective amine and TEA, with heating if required. **nd** = not detected. Yields: a: 73%, b: 85%, c: 63%, f: 37%, g: 91%, l: 45%, o: 29%, p: 22%.

For investigation of a pyrimidine-only scaffold, the commercially available compound 2-chloro-4-pyrrolidin-1-yl pyrimidine was reacted with pyrrolidine or piperidine to give **4h** and **4i** respectively. Surprisingly, unlike all previously attempted reaction with proline, it was possible to isolate the proline-substituted compound **4j**.

From 2,4-dichloro-5-trifluoromethyl pyrimidine, either of the 2- or 4- mono-substituted pyrrolidine compounds **4k.1** and **4k.2** could be generated, but the di-substituted compound was not attained, despite a number of attempts to optimise reaction conditions. Likewise, attempts to make 2,4-di-pyrrolidin-1-yl-thieno[3,2-*d*]pyrimidine were also unsuccessful, however the mono-substituted compound **4l** could be isolated, **Figure 4.4**. Overall, 14 compounds were purified in high enough yields for full characterisation and further testing, including two novel compounds: **4f** and **4k.2**.
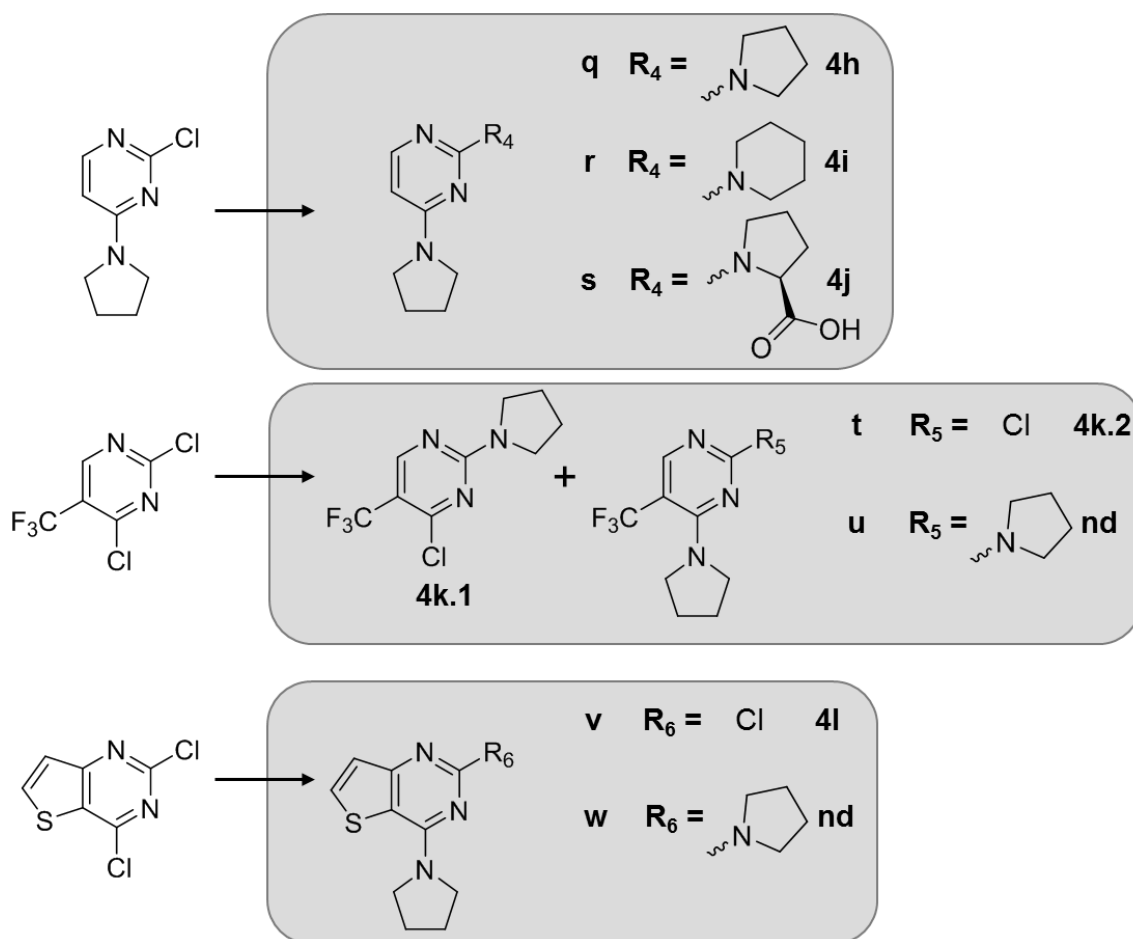
Figure 4.4 Synthetic scheme for pyrimidine-based compounds. All reactions carried out in ethanol with 1.5-2 equivalence of the respective amine and TEA, and heating with microwave radiation if required; nd = not detected. Yields: q: 84%, r: 60%, s: 31% t: 86% (**4k.2**) and 14% (**4k.1**); v: 78%.

4.3   Testing analogues of 3a with MtGlgB

Fourteen of the synthesised compounds were tested for inhibition of MtGlgB at a concentration of 50 µM, using the amylose-iodine assay described in Chapter 3. The percentage inhibition observed was normalised to that of **3a**, to more easily make comparisons. Previous ITC analyses indicated that the actual binding affinities may be higher than the percentage inhibition seen in the MtGlgB assay. However, none of the synthesised compounds showed higher levels of inhibition than **3a**, suggesting that their binding affinities would also be worse than that of **3a**.

Broadly speaking, compounds that retained a quinazoline scaffold showed some level of inhibition, whereas those without this did not inhibit MtGlgB, with the notable exceptions of **4d** and **4k.2**, which had lower and higher levels of inhibition, respectively. Comparing **3a** and **4d**, just switching the 2-pyrrolidine group for piperidine completely abolished inhibition. It was not possible to synthesise other quinazoline compounds with a 2-pyrrolidine substitution and so it is difficult to draw conclusions, although this could indicate that this group has the biggest contribution to the inhibition in a quinazoline scaffold. Most of the pyrimidine-based compounds showed no inhibition apart from **4k.2**, which had higher inhibition than its isomer **4k.1**. Again, this suggests the 2-pyrrolidine substitution is critical for inhibition, although compound **4h** with a 2-pyrrolidine group showed no inhibition.

The order of relative inhibition **4a** > **4b** > **4c.1** indicates that in the presence of a 2-chloro group, the level of inhibition with respect to the 4-position group is pyrrolidine > piperidine > *n*-butylamine. However, this trend is not seen in the di-substituted compounds, for example **4f**, in which the pyrrolidine groups were replaced with 4-*n*-butylamine and 2-piperidine, had a relatively high inhibition. The other 4-*n*-butylamine compounds **4c.1** and **4g**, and the other 2-piperidine compounds **4d** and **4e**, had lower levels of inhibition. The combination and relative position of the amine groups is important and it does not seem possible to consider the amine groups at each position in isolation.

Overall, the trends between the identity of the amine groups were difficult to distinguish and it seems that more chemical space needs to be explored before trends can be clearly established.
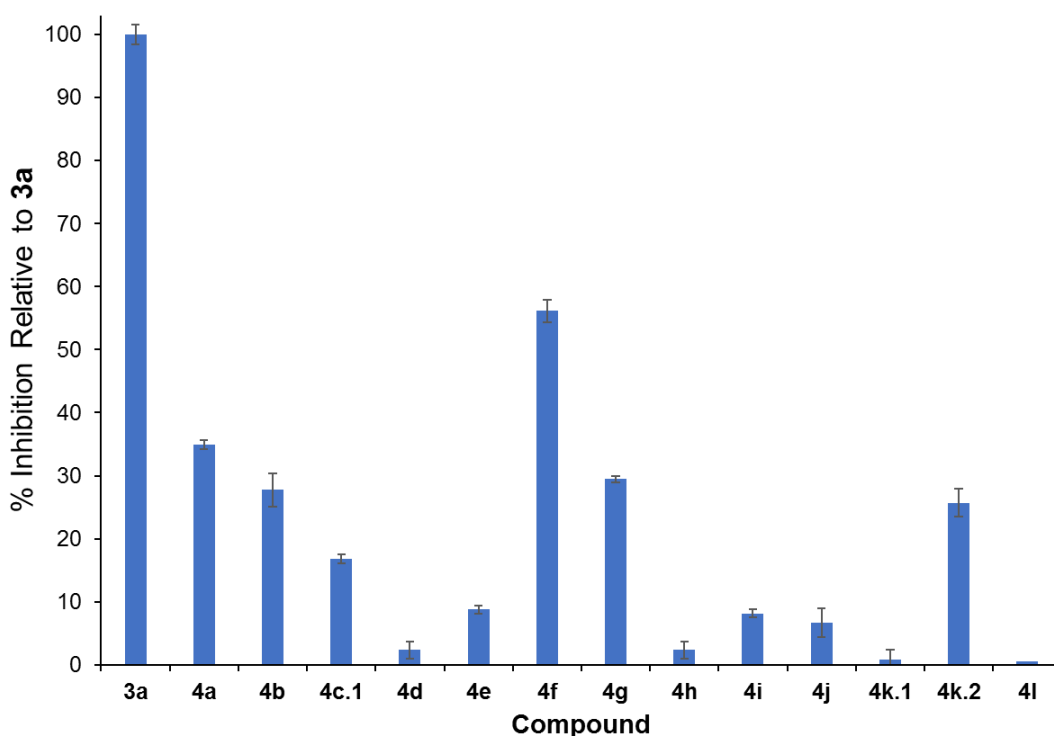
Figure 4.5 Inhibition of MtGlgB by analogues of 3a. Percentage inhibition is normalised to that of **3a** demonstrating that none of the structural modifications led to an improved efficacy. Compounds were tested at 50 µM with the 96-well plate format amylose-iodine assay detailed in section **3.2.2**, with a minimum of three repeats for each compound.

4.4   Testing effect of 3a on *M. smegmatis* growth

M. smegmatis is a faster growing, non-pathogenic relative of *M. tuberculosis*, that is a commonly used surrogate for assessing the *in vivo* activity of potential tuberculosis therapeutics.[6] A *M. tuberculosis ΔglgB* strain is not viable due to the lethal accumulation of M1P. On the other hand, a *M. smegmatis, ΔglgB* strain accumulates M1P, but is viable in normal growth conditions, unless the growth medium is supplemented with trehalose.[9] Therefore, to monitor inhibition of MsGlgB in *M. smegmatis*, it is necessary to add trehalose to the growth medium. This trehalose dependence is useful, as it provides a way to distinguish between growth inhibition specific for the GlgB target and off-target effects, simply by comparing growth in the presence and absence of trehalose.

4.4.1   Developing a *M. smegmatis* assay to test compounds

In the first instance, I explored the agar disk diffusion method,[10] as a potentially quick and easy method to determine the effects of the compounds on mycobacterial growth. However, I observed inconsistent and/or hazy zones of inhibition with positive controls and therefore

this method was not pursued further. Instead, I opted for a liquid culture 96-well plate assay to continuously monitor *M. smegmatis* growth over a number of days in the presence of the compounds. This can be followed by the addition of the redox-sensitive dye resazurin, which turns from blue to pink when it is reduced in the presence of viable cells.[11-12] Hence, by comparing growth and colour, compounds that are bactericidal can be distinguished from those that are bacteriostatic. To optimise the published protocol[13] for my purposes I varied inoculum density, incubation time, DMSO percentage, trehalose concentration and incubation time with resazurin. Each plate included sterility controls and growth controls, comprising medium with and without inoculum.

4.4.2   Three compounds were non-selective inhibitors of *M. smegmatis* growth at high concentrations

Compounds were initially tested at a concentration of 500 µM in the presence and absence of trehalose. Each condition was tested at least in triplicate on every plate, and repeated on 2-3 different plates, with a new mycobacterial inoculum for each plate. Compounds **4a**, **4b**, **4h** and **4l** consistently showed no inhibition. Compounds **4c.1** and **4e** initially showed some growth inhibition, but this was not consistent in repeat experiments. This could indicate that at this concentration they are borderline inhibitors and testing at higher concentrations could determine if this were the case. Compounds **3a**, **4g** and **4f** were consistently bactericidal, however they were not selective for GlgB as there was no difference in growth inhibition with and without trehalose. The approximate MICs of these compounds was established using serial two-fold dilutions from 500 – 4 µM, according to the standard protocol.[13] These were 125 µM for **4f**, 125 µM for **3a** and 250 µM for **4g**, **Figure 4.6**.

4.4.3   Compounds did not inhibit *E. coli* growth at the same concentrations

To determine whether these compounds were effective against Gram-negative bacteria, I attempted to determine the MICs with *E. coli* ATCC 25922, a reference strain for antibiotic susceptibility testing.[14] For all three compounds, growth was visible at the highest concentration tested, therefore the *E. coli* MICs with this strain were all ≥ 500 µM. However, there was a visible decrease in growth that correlated with increasing compound concentration, indicating that the compounds impeded *E. coli* growth by up to 20 - 30% at 500 µM. Whilst these preliminary results indicate that these compounds may be more effective against mycobacteria than *E. coli*, further work is required to confirm this.
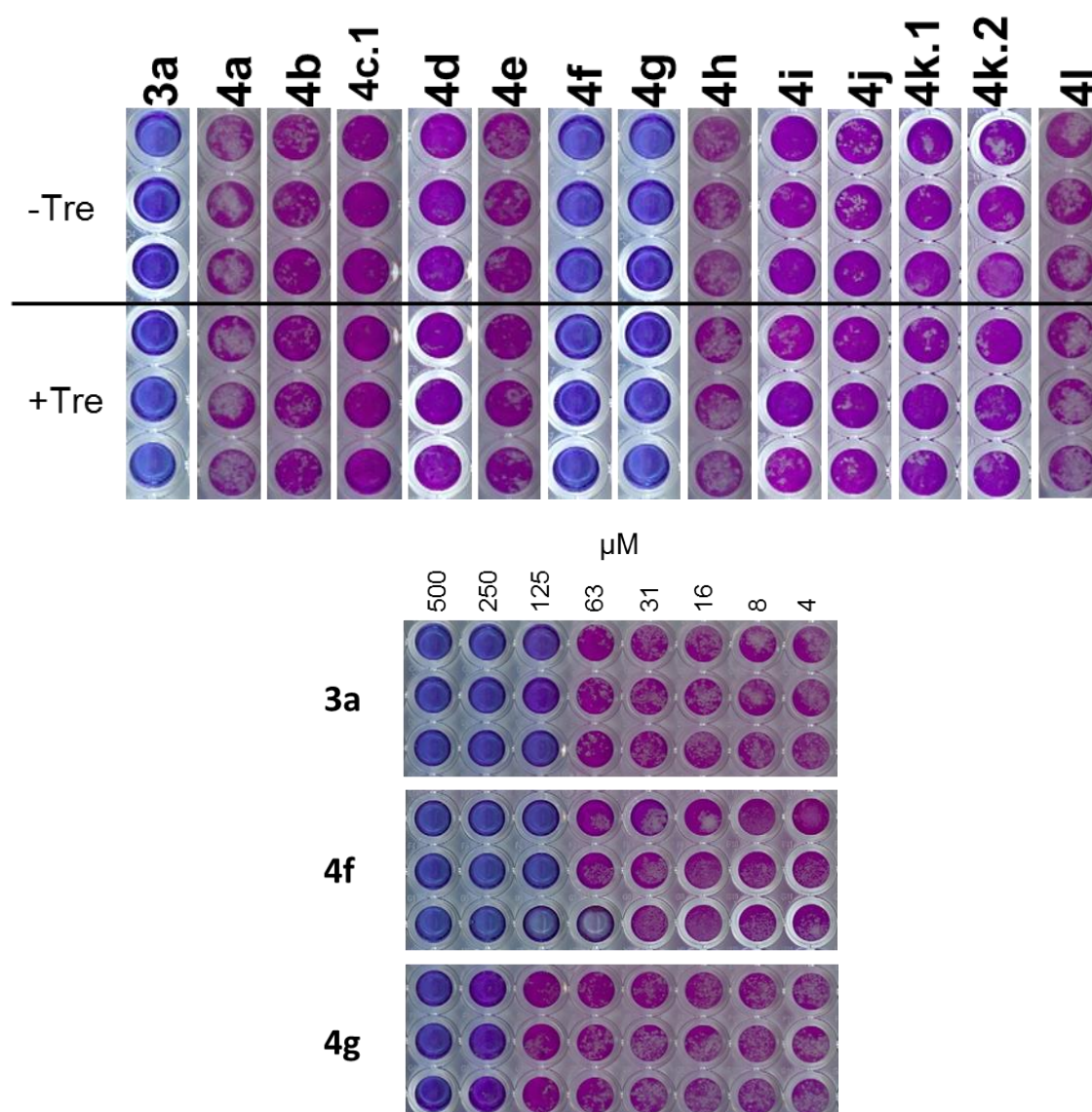
Figure 4.6 *M. smegmatis* growth assays with synthesised compounds. Growth assays are shown in the top panel, with 500 µM of each compound added to growth media lacking trehalose (-Tre) or supplemented with 500 µM trehalose (+Tre). Since the *M. smegmatis* glgB knockout is lethal only in the presence of trehalose, compounds that inhibit growth by targeting GlgB will be effective in the +Tre condition only. Colorimetric read-outs were determined using a resazurin assay. Pink wells contain viable bacteria and blue wells no viable cells. Compounds **3a**, **4f** and **4g** showed non-selective inhibition and the MICs in normal growth media were determined by two-fold serial dilutions, bottom panel. Analysis of growth curves confirmed these were bactericidal, rather than bacteriostatic. Images are representative of at least three biological repeats.

4.5 Crystal soaks of MsGlgB with synthesised compounds

Crystal soaks of MsGlgB and D416A-MsGlgB were carried out with all the compounds that showed > 10% relative inhibition in the *in vitro* assay. As with **3a**, soaks with a mixture of compound in DMSO at ~50 mM and precipitant solution were used. After 3 h, crystals were frozen and diffraction data were collected, however no additional electron density corresponding to any of the compounds was observed.

4.6 Discussion

In this chapter I aimed to address structure-activity relationships of **3a** with a selection of modified compounds. Compound **3a** was identified as an inhibitor of MtGlgB from a high-throughput screen and validated by ITC, which demonstrated a K_d of 3.1 ± 0.7 µM and a 1:1 binding ratio. As the interactions with the protein and the mechanism of inhibition were unknown, my approach was to explore the effect of varying the main structural features. I defined these as a pyrimidine-containing quinazoline core and amine side groups located at the 2 and 4 positions of the pyrimidine ring, **Figure 4.2**. I aimed to make compounds that retained the pyrimidine ring and the position of the amine group substitutions, but had variations in both the type of amine groups and substitutions on the pyrimidine ring.

Substitutions of 2,4-dichloro quinazoline provided a straightforward route to target compounds, taking advantage of the differential activity of each position to add different amine groups to the 2 and 4-positions. The greater rate of reaction at the 4-position compared with that at the 2-position with respect to nucleophilic displacement can be explained by the resonance forms of the respective reaction intermediates, **Figure 4.7**. The intermediate formed by nucleophilic substitution at the 4-position is stabilised by the resultant resonance structures. The resonance structures of the intermediate formed by nucleophilic substitution at the 2-position give less stability due to the *o*-quinoid resonance structure.[8]
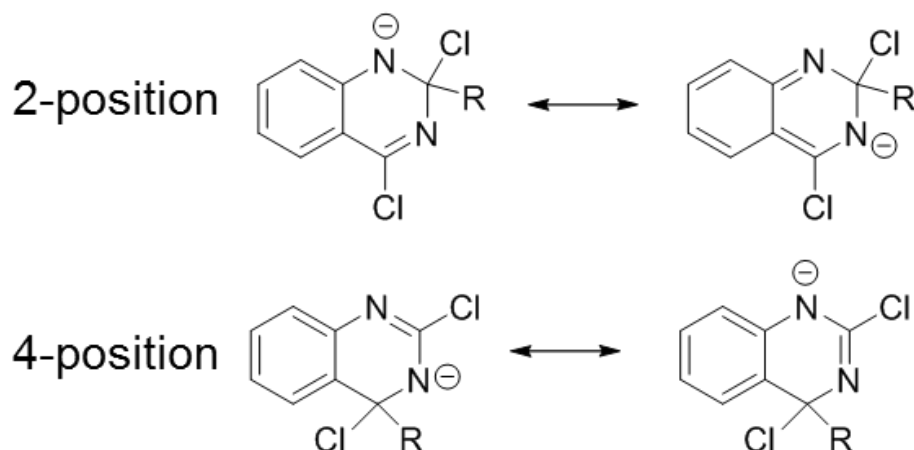
Figure 4.7 Resonance forms in nucleophilic substitution of 2,4-dichloro quinazoline
The resonance forms in the transition states in nucleophilic substitutions of 2,4-dichloro quinazoline at the 2 or 4 positions. The rate of substitution at the 4-position is much greater as these resonance forms (bottom panel) provide more stability than the *o*-quinoid structure (top right) in the 2-position substitution resonance forms. R= nucleophile. Adapted from *The chemistry of heterocyclic compounds; part 1 fused pyrimidines.*[8]

The ease of substitution with pyrrolidine and piperidine at the 4-position of 2,4-dichloro quinazoline to make **4a** and **4b** was expected due to the high nucleophilicities of these secondary amines.[15] The slower 4-position substitution reaction with the primary amine *n*-butylamine to make **4c.1** can also be explained by its lower nucleophilicity. However, it is surprising that reactions with proline or piperazine in similar conditions were also slow, as these would be expected to react similarly to the other secondary amines. The formation of side-products in these reactions may be attributed to the additional functional groups in the nucleophiles. More controlled reactions may be achieved by the use of protecting groups such as silyl esters for the free carboxylic acid in proline or *tert*-butyloxycarbonyl (BOC) for the amino group in piperazine.[16]

The substitution of **4a**, **4b** and **4c.1** at the 2-position required more forcing conditions, as expected. However, it is surprising that substitution with piperidine tended to be achieved more readily than that of pyrrolidine, as these have similar nucleophilicities.[15] It is perhaps also surprising that the 2-position substitution of pyrimidine was achieved more easily as the transition state would be less stabilised in these compounds. The inconsistency in the reactivity of the amine groups to make the di-substituted compounds may be explained by other variables that were not fully considered at the time, such as the pH of the reaction mixture. Whilst there is certainly further scope for optimising the reaction conditions, the *in*

vitro and *in vivo* results with the compounds that could be successfully isolated indicate that future work should explore different structural modifications.

The synthesised compounds were tested against MtGlgB using the amylose-iodine assay described in Chapter 3. None of these showed greater inhibition than **3a**, indicating that the interaction of **3a** and MtGlgB is highly specific for this particular configuration of amine groups and quinazoline core. It appears that the changes explored here were therefore too great and future synthetic work should focus on adding groups to **3a**, rather than modifications. Substitutions of the benzene ring, at the 5, 6, 7 and/or 8 positions are the obvious routes to explore. It was difficult to distinguish clear trends from the set of synthesised compounds. Broadly, the quinazoline scaffold and the 2-pyrrolidine group appeared to contribute to higher inhibition values, indicating that this edge of **3a** is the most important for binding. However, there were exceptions to these trends, suggesting that other, as yet undetermined factors are involved in compound binding and inhibition.

I also wanted to address the question: How do **3a** and its analogues behave *in vivo*? I aimed to gain further insight into the activity of **3a** and synthesised compounds by testing their effect on *M. smegmatis* growth. The trehalose-sensitivity of *M. smegmatis ΔglgB* meant that a simple *in vivo* assay could be established, which can be used to determine whether growth inhibition is specific for the GlgB target. This approach has not been explored previously and will be useful for other researchers investigating MtGlgB inhibitors. An alternative method for the same approach could be to compare growth inhibition in *M. smegmatis* WT and *M. smegmatis ΔglgB* strains. It is usually difficult to distinguish target-specific growth inhibition and this can be an erroneous assumption of inhibition studies, as the Dhkar *et al.*[17] work discussed in Chapter 3 demonstrates. Future work, could include testing compounds with *M. tuberculosis*, as whilst *M. smegmatis* is a good model for this target, there could be some differences between compound uptake between the species.

Testing the non-specific inhibitors with a clinical *E. coli* indicated that at the same concentration of compounds, growth inhibition was more pronounced in *M. smegmatis*. This is perhaps unexpected, due to the greater barrier for cellular uptake in mycobacteria compared to *E. coli*.[18] Whilst this could be suggestive of mycobacteria-specific inhibition, these results are preliminary and further testing on other bacterial strains in required before any conclusions can be made.

In addition to exploring more chemical space to improve the potency of **3a**, the other target for further investigation of this compound is to achieve selectivity for GlgB in the mycobacterial cell. The other target or targets are unknown, although similar sets of

compounds have been shown to interfere with folate acid biosynthesis,[19-20] which could be the actual target here.[21] If this is established, future medicinal chemistry efforts can be focussed on exploiting differences in the binding mode of **3a**, to add groups that obstruct binding in unwanted targets or create additional interactions that favour binding to GlgB.

4.7 Summary

I synthesised a selection of analogues of **3a**, which were modified in the quinazoline core and the amine side groups. Fourteen compounds were tested *in vitro* with MtGlgB. Results suggested that one edge of **3a** was important for inhibition, however no synthesised compound had improved potency. A *M. smegmatis* assay was established, which could determine whether growth inhibition was specific for the GlgB target. Three compounds inhibited *M. smegmatis* growth at high concentrations, but not via the GlgB target. Future synthetic work will aim to increase both potency and target selectivity.

4.8   Materials and methods

4.8.1   Synthesis of 3a analogues

4.8.1.1 General methods

2,4-Dichloro quinazoline was purchased from Key Organics. All other solvents and reagents were purchased from Sigma Aldrich or Fisher Scientific and used without further purification. Analytical thin layer chromatography was carried out using Merck aluminium backed sheets coated with 60-F254 silica gel and spots were visualised with UV light. Compounds were purified by automated flash column chromatography using a Biotage Isolera™ Spektra One system with KP-Sil™ (Uppsala, Sweden). NMR spectra were recorded at 22 °C on a Bruker Avance III spectrometer at 400 MHz (^{1}H), 100 MHz (^{13}C) or 376 MHz (^{19}F). Chemical shifts (δ) are reported in ppm using residual solvent signals from deuterated solvents as references (CDCl$_3$: $\delta\,^1$H 7.26 and $\delta\,^{13}$C 77.0). Assignments were made with the aid of COSY and HSQC experiments. Nominal m/z values are reported in Daltons. After classical work-up of the reaction mixtures, organic layers were dried over MgSO$_4$, filtered and then concentrated under vacuum.

4.8.1.2 2-chloro-4-(pyrrolidin-1-yl)quinazoline (4a)

4a
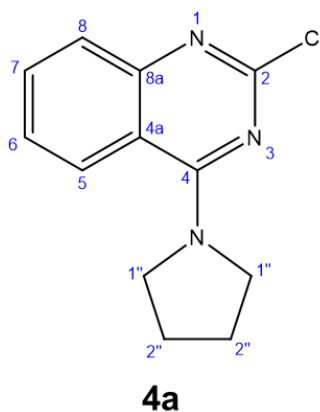
To 2,4-dichloroquinazoline (1 g, 5.02 mmol) dissolved in THF (20 mL) were added pyrrolidine (495 µL, 6.03 mmol) and triethylamine (840 µL, 6.03 mmol). The solution was stirred for 3 h at room temperature, at which point TLC (4:1 hexane/EtOAc) showed conversion to product. The reaction mixture was poured on to ice and extracted three times with CH$_2$Cl$_2$ (3 × 30 mL), then the combined extracts were washed with water and concentrated. The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give **4a** (850 mg, 73%) as a colourless solid. R_f 0.52 (4:1 Hexane/EtOAc);

^{1}H NMR (400 MHz CDCl$_3$): δ 8.13 (dd, 1H, $^{3}J_{7,8}$ 8.5 Hz, $^{4}J_{6,8}$ 0.8 Hz, H-8), 7.74 (dd, 1H, $^{3}J_{5,6}$ 8.4 Hz, $^{4}J_{5,7}$ 1.0 Hz, H-5), 7.67 (ddd, 1H, $^{3}J_{7,8}$ 8.3 Hz, $^{3}J_{6,7}$ 6.9 Hz, $^{4}J_{5,7}$ 1.4 Hz, H-6), 7.36 (ddd, 1H, $^{3}J_{5,6}$ 8.5 Hz, $^{3}J_{6,7}$ 6.9 Hz, $^{4}J_{6,8}$ 1.5 Hz, H-7), 3.99-3.92 (m, 4H, 4 x H-1''), 2.10-2.03 (m, 4H, 4 x H-2''); ^{13}C NMR (100 MHz, CDCl$_3$): δ 160.8 (C-2), 156.8 (C-4), 153.0 (C-8a), 132.7 (C-6), 127.5 (C-5), 125.4 (C-8), 124.6 (C-7), 115.0 (C-4a), 51.3 (C-1''), 25.6 (C-2''); MS (ESI): [M + H]$^{+}$ 234, found 234.4.

4.8.1.3 2-chloro-4-(piperidin-1-yl)quinazoline (4b)
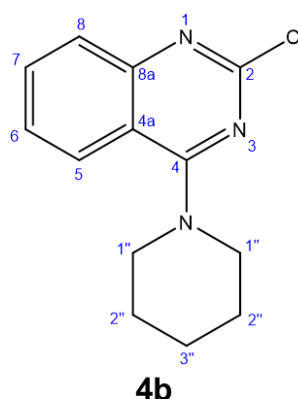
4b

To 2,4-dichloroquinazoline (50 mg, 0.251 mmol) dissolved in ethanol (1 mL) were added piperidine (30 μL, 0.301 mmol) and triethylamine (41 μL, 0.301 mmol). The solution was stirred for 3 h at room temperature, at which point TLC (4:1 hexane/EtOAc) showed conversion to product. The reaction mixture was poured on to ice and extracted three times with CH$_2$Cl$_2$ (3 × 2 mL) and the combined extracts washed with water and concentrated *in vacuo*. The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give **4b** (52 mg, 85%) as a colourless solid. R_f 0.46 (4:1 Hexane/EtOAc); ^{1}H NMR (400 MHz CDCl$_3$): δ 8.05 (dd, 1H, $^{3}J_{7,8}$ 8.3 Hz, $^{4}J_{6,8}$ 1.4 Hz, H-8), 7.58 (ddd, 1H, $^{3}J_{7,8}$ 8.3 Hz, $^{3}J_{6,7}$ 7.1 Hz, $^{4}J_{5,7}$ 1.2 Hz, H-7), 7.38 (dd, 1H, $^{3}J_{5,6}$ 8.1 Hz, $^{4}J_{5,7}$ 1.2 Hz, H-5), 7.13 (ddd, 1H, $^{3}J_{5,6}$ 8.1 Hz, $^{3}J_{6,7}$ 7.1 Hz, $^{4}J_{6,8}$ 1.4 Hz, H-6), 3.78-3.69 (m, 4H, 4 × H-1''), 1.77-1.68 (m, 6H, 4 x H-2'', 2 × H-3''); ^{13}C NMR (100 MHz, CDCl$_3$): δ 165.2 (C-2), 151.6 (C-4), 150.2 (C-8a), 134.8 (C-6), 126.2 (C-5), 125.2 (C-8), 122.2 (C-7), 116.6 (C-4a), 46.2 (C-1''), 25.7 (C-2''), 24.5 (C-3''); MS (ESI): [M + H]$^{+}$ 248, found 248.3.

4.8.1.4 *n*-butyl-2-chloro-quinazolin-4-amine (4c.1)
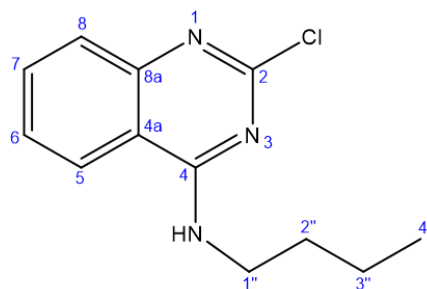
4c.1

To 2,4-dichloroquinazoline (50 mg, 0.251 mmol) dissolved in ethanol (1 mL) were added butylamine (30 µL, 0.301 mmol) and triethylamine (41 µL, 0.301 mmol). The solution was heated at 50 °C for 3 h, at which point TLC (4:1 hexane/EtOAc) showed conversion to product. The reaction mixture was poured on to ice and extracted three times with CH_2Cl_2 (3 × 2 mL) and the combined extracts washed with water and concentrated. The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give **4c.1** (37 mg, 63%) as a colourless solid. R_f 0.35 (4:1 hexane/EtOAc); ^{1}H NMR (400 MHz CDCl$_3$): δ 7.76 (dd, 1H, $^3J_{7,8}$ 8.3 Hz, $^4J_{6,8}$ 1.6 Hz, H-8), 7.72 (ddd, 1H, $^3J_{7,8}$ 8.3 Hz, $^3J_{6,7}$ 6.6 Hz, $^4J_{5,7}$ 1.2 Hz, H-7), 7.67 (dd, 1H, $^3J_{5,6}$ 8.2 Hz, $^4J_{5,7}$ 1.2 Hz, H-5), 7.45 (ddd, 1H, $^3J_{5,6}$ 8.2 Hz, $^3J_{6,7}$ 6.6 Hz, $^4J_{6,8}$ 1.6 Hz, H-6), 5.90 (s, 1H, N-H), 3.71-3.65 (m, 2H, 2 × H-1''), 1.76-1.66 (m, 2H, 2 × H-2''), 1.52-1.41 (m, 2H, 2 × H-3''); 0.99 (t, 3H, 3 x H-4''); ^{13}C NMR (100 MHz, CDCl$_3$): δ 161.0 (C-2), 157.9 (C-4), 150.8 (C-8a), 133.4 (C-6), 127.9 (C-5), 126.1 (C-8), 120.6 (C-7), 113.2(C-4a), 41.4 (C-1''), 31.3 (C-2''), 20.2 (C-3''), 13.8 (C-4''); MS (ESI): [M + H]$^+$ 236, found 236.3.

4.8.1.5 2,4-di-pyrrolidinin-1-yl-quinazoline (3a)
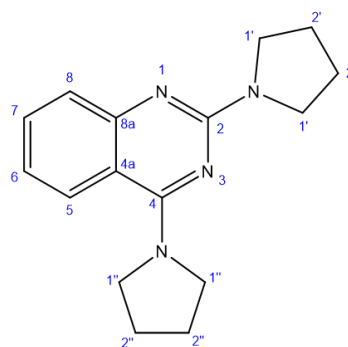
3a

To **4a** (25 mg, 0.214 mmol) in ethanol (1 mL) were added pyrrolidine (13 µL, 0.321 mmol) and triethylamine (22 µL, 0.321 mmol). The solution was heated at 100 °C for 3 h using microwave radiation, at which point TLC (2:1 hexane/EtOAc) showed conversion to product. The reaction mixture was poured on to ice and extracted three times with CH_2Cl_2 (3 × 2 mL) and the combined extracts concentrated *in vacuo*. The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give **3a** (10 mg, 37%) as a colourless solid. R_f 0.42 (2:1 hexane/EtOAc); ^{1}H NMR (400 MHz CDCl$_3$): δ 7.96 (m, 1H, H-8), 7.55-7.48 (m, 1H, H-5), 7.45 (ddd, 1H, $^3J_{7,8}$ 7.9 Hz, $^3J_{6,7}$ 6.5 Hz, $^4J_{5,7}$ 1.2 Hz, H-6), 6.96 (ddd, 1H, $^3J_{5,6}$ 8.2 Hz, $^3J_{6,7}$ 6.5 Hz, $^4J_{6,8}$ 1.3 Hz, H-7), 3.91-3.86 (m, 4H, 4 x H-1'), 3.69-3.63 (m, 4H, 4 x H-1''), 2.02-1.98-3 (m, 4H, 4 x H-2'), 1.98-1.93 (m, 4H, 4 x H-2''); ^{13}C NMR (100 MHz, CDCl$_3$): δ 160.4 (C-2), 153.5 (C-4), 148.2 (C-8a), 138.7 (C-6), 131.8 (C-5), 125.5 (C-8), 119.2 (C-7), 112.0 (C-4a), 50.6 (C-1'), 46.6 (C-1''), 25.8 (C-2'), 25.6 (C-2''); MS (ESI): [M + H]$^+$ 269, found 269.3.

4.8.1.6 2-(piperidin-1-yl)-4-(pyrrolidin-1-yl)quinazoline (4d)

4d
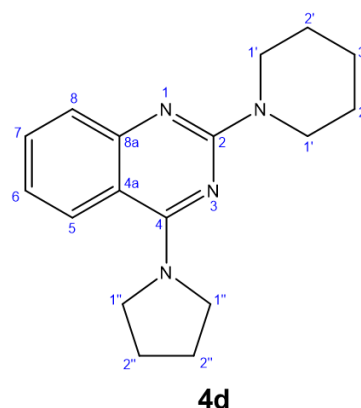
To **4a** (50 mg, 0.214 mmol) in ethanol (2 mL) were added piperidine (32 µL, 0.321 mmol) and triethylamine (45 µL, 0.321 mmol). The solution was heated at reflux for 18 h, at which point TLC (3:1 hexane/EtOAc) showed conversion to product. The reaction mixture was poured on to ice and extracted three times with CH_2Cl_2 (3 × 30 mL) and the combined extracts were concentrated *in vacuo*. The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give **4d** (55 mg, 91%) as a colourless solid. R_f 0.73 (3:1 hexane/EtOAc); ^{1}H NMR (400 MHz CDCl$_3$): δ 7.92 (d, 1H, $^3J_{7,8}$ 8.4 Hz, H-8), 7.49-7.41 (m, 2H, H-5, H-7), 6.99-6.93 (m, 1H, H-6), 3.89-3.82 (m, 8H, 4 x H-1', 4x H-1''), 2.03-1.95 (m, 4H, 4 x H-2''), 1.68-1.57 (m, 6H, 4 x H-2', 2 x H-3'); ^{13}C NMR (100 MHz, CDCl$_3$): δ 160.5 (C-2), 158.7 (C-4), 154.5 (C-8a), 131.6 (C-6), 125.7 (C-5), 125.4 (C-8),

119.3 (C-7), 112.1 (C-4a), 50.6 (C-1'), 44.9 (C-1''), 26.1 (C-2'), 25.7 (C-2''), 25.1 (C-3'); MS (ESI): [M + H]$^+$ 283, found 283.1.

4.8.1.7 2,4-di-piperidin-1-yl-quinazoline (4e)

4e
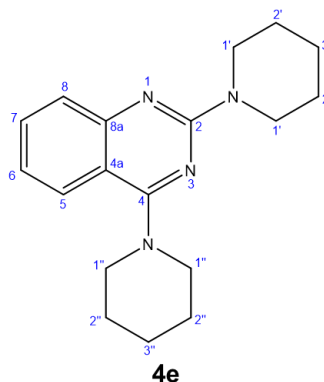
To **4b** (25 mg, 0.101 mmol) dissolved in ethanol (1 mL) were added piperidine (12 μL, 0.121 mmol) and triethylamine (17 μL, 0.121 mmol). The solution was heated in with microwave radiation at 100 °C for 1 h, at which point a crystalline precipitate had formed. The precipitate was removed by filtration and washed three times with ethanol (3 × 2 mL) to give pure **4e** (13 mg, 45%) as a colourless solid. R_f 0.72 (4:1 hexane/EtOAc); ^{1}H NMR (400 MHz CDCl$_3$): δ 7.69-7.62 (m 1H, H-8), 7.50-7.44 (m, 2H, H-5, H-7), 7.03-6.97 (m, 1H, H-6), 3.92-3.81 (m, 4H, 4 × H-1'), 3.63-3.54 (m, 4H, 4 × H-1''), 1.81-1.69 (m, 6H, 4 x H-2'', 2 × H-3''), 1.69-1.56 (m, 6H, 4 x H-2'', 2 × H-3''); ^{13}C NMR (100 MHz, CDCl$_3$): δ 166.1 (C-2), 158.6 (C-4), 154.6 (C-8a), 132.0 (C-6), 126.0 (C-5), 125.3 (C-8), 119.9 (C-7), 112.0 (C-4a), 51.1 (C-1'), 45.0 (C-1''), 26.0 (C-2'), 25.9(C-2''), 25.1 (C-3'), 25.0 (C-3''); MS (ESI): [M + H]$^+$ 297, found 297.5.

4.8.1.8 *n*-butyl-2-(piperidin-1-yl) quinazolin-4-amine (4f)

4f
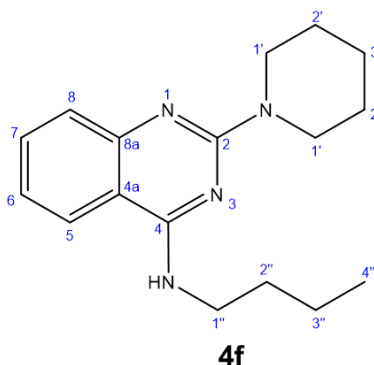
To **4c.1** (20 mg, 0.0851 mmol) dissolved in ethanol (1 mL) were added piperidine (10 μL, 0.102 mmol) and triethylamine (14 μL, 0.102 mmol). The solution was heated with microwave radiation at 100 °C for 1 h, at which point TLC (2:1 hexane/EtOAc) showed

conversion to product. The reaction mixture was poured on to ice and extracted three times with CH_2Cl_2 (3 × 2 mL) and the combined extracts were washed with water and concentrated. The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give **4f** (7 mg, 29%) as a colourless solid. R_f 0.42 (2:1 hexane/EtOAc); ^{1}H NMR (400 MHz $CDCl_3$): δ 7.45-7.43 (m, 3H, H-5, H-7, H-8), 7.05-6.98 (m, 1H, H-6), 5.60 (s, 1H, N-H), 3.91-3.83 (m, 4H, 4 × H-1'), 3.59 (ddd, 2H, $^3J_{1'',2''}$ 7.1 Hz, $^3J_{1'',2''}$ 7.1 Hz, $^2J_{1''a,1''b}$ 5.6 Hz, H-1''), 1.73-1.59 (m, 8H, 4 × H-2', 2 x 2'', 2 × H-3') 1.45 (sxt, 2H, H-3''), 0.98 (t, 3H, H-4''); ^{13}C NMR (100 MHz, $CDCl_3$): δ 159.7 (C-2), 158.4 (C-4) 132.4 (C-8a), 127.9 (C-6), 126.0 (C-5), 125.0 (C-8), 120.8 (C-7), 110.2 (C-4a), 45.1 (C-1'), 40.9 (C-1''), 31.4 (C-2'), 26.0 (C-2''), 25.0 (C-3') 20.3 (C-2''), 13.9 (C-4''); MS (ESI): [M + H]$^+$ 285, found 285.5.

4.8.1.9 N^2,N^4-dibutylquinazoline-2,4-diamine (4g)

4g
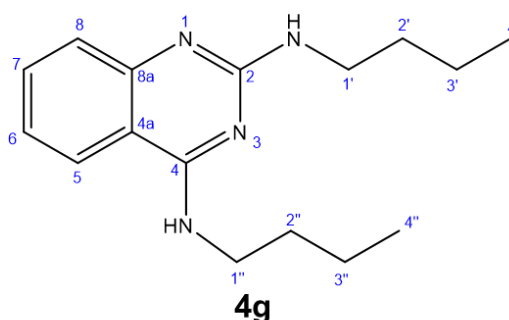
To **4c.1** (20 mg, 0.0851 mmol) dissolved in ethanol (1 mL) were added *n*-butylamine (10 μL, 0.102 mmol) and triethylamine (14 μL, 0.102 mmol). The solution was heated with microwave radiation at 100 °C for 3 h, at which point TLC (3:1 hexane/EtOAc) showed conversion to product. The reaction mixture was poured on to ice and extracted with CH_2Cl_2 (3 × 2 mL) and the combined extracts were washed with water and concentrated. The crude residue was purified by column chromatography on silica gel (3:1 hexane/EtOAc) to give **4f** (5 mg, 22%) as a colourless solid. R_f 0.36 (3:1 hexane/EtOAc); ^{1}H NMR (400 MHz $CDCl_3$): δ 8.05 (d, 1H, $^3J_{7,8}$ 7.9 Hz, H-8), 7.62-7.51 (m, 1H, H-7), 7.40-7.35 (m, 1H, H-6), 7.15-7.12 (m, 1H, H-6), 3.88-3.83 (m, 1H, 1 × H-1'), 3.73-3.66 (m, 1H, 1 × H-1'), 3.60-3.53 (m, 1H, 1 × H-1''), 3.44-3.40 (m, 1H, 1 × H-1''), 1.75-1.59 (m, 8H, 2 × H-2', 2 × H-2'', 2 × H-3', 2 × H-3''), 1.49-1.36 (m, 3H, 3 × H-4'), 1.00-0.89 (m, 3H, 3 × H-4''); ^{13}C NMR (100 MHz, $CDCl_3$): δ 166.1 (C-2), 158.6 (C-4), 154.6 (C-8a), 132.0 (C-6), 126.0 (C-5), 125.3 (C-8), 119.9 (C-7), 112.0 (C-4a), 46.2 (C-1'), 41.0 (C-1''), 31.4 (C-2'), 26.0 (C-2''), 24.9 (C-3'), 24.4 (C-3''), 20.3 (C-4'), 13.8 (C-4''); MS (ESI): [M + H]$^+$ 273, found 273.4.

4.8.1.10 2,4-dipyrrolidin-1-yl pyrimidine (4h)

4h
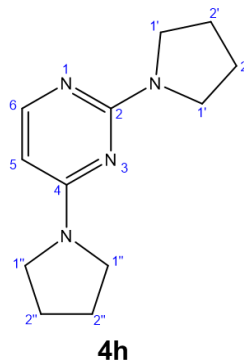
To **2-chloro-4-pyrrolidin-1-yl pyrimidine** (13 mg, 0.0710 mmol) in ethanol (1 mL) were added pyrrolidine (9 µL, 0.107 mmol) and triethylamine (15 µL, 0.107 mmol). The solution was heated with microwave radiation for 100 °C for 3 h at which point TLC (4:1 hexane/EtOAc) showed conversion to product. The reaction mixture was poured on to ice and extracted three times with CH_2Cl_2 (3 × 2 mL) and the combined extracts were concentrated *in vacuo*. The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give **4h** (13 mg, 84%) as a solid. R_f 0.73 (4:1 hexane/EtOAc); ^{1}H NMR (400 MHz $CDCl_3$): δ 7.89 (d, 1H, $^3J_{5,6}$ 5.9 Hz, H-6), 5.62 (d, 2H, $^3J_{5,6}$ 5.9 Hz H-5), 3.58-3.51 (m, 4H, 4 × H-1'), 3.53 – 3.33 (m, 4H, 4 × H-1''), 2.00-1.89 (m, 8H, 4 × H-2', 4 x H-2''); ^{13}C NMR (100 MHz, $CDCl_3$): δ 160.5 (C-2), 160.4 (C-4), 155.5 (C-6), 92.8 (C-5), 46.3 (C-1'), 45.9 (C-1''), 25.6 (C-2'), 25.3 (C-2''); MS (ESI): [M + H]$^+$ 219, found 219.4.

4.8.1.11 2-piperidin-1-yl-4-pyrrolidin-1-yl pyrimidine (4i)
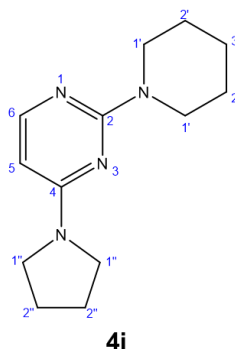
4i

To **2-chloro-4-pyrrolidin-1-yl pyrimidine** (50 mg, 0.273 mmol) in ethanol (2 mL) were added piperidine (41 µL, 0.410 mmol) and triethylamine (57 µL, 0.410 mmol). The solution was heated at reflux for 18 h at which point TLC (3:1 hexane/EtOAc) showed conversion to

product. The reaction mixture was diluted with water and extracted three times with CH_2Cl_2 (3 x 3 mL) and the combined extracts washed with water and brine, dried over $MgSO_4$ and concentrated *in vacuo.* The crude residue was purified by column chromatography on silica gel (3:1 hexane/EtOAc) to give **4i** (38 mg, 60%) as a colourless solid. R_f 0.66 (3:1 hexane/EtOAc); ^{1}H NMR (400 MHz $CDCl_3$): δ 7.88 (d, 1H, $^3J_{5,6}$ 5.8 Hz, H-6), 5.61 (d, 1H, $^3J_{5,6}$ 5.8 Hz, H-5), 3.77-3.69 (m, 4H, 4 x H-1'), 3.59–3.23 (m, 4H, 4x H-1''), 2.00-1.88 (m, 4H, 4 x H-2')) 1.67-1.53 (m, 6H, 4 x H-2'', 2 x H-3''); ^{13}C NMR (100 MHz, $CDCl_3$): δ 161.5 (C-2), 160.5 (C-4), 155.4 (C-6), 93.1 (C-5), 45.9 (C-1'), 44.7 (C-1''), 25.9 (C-2'), 25.2 (C-2''), 25.1 (C-3'); MS (ESI): [M + H]+ 233, found 233.1.

4.8.1.12 (2S)-1-(4-pyrrolidin-1-ylpyrimidin-2-yl)pyrrolidine-2-carboxylic acid (4j)

4j
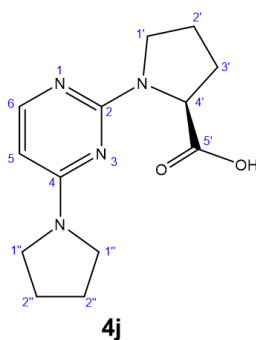
To **2-chloro-4-pyrrolidin-1-yl pyrimidine** (50 mg, 0.273 mmol) in ethanol (2 mL) were added proline (38 mg, 0.327 mmol) and triethylamine (46 µL, 0.327 mmol). The solution was heated at reflux for 18 h at which point TLC (3:1 hexane/EtOAc) showed conversion to product. The reaction mixture was diluted with water and extracted three times with CH_2Cl_2 (3 × 3 mL) and the combined extracts washed with water, dried over $MgSO_4$ and concentrated *in vacuo.* The crude residue was purified by column chromatography on silica gel (3:1 hexane/EtOAc) to give **4j** (22 mg, 31%) as a colourless solid. R_f 0.05 (3:1 hexane/EtOAc); ^{1}H NMR (400 MHz $CDCl_3$): δ 7.46 (d, 1H, $^3J_{5,6}$ 7.5 Hz, H-6), 5.61 (d, 1H, $^3J_{5,6}$ 7.5 Hz, H-5), 4.47-4.04 (m, 1H, 4 × H-4'), 3.56–3.39 (m, 4H, 2 × H-1', 2 × H-2'), 3.12-3.01 (m, 4H, 4 × H-1''), 2.36-2.23 (m, 2H, H-3'), 1.97-1.75 (m, 4H, 4 × H-2'''); ^{13}C NMR (100 MHz, $CDCl_3$): δ 173.1 (C-5'), 158.7 (C-2), 150.4 (C-4), 141.8 (C-6), 95.9 (C-5), 60.8 (C-4'), 58.2 (C-3'), 48.3 (C-1'), 47.4 (C-1''), 29.6 (C-2'), 25.5 (C-2''); MS (ESI): [M + H]+ 263, found 263.4.

4.8.1.13 4-chloro-2-(pyrrolidin-1-yl)-5-(trifluoromethyl)pyrimidine (4k.1)
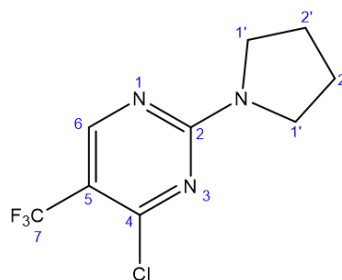
4k.1

To **2-chloro-4-(pyrrolidin-1-yl)-5-(trifluoromethyl)** (62 μL, 0.461 mmol) in methanol (0.5 mL) on ice were added pyrrolidine (39 μL, 0.461 mmol) and triethylamine (63 μL, 0.461 mmol). The solution was allowed to warm to room temperature and stirred overnight, at which point TLC (100% hexane) showed conversion to product. The reaction mixture was extracted three times with CH_2Cl_2 (3 × 30 mL) and the combined extracts washed with water and concentrated *in vacuo*. The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give two products: firstly **4k.1** (100 mg, 86%) as a colourless solid. R_f 0.42 (9:1 hexane/EtOAc) ^{1}H NMR (400 MHz CDCl$_3$): δ 8.37 (s, 1H, H-6), 3.70-3.64 (m, 4H, 4 x H-1''), 2.06–1.94 (m, 4H, 4x H-2''); ^{13}C NMR (100 MHz, CDCl$_3$): δ 162.0 (C-2), 157.4 (C-6), 125.3 (C-4), 122.6 (C-5), 105.2 (C-7), 49.6 (C-1'), 25.2 (C-2'); ^{19}F NMR (MHz CDCl$_3$) δ 53.0 (3 x F); MS (ESI): [M + H]$^+$ 252, found 252.2.

4.8.1.14 2-chloro-4-(pyrrolidin-1-yl)-5-(trifluoromethyl)pyrimidine (4k.2)
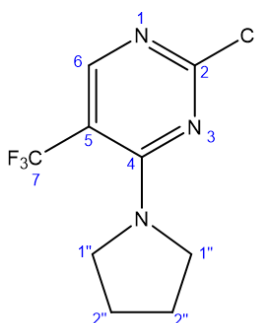
4k.2

The second product was **4k.2** (16 mg, 14%) as a colourless solid. R_f 0.46 (9:1 hexane/EtOAc); ^{1}H NMR (400 MHz CDCl$_3$): δ 8.43 (s, 1H, H-6), 3.65-3.58 (m, 4H, 4 × H-1''), 2.07–1.97 (m, 4H, 4 × H-2''); ^{13}C NMR (100 MHz, CDCl$_3$): δ 160.1 (C-2), 158.6 (C-4), 156.8 (C-6), 124.6 (C-5), 110.1 (C-7), 47.3 (C-1''), 25.4 (C-2''); 19F NMR (MHz CDCl$_3$) δ 61.1 (3 × F); MS (ESI): [M + H]$^+$ 252, found 252.1.

4.8.1.15 Synthesis of 2-chloro-4-pyrrolidin-1-yl-thieno[3,2-*d*]pyrimidine (4l)

4l
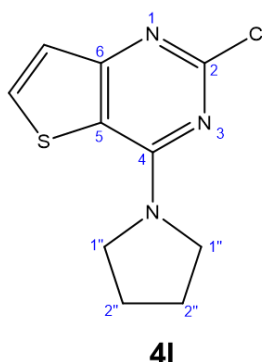
To **2,4-dichlorothieno[3,2-*d*]pyrimidine** (87.5 mg, 0.427 mmol) in ethanol (4 mL) was added pyrrolidine (77 µL, 0.939 mmol). The solution was stirred at 22 °C for 18 h, at which point a crystalline precipitate could be observed. The precipitate was removed by filtration and washed with ethanol (3 × 5 mL). The crude residue was purified by column chromatography on silica gel (4:1 hexane/EtOAc) to give **4l** (mg, 78%) as a colourless solid. R_f 0.53 (4:1 hexane/EtOAc); ^{1}H NMR (400 MHz CDCl$_3$): δ 7.74 (d, 1H, H-5), 7.30 (d, 1H, H-6), 4.00-3.74 (m, 4H, 4 × H-1"), 2.15-1.97 (m, 4H, 4 × H-2"); ^{13}C NMR (100 MHz, CDCl$_3$): δ 161.4 (C-2), 157.4 (C-4), 157.2 (C-8a), 133.0 (C-4a), 124.3 (C-8), 114.0 (C-5), 48.1 (C-1"), 25.7 (C-2"); MS (ESI): [M + H]$^+$ 240, found 240.3.

4.8.2   Compound preparation

Stock solutions of synthesised or purchased compounds were prepared in DMSO and diluted to achieve a final concentration of 50 µM with 10% DMSO for *in vitro* assays or 500 µM with 1% DMSO for *M. smegmatis* assays.

4.8.3   *M. smegmatis* growth and assay conditions

The *M. smegmatis* mc^2 155 wild type strain was used for growth assays. Cultures were grown aerobically at 37 °C in Middlebrook 7H9 liquid media or on 7H10 solid media supplemented with 0.5% (v/v) glycerol and 0.05% (v/v) tyloxapol and containing 10% (v/v) ADS enrichment (5%, w/v, bovine serum albumin fraction V (BSA); 2%, w/v, glucose; 0.85%, w/v, sodium chloride). Freshly streaked 7H10 agar plates were used to generate starter cultures of *M. smegmatis* wildtype or mutant strains in 7H9 liquid media. Cultures were grown with or without the addition of 0.5 mM trehalose until they reached logarithmic phase, at which point they were diluted to the desired density for initiating experiments.

For assays, 150 μL *M. smegmatis* cultures with or without compounds of interest were incubated at 37 °C with continuous orbital shaking (Bioanalyzer, EON) in a transparent, flat-bottom 96-well plate (Corning) over a period of up to 96 hours. External wells were not used for growth measurements and instead contained 250 μL of dH_2O or medium to allow for evaporation. Initial cell densities of OD_{600} ~0.002 were used and growth was monitored by hourly OD_{600} measurements. Control wells contained medium only (sterility control), bacteria only (growth control) or rifampicin (bactericidal control). A final concentration of 1% DMSO in all wells was achieved by dilution of compounds stock solutions or the addition of DMSO to control wells. In each plate compounds were tested in triplicate at 500 μM with and without the addition of 0.5 mM trehalose. No difference in growth was observed resulted from trehalose addition in the absence of compounds. To determine MICs serial two-fold dilutions were prepared to give final concentrations of 500 – 4 μM.

A 0.01% w/v solution of resazurin was prepared from the sodium salt (Sigma) in distilled water and filter sterilised. After growth of *M. smegmatis* cultures 30 μL resazurin solution was added and the plate was incubated at 37 °C for four hours before visualisation. A colour change from blue to pink was indicative of viable *M. smegmatis* cells. The MIC was defined as the lowest compound concentration with no bacterial growth.

4.8.4 Antibiotic susceptibility testing in *E. coli*

E. coli ATCC 25922 strain was provided by the A. Maxwell group (JIC). MICs were determined according to the standard protocol.[22] Briefly, *E. coli* was grown for 16 h at 37 °C on an LB-agar plate, after which point LB liquid medium was inoculated with a single colony and grown to the desired OD_{600}. A final inoculum of 10^5 cfu/mL was prepared based on the previously established conversion from OD_{600} to cfu/mL for this strain (Natassja Bush). A two-fold dilution series of each compound was prepared in a sterile U-bottom 96-well plate (Corning), to achieve final concentrations of 500 - 4 μM and 1% DMSO upon addition of 1:1 v/v organism suspension. Plates were covered with a lid, incubated at 37 °C for 18 h and then inspected by eye and OD_{600} measurement. Each concentration was tested at least in triplicate and control wells of inoculated and uninoculated were included in each plate.

4.9   References

1. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **1997,** *23* (1), 3-25.

2. Hopkins, A. L.; Keseru, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H., The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.* **2014,** *13* (2), 105-121.

3. Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L., Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012,** *4* (2), 90-98.

4. Selvam, T. P.; Kumar, P. V., Quinazoline marketed drugs: a review. *Research in Pharmacy* **2011**.

5. Chaturvedi, V.; Dwivedi, N.; Tripathi, R. P.; Sinha, S., Evaluation of *Mycobacterium smegmatis* as a possible surrogate screen for selecting molecules active against multi-drug resistant *Mycobacterium tuberculosis. The Journal of General and Applied Microbiology* **2007,** *53* (6), 333-337.

6. Shiloh, M. U.; DiGiuseppe Champion, P. A., To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis? *Curr. Opin. Microbiol.* **2010,** *13* (1), 86-92.

7. Wang, D.; Gao, F., Quinazoline derivatives: synthesis and bioactivities. *Chem. Cent. J.* **2013,** *7* (1), 95.

8. Armarego, W. L. F., *The Chemistry of Heterocyclic Compounds, Fused Pyrimidines*. Wiley: 2009.

9. Kalscheuer, R.; Syson, K.; Veeraraghavan, U.; Weinrick, B.; Biermann, K. E.; Liu, Z.; Sacchettini, J. C.; Besra, G.; Bornemann, S.; Jacobs, W. R., Self-poisoning of *Mycobacterium tuberculosis* by targeting GlgE in an α-glucan pathway. *Nat. Chem. Biol.* **2010,** *6* (5), 376-384.

10. Flores, A. R.; Parsons, L. M.; Pavelka, Martin S., Genetic analysis of the β-lactamases of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* and susceptibility to β-lactam antibiotics. *Microbiology* **2005,** *151* (2), 521-532.

11. O'Brien, J.; Wilson, I.; Orton, T.; Pognan, F., Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur. J. Biochem.* **2000,** *267* (17), 5421-5426.

12. Palomino, J.-C.; Martin, A.; Camacho, M.; Guerra, H.; Swings, J.; Portaels, F., Resazurin microtiter assay plate: Simple and inexpensive method for detection of drug resistance in *Mycobacterium tuberculosis. Antimicrob. Agents Chemother.* **2002,** *46* (8), 2720-2722.

13. Taneja, N. K.; Tyagi, J. S., Resazurin reduction assays for screening of anti-tubercular compounds against dormant and actively growing *Mycobacterium tuberculosis, Mycobacterium bovis* BCG and *Mycobacterium smegmatis. J. Antimicrob. Chemother.* **2007,** *60* (2), 288-293.

14. Minogue, T. D.; Daligault, H. A.; Davenport, K. W.; Bishop-Lilly, K. A.; Broomall, S. M.; Bruce, D. C.; Chain, P. S.; Chertkov, O.; Coyne, S. R.; Freitas, T.; Frey, K. G.; Gibbons, H. S.; Jaissle, J.; Redden, C. L.; Rosenzweig, C. N.; Xu, Y.; Johnson, S. L., Complete genome assembly of *Escherichia coli* ATCC 25922, a serotype O6 reference strain. *Genome Announcements* **2014,** *2* (5), e00969-14.

15. Brotzel, F.; Chu, Y. C.; Mayr, H., Nucleophilicities of primary and secondary amines in water. *J. Org. Chem.* **2007,** *72* (10), 3679-88.

16. Lindhorst, T. K., *Essentials of carbohydrate chemistry and biochemistry*. Wiley-VCH: Weinheim, 2007.

17. Dkhar, H. K.; Gopalsamy, A.; Loharch, S.; Kaur, A.; Bhutani, I.; Saminathan, K.; Bhagyaraj, E.; Chandra, V.; Swaminathan, K.; Agrawal, P.; Parkesh, R.; Gupta, P., Discovery of *Mycobacterium tuberculosis* α-1,4-glucan branching enzyme (GlgB) inhibitors by structure- and ligand-based virtual screening. *J. Biol. Chem.* **2015,** *290* (1), 76-89.

18. Jackson, M.; McNeil, M. R.; Brennan, P. J., Progress in targeting cell envelope biogenesis in *Mycobacterium tuberculosis. Future Microbiol.* **2013,** *8* (7), 855-875.

19. Van Horn, K. S.; Zhu, X.; Pandharkar, T.; Yang, S.; Vesely, B.; Vanaerschot, M.; Dujardin, J.-C.; Rijal, S.; Kyle, D. E.; Wang, M. Z.; Werbovetz, K. A.; Manetsch, R., Antileishmanial activity of a series of N^2,N^4-disubstituted quinazoline-2,4-diamines. *J. Med. Chem.* **2014,** *57* (12), 5141-5156.

20. Zhu, X.; Van Horn, K. S.; Barber, M. M.; Yang, S.; Wang, M. Z.; Manetsch, R.; Werbovetz, K. A., SAR refinement of antileishmanial N^2,N^4-disubstituted quinazoline-2,4-diamines. *Biorg. Med. Chem.* **2015,** *23* (16), 5182-5189.

21. Minato, Y.; Thiede, J. M.; Kordus, S. L.; McKlveen, E. J.; Turman, B. J.; Baughn, A. D., *Mycobacterium tuberculosis* folate metabolism and the mechanistic basis for para-aminosalicylic acid susceptibility and resistance. *Antimicrob. Agents Chemother.* **2015,** *59* (9), 5097-106.

22. Andrews, J. M., Determination of minimum inhibitory concentrations. *J. Antimicrob. Chemother.* **2001,** *48 Suppl 1*, 5-16.

5. Investigation of 4-deoxy trehalose analogues as precursors for GlgE inhibitors

5.1 Introduction

M. tuberculosis GlgE is an essential maltosyltransferase that is an attractive drug target for TB therapeutics.[1-2] GlgE inhibitors reported to date function as substrate or transition state mimics. These show moderate inhibition *in vitro* and binding at the GlgE M1P donor site has been confirmed by crystallographic studies.[3-8] However, M1P-based compounds are not expected to exert an effect *in vivo* because *M. tuberculosis* lacks the capacity to import maltose and its derivatives. There are just five carbohydrate importers predicted from the genome sequence of *M. tuberculosis*.[9] Three of these have been characterised and their specificities are limited to trehalose, glycerophosphocholine, and amino sugars, respectively.[10-12] Therefore, unlike most other carbohydrates, trehalose is readily incorporated into mycobacteria. This is mediated by the LpqY-SugABC transporter, which can also tolerate modified substrates, whereby the uptake of trehalose analogues by both *M. tuberculosis* and *M. smegmatis* has been documented.[13-15]

In the first step of the GlgE pathway, TreS isomerises trehalose to maltose, which is then phosphorylated by Pep2 to make M1P, with the two enzymes functioning as a hetero-octameric complex.[1, 16-17] A trehalose analogue that had been incorporated into the cytosol could therefore be converted into a M1P analogue, with the potential to inhibit GlgE. Blocking GlgE not only stops production of the α-glucan end product and therefore reduces *M. tuberculosis* virulence, but also causes the toxic accumulation of M1P via the TreS-Pep2 or GlgM routes.[18] How M1P elicits a pleiotropic stress response in the cell remains unknown, but it is possible that accumulation of the resultant M1P analogue could also be toxic. Thus, trehalose analogues could serve as precursors for TB antibiotics with up to three potential modes of action.

My approach for the investigation of this hypothesis consisted of three parts: the synthesis of trehalose analogues, testing whether these were taken up into the mycobacterial cytosol, and determining whether these were processed into the GlgE pathway. If this could be established as a proof of concept, the next step would be modifying the trehalose analogue such that the resultant M1P analogue could inhibit GlgE. However, in the first instance, I focused on analogues that were likely to be processed by TreS. Previously, it was shown that TreS could tolerate a number of maltose analogues as substrates in the trehalose

synthesis direction, indicating some degree of substrate promiscuity.[16, 19] Another study investigated the incorporation of asymmetrical azido trehalose analogues into *M. smegmatis* cells via labelling with a fluorescent probe. Unlike the other azido analogues, the incorporation of 4-azido trehalose by *M. smegmatis* was greatly reduced in a *ΔtreS* strain, suggesting that it could be processed by TreS *in vivo*.[14] In principle, a 4-deoxy trehalose analogue would be converted into a 4-deoxy M1P analogue in the GlgE pathway. This could act as a donor for GlgE, but once incorporated into a malto-oligosaccharide, the non-reducing end is blocked and it cannot be further extended, impairing the production of α-glucan as a result. Therefore, 4-azido trehalose and other 4-deoxy trehaloses were good candidates for exploratory studies into this approach.

In this chapter I address the following questions: 'Can a synthetic route to 4-deoxy trehalose analogues be established?' 'Are 4-deoxy trehalose compounds taken up in mycobacterial cells?' And, 'Can 4-deoxy trehalose analogues be processed in the GlgE pathway?'.

5.2   4-Deoxy trehalose analogues

5.2.1   Selection of 4-deoxy trehalose analogues

The 4-azido trehalose analogue (**5g**) was of interest as it was found to be taken up by *M. smegmatis* in a TreS-dependent fashion. Additionally, the azido group offers a handle for the facile conjugation of other groups via click chemistry, which could be useful in subsequent localisation or imaging studies.[14] Other 4-position modifications were considered that would be minimally sterically perturbing, but offer insight into the effect of different functionalities. Therefore, methoxy (**4-OMe**) and fluoro- (**5i**) substituents were targeted, with the latter also being of interest for ^{18}F-PET imaging of *M. tuberculosis* infection.[20] To investigate how stereochemistry at the 4-position affects uptake and TreS activity, the axial fluoro- analogue (**5e**) was also targeted, along with the galactose-type analogue (**5c**). Target compounds are shown in **Figure 5.1**.
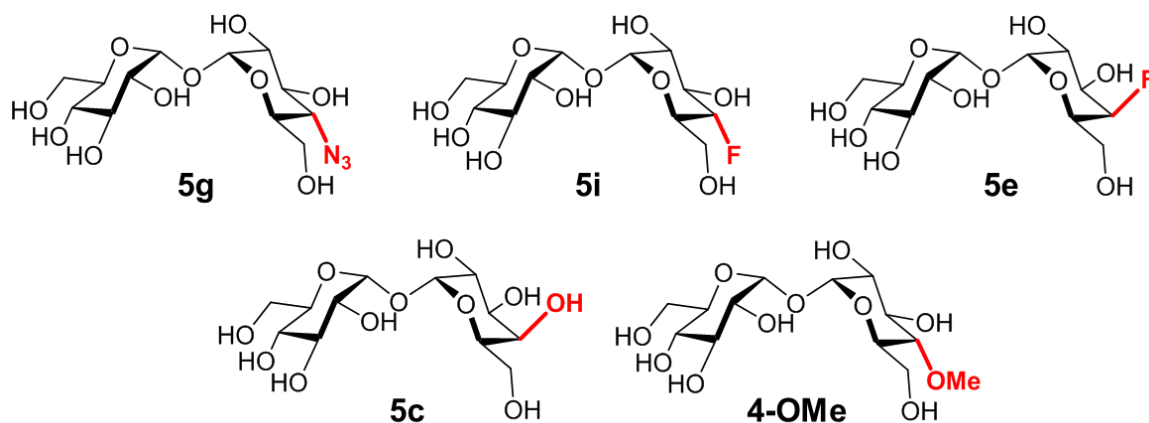
Figure 5.1 4-Deoxy trehalose target compounds. Modifications at the 4-position are highlighted in red. Compound numbering is in accord with **Figure 5.2**.

5.2.2   Synthesis of 4-deoxy trehalose analogues

Asymmetric trehalose analogues have typically been generated via the coupling of glucose derivatives, both chemically[21] and enzymatically.[22-23] The former inevitably involves lengthy protection/deprotection strategies, whilst the latter is particularly low-yielding for 4-position analogues. Therefore, I used an alternative chemical route for this work, with a trehalose starting material and by taking advantage of the differential reactivity of trehalose hydroxyls with respect to esterification, namely: 6,6' > 2,2' > 3,3' > 4,4'.[24] As had been previously reported, it was possible to isolate hepta-benzoylated trehalose with a free 4-OH in 43% yield (lit. 58%),[25] by optimisation of reaction conditions and careful separation of product

mixtures. The number and position of protecting groups were confirmed by MALDI mass spectrometry and full assignment of ^{1}H NMR spectra, respectively.

With compound **5a** in hand, it was possible to generate most of the desired 4-deoxy trehalose analogues in two or three steps, as summarised in **Figure 5.2**. Reaction with diethylaminosulphur trifluoride (DAST) and dimethylaminopyridine (DMAP) achieved fluoro substitution with inversion of stereochemistry in 52% yield to give **5d**, and subsequent deprotection gave the desired axial fluoro-trehalose analogue **5e**. Inversion of the 4-hydroxyl stereochemistry of **5a** was achieved via a Lattrell-Dax epimerisation, aided by the equatorial orientation of the neighbouring 3-OBz.[26] This involved conversion to a triflate leaving group and subsequent KNO_2 substitution and hydrolysis to give **5b** in 67% yield (lit. 72%).[25] This could be either deprotected directly to give target compound **5c**, or reacted further to generate equatorial 4-deoxy trehalose analogues.

An equatorial azide group was added via a further triflate leaving group and NaN_3, in the presence of 15-crown-5, producing **5f** in 79% yield (lit. 73%)[25], which was deprotected to give target compound **5g**. Reaction of **5b** with DAST to generate the equatorial fluoro analogue was more challenging due to a competing elimination reaction. Multiple purification steps were required to eventually isolate purified **5h** in 37% yield, which was deprotected to give target compound **5i**.

Synthesis of **4-OMe** was attempted using the Purdie methylation with MeI and Ag_2O, with an aim to avoid the highly basic conditions of more commonly used methylation procedures that would remove the alkali-labile benzoyl protecting groups. However, a complex mixture of differently benzoylated and methylated products was formed in a slow reaction, from which it was not possible to isolate the target compound. The addition of dimethyl sulphide significantly increased the rate of starting material consumption, but did not aid isolation of **4-OMe**, so the synthesis of **4-OMe** was not pursued further.

The identity and orientation of novel functional groups were confirmed by 1D and 2D ^{1}H, ^{13}C and ^{19}F NMR spectroscopic analyses. Interestingly, no coupling was observed between H-4 and H-5 in **5d** or **5e** and between F and H-5 in **5h** or **5i**, indicating that the 4-equatorial position in these compounds lies nearly orthogonal to the H5 proton, resulting in *J* coupling values of zero.
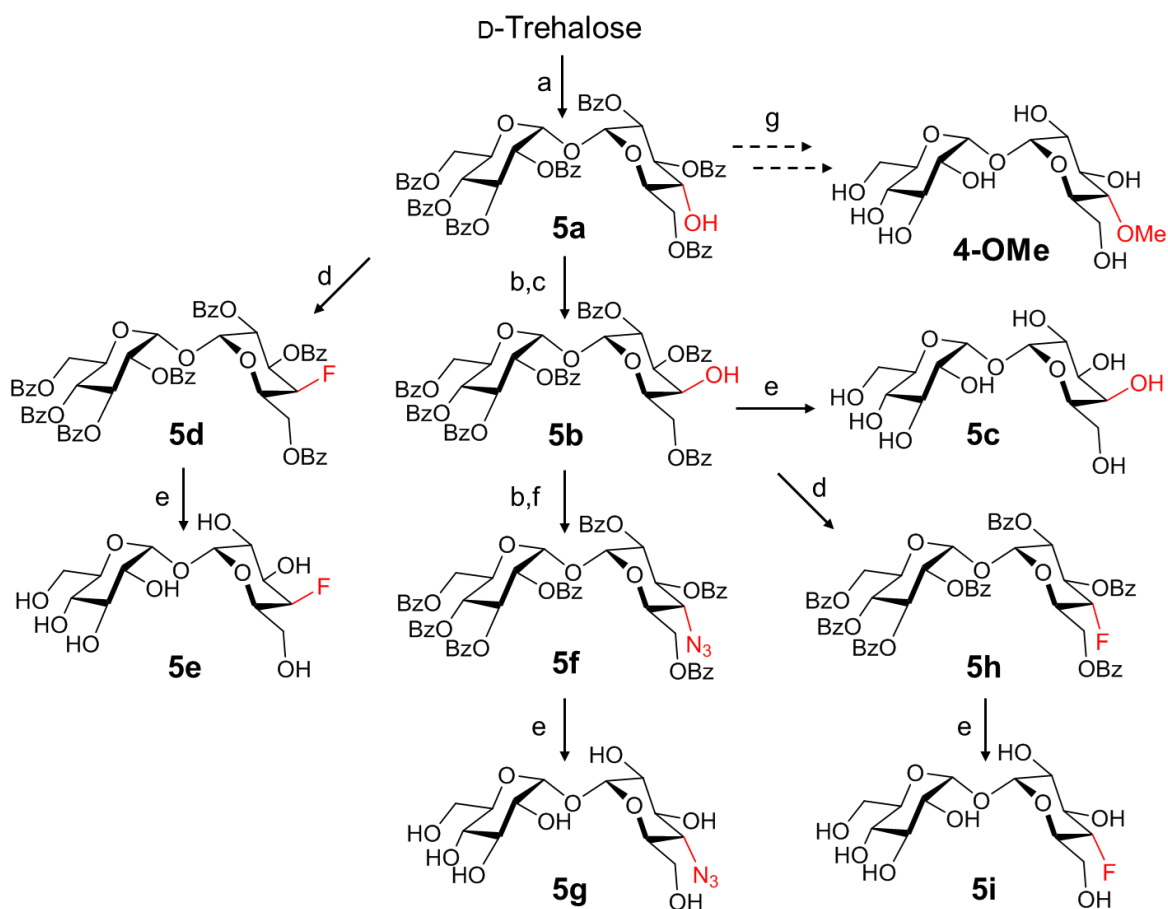
Figure 5.2 Scheme of 4-deoxy trehalose analogue synthesis. Modifications at the 4-position are highlighted in red. Reagents and conditions: (a) BzCl, pyridine, 3 h, 0 °C – RT; (b) Tf$_2$O, CH$_2$Cl$_2$, pyridine, 0 °C-RT, 0.5 h; (c) KNO$_2$, DMF, RT, 3 h; (d) DAST, DMAP,, CH$_2$Cl$_2$, 0 °C – RT, 18 h; e) NaOMe/MeOH, 18 h; (f) NaN$_3$, 15-crown-5, DMF 60 °C, 6 h; (g) Ag$_2$O, MeI, Me$_2$S, RT, 5 h. Yields: **5a** 43%; **5b** 67%; **5d** 52%; **5f** 79%; **5h** 37%; **5c, 5e, 5g, 5h** quantitative.

5.3   4-Deoxy trehalose analogues are incorporated into *M. smegmatis* cells by the trehalose recycling pathway

5.3.1   Compound 5e inhibits the growth of *M. smegmatis* at high concentrations

To test whether the synthesised 4-deoxy trehalose analogues had any negative effect on mycobacterial growth, I grew *M. smegmatis* for 25 h in the presence of the compounds at concentrations of up to 500 µM. At this concentration, **5g** had no effect on *M. smegmatis* growth, consistent with what had previously been reported,[14] along with compounds **5c** and **5i**, for which growth inhibition studies had not been described. On the other hand, **5e** had

previously been shown to be a substrate for the mycolate transferase Ag85 and showed mild inhibition against *M. tuberculosis*, with an MIC of 100 µg/mL (290 µM).[13] Indeed, there was also a notable reduction of *M. smegmatis* growth with **5e** at 500 µM, **Figure 5.3**. I then attempted to determine the MIC for this compound against *M. smegmatis*, which was > 1 mM, indicating this compound is less potent against *M. smegmatis*. Therefore, as the compounds showed no or very low growth inhibition, I went on to determine whether they were incorporated into mycobacterial cells.
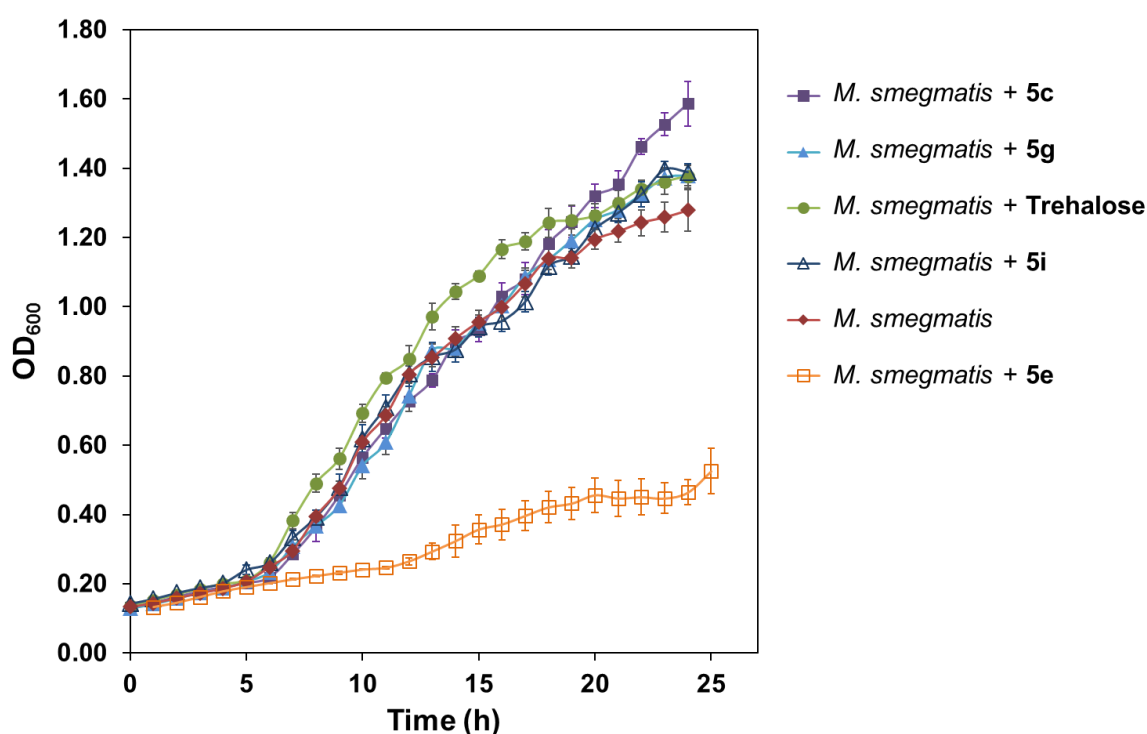
Figure 5.3 Compound 5e at 500 µM partially inhibits *M. smegmatis* growth. The growth of *M. smegmatis* WT over 25 h was measured by optical density at 600 nm with or without 500 µM trehalose or 4-deoxy trehalose analogues **5c**, **5e**, **5g** or **5i**. Only compound **5e** negatively affected *M. smegmatis* growth. Data are representative of at least two independent experiments

5.3.2 4-Deoxy trehalose analogues are incorporated into *M. smegmatis* cells

To investigate the incorporation of the synthesised analogues into mycobacterial cells, I grew *M. smegmatis* WT in the presence of each compound. The previous study with **5g** showed that 250 µM was required to detect cell labelling, ten-fold higher than for 2-azido trehalose or 6-azido trehalose.[14] Therefore this concentration was used for uptake experiments and this was sufficient to allow reasonable *M. smegmatis* growth with **5e**. I then lysed the cells and analysed the water-soluble extracts by high performance anion exchange chromatography (HPAEC) with pulsed amperometric detection (PAD) comparing

these to standards for each compound, **Figure 5.4**. As expected, all extracts, including those from untreated cells contained a peak corresponding to endogenous trehalose. Compound **5g** was observed in cytosolic extracts as expected, and the uptake of **5c** and **5e** was also observed, with peaks that were clearly larger than that of trehalose. On the other hand, **5i** was only detected at very low amounts, far less than that of trehalose. A study subsequent to this work has reported no uptake of **5i** in *M. smegmatis* at 25 µM,[20] demonstrating that uptake is concentration dependent. Overall, all analogues were incorporated into *M. smegmatis* cells, with **5i** at much lower levels.

5.3.3   4-Deoxy trehalose analogues are incorporated via the trehalose recycling pathway

Trehalose can be incorporated into the mycobacterial cell either by being anchored to the cell wall via the periplasmic Ag85 route or by being taken into the cytosol by the LpqY-SugABC mediated recycling pathway.[10] Previous work demonstrated that labelling with **5g** was absent in a *M. smegmatis ΔsugC* strain, showing incorporation via the recycling pathway.[14] To determine whether this held true for the other trehalose analogues, I repeated the growth and extractions with the *M. smegmatis ΔsugC* strain, as in section **5.3.2.** For this strain, I saw an additional peak in the extracts from untreated cells, running at a longer retention time than trehalose. However, this was consistent across all samples and did not affect subsequent analyses. For all the analogues, uptake was negated in the *ΔsugC* strain, **Figure 5.4**, demonstrating that these are all incorporated via the trehalose recycling pathway. This conclusively demonstrates that all synthesised compounds were imported into the mycobacterial cytosol, where they can potentially be processed by the enzymes of the GlgE pathway, fulfilling the first criterion for targeting GlgE by the approach set out in this chapter.
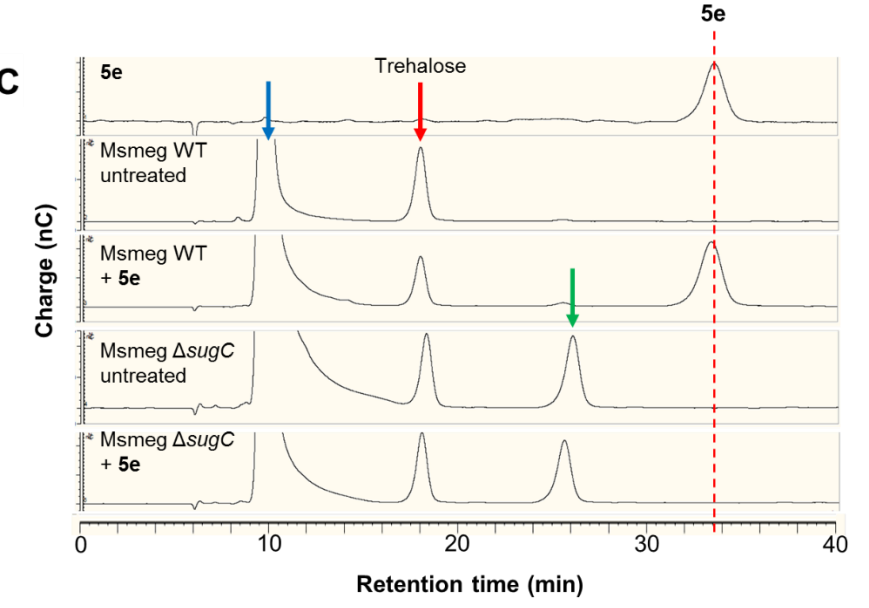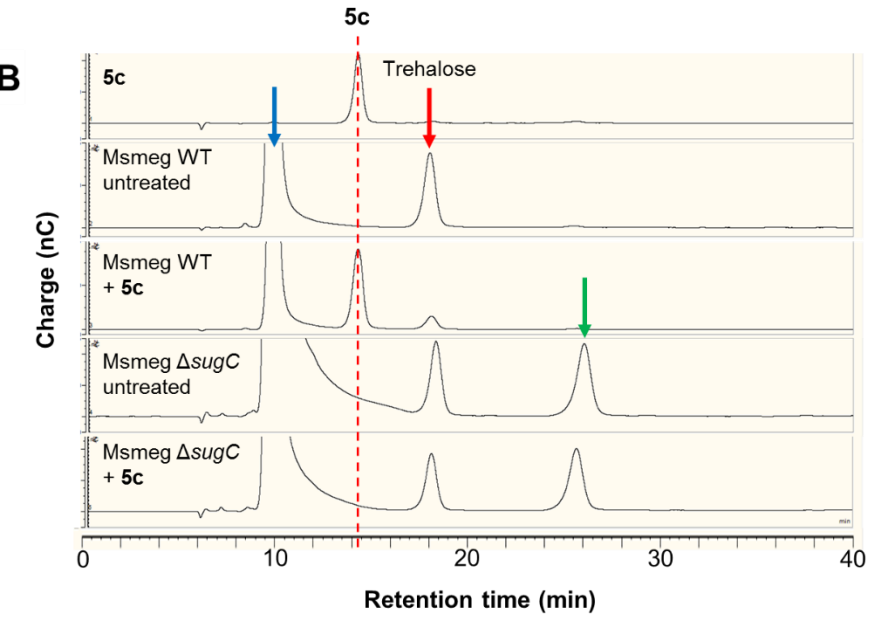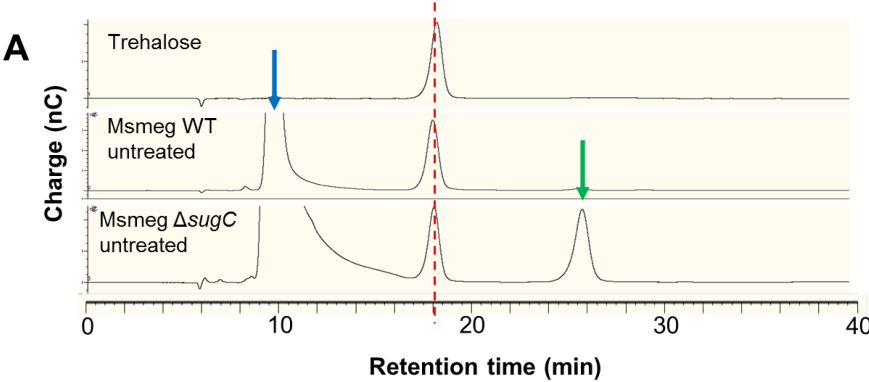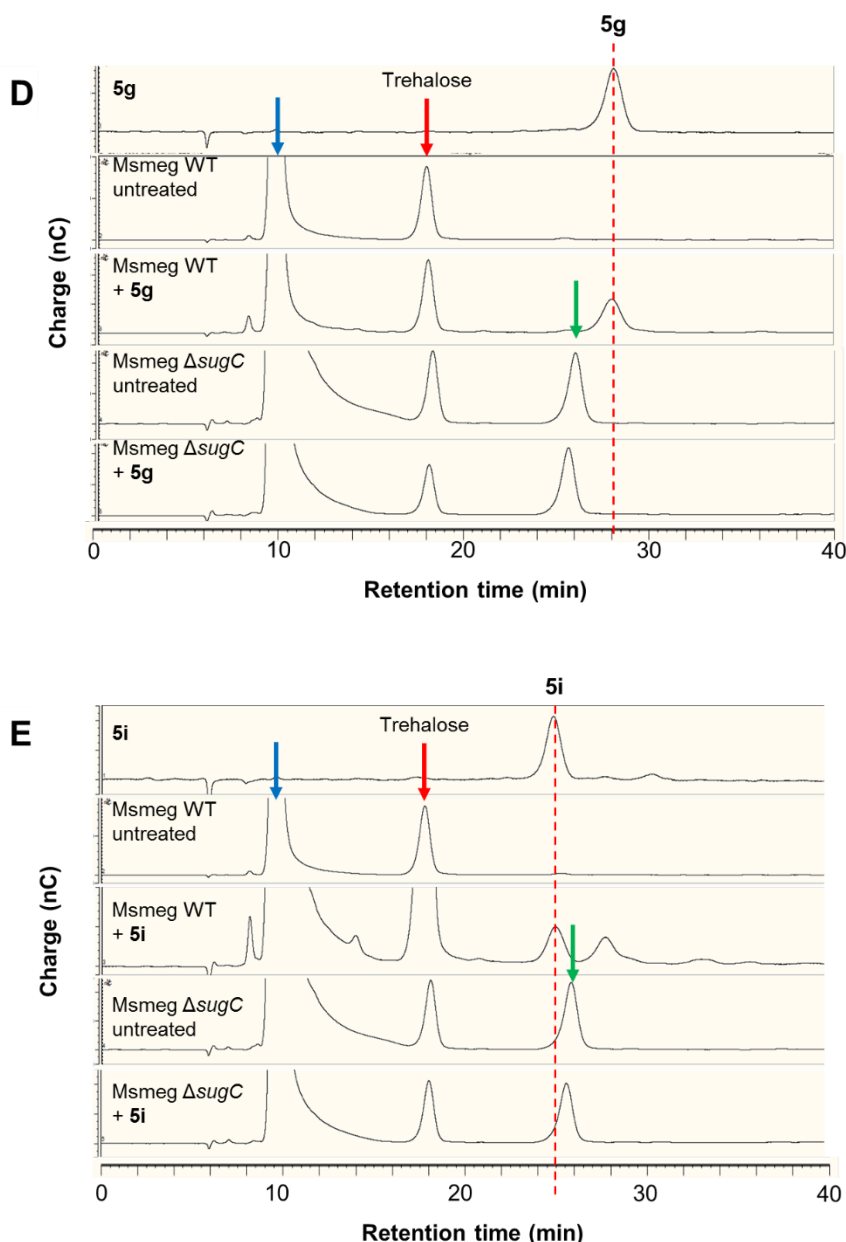
Figure 5.4 4-Deoxy trehalose analogues are incorporated into *M. smegmatis* cells via the trehalose recycling pathway. HPAEC-PAD analyses of cytosolic extracts from *M. smegmatis* WT or *M. smegmatis ΔsugC* strains after incubation with 250 µM of 4-deoxy trehalose analogues. Chromatograms are scaled to the trehalose or analogue peak as appropriate. **A)** Extracts from untreated *M. smegmatis* WT and *M. smegmatis ΔsugC* show a peak corresponding to trehalose running at 18.0 min retention time, marked with a red arrow in the following panels. A large peak indicated with a blue arrow at a lower retention time likely corresponds to glucose and other reducing sugars that are poorly retained in the analytical column. Indicated with a green arrow is a peak of unknown identity running at a higher retention time, that was consistently observed in *M. smegmatis ΔsugC* extracts but this did not interfere with subsequent analyses. **B)** Incubation with **5c**; **C)** Incubation with **5e**; **D)** Incubation with **5g; E)** Incubation with **5i. B-D** show clear LpqY-SugABC dependent uptake. **E** shows a very low level of LpqY-SugABC dependent uptake. For all chromatograms dashed red lines indicate the alignments with standards. Data are representative of at least two independent experiments.

5.3.4 Axial 4-deoxy trehalose analogues accumulate at higher levels relative to trehalose in the cell

In the analysis of cytosolic extracts shown in **Figure 5.4**, the size of the chromatogram peaks for each analogue varies in comparison to that of trehalose. However, direct comparison is not valid as modifications to trehalose not only change ionisation and therefore, retention time in HPEAC, but also affect the capacity for oxidation, and hence the signal seen by PAD.[27] To roughly quantify the level of accumulation relative to endogenous trehalose, 1:1 injections of each compound with trehalose were analysed and compared to cytosolic extractions, as shown in **Table 5.1**. The level of detection of **5c** was similar to that of trehalose, as would be expected for these diastereoisomers. The signal seen for **5g** was slightly less than that of trehalose, indicating the azide group moderately decreased the signal seen for this compound. The signals for the fluoro derivatives **5e** and **5i** were 3.0 and 4.7 times less than that of trehalose, corresponding with the decrease in oxidation capacity resulting from a fluoro group substitution.

Table 5.1 4-Deoxy trehalose analogues accumulate at different levels relative to trehalose. Each analogue was mixed in a 1:1 molar equivalence with trehalose. The resultant chromatogram signals were compared by determining the areas under each peak. The observed trehalose:analogue ratio varied depending on the identity of the compound, with the signal for trehalose appearing 1.3 – 4.7 fold increased relative to the analogue. This was used to adjust the ratio of trehalose:analogue observed in cytosolic extracts to give a more accurate estimate of the amount of each analogue relative to trehalose in the cell. Axial analogues **5c** and **5e** were both accumulated at approximately ten-fold higher levels than that of trehalose. Equatorial azido analogue **5g** accumulated at approximately equal levels to that of trehalose. There was a three-fold excess of trehalose relative to the equatorial fluoro analogue **5i**. Ratios are given to one decimal place.

Compound	Retention Time (min)	Trehalose: Analogue Ratio		
		1:1 Injection of Standards	Cytosolic Extracts	Adjusted
5c	14.1	1.3	0.2	0.1
5e	33.7	3.0	0.4	0.1
5g	28.0	1.6	1.4	0.9
5i	25.2	4.7	13.5	2.9

Even when taking into account the lower signal observed for the fluoro analogues, **5i** was present at a level three-fold less than that of trehalose. The level of **5g** was approximately equal relative to trehalose. Interestingly, levels of both axial analogues were about ten-fold higher than that of endogenous trehalose. If the endogenous levels of trehalose remained

unchanged in these experiments, these differences could be attributed to changes in the efficiency of uptake or different levels of incorporation into trehalose metabolic pathways for each compound. However, as the presence of the compounds may have secondary effects on trehalose metabolism, for example in stress response pathways or inhibition of other trehalose-utilising pathways, it is not possible to make further conclusions from these experiments. The wider metabolic picture is clearly an important further consideration for the utilisation of trehalose analogues as precursors to GlgE inhibitors. Having established that the synthesised analogues are incorporated into the mycobacterial cytosol, the next step was to test whether they could be processed into the GlgE pathway.

5.4 *M. tuberculosis* TreS does not isomerise 4-deoxy trehalose analogues

To investigate the possibility of incorporation of the synthesised analogues into the GlgE pathway, I investigated whether they could be processed by recombinant *M. tuberculosis* TreS. The interconversion of trehalose and maltose by recombinant *M. tuberculosis* TreS had previously been studied in-depth with ^{1}H NMR spectroscopy.[16] When starting with a trehalose substrate, conversion to maltose can be monitored by the depletion of the trehalose H-1 peak at 5.10 ppm and the build-up of the α-maltose H-1' and H-1 peaks at 5.32 ppm and 5.14 ppm, respectively. The latter peak overlaps with the H-1 peak of α-glucose that is released in a hydrolytic side reaction. The picture is further complicated by the mutarotation of α-maltose and α-glucose to give the corresponding β-anomers, as illustrated in **Figure 5.5**. Following the previously established protocol, 10 mM of **5c**, **5e**, **5g** or **5i** were incubated with 2 µM recombinant *M. tuberculosis* TreS and ^{1}H NMR spectra recorded over time.
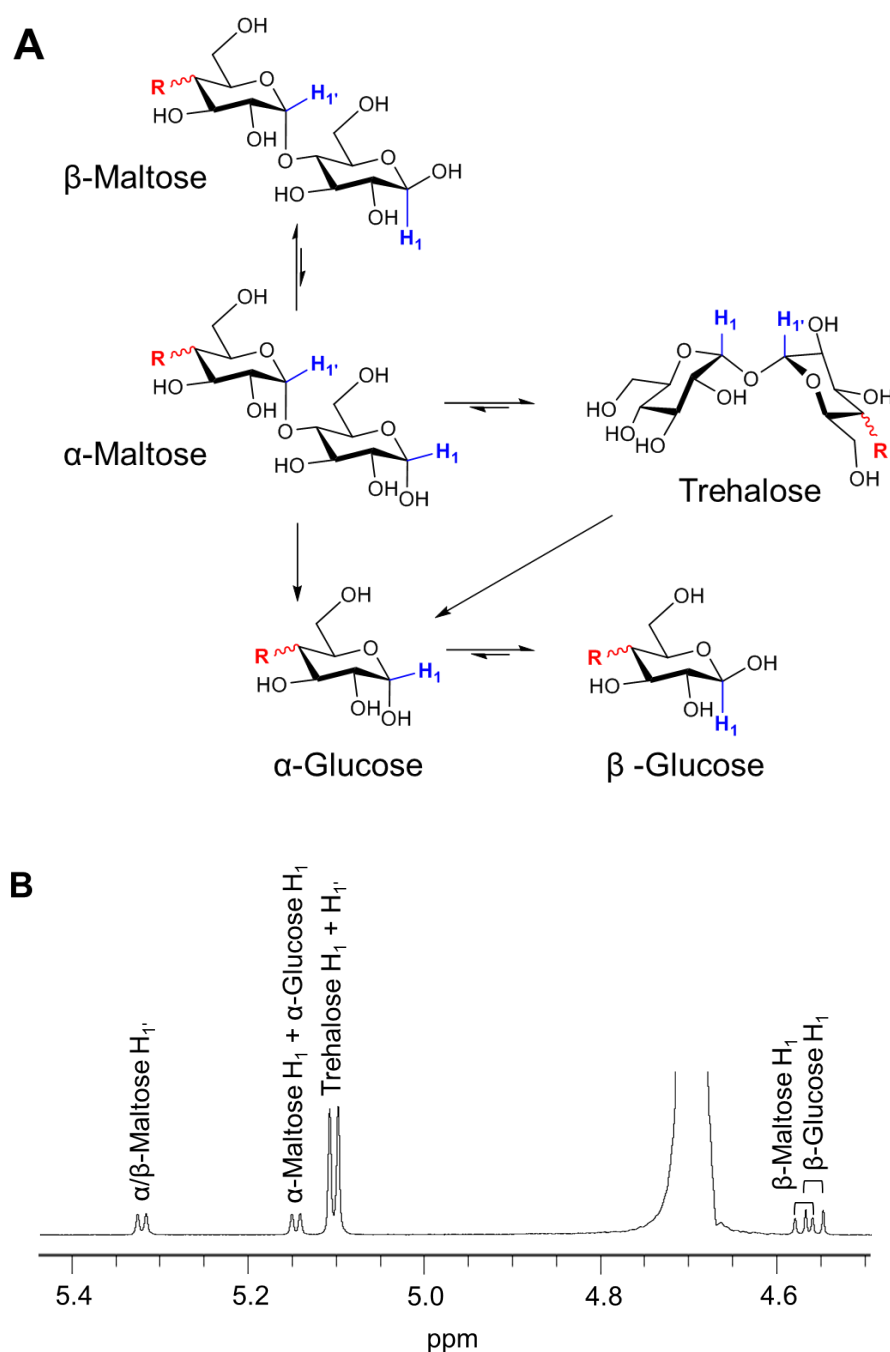
Figure 5.5 Monitoring the reaction of TreS by ^{1}H NMR spectroscopy. A) The species present in a TreS-catalysed reaction. TreS catalyses the interconversion between trehalose and α-maltose, as well as a slower hydrolysis side reaction that releases α-glucose. The equilibrium position favours the formation of trehalose, however the flux *in vivo* is in the α-maltose-direction. Relatively slow, non-enzymatic mutarotation of α-maltose and α-glucose gives the corresponding β-anomers. Potential 4-deoxy trehalose modifications are highlighted in red; R = OH for natural substrates. Anomeric protons are highlighted in blue. Adapted from Miah *et. al.* 2013.[16] **B)** ^{1}H NMR spectrum of TreS reaction products. The ^{1}H NMR spectrum in the region 4.5 - 5.4 ppm after trehalose has been incubated with TreS. Peak labels correspond to the protons highlighted in A.
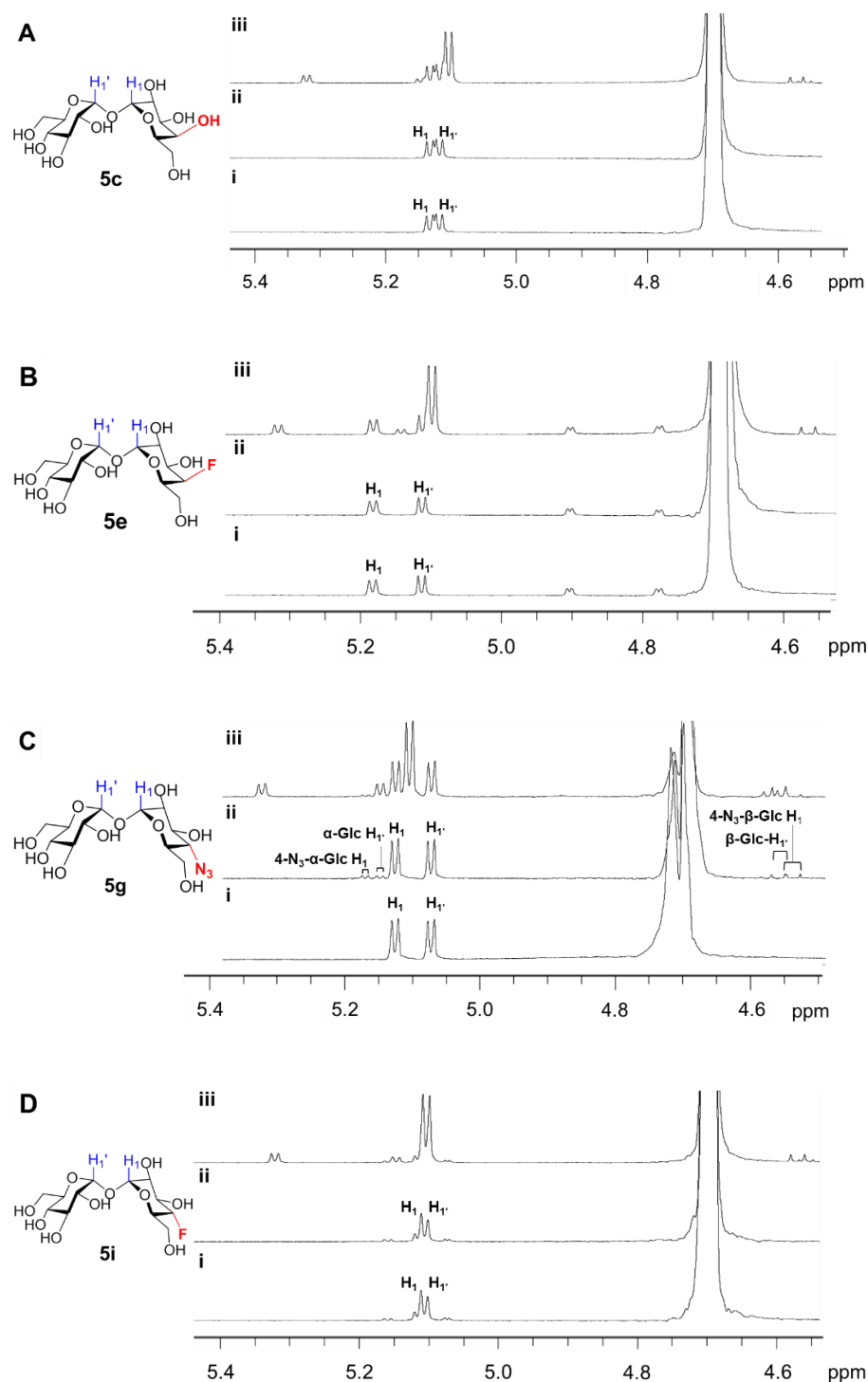
Figure 5.6 4-Deoxy trehalose analogues are not isomerised to maltose analogues by *M. tuberculosis* TreS. Representative ^{1}H NMR spectra recorded (i) before addition of TreS, (ii) 40 h after addition of TreS and (iii) 52 h after TreS-addition, 6 h after addition 10 mM trehalose. For **5c**, **5e** and **5i** spectra were identical after 40 h and up to one week (data not shown) after TreS-addition. For **5g**, peaks corresponding to hydrolytic products were first observed after 24 h. The addition of 10 mM trehalose was followed by the immediate appearance of α-maltose and α-glucose peaks, along with the corresponding β anomers, **Figure 5.5**. This demonstrates the activity of TreS was not impaired and suggests it cannot use 4-deoxy trehalose analogues as substrates.

I expected that the corresponding maltose analogues would form more slowly from the 4-deoxy trehalose analogues, as the formation of an α-1,4 bond would be impeded when the analogue orientation was such that the 4-position modification obstructed bond formation. However, conversion to the corresponding maltose analogue products was not observed even after incubation periods of greater than 1 week. After 46 h, addition of 10 mM trehalose showed rapid conversion to maltose, demonstrating that TreS activity was not obviously impeded and trehalose was the preferred substrate. For **5g** a small amount of hydrolysis to glucose and 4-azido glucose was observed after 24 h. This suggests that *M. tuberculosis* TreS does not tolerate modifications at the 4-position of trehalose and therefore this type of trehalose analogue is not suitable to use as a precursor to a GlgE inhibitor.

5.5   Discussion

This chapter describes the exploration of trehalose-based precursors of GlgE inhibitors with 4-deoxy trehalose analogues. I aimed to synthesise a selection of trehalose analogues, test whether these were imported into the mycobacterial cytosol, and determine whether these were processed in the GlgE pathway, with a view to then modifying the compounds to achieve GlgE inhibition. To this end, I synthesised four 4-deoxy trehalose analogues and demonstrated that these were incorporated to the *M. smegmatis* cytosol via the trehalose recycling pathway. However, the synthesised analogues were not processed by recombinant *M. tuberculosis* TreS, demonstrating that this type of analogue is not incorporated into the GlgE pathway and therefore, is not suitable for developing further as a GlgE inhibitor.

The synthesis of the trehalose analogues was achieved by modifying a previously established synthetic route to 4-azido trehalose,[25] which was readily adapted to make other 4-deoxy trehalose analogues. These are not accessible by the chemoenzymatic method for making trehalose analogues established by the Swarts group[22-23] and so this work provides a versatile route to this type of trehalose analogue that is likely to be valuable for other researchers. 4-Deoxy trehalose analogues could prove useful not only for mycobacterial research, but also in the investigation of the diverse roles of trehalose in other organisms, including as a stress protectant, energy store and carbon source.[28]

These small modifications to the natural trehalose scaffold led to markedly different uptake and inhibition profiles in *M. smegmatis*. All four analogues were taken up via LpqY-SugABC, which is specific for α,α-trehalose over other disacchardies, but able to tolerate a range of asymmetric trehalose modifications.[10, 14, 20] This work shows for the first time that LpqY-

SugABC can accommodate trehalose analogues with different stereochemistry at the 4-position than that of the natural substrate. However, modifications at the 4-position are generally not tolerated that well, because ten-fold higher concentrations are required for their uptake, compared to trehalose analogues altered at the 2 or 6 positions.[14, 20] There is also some variation in LpqY-SugABC uptake across different mycobacterial species; for example **5g** was not taken up *M. bovis*.[14] The structure of LpqY-SugABC, which has not yet been determined, could shed more light on substrate constraints and how these might differ across species. It is also important to note that *M. tuberculosis* and *M. smegmatis* differ in their carbohydrate uptake capabilities. There are just five carbohydrate uptake permeases in *M. tuberculosis* versus 28 in *M. smegmatis*, most likely reflecting the limited nutritional environment of the phagosome versus the soil.[9]

I showed that at 500 µM, **5e** partially inhibited *M. smegmatis* growth, reflecting the inhibition seen in *M. tuberculosis*, but with a higher MIC of >1 mM in *M. smegmatis*, compared to 290 µM in *M. tuberculosis*. Compound **5e** is a substrate for recombinant Ag85 *in vitro*, suggesting that the growth inhibition of **5e** could be due to interference in trehalose mycolate production.[13] My work shows that whilst **5e** is not processed by TreS, it accumulates in the cytosol at levels approximately ten-fold higher than trehalose, and therefore another mechanism of growth inhibition is possible. The increase in the relative amount of 4-deoxy trehalose analogues with axial substituents versus equatorial ones could mean these are more efficiently imported via LpqY-SugABC. Alternatively, they may serve as worse substrates for trehalose metabolising enzymes other than TreS, such as trehalases. It cannot be ruled out that the presence of the axial analogues decreases the cellular levels of trehalose, for example by inhibiting trehalose production. The metabolic profiles could be investigated in more detail with a chromatography method combined with mass spectrometry, such as GC-MS, and the use of internal standards and calibration curves for more accurate quantification.

A more detailed analysis of the TreS mechanism gives insight into the lack of activity with 4-deoxy trehalose analogues, **Figure 5.7**. The interconversion of trehalose and α-maltose by TreS follows a two-step double displacement mechanism with retention of configuration, as seen across the GH13 enzyme family.[16, 19] In the trehalose to maltose direction, the binding of trehalose is followed by the formation of a covalent enzyme intermediate with the nucleophilic aspartate residue and the glucose unit in the -1 subsite. The glucose unit in the +1 subsite is mostly retained in the active site, with hydrolysis occurring in approximately 10% of reactions. The +1 glucose moiety is repositioned to allow attack by the 4-OH, to form an α-1,4 bond and give maltose.[19] Therefore, the conversion of 4-deoxytrehalose into

4-deoxymaltose could only be achieved when the trehalose analogue binds such that the modified 4 position is located in the -1 subsite.
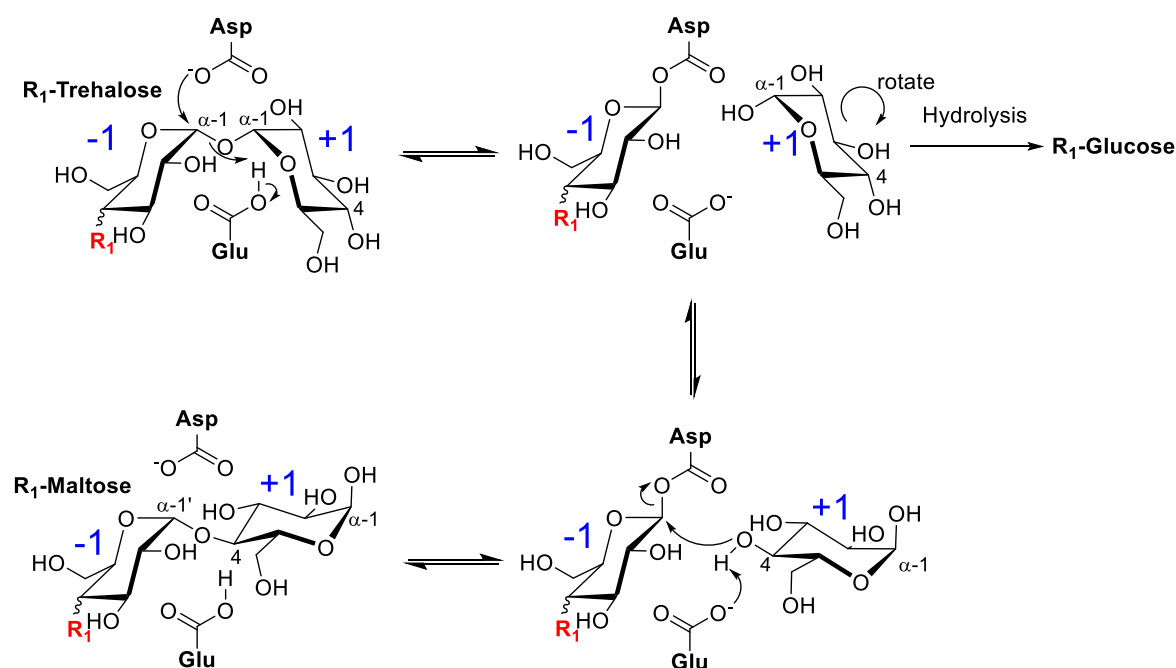
Figure 5.7 The catalytic mechanism of *M. tuberculosis* TreS. The *M. tuberculosis* TreS catalytic residues Asp (nucleophile) and Glu (acid/base) are labelled. The 4-position substitution in the disaccharide is indicated by R_1 in red, with this being an equatorial hydroxyl in the natural substrate/product. The -1 donor and +1 acceptor subsites are labelled in blue. Trehalose binds at the active site, which allows attack by the nucleophile (top left panel) to form a covalent intermediate (top right panel). Attack by a water nucleophile leads to hydrolysis, which occurs in approximately 10% of reactions to release a glucose side product. Rotation of the glucose in the +1 subsite orientates the 4-OH for nucleophilic attack of the covalent intermediate (bottom right panel), forming an α-1,4 glycosidic bond and releasing maltose (bottom left panel). The conversion of maltose to trehalose proceeds by the reverse pathway.

The lack of substrate-bound TreS structures mean that the environments of the subsites are unknown. It could be argued that binding with the modified 4-position in the +1 subsite is more likely to be tolerated because this site would be inherently less restrictive and allow for rotation. Binding in this orientation could promote hydrolysis due to the lack of the 4-OH nucleophile for α-1,4 bond formation, but hydrolysis was only seen in small amounts for **5g**. Coupled with the lack of significant TreS inhibition by the 4-deoxy trehalose analogues, this suggests that **5c**, **5e** and **5i** are unable to bind well to TreS. For the galactose-type analogues **5c** and **5e** this can be attributed to the change in stereochemistry. Compound **5i** might be expected to be more tolerated by TreS, at least in the +1 site, and the lack of activity could be due to the observed distortion of C-F bond, demonstrated by the altered J

values (Section **5.2.2**). Crystal structures of TreS in an active state bound to trehalose or maltose substrates would reveal more details about the environment of the active site. The currently available TreS structures are of the enzyme in an inactive state with a leucine residue apparently blocking substrate binding,[17, 29] precluding modelling of the substrates into the active site.

It was previously reported that the labelling of mycobacterial cells with **5g** was decreased in a *M. smegmatis* Δ*treS* mutant strain.[14] One explanation put forward for this was that **5g** was incorporated into the cells via the GlgE pathway, hence when this is disrupted in a Δ*treS* knockout strain less of **5g** is incorporated into the mycobacterial cell. However, the findings presented in this chapter demonstrate that **5g** is not processed along the GlgE pathway, but instead I showed that TreS has a low level of hydrolytic activity with **5g**. An alternative explanation for the original observation is that **5g** is outcompeted by trehalose in the Δ*treS* mutant strain. Experiments in the original paper indeed demonstrated that **5g** is most sensitive to trehalose competition. Additionally, there are elevated levels of trehalose in the Δ*treS* strain. Thus, **5g** is outcompeted by the natural substrate and hence the cell labelling with this compound is depleted.[14]

The next steps for following up this work are the synthesis of trehalose analogues with modifications at different positions and testing whether these are converted to the corresponding maltose analogues by *M. tuberculosis* TreS. This is necessary to validate the approach of using trehalose analogues as precursors to GlgE inhibitors. The obvious candidates are trehalose analogues modified at the 2 and/or 6 positions, because trehalose analogues modified at the 3 position are tolerated less well by LpqY-SugABC.[14, 20] The 1 and 5 position analogues are synthetically accessible, but more likely to interfere with the mechanism of TreS. If any of these were successful, the analogues could then be tested with Pep2, either sequentially or with the TreS/Pep2 complex, to test for formation of the corresponding M1P analogue. The structures could be modified to achieve an end product capable of inhibiting GlgE and establish SAR, before testing compounds *in vivo*.

5.6   Summary

I explored the idea of trehalose analogues acting as precursors for GlgE inhibitors, using four synthesised 4-deoxy trehalose analogues. I showed that in *M. smegmatis* the 4-deoxy trehalose analogues were taken up via the LpqY-SugABC mediated trehalose recycling pathway, at different levels depending on the functional group and stereochemistry at the 4-position. However, *M. tuberculosis* TreS was not tolerant of these 4-position

modifications, indicating that 4-deoxy trehalose analogues are not suitable for this overall approach.

5.7   Materials and methods

5.7.1   Synthesis of 4-deoxy trehalose analogues

5.7.1.1 General methods

All solvents used in organic reactions were anhydrous and purchased from Sigma Aldrich and used as supplied. Anhydrous D-trehalose was purchased from Tokyo Chemical Industry. All other solvents and reagents were purchased from Sigma Aldrich or Fisher Scientific and used without further purification. Analytical TLC was carried out using Merck aluminium backed sheets coated with 60-F254 silica gel. Spots were visualised by dipping the TLC plate in ethanolic 5% H_2SO_4 solution followed by heating with a heat gun. Compounds were purified by automated flash column chromatography using a Biotage Isolera™ Spektra One system with KP-Sil™, with Biotage SNAP and SNAP Ultra cartridges (Uppsala, Sweden). Optical rotations were measured at ambient temperature on a Perkin Elmer model 141 polarimeter using a sodium lamp. NMR spectra were recorded at 22 °C on a Bruker Avance III spectrometer at 400 MHz (^{1}H), 100 MHz (^{13}C) or 376 MHz (^{19}F). Chemical shifts (δ) are reported in ppm using residual solvent signals from deuteriated solvents as references (CDCl$_3$: δ ^{1}H 7.26 and δ ^{13}C 77.0). Assignments were made with the aid of COSY and HSQC experiments. Chemical shifts of ^{1}H-coupled ^{19}F NMR signals recorded in CDCl$_3$ are reported with respect to external CFCl$_3$ at δ ^{19}F 0 ppm. High-resolution mass spectra (HRMS) were obtained using a Synapt G2 HDMS mass spectrometer (Waters Corp., Manchester UK) using positive electrospray ionisation. Nominal and exact *m/z* values are reported in Daltons. After work-up of the reaction mixtures, organic layers were dried over MgSO$_4$, filtered and then concentrated under vacuum.

5.7.1.2 2,3,6-tri-*O*-benzoyl-α-D-glucopyranosyl-(1→1')-2',3',4',6'-tetra-*O*-benzoyl-α-D-glucopyranoside (5a)

5a
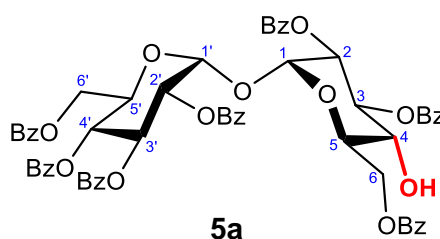
Anhydrous trehalose (5.00 g, 14.6 mmol) was dissolved in pyridine (300 mL) and the solution was stirred and cooled on ice. Benzoyl chloride (13.6 mL, 117 mmol) was added drop-wise. The mixture was allowed to warm to room temperature and was stirred for a further 6 h. At this point, TLC (4:1 toluene/EtOAc) showed three products: R_f 0.71, R_f 0.43 and R_f 0.16. The mixture was poured onto 1 M HCl (300 mL) and then extracted with CH_2Cl_2 (3 × 300 mL). The extract was washed successively with $NaHCO_3$ (1 L) and H_2O (1 L) and then concentrated. The crude residue was purified by column chromatography on silica gel (4:1 toluene/EtOAc) to afford **5a** (6.68 g, 43%) as a colourless solid: $[\alpha]_D^{20}$ + 231.4 (c 1.0, CH_2Cl_2), lit.[25] + 225 (c 1.97, CH_2Cl_2); R_f 0.43 (4:1 toluene/EtOAc); ^{1}H NMR (400 MHz $CDCl_3$): δ 8.08-7.98 (m, 8H, o-Bz), 7.93-7.88 (m, 4H, o-Bz), 7.86-7.82 (m, 2H, o-Bz), 7.60-7.51 (m, 5H, m-Bz, p-Bz), 7.49-7.36 (m, 12H, m-Bz, p-Bz), 7.34-7.26 (m, 4H, m-Bz, p-Bz), 6.25 (dd, 1H, $^3J_{2',3'}$ 10.1 Hz, $^3J_{3',4'}$ 10.0 Hz, H-3'), 5.93 (dd, 1H, $^3J_{2,3}$ 10.0 Hz, $^3J_{3,4}$ 9.6 Hz, H-3), 5.69-5.61 (m, 3H, H-1, H-1', H-4'), 5.51-5.43 (two overlapping dd, 2H, $^3J_{2',3'}$ 10.1 Hz, $^3J_{2,3}$ 10.0 Hz, $^3J_{1,2}$ 3.9 Hz, $^3J_{1',2'}$ 3.7 Hz, H-2, H-2'), 4.38 - 4.31 (m, 1H, H-5'), 4.21 (dd, 1H, $^2J_{6a,6b}$ 12.4 Hz, $^3J_{5,6b}$ 4.0 Hz, H-6b), 4.06-4.01 (m, 1H, H-5), 3.97 (dd, 1H, $^2J_{6a',6b'}$ 12.5 Hz, $^3J_{5',6b'}$ 2.7 Hz, H-6b'), 3.89-3.75 (m, 3H, H-4, H-6a, H-6a'), 3.34 (d, 1 H, $^3J_{4,OH}$ 4.6 Hz, D_2O-exchangable, OH); ^{13}C NMR (100 MHz, $CDCl_3$): δ 167.2, 166.9, 165.8, 165.6, 165.4, 165.0 (7C, C=O), 134.1, 133.8, 133.6, 133.5, 133.4, 133.3, 133.1 (7C, p-Bz), 130.0, 129.9, 129.9, 129.8, 129.8, 129.8, 129.5, 129.4, 129.2, 129.1, 128.9, 128.7, 128.7, 128.5, 128.5, 128.4, 128.4 (35C, i-Bz, m-Bz, o-Bz), 93.0 (C-1), 92.8 (C-1'), 73.6 (C-3), 71.2 (C-2'), 71.1 (C-5), 70.7 (C-2), 70.2 (C-3), 69.2 (C-4), 68.9 (C-4'), 68.6 (C-5'), 62.5 (C-6), 62.0 (C-6'); NMR spectra were in agreement with published values;[25] HRMS, ESI-TOF/MS positive (m/z): calcd. for $C_{61}H_{50}O_{18}$ [M+H]+: 1071.3070, found: 1071.3066.

5.7.1.3 2,3,6-tri-*O*-benzoyl-α-D-galactopyranosyl-(1→1)-2,3,4,6-tetra-*O*-benzoyl-α-D-glucopyranoside (5b)

5b
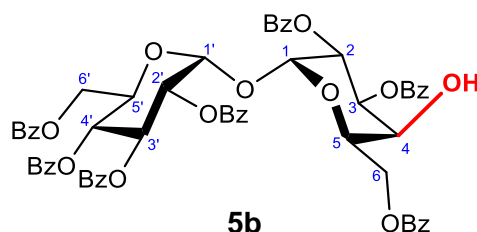
To a solution of **5a** (4.00 g, 3.74 mmol) dissolved in dry CH_2Cl_2 (50 mL) was added dry pyridine (3.00 mL, 37.4 mmol). The solution was cooled on ice and triflic anhydride (1.27 mL, 7.47 mmol) was added drop-wise. The reaction was allowed to reach room temperature and was stirred for a further 30 min, at which point TLC (4:1 toluene/EtOAc) showed

conversion to product. The mixture was diluted with CH_2Cl_2 (50 mL) and then successively washed with 1 M HCl (100 mL), 10% $NaHCO_3$ (100 mL) and H_2O (100 mL) and concentrated to give 4-deoxy-4-O-trifluoromethylsulphonyl-2,3,6-tri-O-benzoyl-α-D-glucopyranosyl-(1→1')-2',3',4',6'-tetra-O-benzoyl-α-D-glucopyranoside (**5a.1**) (3.82 g) as a crude yellow solid: R_f 0.69 (4:1 toluene/EtOAc). This was used directly without further purification.

To a solution of **5a.1** (3.82 g, 3.17 mmol) dissolved in DMF (50 mL) was added potassium nitrite (1.35 g, 15.9 mmol). The suspension was stirred at ambient temperature for 3 h at which point TLC (4:1 toluene/EtOAc) showed conversion to product. The mixture was then diluted with CH_2Cl_2 (50 mL), washed with brine (2 × 100 mL) and H_2O (100 mL) and concentrated to give the crude residue, which was purified by silica column chromatography (4:1 toluene/EtOAc) to give **5b** (2.68 g, 67% over 2 steps) as a colourless solid: $[\alpha]_D^{20}$ + 194.6 (*c* 1.0, CH_2Cl_2), lit.[25] + 151 (*c* 1.85, CH_2Cl_2); R_f 0.45 (4:1 toluene/EtOAc); ^1H NMR (400 MHz $CDCl_3$): δ 8.09-8.03 (m, 4H, o-Bz), 7.99-7.91 (m, 4H, o-Bz), 7.90-7.86 (m, 2H, o-Bz), 7.81-7.76 (m, 4H, o-Bz), 7.60-7.49 (m, 4H, m-Bz, p-Bz), 7.45-7.26 (m, 17H, m-Bz, p-Bz), 6.25 (dd, 1H, $^3J_{2',3'}$ 10.1 Hz, $^3J_{3',4'}$ 9.9 Hz, H-3'), 5.92 (dd, 1H, $^3J_{2,3}$ 10.7 Hz, $^3J_{1,2}$ 3.1 Hz, H-2), 5.89 (dd, 1H, $^3J_{2,3}$ 10.7 Hz, $^3J_{3,4}$ 2.4 Hz, H-3), 5.75 (d, 1H, $^3J_{1',2'}$ 3.8 Hz, H-1'), 5.72 (d, 1H, $^3J_{1,2}$ 3.0 Hz, H-1), 5.65 (dd, 1H, $^3J_{3',4'}$ 9.9 Hz, $^3J_{4',5'}$ 9.9 Hz, H-4'), 5.47 (dd, 1H, $^3J_{2',3'}$ 10.1 Hz, $^3J_{1',2'}$ 3.8 Hz, H-2'), 4.37-4.20 (m, 4H, H-4, H-5',H-6a, H-6b), 4.07-4.00 (m, 2H, H-5, H-6b'), 3.94 (dd, 1H, $^2J_{6a',6b'}$ 12.4 Hz, $^3J_{5',6a'}$ 5.0 Hz, H-6a'), 2.66 (d, 1H, $^3J_{4,OH}$ 3.6 Hz, D_2O-exchangable, OH); ^{13}C NMR (100 MHz, $CDCl_3$): δ 166.2, 165.9, 165.8, 165.7, 165.6, 165.4, 165.0 (7C, C=O), 133.7, 133.6, 133.5, 133.3, 133.3, 133.2 (7C, p-Bz), 129.9, 129.9, 129.8, 129.8, 129.7, 129.4, 129.3, 129.2, 129.1, 128.9, 128.7, 128.7, 128.7, 128.6, 128.4, 128.3, 128.3, (35C, i-Bz, m-Bz, o-Bz), 92.9 (C-1'), 92.2 (C-1), 71.3 (C-2'), 70.9 (C-3), 70.3 (C-3'), 69.0 (C-4'), 68.6 (C-5), 68.5 (C-5'), 68.1 (C-2), 67.4 (C-4), 62.5 (C-6), 62.2 (C-6'); NMR spectra were in agreement with published values[25]; HRMS, ESI-TOF/MS positive (*m/z*): calcd. for $C_{61}H_{50}O_{18}$ [M+H]$^+$: 1071.3070, found: 1071.3070.

5.7.1.4 α-D-galactopyranosyl-(1→1)- α-D-glucopyranoside (5c)

5c
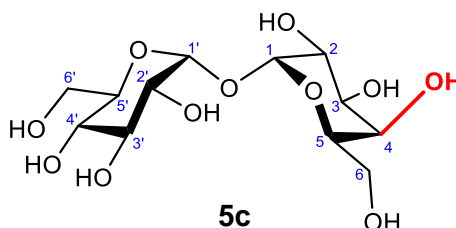
To a solution of **5b** (235 mg, 0.219 mmol) in anhydrous methanol (2 mL) was added freshly prepared 1 M NaOMe in MeOH (1.54 mL, 1.54 mmol). The reaction mixture was stirred overnight at room temperature at which point the reaction was complete by TLC. The solution was neutralised, titurated with hexane (3 × 2 mL) and passed through quaternary ammonium resin (BioRad AG 1-X8), which was then washed with water (3 × 2 mL), to give pure **5c** in quantitative yield: R_f 0.56 (1:1 MeOH/H_2O); ^1H NMR (400 MHz, D_2O): δ, 5.15 (d, 1H, $^3J_{1,2}$ 3.9 Hz, H-1), 5.14 (d, 1H, $^3J_{1',2'}$ 3.9 Hz, H-1'), 4.00 (dd, 1H, $^3J_{3,4}$ 6.2 Hz, $^3J_{4,5}$ 6.2 Hz, H-4), 3.97-3.91 (m, 2H, H-3, H-5'), 3.86-3.79 (m, 2H, H-2, H-6a), 3.78-3.63 (m, 5H, H-3', H-5, H-6a, H-6'b, H-6b), 3.57 (dd, 1H, $^3J_{2',3'}$ 10.0 Hz, $^3J_{1',2'}$ 3.8 Hz, H-2'), 3.37 (dd, 1H, $^3J_{2',3'}$ 9.4 Hz, $^3J_{3',4'}$ 9.4 Hz, H-4'); ^{13}C NMR (100 MHz, D_2O): δ 93.3 (C-1'), 93.1 (C-1), 72.5 (C-3'), 72.1 (C-5), 71.3 (C-4), 71.1 (C-2'), 69.7 (C-4'), 69.3 (C-3), 68.9 (C-5'), 67.9 (C-2), 61.2 (C-6'), 60.5 (C-6); NMR spectra were in agreement with published values;[25] HRMS, ESI-TOF/MS positive (*m*/*z*): calcd. for $C_{12}H_{22}O_{11}$ [M+Na]$^+$: 365.1054, found: 365.10656.

5.7.1.5   4-deoxy-4-fluoro-2,3,6-tri-*O*-benzoyl-α-D-galactopyranosyl-(1→1)-2,3,4,6-tetra-*O*-benzoyl-α-D-glucopyranoside (5d)

5d
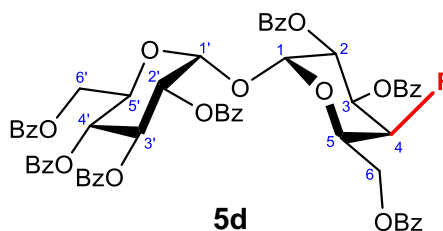
Compound **5a** (805 mg, 0.752 mmol) and DMAP (193 mg, 1.58 mmol) were dissolved in dry CH_2Cl_2 (15 mL) and cooled on ice. DAST (199 μL, 1.50 mmol) was added drop-wise and the solution was stirred for 18 h at room temperature. After this period, TLC (10:1 toluene/EtOAc) showed conversion to product. The reaction was quenched with H_2O (20 mL) and extracted twice with CH_2Cl_2 (2 × 20 mL). The combined organic phases were then washed with H_2O (40 mL), and concentrated. The crude residue was purified by column chromatography on silica gel (10:1 toluene/EtOAc) to give **5d** (423 mg, 52%) as a colourless solid: $[\alpha]_D^{20}$ +206.9 (*c* 1.0, CH_2Cl_2); R_f 0.53 (10:1 toluene/EtOAc); ^1H NMR (400 MHz $CDCl_3$): δ 8.09-8.05 (m, 4H, *o*-Bz), 7.95-7.91 (m, 4 H, *o*-Bz), 7.90-7.86 (m, 2H, *o*-Bz), 7.84-7.78 (m, 4H, *o*-Bz), 7.61-7.50 (m, 4H, *m*-Bz, *p*-Bz), 7.47-7.26 (m, 17H, *m*-Bz, *p*-Bz), 6.23 (dd, 1H, $^3J_{2',3'}$ 10.1 Hz, $^3J_{3',4'}$ 9.8 Hz, H-3'), 5.94 (ddd, 1H, $^3J_{3,F}$ 26.5 Hz, $^3J_{2,3}$ 10.7 Hz, $^3J_{3,4}$ 2.4 Hz, H-3), 5.89 (ddd, 1H, $^3J_{2,3}$ 10.7 Hz, $^3J_{1,2}$ 3.8 Hz, $^4J_{2,F}$ 1.1 Hz, H-2), 5.76-5.72 (two overlapping d, 2H, $^3J_{1',2'}$ 3.9 Hz, $^3J_{1,2}$ 3.8 Hz, H-1, H-1'), 5.65 (dd, 1H, $^3J_{3',4'}$ 10.0 Hz, $^3J_{4',5'}$ 10.2 Hz, H-4'), 5.46 (dd, 1H, $^3J_{2',3'}$ 10.1 Hz, $^3J_{1',2'}$ 3.9 Hz, H-2'), 5.09 (dd, 1H, $^2J_{4,F}$ 50.4 Hz, $^3J_{3,4}$ 2.4 Hz, $^3J_{4,5}$

0 Hz, H-4), 4.35-4.18 (two overlapping ddd, 2H, $^3J_{5,F}$ 27.6 Hz, $^3J_{4',5'}$ 10.2 Hz, $^3J_{5,6a}$ 6.9 Hz, $^3J_{5,6b}$ 6.7 Hz, $^3J_{5',6a'}$ 4.9 Hz, $^3J_{5',6b'}$ 2.5 Hz, $^3J_{4,5}$ 0 Hz, H-5,H-5'), 4.15-4.06 (m, 2H, H-6a, H-6b), 4.03 (dd, 1H, $^2J_{6a',6b'}$ 12.4 Hz, $^3J_{5',6b'}$ 2.5 Hz, H-6b'), 3.92 (dd, 1H, $^2J_{6a',6b'}$ 12.4 Hz, $^3J_{5',6a'}$ 4.9 Hz, H-6a'); ^{13}C NMR (100 MHz, CDCl$_3$): δ 165.9, 165.7, 165.4, 165.4, 165.4, 165.0 (7C, C=O), 133.8, 133.8, 133.7, 133.6, 133.3, 133.3, 133.2 (7C, *p*-Bz), 130.0, 129.9, 129.9, 129.8, 129.8, 129.7, 129.5, 129.3, 129.1, 129.1, 129.0, 128.8, 128.7, 128.6, 128.4, 128.3, 128.3 (35C, *i*-Bz, *m*-Bz, *o*-Bz), 92.9 (C-1'), 92.5 (C-1), 86.7 (d, 1C, $^1J_{C-4,F}$ 187.2 Hz, C-4), 71.3 (C-2'), 70.2 (C-3'), 68.9 (C-2), 68.8 (d, 1C, $^2J_{C-3,F}$ 18.3 Hz, C-3), 68.7 (C-4'), 67.9 (C-5'), 67.8 (d, 1C, $^2J_{C-5,F}$ 18.3 Hz, C-5), 62.1 (C-6'), 61.5 (d, 1C, $^3J_{C-6,F}$ 5.7 Hz, C-6); ^{19}F {^{1}H} NMR (376 MHz, CDCl$_3$): δ -218.8, $^2J_{4,F}$ 50.4 Hz, $^3J_{5,F}$ 27.6 Hz, $^3J_{3,F}$ 26.5 Hz; HRMS, ESI-TOF/MS positive (*m/z*): calcd. for C$_{61}$H$_{50}$O$_{18}$ [M+H]$^+$: 1073.3026, found: 1073.3026.

5.7.1.6 4-deoxy-4-fluoro-α-ᴅ-galactopyranosyl-(1→1)-α-ᴅ-glucopyranoside (5e)

5e
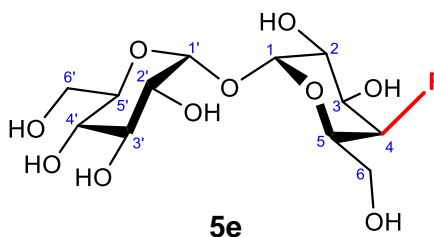
Compound **5d** (255 mg, 0.238 mmol) was deprotected as above to give pure **5e** in quantitative yield: R_f 0.75 (1:1 MeOH/H$_2$O); ^{1}H NMR (400 MHz, D$_2$O): δ 5.12 (d, 1H, $^3J_{1,2}$ 3.8 Hz, H-1), 5.05 (d, 1H, $^3J_{1',2'}$ 3.8 Hz, H-1'), 4.77 (dd, 1H, $^2J_{4,F}$ 50.6 Hz, $^3J_{3,4}$ 2.6 Hz, $^4J_{4,5}$ 0 Hz, H-4), 4.06-3.91 (m, 2H, H-3, H-5), 3.81 (dd, 1H, $^3J_{2,3}$ 9.8 Hz, $^3J_{1,2}$ 3.8 Hz, H-2), 3.73-3.57 (m, 6H, H-3', H-5', H-6a, H-6b, H-6'a, H-6'b;), 3.49 (dd, 1H, $^3J_{2',3'}$ 10.0 Hz, $^3J_{1',2'}$ 3.8 Hz, H-2'), 3.29 (dd, 1H, $^3J_{3',4'}$ 9.5 Hz, $^3J_{4',5'}$ 9.5 Hz, H-4'); ^{13}C NMR (100 MHz, D$_2$O): δ 93.5 (C-1, C-1'), 90.4 (d, 1C, $^1J_{C-4,F}$ 177.7 Hz, C-4), 72.5 (C-3'), 72.2 (C-5'), 71.0 (C-2'), 70.1 (d, 1C, $^2J_{C-5,F}$ 17.7 Hz, C-5), 69.6 (C-4'), 67.9 (C-2), 67.7 (d, 1C, $^2J_{C-3,F}$ 18.4 Hz, C-3), 60.5 (C-6'), 60.1 (d, 1C, $^3J_{C-6,F}$ 5.9 Hz, C-6); ^{19}F {^{1}H} NMR (376 MHz, D$_2$O): δ -219.6, ddd, $^2J_{4,F}$ 50.6 Hz, $^3J_{3,F}$ 30.4 Hz, $^3J_{5,F}$ 30.4 Hz; NMR spectra were in agreement with published values;[13] HRMS, ESI-TOF/MS positive (*m/z*): calcd. for C$_{12}$H$_{21}$FO$_{10}$ [M+Na]$^+$: 367.1011, found: 367.1013.

5.7.1.7 4-azido-4-deoxy-2,3,6-tri-*O*-benzoyl-α-D-glucopyranosyl-(1→1')-2',3',4',6'-tetra-*O*-benzoyl-α-D-glucopyranoside (5f)

5f
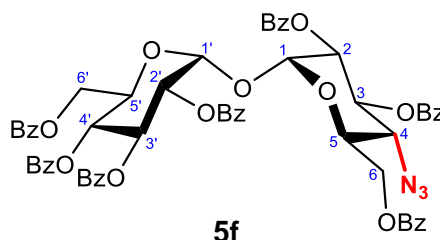
A solution of **5b** (866 mg, 0.809 mmol) and pyridine (0.651 mL, 8.09 mmol) dissolved in CH_2Cl_2 (12 mL) was cooled on ice and triflic anhydride (0.274 mL, 1.62 mmol) was added drop-wise. The reaction was allowed to reach room temperature and was stirred for a further 30 min, at which point TLC (4:1 toluene/EtOAc) showed conversion to product. The mixture was diluted with CH_2Cl_2 (13 mL) and then successively washed with 1 M HCl (25 mL), 10% $NaHCO_3$ (25 mL) and H_2O (25 mL) and concentrated to give 4-deoxy-4-*O*-trifluoromethylsulphonyl-2,3,6-tri-*O*-benzoyl-α-D-galatopyranosyl-(1→1)-2,3,4,6-tetra-*O*-benzoyl-α-D-glucopyranoside (**5b.1**) (787 mg) as a dark yellow solid: R_f 0.75 (4:1 Toluene/EtOAc). This was used directly without further purification.

To a solution of **5b.1** (787 mg, 0.654 mmol) dissolved in DMF (5 mL) was added dicyclopentano 15-crown-5 (13 µL, 65.4 µmol) and sodium azide (213 mg, 3.27 mmol). The suspension was stirred at 60 °C overnight at which point TLC (10:1 toluene/EtOAc) revealed a single product. The mixture was cooled to room temperature, diluted with CH_2Cl_2 (5 mL), washed with H_2O (3 × 10 mL), concentrated, and the crude residue was purified by silica column chromatography (10:1 toluene/EtOAc) to give **5f** (702 mg, 79% over 2 steps) as a colourless solid: $[\alpha]_D^{20}$ + 224.8 (*c* 1.0, CH_2Cl_2), lit.[25] + 259 (*c* 1.7, CH_2Cl_2); R_f 0.57 (10:1 toluene/EtOAc); ^{1}H NMR (400 MHz CDCl3): δ 8.07-7.97 (m, 8H, *o*-Bz), 7.94-7.89 (m, 4H, *o*-Bz), 7.86-7.83 (m, 2H, *o*-Bz), 7.69-7.52 (m, 5H, *m*-Bz, *p*-Bz), 7.51-7.37 (m, 12H, *m*-Bz, *p*-Bz), 7.33-7.28 (m, 4H, *m*-Bz, *p*-Bz), 6.24 (dd, 1H, [3]$J_{2',3'}$ 10.1 Hz, [3]$J_{3',4'}$ 9.9 Hz, H-3'), 6.14 (dd, 1H, [3]$J_{2,3}$ 10.1 Hz, [3]$J_{3,4}$ 10.0 Hz, H-3), 5.70-5.63 (m, 3H, H-1, H-1', H-4'), 5.47 (dd, 1H, [3]$J_{2',3'}$ 10.1 Hz, [3]$J_{1',2'}$ 3.8 Hz, H-2'), 5.38 (dd, 1H, [3]$J_{2,3}$ 10.1 Hz, [3]$J_{1,2}$ 3.8 Hz, H-2), 4.30 (ddd, [3]$J_{4',5'}$ 10.2 Hz, [3]$J_{5',6a'}$ 4.4 Hz, [3]$J_{5',6b'}$ 2.8 Hz, H-5'), 4.04-3.91 (m, 4H, H-5, H-6a, H-6b, H-6b'), 3.87-3.78 (two overlapping dd, 2H, [2]$J_{6a',6b'}$ 12.4 Hz, [3]$J_{4,5}$ 10.2 Hz, [3]$J_{3,4}$ 10.0 Hz, [3]$J_{5',6a'}$ 4.4 Hz, H-4, H-6a').^{13}C NMR (100 MHz, CDCl3): δ 165.8, 165.8, 165.5, 165.4, 165.4, 164.9 (7C, C=O), 134.1, 133.9, 133.6, 133.5, 133.4, 133.2, 133.1 (7C, *p*-Bz), 129.9, 129.9, 129.8, 129.5, 129.1, 129.0, 128.9, 128.7, 128.6, 128.5, 128.4, 128.4, 128.2 (35C, *i*-Bz, *m*-Bz, *o*-Bz), 93.1 (C-1'), 93.0 (C-1), 71.1 (2C, C-2, C-2'), 70.9 (C-3), 70.0 (C-3'), 68.9 (C-5), 68.7 (C-4'), 68.6

(C-5'), 62.3 (C-6), 61.8 (C-6'), 60.5 (C-4) NMR spectra were in agreement with published values;[25] HRMS, ESI-TOF/MS positive (m/z): calcd. for $C_{61}H_{50}O_{18}$ $[M+H]^+$: 1096.3134, found: 1096.3134.

5.7.1.8 4-azido-4-deoxy- α-D-glucopyranosyl-(1→1')-α-D-glucopyranoside (5g)
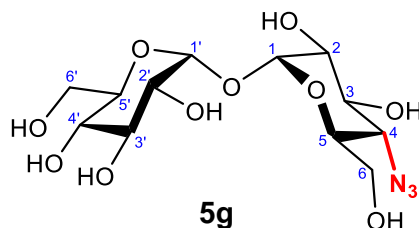
Compound **5f** (350 mg, 0.319 mmol) was deprotected as above to give pure **5g** in quantitative yield: R_f 0.74 (1:1 MeOH/H$_2$O); ^{1}H NMR (400 MHz, D$_2$O): δ 5.14 (d, 1H, $^3J_{1,2}$ 3.8 Hz, H-1), 5.08 (d, 1H, $^3J_{1',2'}$ 3.8 Hz, H-1'), 3.92 (dd, 1H, $^3J_{2,3}$ 9.7 Hz, $^3J_{3,4}$ 9.7 Hz, H-3), 3.80-3.66 (m, 7H, H-3', H-5, H-5', H-6a, H-6'a, H-6b, H-6'b), 3.64 (dd, 1H, $^3J_{2,3}$ 9.9 Hz, $^3J_{1,2}$ 3.8 Hz, H-2), 3.55 (dd, $^3J_{2',3'}$ 10.0 Hz, $^3J_{1',2'}$ 3.8 Hz, H-2'), 3.41 (dd, 1H, $^3J_{3,4}$ 10.1 Hz, $^3J_{4,5}$ 10.1 Hz H-4), 3.36 (dd, 1H, $^2J_{3',4'}$ 9.4 Hz, $^3J_{4',5'}$ 9.4 Hz, H-4').^{13}C NMR (100 MHz, D$_2$O): δ 93.4 (C-1), 93.4 (C-1'), 72.5 (C-5), 72.2 (C-5'), 71.7 (C-3), 71.0 (C-2), 70.9 (C-2'), 70.7 (C-3'), 69.6 (C-4'), 61.9 (C-4), 60.6 (C-6), 60.5 (C-6'); NMR spectra were in agreement with published values;[14] HRMS, ESI-TOF/MS positive (m/z): calcd. for $C_{12}H_{21}N_3O_{10}$ $[M+Na]^+$: 390.1119, found: 390.1117.

### 5.7.1.9 4-deoxy-4-fluoro-2,3,6-tri-*O*-benzoyl-α-D-glucopyranosyl-(1→1')-2',3',4',6'-tetra-*O*-benzoyl-α-D-glucopyranoside (5h)
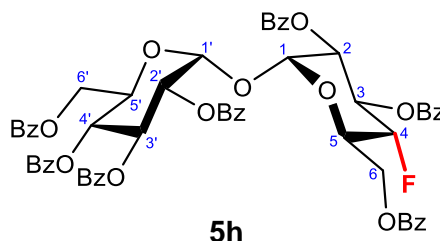
Compound **5b** (769 mg, 0.718 mmol) and DMAP (184 mg, 1.51 mmol) were dissolved in dry CH$_2$Cl$_2$ (15 mL) and cooled on ice. DAST (190 µL, 1.44 mmol) was added drop-wise and the solution was stirred for 18 h at room temperature. After this period, TLC (10:1 toluene/EtOAc) showed conversion to product. The reaction was quenched with H$_2$O (20 mL) and extracted twice with CH$_2$Cl$_2$ (2 × 20 mL). The combined organic phases were

washed with H_2O (40 mL) and concentrated. The crude residue was purified by column chromatography on silica gel (10:1 toluene/EtOAc) to give **5h** (285 mg, 37%) as a colourless solid: $[\alpha]_D^{20}$ + 246.0 (*c* 1.0, CH_2Cl_2); R_f 0.61 (10:1 Toluene/EtOAc); ^{1}H NMR (400 MHz CDCl$_3$): δ 8.10-8.02 (m, 6H, *o*-Bz), 8.01-7.90 (m, 6 H, *o*-Bz), 7.86-7.82 (m, 2H, *o*-Bz), 7.58-7.27 (m, 21H, *m*-Bz, *p*-Bz), 6.39-6.28 (m, 2H, H-3, H-3'), 5.76-5.69 (m, 3H, H-1, H-1', H-4'), 5.56 (dd, 1H, $^3J_{2,3}$ 10.3 Hz, $^3J_{1,2}$ 3.8 Hz, H-2), 5.46 (dd, 1H, $^3J_{2',3'}$ 10.3 Hz, $^3J_{1',2'}$ 3.8 Hz, H-2'), 4.77 (ddd, 1H, $^2J_{4,F}$ 50.6 Hz, $^3J_{4,5}$ 9.5 Hz, $^3J_{3,4}$ 9.5 Hz, H-4), 4.38-4.28 (m, 2H, H-5, H-5'), 4.07-4.00 (m, 3H, H-6a, H-6b, H-6b'), 3.90 (dd, 1H, $^2J_{6a',6b'}$ 12.5 Hz, $^3J_{5',6a'}$ 4.6 Hz, H-6a'); ^{13}C NMR (100 MHz, CDCl$_3$): δ 165.9, 165.6, 165.5, 165.4, 165.4, 165.0 (7C, C=O), 134.1, 134.0, 133.6, 133.6, 133.4, 133.3, 133.2 (7C, *p*-Bz), 129.9, 129.8, 129.8, 129.5, 129.2, 129.1, 129.1, 128.9, 128.8, 128.8, 128.6, 128.5, 128.5, 128.4, 128.4, 128.3, 128.3 (35C, *i*-Bz, *m*-Bz, *o*-Bz), 93.0 (C-1'), 92.6 (C-1), 87.0 (d, 1C, $^1J_{C-4,F}$ 188.4 Hz, C-4), 71.1 (C-2'), 70.8 (d, 1C, $^3J_{C-2,F}$ 8.3 Hz, C-2), 70.5 (d, 1C, $^2J_{C-3,F}$ 20.3 Hz, C-3), 70.2 (C-3'), 68.8 (C-4'), 68.8 (C-5'), 68.0 (d, 1C, $^2J_{C-5,F}$ 23.4 Hz, C-5), 61.9 (C-6'), 61.7 (C-6); ^{19}F {^{1}H} NMR (376 MHz, CDCl$_3$): δ -197.5, $^2J_{4,F}$ 50.6 Hz, $^3J_{3,F}$ 14.0 Hz, $^3J_{5,F}$ 0 Hz. HRMS, ESI-TOF/MS positive (*m/z*): calcd. for $C_{61}H_{50}O_{18}$ [M+H]$^+$: 1073.3026, found: 1073.3036.

5.7.1.10   4-deoxy-4-fluoro-α-D-glucopyranosyl-(1→1')-α-D-glucopyranoside (5i)

5i
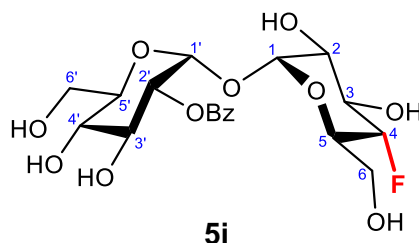
Compound **5h** (180 mg, 0.168 mmol) was deprotected as above to give pure **5i** in quantitative yield: R_f 0.77 (1:1 MeOH/H$_2$O); ^{1}H NMR (400 MHz, D$_2$O): δ 5.10-4.99 (m, 2H, H-1, H-1'), 4.36-3.94 (m, 2H, H-3, H-4), 3.92-3.83 (m, 1H, H-5), 3.74-3.57 (m, 6H, H-3', H-5', H-6a, H-6'a, H-6b, H-6'b), 3.54 (dd, 1H, $^3J_{2,3}$ 10.1 Hz, $^3J_{1,2}$ 3.6 Hz, H-2), 3.49 (dd, 1H, $^3J_{2',3'}$ 9.9 Hz, $^3J_{1',2'}$ 3.8 Hz, H-2'), 3.29 (dd, 1H, $^3J_{3',4'}$ 9.5 Hz, $^3J_{4',5'}$ 9.5 Hz, H-4'); ^{13}C NMR (100 MHz, D$_2$O): δ 93.5 (C-1), 93.2 (C-1'), 89.2 (d, 1C, $^1J_{C-4,F}$ 179.9 Hz, C-4), 72.5 (C-3'), 72.2 (C-5'), 71.0 (C-2'), 70.8 (d, 1C, $^2J_{C-3,F}$ 18.2 Hz, C-3), 70.5 (d, 1C, $^3J_{C-2,F}$ 8.2 Hz, C-2), 69.7 (C-4'), 69.6 (d, 1C, $^3J_{C-5,F}$ 24.4 Hz, C-5), 60.5 (C-6), 59.8 (C-6'); ^{19}F {^{1}H} NMR (376 MHz, D$_2$O): δ -198.2, $^2J_{4,F}$ 50.9 Hz, $^3J_{3,F}$ 15.6 Hz, $^3J_{5,F}$ 0 Hz. NMR spectra were in agreement with published values;[20] HRMS, ESI-TOF/MS positive (*m/z*): calcd. for $C_{12}H_{21}FO_{10}$ [M+Na]$^+$: 367.1011, found: 367.1006.

5.7.2 Uptake of 4-deoxy trehalose analogues by *M. smegmatis*

5.7.2.1 *M. smegmatis* strains and growth conditions

M. smegmatis mc^2 155 wild type strain was used as previously and *M. smegmatis ΔsugC* (Hyg^r) mutant strain was generously provided by R. Kalscheuer.[22] Strains were grown aerobically at 37 °C in Middlebrook 7H9 liquid media or on 7H10 as described in Chapter 4, with the addition of hygromycin (50 mg l^{-1}) for appropriate antibiotic selection. Growth curves and MIC determination experiments were carried out in 96-well plates as described in Chapter 4.

5.7.2.2 Extraction and analysis of small-molecule metabolites

M. smegmatis wildtype or *ΔsugC* mutant were cultured in the presence or absence of 4-deoxytrehalose analogues (250 µM) in 2.0 mL 7H9 liquid medium for 18 h, at which point culture densities measured by OD$_{600}$ were all between 0.9 and 1.2. Cells were centrifuged (1500 *g*, 3 min) and washed (PBSB, 1.0 mL) three times, then resuspended in 2.0 mL 70% EtOH and lysed by mechanical disruption using 0.1 mm zirconia/silica beads on a FastPrep bead-beating instrument. Supernatants were cleared by centrifugation (16,000 *g*, Supernatants (16,000 *g*, 10 10 min, 4°C), dried in a speed-vac, and resuspended in 500 µL 18 MΩ H2O. Small-molecule metabolites were collected via filtration through a spin concentrator with a 10 kDa molecular weight cut-off (Amicon).

HPAEC-PAD was performed on a Dionex ICS3000 system with a CarboPac MA-1 analytical column (4 mm × 250 mm) and MA-1 guard column (4 mm × 50 mm) kept at 30 °C. Pulsed amperometry with standard quadrupole waveform was used for detection. The system was equipped with an autosampler kept at 10 °C that was set up to inject 10 µL sample volumes. Isocratic elution was performed using H2O (76%) and 2 M NaOH (24%) for a final eluent concentration of 480 mM NaOH at a flow rate of 0.4 mL/min. Standards of trehalose and each analogue were prepared at 10 µg/mL, Chromeleon 7 software (Dionex) was used for data processing.

5.7.3 *M. tuberculosis* TreS-catalysed reactions

Recombinant *M. tuberculosis* TreS was produced by F. Miah, as described previously.[30] Enzymatic reactions were prepared with 2 µM recombinant TreS, 10 mM trehalose/4-deoxytrehalose analogue in 90 mM citrate buffer, pH 6.7, in a total volume of 600 µL with 10% D$_2$O. Reactions were monitored for up to 96 h by ^{1}H-NMR spectra with suppression of

the water signal at 4.70 ppm. Signals at 5.32 ppm (d, α/β-maltose H-1), 5.24 ppm (d, α-maltose + α-glucose H-1'), 5.10 ppm (d, trehalose H-1, H-1'), 4.47 ppm (d, β-maltose H-1') and 4.56 ppm (d, β-glucose H-1') were used to follow the consumption/formation of the natural substrates.

5.8   References

1. Kalscheuer, R.; Syson, K.; Veeraraghavan, U.; Weinrick, B.; Biermann, K. E.; Liu, Z.; Sacchettini, J. C.; Besra, G.; Bornemann, S.; Jacobs, W. R., Self-poisoning of *Mycobacterium tuberculosis* by targeting GlgE in an α-glucan pathway. *Nat. Chem. Biol.* **2010,** *6* (5), 376-384.

2. Kalscheuer, R.; Jacobs, W. R., Jr., The significance of GlgE as a new target for tuberculosis. *Drug News Perspect.* **2010,** *23* (10), 619-24.

3. Lindenberger, J. J.; Kumar Veleti, S.; Wilson, B. N.; Sucheck, S. J.; Ronning, D. R., Crystal structures of *Mycobacterium tuberculosis* GlgE and complexes with non-covalent inhibitors. *Sci. Rep.* **2015,** *5*, 12830.

4. Thanna, S.; Lindenberger, J. J.; Gaitonde, V. V.; Ronning, D. R.; Sucheck, S. J., Synthesis of 2-deoxy-2,2-difluoro-α-maltosyl fluoride and its X-ray structure in complex with *Streptomyces coelicolor* GlgEI-V279S. *Org. Biomol. Chem.* **2015,** *13* (27), 7542-7550.

5. Veleti, S. K.; Lindenberger, J. J.; Ronning, D. R.; Sucheck, S. J., Synthesis of a *C*-phosphonate mimic of maltose-1-phosphate and inhibition studies on *Mycobacterium tuberculosis* GlgE. *Biorg. Med. Chem.* **2014,** *22* (4), 1404-1411.

6. Veleti, S. K.; Lindenberger, J. J.; Thanna, S.; Ronning, D. R.; Sucheck, S. J., Synthesis of a poly-hydroxypyrolidine-based inhibitor of *Mycobacterium tuberculosis* GlgE. *J. Org. Chem.* **2014,** *79* (20), 9444-9450.

7. Veleti, S. K.; Petit, C.; Ronning, D. R.; Sucheck, S. J., Zwitterionic pyrrolidene-phosphonates: inhibitors of the glycoside hydrolase-like phosphorylase *Streptomyces coelicolor* GlgEI-V279S. *Org. Biomol. Chem.* **2017,** *15* (18), 3884-3891.

8. Syson, K.; Stevenson, C. E. M.; Rashid, A. M.; Saalbach, G.; Tang, M.; Tuukkanen, A.; Svergun, D. I.; Withers, S. G.; Lawson, D. M.; Bornemann, S., Structural insight into how *Streptomyces coelicolor* maltosyl transferase GlgE binds α-maltose 1-phosphate and forms a maltosyl-enzyme intermediate. *Biochemistry* **2014,** *53* (15), 2494-2504.

9. Titgemeyer, F.; Amon, J.; Parche, S.; Mahfoud, M.; Bail, J.; Schlicht, M.; Rehm, N.; Hillmann, D.; Stephan, J.; Walter, B.; Burkovski, A.; Niederweis, M., A genomic view of sugar transport in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *J. Bacteriol.* **2007,** *189* (16), 5903-5915.

10. Kalscheuer, R.; Weinrick, B.; Veeraraghavan, U.; Besra, G. S.; Jacobs, W. R., Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **2010,** *107* (50), 21761-21766.

11. Jiang, D.; Zhang, Q.; Zheng, Q.; Zhou, H.; Jin, J.; Zhou, W.; Bartlam, M.; Rao, Z., Structural analysis of *Mycobacterium tuberculosis* ATP-binding cassette transporter subunit UgpB reveals specificity for glycerophosphocholine. *FEBS J.* **2014,** *281* (1), 331-341.

12. Fullam, E.; Prokes, I.; Fütterer, K.; Besra, G. S., Structural and functional analysis of the solute-binding protein UspC from *Mycobacterium tuberculosis* that is specific for amino sugars. *Open Biol.* **2016,** *6* (6).

13. Backus, K. M.; Boshoff, H. L.; Barry, C. S.; Boutureira, O.; Patel, M. K.; D'Hooge, F.; Lee, S. S.; Via, L. E.; Tahlan, K.; Barry, C. E.; Davis, B. G., Uptake of unnatural trehalose

analogs as a reporter for *Mycobacterium tuberculosis. Nat. Chem. Biol.* **2011,** *7* (4), 228-235.

14. Swarts, B. M.; Holsclaw, C. M.; Jewett, J. C.; Alber, M.; Fox, D. M.; Siegrist, M. S.; Leary, J. A.; Kalscheuer, R.; Bertozzi, C. R., Probing the mycobacterial trehalome with bioorthogonal chemistry. *J. Am. Chem. Soc.* **2012,** *134* (39), 16123-16126.

15. Wolber, J. M.; Urbanek, B. L.; Meints, L. M.; Piligian, B. F.; Lopez-Casillas, I. C.; Zochowski, K. M.; Woodruff, P. J.; Swarts, B. M., The trehalose-specific transporter LpqY-SugABC is required for antimicrobial and anti-biofilm activity of trehalose analogues in *Mycobacterium smegmatis. Carbohydr. Res.* **2017**.

16. Miah, F.; Koliwer-Brandl, H.; Rejzek, M.; Field, R. A.; Kalscheuer, R.; Bornemann, S., Flux through trehalose synthase flows from trehalose to the α-anomer of maltose in mycobacteria. *Chem. Biol.* **2013,** *20* (4), 487-493.

17. Roy, R.; Usha, V.; Kermani, A.; Scott, D. J.; Hyde, E. I.; Besra, G. S.; Alderwick, L. J.; Futterer, K., Synthesis of α-glucan in mycobacteria involves a hetero-octameric complex of trehalose synthase TreS and maltokinase Pep2. *ACS Chem. Biol.* **2013,** *8* (10), 2245-2255.

18. Koliwer-Brandl, H.; Syson, K.; van de Weerd, R.; Chandra, G.; Appelmelk, B.; Alber, M.; Ioerger, T. R.; Jacobs, W. R., Jr.; Geurtsen, J.; Bornemann, S.; Kalscheuer, R., Metabolic network for the biosynthesis of intra- and extracellular α-glucans required for virulence of *Mycobacterium tuberculosis. PLoS Path.* **2016,** *12* (8), e1005768.

19. Zhang, R.; Pan, Y. T.; He, S. M.; Lam, M.; Brayer, G. D.; Elbein, A. D.; Withers, S. G., Mechanistic analysis of trehalose synthase from *Mycobacterium smegmatis. J. Biol. Chem.* **2011,** *286* (41), 35601-35609.

20. Rundell, S. R.; Wagar, Z. L.; Meints, L. M.; Olson, C. D.; O'Neill, M. K.; Piligian, B. F.; Poston, A. W.; Hood, R. J.; Woodruff, P. J.; Swarts, B. M., Deoxyfluoro-D-trehalose (FDTre) analogues as potential PET probes for imaging mycobacterial infection. *Org. Biomol. Chem.* **2016,** *14* (36), 8598-8609.

21. Pratt, M. R.; Leigh, C. D.; Bertozzi, C. R., Formation of 1,1-α,α-glycosidic bonds by intramolecular aglycone delivery. A convergent synthesis of trehalose. *Org. Lett.* **2003,** *5* (18), 3185-3188.

22. Urbanek, B. L.; Wing, D. C.; Haislop, K. S.; Hamel, C. J.; Kalscheuer, R.; Woodruff, P. J.; Swarts, B. M., Chemoenzymatic synthesis of trehalose analogues: rapid access to chemical probes for investigating mycobacteria. *ChemBioChem* **2014,** *15* (14), 2066-2070.

23. Meints, L. M.; Poston, A. W.; Piligian, B. F.; Olson, C. D.; Badger, K. S.; Woodruff, P. J.; Swarts, B. M., Rapid one-step enzymatic synthesis and all-aqueous purification of trehalose analogues. *J. Vis. Exp.* **2017,** (120), e54485.

24. Garcia, R. C.; Hough, L.; Richardson, A. C., Syntheses of hepta-pivalates, hexa-pivalates, and pentapivalates of trehalose by selective pivaloylation *Carbohydr. Res.* **1990,** *200*, 307-317.

25. Bassily, R. W.; Elsokkary, R. I.; Silwanis, B. A.; Nematalla, A. S.; Nashed, M. A., An improved synthesis of 4-azido-4-deoxy-α,α-trehalose and 4-amino-4-deoxy-α,α-trehalose and their epimers. *Carbohydr. Res.* **1993,** *239*, 197-207.

26. Dong, H.; Pei, Z. C.; Ramstrom, O., Stereospecific ester activation in nitrite-mediated carbohydrate epimerization. *J. Org. Chem.* **2006,** *71* (8), 3306-3309.

27. Rohrer, J., Analysis of carbohydrates by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAE-PAD). *Thermofisher Technical Note No. 20* **2013**.

28. Iturriaga, G.; Suarez, R.; Nova-Franco, B., Trehalose metabolism: from osmoprotection to signaling. *Int. J. Mol. Sci.* **2009,** *10* (9), 3793-3810.

29. Caner, S.; Nguyen, N.; Aguda, A.; Zhang, R.; Pan, Y. T.; Withers, S. G.; Brayer, G. D., The structure of the *Mycobacterium smegmatis* trehalose synthase reveals an unusual active site configuration and acarbose-binding mode. *Glycobiology* **2013,** *23* (9), 1075-1083.

30. Miah, F.; Bibb, M. J.; Barclay, J. E.; Findlay, K. C.; Bornemann, S., Developmental delay in a *Streptomyces venezuelae glgE* null mutant is associated with the accumulation of α-maltose 1-phosphate. *Microbiology* **2016,** *162* (7), 1208-1219.

6.   Discussion and Future Work

The focus of this thesis was the development of potential anti-tubercular therapeutics by targeting *M. tuberculosis* α-glucan biosynthesis. The overall aims were to:

1) Develop the fundamental understanding of mycobacterial α-glucan synthesis with structural studies of GlgB, and investigate α-glucan synthesis by GlgE and GlgB *in vitro*.

2) Identify a small-molecule inhibitor of GlgB using high-throughput screening, and investigate SAR with a selection of modified compounds, along with *in vivo* activity.

3) Investigate trehalose analogues as a route to GlgE inhibition *in vivo*.

6.1   Mycobacterial α-glucan synthesis

Compared to other biologically important carbohydrate molecules, α-glucans are relatively simple, consisting of just one type of monosaccharide subunit and two types of glycosidic linkage.[1] Furthermore, α-glucans are generated from an activated sugar substrate by the action of just two enzymes in bacteria.[2] However, the details of how they control the α-glucan polysaccharide architecture were not fully understood. I showed that GlgE and GlgB were sufficient to synthesise actinomycete α-glucan *in vitro*, with similar branch lengths to α-glucan extracted from actinomycete cells. This implies that inherent properties of GlgE/GlgB are responsible for the distinct architecture, ruling out other factors, such as biophysical constraints in the cellular environment or the involvement of other enzymes that edit the structure. This contrasts with the synthesis of starch, which involves multiple isoforms of starch synthases, branching enzymes and isoamylases, each with distinct effects on glucan formation.[3]

Not only were the average branch lengths shorter than those of mammalian and *E. coli* glycogen, but there was also a narrower distribution of DPs. Additionally, the A:BC chain ratio was lower, with a corresponding lower number of branches per B/C chain.[4] This suggests that α-glucan derived from the GlgE pathway has an overall less diverse structure than that from classical glycogen synthesis. It is not yet known whether these differences have any biological consequence. It is possible to speculate that in *M. tuberculosis* the distinct architecture could play a role in α-glucan export, and recognition by human immune receptors. However, a previous study showed that the recognition of α-glucan by DC-SIGN was a general feature mediated by internal glucosyl residues, and similar binding responses for *M. tuberculois* α-glucan and bovine glycogen were observed.[5] It remains to be seen whether the structural differences between α-glucans have a bearing on interactions with other immune receptors. Ongoing studies by J. Codee and co-workers aim to chemically

synthesise a library of α-glucan fragments,[6] which could be used to assess the effect of branch length on the binding affinity to DC-SIGN and other immune receptors.

It was predicted that the structural features of mycobacterial α-glucan are controlled by the activity of the branching enzyme GlgB. A low quality crystal structure of MtGlgB was previously published, however no ligand-bound structures existed for this enzyme.[7] I solved the structure of MsGlgB and confirmed that the three-dimensional structures of the mycobacterial GlgB homologues were very similar, validating the latter as model for the former. I identified ten oligosaccharide binding sites in MsGlgB, including the donor site and a likely acceptor site. The residues that interacted with the oligosaccharide in the donor site were mainly conserved with those of the cyanobacterial branching enzyme, the only other branching enzyme structure with an oligosaccharide-bound at the active site.[8] Considering the relatively distant evolutionary relationship of these two proteins, according to a maximum likelihood tree of GH13_9 branching enzymes previously published by Suzuki *et al.*,[9] the donor site may be conserved across a number of other species. A more detailed comparison of branching enzyme sequences across actinobacteria, cyanobacteria and proteobacteria may indicate whether this is indeed the case.

Considering the details of the branching mechanism, the contribution of the oligosaccharide-bound at the donor site to the final branch lengths depends on the action of the polymerase. Studies comparing the action of mycobacterial GlgE on different α-glucan substrates showed that it prefers to extend A chains,[4] which was consistent with the architecture of the enzyme's acceptor site.[10] Therefore, the length of newly created A chains will normally be transitory and bear little influence on the final branch lengths. This means that the length of the sugar accommodated in the branching enzyme donor site does not represent the final branch lengths. The low A:BC chain ratio of mycobacterial α-glucan demonstrates that most of the chains in the molecule are B chains. Furthermore, studies with labelled oligosaccharides demonstrated that mycobacterial GlgB transfer was strictly intrachain,[4] indicating that B chains are mainly the residual portion of cleaved chains. Overall, this suggests that most branch lengths are controlled by the GlgB acceptor site, **Figure 6.1**.
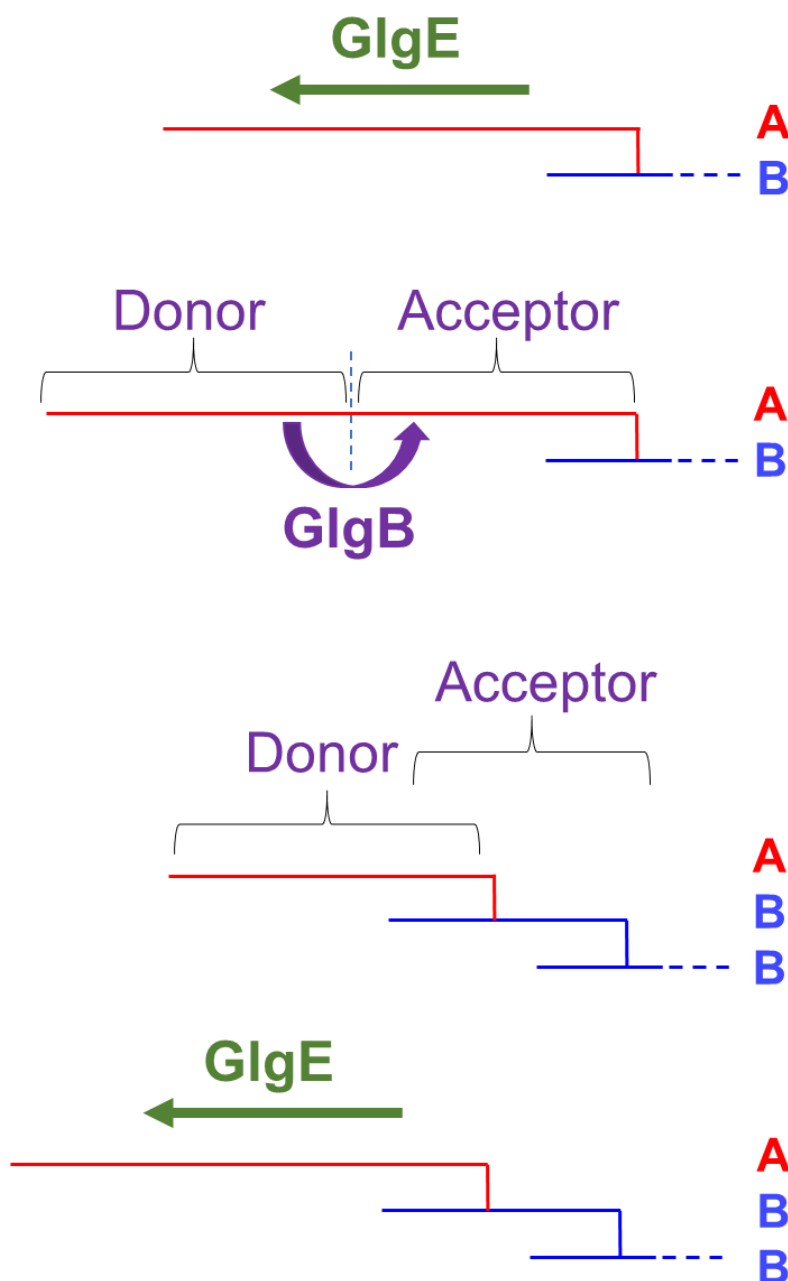
Figure 6.1 Generation of mycobacterial α-glucan by GlgE and GlgB. GlgE prefers to extend A chains, which are then branched by GlgB via the cleavage and transfer of the non-reducing end on to a 6-OH in the same chain. This creates a new A chain which can then be extended again by GlgE. The length of the chain in the GlgB donor site is therefore transitory, whereas the chain in the acceptor site is unlikely to change. Since most of the chains are B chains the distance between the cleavage point of the GlgB enzyme and an existing branch is likely to control the overall branch lengths.

I identified a distal acceptor site for GlgB, from which the complete acceptor site could be predicted. The relative orientation of oligosaccharides at this site was suggestive of a distal branchpoint, indicating that the enzyme can coordinate a branched structure, which is likely, given that the enzyme would typically act on a branched substrate. How the protein controls the relative locations of an existing branchpoint and a newly forming one is integral to the resultant branch lengths. However, without oligosaccharides bound in the proximal acceptor subsites it is not possible to draw further conclusions at this stage.

Considering further the intrachain nature of transfer in the branching mechanism, this necessitates a significant shift in the position of the acceptor oligosaccharide of 2-3 pyranose residues, given average branch lengths of DP 7-8 and average internal chain lengths of DP ~5. This could be facilitated by conformational changes in the protein, or the binding of the acceptor oligosaccharide in a conformation such that the 4-OH of the cleaved chain and the 6-OH that forms the new branch are close in space. Therefore, there may be alternative protein conformations that have not yet been detected by the crystallographic studies presented here. If the acceptor oligosaccharide adopts different conformations it may be challenging to determine these using crystallography, as the electron density for any one conformation may be weak. A branched oligosaccharide may help reveal acceptor subsites, but this work demonstrates that high concentrations of oligosaccharides of a defined structure are required to obtain strong electron density for bound ligands. The aforementioned synthetic fragments of α-glucans[6] could therefore also be useful in this regard.

Future studies could employ site-directed mutagenesis to determine which residues affect GlgB branch length specificity. Since I showed that α-glucan can readily be generated *in vitro*, this provides a system to quickly test the effect of mutagenesis on branch lengths in the context of the whole polymer. These results could be relevant to engineering the specificity of starch branching enzymes, which is a key objective in the field of starch research, as starches with specific branch lengths are desirable for different industrial applications.[11] However, starch synthesis is a more complex process that involves the coordination of multiple starch synthases, branching enzymes and debranching enzymes[12] and hence, it may not be possible to directly translate the findings across kingdoms.

6.2   Targeting *M. tuberculosi*s GlgB and GlgE

M. tuberculosis GlgB and GlgE are genetically validated drug targets, which can be targeted to both block the synthesis of an α-glucan virulence factor and to cause a lethal accumulation of M1P.[13-14] This dual mechanism of action is distinct from all known classes

of TB drugs,[15] making them attractive targets for the development of novel therapeutics that may be effective against multi-drug resistant strains of the bacterium.

6.2.1 Targeting GlgB

Here, I identify a potent MtGlgB inhibitor that represents the first validated, small-molecule inhibitor of an α-1,6 branching enzyme. Inhibitors of MtGlgB had previously been proposed from molecular docking experiments,[16] however the results presented here suggest that these compounds do not actually bind to the protein. This illustrates the importance of counter-screens and biophysical methods to validate hit compounds.[17] I synthesised a selection of analogues of **3a** to investigate its SAR and, whilst these studies indicated that one edge of the compound was important for inhibition, the relationship between structure and potency was unclear. More chemical space must be explored before SAR can emerge. A more diverse library of structural modifications may also help identify a compound with improved potency. The addition of groups at other positions of the quinazoline rings is one option that can be investigated.

Beyond this, generating selectivity for the *in vivo* target is the next priority for compound development. I showed that the original compound and some of the synthesised compounds inhibited the growth of *M. smegmatis* at concentrations of 125-500 μM, however this was not the result of GlgB inhibition. This suggests that the compound has another target *in vivo*. Similar sets of compounds have been shown to target folic acid biosynthesis,[18-19] which could be the actual target here.[20] This could be explored further by sequencing the genomes of spontaneously resistant strains and by transcriptomic analyses of cells that have been treated at sub-lethal concentrations. Medicinal chemistry efforts could then focus on modifications that would determine selectivity for the GlgB target over the other, for example because GlgB may be able to accommodate a bulky side group that other putative targets cannot. It is also important to test the compounds *in M. tuberculosis* cells, as there could be some differences in uptake and potency across the species.

For future therapeutic application of MtGlgB inhibitors, one important consideration is selectivity, due to the existence of a human GlgB homologue.[21] Whilst the *M. tuberculosis* and human GlgB enzymes only share 27% sequence identity, some catalytic residues are conserved so selectivity would have to be tuned to the mycobacterial enzyme. Historically, the existence of a human homologue was often a criterion used to rule out a potential bacterial drug target.[22-24] However, there are now numerous examples of bacterial enzyme inhibitors that are highly selective for their target over a human homologue,[25-27] even in cases of high sequence identity.[28] On the other hand, a non-selective small-molecule branching enzyme inhibitor could also prove to be a useful tool for the investigation and manipulation of glycogen or starch synthesis.

Both goals of improving potency and selectivity would be greatly aided by the identification of the binding site of the compound. Here, I attempted this by crystallographic studies, but did not detect any additional electron density that suggested binding. Future work could explore advance modelling simulations or STD-NMR studies to try to identify this by other means. Only with the nature of the protein-inhibitor interactions known, can inhibitors with improved efficacy be developed using a rational approach.

There are a number of options for identifying additional GlgB inhibitors, including screening more diverse libraries of compounds with the high-throughput assay developed in this work. However, the limited commercial availability of amylose-2800 could restrict the use of this assay in the future, necessitating the development of alternatives. One option is the use of synthetic substrates of a defined DP, which could be chemically modified to improve solubility. Hypothetically these could be installed with fluorophore groups that could give a differential read-out upon the creation of a branchpoint. Other studies have described the use of peptides to modify the activity of branching enzymes,[29] which may be a further route for exploration here.

6.2.2 Targeting GlgE

By industrial standards, GlgE is a more attractive drug target because there is no close human homologue.[22] However, efforts to identify small-molecule inhibitors from high-throughput screening have been unsuccessful to date. Inhibitors based on substrate and transition state mimetics have been developed, however these are moderate binders, with the best reported thus far having a K_i value of 45 µM.[30] Furthermore, they cannot exert an effect *in vivo*, as *M. tuberculosis* lacks the capacity to import maltose-based compounds.[31] Screening more diverse chemical libraries may yet yield inhibitors that are more drug-like and can therefore cross the mycobacterial membranes without the need for specific importers. Alternatively, it may be possible to rationally design an inhibitor that mimics features of M1P required for binding to GlgE, but avoids a carbohydrate scaffold.

Here I investigated trehalose analogues as precursors of GlgE inhibitors, using 4-deoxy trehalose analogues. Whilst analogues were readily synthesised and incorporated into mycobacterial cells, they were not processed by *M. tuberculosis* TreS and so were not suitable for the development of GlgE inhibitors. Future work will explore whether trehalose analogues with modifications at the 2- and 6- positions are isomerised to the corresponding maltose analogues by *M. tuberculosis* TreS. Trehalose metabolism in mycobacteria has already attracted attention as a target for TB therapeutics due to the lack of trehalose in mammalian systems and its integral role as a component of the mycomembrane, in the form of trehalose mycolates.[32] Previous studies have shown that trehalose analogues can be used for labelling cells via the attachment of fluorophores[33-34] or the incorporation of [18]F.[35]

Very recently, trehalose analogues with modifications at the 2,5 and 6 positions were shown to disrupt biofilm formation in *M. smegmatis*, which is linked to mycobacterial drug tolerance.[36] However, whilst some analogues show promise as diagnostic tools, overall trehalose analogues were weak inhibitors of mycobacterial growth, with MIC values of ≥ 100 µg/mL.[33, 36] Furthermore, these studies support the hypothesis that extracellular trehalose is preferentially incorporated into trehalose mycolates, rather than other trehalose metabolic pathways. A trehalose analogue that will be converted to a GlgE inhibitor *in vivo* is therefore a challenging prospect, but given the paucity of GlgE inhibitors, it is an approach that merits further investigation.

6.3   Future work on the GlgE pathway

Inhibition of GlgE or GlgB in *M. tuberculosis* is lethal via the build-up of M1P, however the mechanism of M1P toxicity remains unknown. Similarly, the intracellular accumulation of galactose-1-phosphate is toxic in both yeast and humans, but the mechanism of toxicity has also not been determined.[37-38] The pleiotropic stress responses observed upon M1P accumulation in *M. tuberculosis* include inhibition of respiration, induction of the stringent response and DNA damage, with the latter most likely to be the main cause of cell death.[13] These phenomena could be a direct result of the increased levels of M1P, because M1P is an activated sugar that can react with other nucleophiles in the cell and disrupt metabolic processes. Alternatively, the observed effects could be an indirect consequence of M1P accumulation, for example by causing an imbalance in phosphate metabolism. However, a recent study demonstrated that phosphate starvation in *M. tuberculosis* upregulates the production of α-glucan,[39] suggesting that phosphate limitation is not the underlying mechanism.

In *M. smegmatis*, the effect of removing either the *glgE* or *glgB* gene is less pronounced and is only detrimental to growth in the presence of exogenous trehalose and the resultant M1P accumulation is bacteriostatic, rather than bactericidal. In a *S. venezualae ΔglgE* strain, M1P accumulated only when it was grown on a maltose carbon source. This was also non-lethal, but was associated with a developmental delay phenotype, longer spores and reduced tolerance of these spores to stress.[40] Ongoing studies with *Pseudomonas aeruginosa* and *Pseudomonas syringae ΔglgE* strains show that M1P accumulation occurs when they are grown on glucose media. In *P. syringae*, this is associated with a reduced tolerance to osmotic stress (S. Woodcock, personal communication with permission).

Further work is required to determine how M1P accumulation results in these phenotypes and whether the underlying mechanisms are the same or vary between species. The microarray analyses used to characterise the stress responses of *M. tuberculosis* could be

applied to the other *ΔglgE* strains discussed above, to allow for comparison of upregulated and down regulated genes between species. If it could be established why M1P accumulation elicits such wide-ranging deleterious effects in *M. tuberculosis* but not in other species, it opens the possibility of targeting this mechanism by means other than GlgE or GlgB small-molecule inhibitors.

Another question that remains unanswered is how an α-glucan molecule 100s of kDa in size is exported through the multiple layers of mycobacterial cell envelope to the capsule. Other mycobacterial virulence factors are secreted via membrane vesicles; however, this was shown not to be the case for α-glucan.[41] In *E. coli*, capsular polysaccharides are exported by multiprotein complexes spanning the cell envelope,[42] but equivalent homologues have not been found in mycobacteria. Improved annotation of the *M. tuberculosis* genome may aid the discovery of the α-glucan export system.

6.4   Conclusions

This thesis represents a significant step forward in both the fundamental understanding of mycobacterial α-glucan synthesis and the targeting of GlgE and GlgB to develop anti-tubercular therapeutics. I provided direct evidence that GlgE and GlgB are necessary and sufficient to generate α-glucan and established that the distinct branch lengths seen *in vivo* are matched by those *in vitro*. I validated MsGlgB as a structural model of MtGlgB, which enabled the generation of high-resolution and substrate-bound structures. I identified ten oligosaccharide binding sites in MsGlgB, including the donor site and a likely acceptor site, giving unprecedented insight into substrate binding in α-1,6 branching enzymes.

I then went on to investigate GlgB and GlgE as drug targets. Firstly, I developed a high-throughput screening assay for GlgB and used this to identify an inhibitor, which binds in a 1:1 molar ration with a K_d of 3.1 μM. I investigated SAR with a selection of modified compounds, which demonstrated that the aromatic face of the compound was most important for inhibition. I developed an assay to determine target selectivity in *M. smegmatis* and established that whilst some of the compounds inhibit *M. smegmatis* growth, this is not via GlgB inhibition, suggesting they have other targets in the cell. I went on to explore an alternative approach to generating GlgE inhibitors, via the *in vivo* conversion of trehalose analogues into GlgE substrate mimics. 4-Deoxy trehalose analogues were synthesised and taken up into the mycobacterial cytosol by the LpqY-SugABC transporter, but this trehalose modification was not tolerated in the GlgE pathway.

Typically, it takes a minimum of 10 years to go from an initial hit to a marketable drug, with many potential pitfalls along the way.[17] The work presented here therefore represents just a snapshot of the early stages of the drug discovery process. An in-depth understanding of

the pathways and enzymes targeted, along with tools such as X-ray crystallography, provide strong foundations for progress, and to deal with challenges down the line. With the alarming rise in MDR and XDR-TB,[43] it is of vital importance that novel targets are explored now to create the new TB therapeutics of the future.

6.5 References

1. Dumitriu, S., *Polysaccharides: Structural diversity and functional versatility, Second edition*. CRC Press: Boca Raton, 2004.

2. Preiss, J., Glycogen Biosynthesis. In *Encyclopedia of Microbiology (Third Edition)*, Schaechter, M., Ed. Academic Press: Oxford, 2009; pp 145-158.

3. Pfister, B.; Sánchez-Ferrer, A.; Diaz, A.; Lu, K.; Otto, C.; Holler, M.; Shaik, F. R.; Meier, F.; Mezzenga, R.; Zeeman, S. C., Recreating the synthesis of starch granules in yeast. *eLife* **2016,** *5*, e15552.

4. 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., Assembly of α-glucan by GlgE and GlgB in mycobacteria and streptomycetes. *Biochemistry* **2016,** *55* (23), 3270-3284.

5. Geurtsen, J.; Chedammi, S.; Mesters, J.; Cot, M.; Driessen, N. N.; Sambou, T.; Kakutani, R.; Ummels, R.; Maaskant, J.; Takata, H.; Baba, O.; Terashima, T.; Bovin, N.; Vandenbroucke-Grauls, C.; Nigou, J.; Puzo, G.; Lemassu, A.; Daffe, M.; Appelmelk, B. J., Identification of mycobacterial α-glucan as a novel ligand for DC-SIGN: Involvement of mycobacterial capsular polysaccharides in host immune modulation. *J. Immunol.* **2009,** *183* (8), 5221-5231.

6. Wang, L.; Overkleeft, H.; va der Marel, G.; Codee, J., Synthesis of *Mycobacterium tuberculosis* derived α-glucans. In *19th European Carbohydrate Symposium*, Barcelona, 2017.

7. Pal, K.; Kumar, S.; Sharma, S.; Garg, S. K.; Alam, M. S.; Xu, H. E.; Agrawal, P.; Swaminathan, K., Crystal structure of full-length *Mycobacterium tuberculosis* H37Rv glycogen branching enzyme: Insights of N-terminal β-sandwich in subtrate specificity and enzymatic activity *J. Biol. Chem.* **2010,** *285* (27), 20897-20903.

8. Hayashi, M.; Suzuki, R.; Colleoni, C.; Ball, S. G.; Fujita, N.; Suzuki, E., Bound substrate in the structure of cyanobacterial branching enzyme supports a new mechanistic model. *J. Biol. Chem.* **2017**.

9. Suzuki, E.; Suzuki, R., Distribution of glucan-branching enzymes among prokaryotes. *Cell. Mol. Life Sci.* **2016**, 1-18.

10. Syson, K.; Stevenson, C. E. M.; Miah, F.; Barclay, J. E.; Tang, M.; Gorelik, A.; Rashid, A. M.; Lawson, D. M.; Bornemann, S., Ligand-bound structures and site-directed mutagenesis identify the acceptor and secondary binding sites of *Streptomyces coelicolor* maltosyltransferase GlgE. *J. Biol. Chem.* **2016,** *291* (41), 21531-21540.

11. Kaur, B.; Ariffin, F.; Bhat, R.; Karim, A. A., Progress in starch modification in the last decade. *Food Hydrocolloids* **2012,** *26* (2), 398-404.

12. Zeeman, S. C.; Kossmann, J.; Smith, A. M., Starch: Its metabolism, evolution, and biotechnological modification in plants. *Annu. Rev. Plant Biol.* **2010,** *61* (1), 209-234.

13. Kalscheuer, R.; Syson, K.; Veeraraghavan, U.; Weinrick, B.; Biermann, K. E.; Liu, Z.; Sacchettini, J. C.; Besra, G.; Bornemann, S.; Jacobs, W. R., Jr., Self-poisoning of *Mycobacterium tuberculosis* by targeting GlgE in an α-glucan pathway. *Nat. Chem. Biol.* **2010,** *6* (5), 376-384.

14. Koliwer-Brandl, H.; Syson, K.; van de Weerd, R.; Chandra, G.; Appelmelk, B.; Alber, M.; Ioerger, T. R.; Jacobs, W. R., Jr.; Geurtsen, J.; Bornemann, S.; Kalscheuer, R., Metabolic network for the biosynthesis of intra- and extracellular α-glucans required for virulence of *Mycobacterium tuberculosis. PLoS Path.* **2016,** *12* (8), e1005768.

15. Zumla, A.; Nahid, P.; Cole, S. T., Advances in the development of new tuberculosis drugs and treatment regimens. *Nat. Rev. Drug Discov.* **2013,** *12* (5), 388-404.

16. Dkhar, H. K.; Gopalsamy, A.; Loharch, S.; Kaur, A.; Bhutani, I.; Saminathan, K.; Bhagyaraj, E.; Chandra, V.; Swaminathan, K.; Agrawal, P.; Parkesh, R.; Gupta, P., Discovery of *Mycobacterium tuberculosis* α-1,4-glucan branching enzyme (GlgB) inhibitors by structure- and ligand-based virtual screening. *J. Biol. Chem.* **2015,** *290* (1), 76-89.

17. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L., Principles of early drug discovery. *Br. J. Pharmacol.* **2011,** *162* (6), 1239-1249.

18. Van Horn, K. S.; Zhu, X.; Pandharkar, T.; Yang, S.; Vesely, B.; Vanaerschot, M.; Dujardin, J.-C.; Rijal, S.; Kyle, D. E.; Wang, M. Z.; Werbovetz, K. A.; Manetsch, R., Antileishmanial activity of a series of N^2,N^4-disubstituted Quinazoline-2,4-diamines. *J. Med. Chem.* **2014,** *57* (12), 5141-5156.

19. Zhu, X.; Van Horn, K. S.; Barber, M. M.; Yang, S.; Wang, M. Z.; Manetsch, R.; Werbovetz, K. A., SAR refinement of antileishmanial N^2,N^4-disubstituted quinazoline-2,4-diamines. *Biorg. Med. Chem.* **2015,** *23* (16), 5182-5189.

20. Minato, Y.; Thiede, J. M.; Kordus, S. L.; McKlveen, E. J.; Turman, B. J.; Baughn, A. D., *Mycobacterium tuberculosis* folate metabolism and the mechanistic basis for para-aminosalicylic acid susceptibility and resistance. *Antimicrob. Agents Chemother.* **2015,** *59* (9), 5097-106.

21. Thon, V. J.; Khalil, M.; Cannon, J. F., Isolation of human glycogen branching enzyme cDNAs by screening complementation in yeast. *J. Biol. Chem.* **1993,** *268* (10), 7509-7513.

22. Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L., Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov.* **2007,** *6* (1), 29-40.

23. Sakharkar, K. R.; Sakharkar, M. K.; Chow, V. T. K., Biocomputational strategies for microbial drug target identification. In *New Antibiotic Targets*, Champney, W. S., Ed. Humana Press: Totowa, NJ, 2008; pp 1-9.

24. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L., How many drug targets are there? *Nat. Rev. Drug Discov.* **2006,** *5* (12), 993-996.

25. Yonath, A., Antibiotics targeting ribosomes: Resistance, selectivity, synergism, and cellular regulation. *Annu. Rev. Biochem.* **2005,** *74* (1), 649-679.

26. Soares da Costa, T. P.; Tieu, W.; Yap, M. Y.; Pendini, N. R.; Polyak, S. W.; Sejer Pedersen, D.; Morona, R.; Turnidge, J. D.; Wallace, J. C.; Wilce, M. C. J.; Booker, G. W.; Abell, A. D., Selective inhibition of Biotin Protein Ligase from *Staphylococcus aureus. J. Biol. Chem.* **2012,** *287* (21), 17823-17832.

27. Arya, T.; Reddi, R.; Kishor, C.; Ganji, R. J.; Bhukya, S.; Gumpena, R.; McGowan, S.; Drag, M.; Addlagatta, A., Identification of the molecular basis of inhibitor selectivity between the human and streptococcal type I methionine aminopeptidases. *J. Med. Chem.* **2015,** *58* (5), 2350-2357.

28. Kasbekar, M.; Fischer, G.; Mott, B. T.; Yasgar, A.; Hyvönen, M.; Boshoff, H. I. M.; Abell, C.; Barry, C. E.; Thomas, C. J., Selective small molecule inhibitor of the *Mycobacterium tuberculosis* fumarate hydratase reveals an allosteric regulatory site. *Proc. Natl. Acad. Sci. USA* **2016,** *113* (27), 7503-7508.

29. Froese, D. S.; Michaeli, A.; McCorvie, T. J.; Krojer, T.; Sasi, M.; Melaev, E.; Goldblum, A.; Zatsepin, M.; Lossos, A.; Álvarez, R.; Escribá, P. V.; Minassian, B. A.; von Delft, F.; Kakhlon, O.; Yue, W. W., Structural basis of glycogen branching enzyme deficiency and pharmacologic rescue by rational peptide design. *Hum. Mol. Genet.* **2015,** *24* (20), 5667-5676.

30. Veleti, S. K.; Petit, C.; Ronning, D. R.; Sucheck, S. J., Zwitterionic pyrrolidene-phosphonates: inhibitors of the glycoside hydrolase-like phosphorylase *Streptomyces coelicolor* GlgEI-V279S. *Org. Biomol. Chem.* **2017,** *15* (18), 3884-3891.

31. Titgemeyer, F.; Amon, J.; Parche, S.; Mahfoud, M.; Bail, J.; Schlicht, M.; Rehm, N.; Hillmann, D.; Stephan, J.; Walter, B.; Burkovski, A.; Niederweis, M., A genomic view of sugar transport in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis. J. Bacteriol.* **2007,** *189* (16), 5903-5915.

32. Thanna, S.; Sucheck, S. J., Targeting the trehalose utilization pathways of *Mycobacterium tuberculosis. MedChemComm* **2016,** *7* (1), 69-85.

33. Backus, K. M.; Boshoff, H. L.; Barry, C. S.; Boutureira, O.; Patel, M. K.; D'Hooge, F.; Lee, S. S.; Via, L. E.; Tahlan, K.; Barry, C. E.; Davis, B. G., Uptake of unnatural trehalose analogs as a reporter for *Mycobacterium tuberculosis. Nat. Chem. Biol.* **2011,** *7* (4), 228-235.

34. Swarts, B. M.; Holsclaw, C. M.; Jewett, J. C.; Alber, M.; Fox, D. M.; Siegrist, M. S.; Leary, J. A.; Kalscheuer, R.; Bertozzi, C. R., Probing the mycobacterial trehalome with bioorthogonal chemistry. *J. Am. Chem. Soc.* **2012,** *134* (39), 16123-16126.

35. Rundell, S. R.; Wagar, Z. L.; Meints, L. M.; Olson, C. D.; O'Neill, M. K.; Piligian, B. F.; Poston, A. W.; Hood, R. J.; Woodruff, P. J.; Swarts, B. M., Deoxyfluoro-D-trehalose (FDTre) analogues as potential PET probes for imaging mycobacterial infection. *Org. Biomol. Chem.* **2016,** *14* (36), 8598-8609.

36. Wolber, J. M.; Urbanek, B. L.; Meints, L. M.; Piligian, B. F.; Lopez-Casillas, I. C.; Zochowski, K. M.; Woodruff, P. J.; Swarts, B. M., The trehalose-specific transporter LpqY-SugABC is required for antimicrobial and anti-biofilm activity of trehalose analogues in *Mycobacterium smegmatis. Carbohydr. Res.* **2017**.

37. Slepak, T.; Tang, M.; Addo, F.; Lai, K., Intracellular galactose-1-phosphate accumulation leads to environmental stress response in yeast model. *Mol. Genet. Metab.* **2005,** *86* (3), 360-371.

38. Isselbacher, K. J.; Anderson, E. p.; Kurahashi, K.; Kalckar, H. m., Congenital galactosemia, a single enzymatic block in galactose metabolism. *Science* **1956,** *123* (3198), 635-636.

39. van de Weerd, R.; Boot, M.; Maaskant, J.; Sparrius, M.; Verboom, T.; van Leeuwen, L. M.; Burggraaf, M. J.; Paauw, N. J.; Dainese, E.; Manganelli, R.; Bitter, W.; Appelmelk, B. J.; Geurtsen, J., Inorganic Phosphate Limitation Modulates Capsular Polysaccharide Composition in Mycobacteria. *J. Biol. Chem.* **2016,** *291* (22), 11787-99.

40. Miah, F.; Bibb, M. J.; Barclay, J. E.; Findlay, K. C.; Bornemann, S., Developmental delay in a *Streptomyces venezuelae glgE* null mutant is associated with the accumulation of α-maltose 1-phosphate. *Microbiology* **2016,** *162* (7), 1208-1219.

41. Prados-Rosales, R.; Baena, A.; Martinez, L. R.; Luque-Garcia, J.; Kalscheuer, R.; Veeraraghavan, U.; Camara, C.; Nosanchuk, J. D.; Besra, G. S.; Chen, B.; Jimenez, J.; Glatman-Freedman, A.; Jacobs, W. R., Jr.; Porcelli, S. A.; Casadevall, A., Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice. *J. Clin. Invest.* **2011,** *121* (4), 1471-83.

42. Whitfield, C., Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli. Annu. Rev. Biochem.* **2006,** *75* (1), 39-68.

43. Zumla, A.; George, A.; Sharma, V.; Herbert, N.; Masham, B., WHO's 2013 global report on tuberculosis: successes, threats, and opportunities. *Lancet* **2013,** *382* (9907), 1765-1767.

7. Appendices

7.1   Appendix 1: Publication

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., **Assembly of α-glucan by GlgE and GlgB in mycobacteria and streptomycetes**. *Biochemistry* **2016**, 55 (23), 3270-3284

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Assembly of α-Glucan by GlgE and GlgB in Mycobacteria and Streptomycetes

Abdul M. Rashid,[†] Sibyl F. D. Batey,[†] Karl Syson,[†] Hendrik Koliwer-Brandl,[#] Farzana Miah,[†] J. Elaine Barclay,[‡] Kim C. Findlay,[‡] Karol P. Nartowski,[§] Yaroslav Z. Khimyak,[§] Rainer Kalscheuer,[#] and Stephen Bornemann*,[†]

[†]Biological Chemistry Department, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom
[#]Institute for Medical Microbiology and Hospital Hygiene, and Institute for Pharmaceutical Biology and Biotechnology, Heinrich-Heine-University Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany
[‡]Cell and Developmental Biology Department, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom
[§]School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

Ⓢ Supporting Information

ABSTRACT: Actinomycetes, such as mycobacteria and streptomycetes, synthesize α-glucan with α-1,4 linkages and α-1,6 branching to help evade immune responses and to store carbon. α-Glucan is thought to resemble glycogen except for having shorter constituent linear chains. However, the fine structure of α-glucan and how it can be defined by the maltosyl transferase GlgE and branching enzyme GlgB were not known. Using a combination of enzymolysis and mass spectrometry, we compared the properties of α-glucan isolated from actinomycetes with polymer synthesized in vitro by GlgE and

GlgB. We now propose the following assembly mechanism. Polymer synthesis starts with GlgE and its donor substrate, α-maltose 1-phosphate, yielding a linear oligomer with a degree of polymerization (∼16) sufficient for GlgB to introduce a branch. Branching involves strictly intrachain transfer to generate a C chain (the only constituent chain to retain its reducing end), which now bears an A chain (a nonreducing end terminal branch that does not itself bear a branch). GlgE preferentially extends A chains allowing GlgB to act iteratively to generate new A chains emanating from B chains (nonterminal branches that themselves bear a branch). Although extension and branching occur primarily with A chains, the other chain types are sometimes extended and branched such that some B chains (and possibly C chains) bear more than one branch. This occurs less frequently in α-glucans than in classical glycogens. The very similar properties of cytosolic and capsular α-glucans from *Mycobacterium tuberculosis* imply GlgE and GlgB are sufficient to synthesize them both.
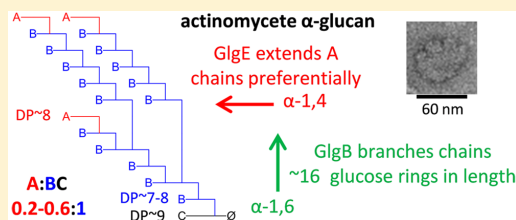
α-Glucans are widespread among bacteria, yeasts, plants, insects, and mammals, and in most cases they are thought to act as a store of carbon and energy. Arguably the best documented α-glucan in bacteria is glycogen, which typically accumulates under nitrogen-limiting growth conditions.[1] Glycogen comprises linear α-1,4-linked malto-oligosaccharides connected by α-1,6-linked branch points giving rise to an arboreal (tree-like) dendrimer and particles tens of nanometers in diameter. In the classical bacterial glycogen pathway, the α-1,4-links are synthesized by glycogen synthase GlgA (E.C. 2.4.1.21), using ADP-glucose as the sugar donor (Figure 1), and the α-1,6-linked branches are generated by GlgB branching enzyme (E.C. 2.4.1.18). About 32% of sequenced bacterial genomes contain the genes coding for the enzymes of the classical pathway.[2]

The alternative GlgE α-glucan pathway has been discovered recently,[3,4] which is about half as widespread among bacteria as the classical pathway.[2] This involves trehalose synthase and maltose kinase that convert trehalose into α-maltose 1-

phosphate (Figure 1), which is the donor for the maltosyl transferase GlgE (E.C. 2.4.99.16) that creates α-1,4-linkages.[5] Bioinformatic analysis suggests that even when organisms possess both the classical and GlgE pathways, they possess only one GlgB enzyme,[2] implying the branching enzyme is shared between both pathways. GlgB branching enzymes associated with these pathways are members of the glycoside hydrolase 13_9 CAZy family.[6,7] Interestingly, the gene for GlgB is often clustered with the gene for GlgE rather than for GlgA in such cases. Some organisms possess only the GlgE pathway, and we have established using genetics that GlgE and GlgB are necessary and sufficient for the biosynthesis of α-glucan in one such organism, the actinomycete *Streptomyces venezuelae*.[8]

The GlgE pathway is widespread among other actinomycetes such as mycobacteria.[2] Indeed, both GlgE and GlgB have been
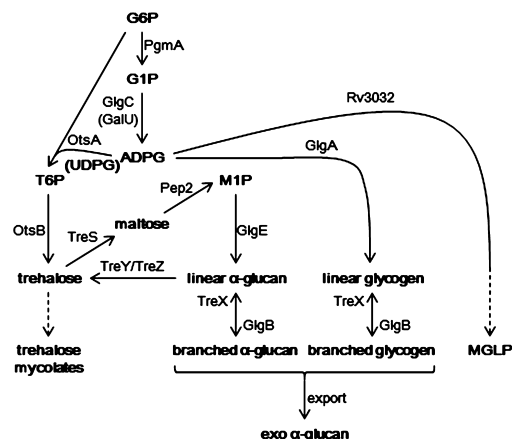
Figure 1. Metabolism of α-glucans thought to exist in *Mycobacterium tuberculosis*. G6P, glucose 6-phosphate; G1P, glucose 1-phosphate; M1P, α-maltose 1-phosphate; T6P, trehalose 6-phosphate; ADPG, ADP-glucose; UDPG, UDP-glucose; MGLP, methylglucose lipopolysaccharide; PgmA, phosphoglucomutase; GlgC, ADP-glucose pyrophosphorylase; GalU, UDP-glucose pyrophosphorylase; OtsA, trehalose 6-phosphate synthase; OtsB, trehalose 6-phosphate phosphatase; TreS, trehalose synthase; Pep2, maltose kinase; GlgE, maltosyl transferase; GlgB, branching enzyme; TreX, debranching enzyme; TreY, maltooligosyltrehalose synthase; TreZ, maltooligosyltrehalose trehalohydrolase; GlgA, glycogen synthase; Rv3032, putative glycosyltransferase.

genetically validated as potential drug targets for the treatment of tuberculosis.[3] Some species, including *Mycobacterium tuberculosis*, possess not only the genes for the GlgE and classical pathways, but also genes for a methyl glucose lipopolysaccharide pathway (Figure 1). This pathway seems to involve Rv3032 in the formation of α-1,4 linkages in the highly modified malto-oligosaccharide product of this pathway.[9] This specialized α-glucan is thought to chaperone fatty acid biosynthesis. Carbohydrate metabolism in mycobacteria is further complicated by the fact that the capsule coating the cells is mainly composed of a glycogen-like α-glucan.[10,11] This polymer might contribute to virulence of *M. tuberculosis* by binding to DC-SIGN to evade the host immune response.[12−15] The capsular and cytosolic α-glucan polymers appear to be very similar, with only subtle differences having been reported.[15,16] Quite how each metabolic pathway contributes to each α-glucan product is not yet clear.

Mycobacterial α-glucans appear to resemble classical glycogen except for a higher degree of branching and correspondingly shorter linear chains.[10,16,17] This implies that the branching enzyme in mycobacteria makes shorter branches than the corresponding enzymes in organisms that synthesize classical glycogen, such as in *Escherichia coli*.[18] To address this issue, we first characterized GlgE-derived α-glucan from *S. venezuelae*, and α-glucans from mycobacteria. We then determined the chain length specificity of actinomycete GlgB branching enzymes and the chain specificity of *M. tuberculosis* GlgE. This enabled us to establish how the GlgE pathway-derived polymer is assembled, which has implications for how the cytosolic and capsular α-glucans of *M. tuberculosis* are synthesized.

■ MATERIALS AND METHODS

Materials. Dextrans (5, 12, 25, 50, 100, 150, 270, 410, and 1400 kDa), amylose from potatoes (average mass >150 kDa), amylopectin from potato, individual malto-oligosaccharides (degree of polymerization, DP2-7), rabbit liver glycogen and pullulanases were purchased from Sigma-Aldrich (Gillingham, United Kingdom). Malto-octaose (DP8) was purchased from Carbosynth (Compton, United Kingdom). Amylose 2800 (mean DP 12.6 ± 5.3, median and mode DP11, DP6-16 contributing at least 5% of total malto-oligosaccharides according to capillary electrophoresis) was purchased from TCI Europe N.V. (Oxford, United Kingdom). α-Maltose 1-phosphate was synthesized as previously described.[19] [U-¹³C]α-maltose 1-phosphate was a gift from John (Jack) Thompson.

Isolation of α-Glucans. *Streptomyces venezuelae.* ATCC 10712 was grown on sterile cellophane discs on MYM Tap agar plates at 30 °C for 30 h. The cells were removed from the cellophane, suspended in water (15 mL), and centrifuged at 4000*g* for 30 min. The cells were washed three more times, suspended in water (10 mL), and disrupted by sonication. The cell debris was pelleted by centrifugation at 30000*g* for 15 min, and the resultant supernatant was washed twice with a 1:1 (v/v) mixture of 0.2 M glycine, pH 10.5, and chloroform (10 mL). The aqueous layer was concentrated to ∼8 mL using a rotary evaporator and then centrifuged at 108000*g* at 4 °C for 2 h. The gelatinous pellet was collected and dissolved in water (2 mL), and the α-glucan was precipitated with ethanol (3 vol) at 4 °C. The solid collected by centrifugation at 4000*g* for 10 min was redissolved in water (8 mL) and centrifuged at 12000*g* for 10 min to remove insoluble material. The ultracentrifugation and ethanol precipitation steps were repeated. Finally the pellet was dissolved in 2 mL of water and freeze-dried to yield the α-glucan as an amorphous white powder.

Mycobacterium smegmatis. Cells were grown on sterile cellophane discs for 72 h at 37 °C on Difco Middlebrook 7H10 agar solid medium supplemented with 10% ADS (5% w/v bovine serum albumin, 2% glucose, and 0.85% NaCl). Cells were resuspended in 10 mL of water and disrupted by sonication. α-Glucan was extracted as described above.

Escherichia coli. BL21(DE3) cells were transformed with a pUC19 vector conferring resistance to carbenicillin. Cells were grown at 37 °C for 22 h in nitrogen-limited M9 medium (1 L) comprising Na₂HPO₄·7H₂O (6 g), KH₂PO₄ (3 g), NaCl (0.5 g), pH 7.4, glucose (3 g), (NH₄)₂SO₄ (0.22 g), MgSO₄·7H₂O (0.25 g), CaCl₂·2H₂O (15 mg), and thiamine (10 mg), supplemented with carbenicillin (100 mg). Cell were harvested by centrifugation and resuspended in 10 mL of water. Glycogen was isolated as described for α-glucan above.

M. tuberculosis. H37Rv cells were grown for 2 weeks in 100 mL cultures in a bovine serum albumin-free minimal medium containing 20 mM glucose, 0.5% glycerol, and 0.05% tyloxapol.[20] Because of the loose attachment of the *M. tuberculosis* capsule layer to the cell surface, treatment with detergent and physical agitation were sufficient to quantitatively strip off the capsule from cells and thus allow the satisfactory separation of capsular and cytosolic α-glucan polymers.[21] The cells were harvested by centrifugation. The supernatant was sterilized by passing twice through a 0.22 μm sterile filter, heat inactivated at 65 °C for 12 h, and concentrated to 1.5 mL by centrifugation (4000*g*) through a 30 kDa cutoff membrane filtration device. The retentate was diluted with water (10 mL) and concentrated to 1.5 mL. The concentrate was treated with

absolute ethanol (5 vol), and the resulting precipitate was collected by centrifugation. The solid was dissolved in water (1 mL) and centrifuged, and the resulting supernatant was freeze-dried to yield capsular α-glucan as an amorphous solid. The cell pellet was washed twice with 20 mM sodium phosphate buffer, pH 7.4, containing 20 mM NaCl to remove all remaining capsular material, resuspended in water, and heated for 4 h at 95 °C. The cell extract was centrifuged to remove debris, and the supernatant was centrifuged at 108000g at 4 °C for 2 h. The gelatinous pellet was dissolved in water (0.5 mL), and the α-glucan was precipitated with ethanol (3 vol). The solid was collected by centrifugation, redissolved in water (0.5 mL), and freeze-dried to yield cytosolic α-glucan as an amorphous power.

NMR Spectroscopy. Solution-state NMR spectra were recorded at 298 K on a Bruker Avance III 400 MHz spectrometer and analyzed using Bruker TopSpin 2.1 (Rheinstetten, Germany). Chemical shifts in D_2O are reported with reference to residual water at δ_H 4.79 or to the methyl resonance of internal acetone at δ_C 29.84 ppm. Assignments were made with the aid of COSY and HSQC experiments. The assignment of the peaks was based on our previous studies.[3,22]

Pulse field gradient DOSY experiments[23] were carried out with up to 5 mg of polymer in 0.6 mL of D_2O at 313 K using TopSpin 2.1. The gradient strength was incremented longitudinally in 16 steps from 2% up to 95% of the maximum gradient strength. Diffusion times and gradient pulse durations were optimized for each experiment in order to achieve a 95% decrease in the resonance intensity at the largest gradient amplitude. Typically, diffusion time was set at 200 μs and the diffusion gradient length at 3.5 ms. The 2D diffusion spectra were analyzed with T1/T2 processing to derive the corresponding diffusion coefficient. The relationship between diffusion coefficient and molecular mass was determined using a series of dextrans (Figure S2).

Molecular radii were calculated from the Stokes−Einstein equation, which relates the diffusion coefficient D to the macromolecular hydrodynamic radius r. The assumption that the macromolecules are spherical is reasonable given the morphology of particles according to electron microscopy (Figure 2).

$$D = \frac{k_B T}{6\pi\eta r}$$

where $k_B = 1.38 \times 10^{-23}$ J K^{-1}, $T = 313$ K, and dynamic viscosity of water $\eta = 0.000653$ N s m^{-2}
(http://www.engineeringtoolbox.com/water-dynamic-kinematic-viscosity-d_596.html)

Solid-state NMR measurements were performed at 293 K using a Bruker 400 MHz Avance III spectrometer equipped with a triple resonance probe at frequencies 400.23 (^{1}H) and 100.64 MHz (^{13}C). Amylose, amylopectin, and rabbit liver glycogen were each packed in 4 mm zirconia rotors, and α-glucan samples were packed into 12 μL CRAMPS rotors 4 mm in diameter. Samples were rotated at a magic angle spinning rate of 10 kHz. ^{1}H−^{13}C cross-polarization magic angle spinning (CP/MAS) spectra were acquired using the following parameters: ^{1}H π/2 pulse length 3.5 μs, ^{13}C π/2 pulse length 3.5 μs, and ^{1}H−^{13}C CP contact time 2 ms. SPINAL64 decoupling was used during signal acquisition with recycle delay of 20 s. The Hartmann−Hahn conditions for ^{1}H−^{13}C CP/MAS NMR experiment were set with hexamethylbenzene. Typically, 256 scans were acquired for amylose, amylopectin,
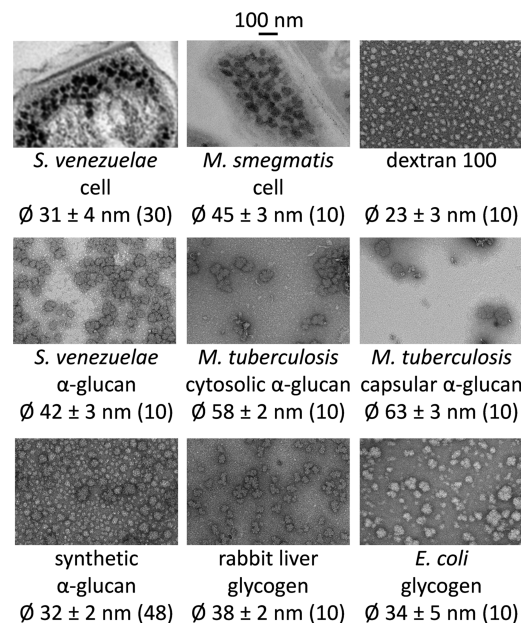
Figure 2. Transmission electron microscopy of α-glucans, glycogen, and dextran. Cells, isolated α-glucans, and dextran 100 were stained using periodic acid-thiocarbohydrazide-silver proteinate (PATAg), uranyl acetate, and phosphotungstic acid, respectively. Representative images are shown. The scale bar represents 100 nm. The diameters (Ø) of the particles (β particles in the case of α-glucans) are quoted as mean ± SE (n). α-Glucan isolated from *M. smegmatis* had a diameter of 35 ± 3 (21). Synthetic α-glucan was produced by treating maltohexaose with *M. tuberculosis* GlgE and GlgB together with α-maltose 1-phosphate. See main text and Table 1 for other measured diameters.

and rabbit liver glycogen and 2048 scans for α-glucan. The ^{13}C chemical shifts were recorded with respect to trimethylsilane.

Transmission Electron Microscopy. Aqueous polymer samples (typically 0.1 mg mL^{-1}) were allowed to settle briefly onto a 400 mesh copper grid with a carbon-coated pyroxylin support film before being negatively stained. Glycogen and α-glucan was stained with 2% aqueous uranyl acetate, pH 4.5, and dextran was stained with 1% phosphotungstic acid. The grids were viewed in a FEI Tecnai 20 transmission electron microscope (Eindhoven, Netherlands) at 200 kV, and digital TIFF images were taken using an AMT XR60B digital camera (Deben, Bury St Edmunds, UK).

S. venezuelae was grown for 5 days on MYM-TAP solid medium at 28 °C,[24] and *M. smegmatis* was grown on Difco Middlebrook 7H10 agar solid medium supplemented with 10% ADS (5% w/v bovine serum albumin, 2% glucose, and 0.85% NaCl) for 3 days at 37 °C. Single colonies of *S. venezuelae* and *M. smegmatis* were positively stained for α-glucan using periodic acid-thiocarbohydrazide-silver proteinate (PATAg) as previously described.[8]

Dynamic Light Scattering. Dynamic light scattering analysis was performed on a DynaPro Titan (Wyatt Technology Corporation, Santa Barbara CA, U.S.A.) equipped with DYNAMICS software (v6) using a 12 μL sample cell at

293 K. The sample concentration was typically 1 mg mL^{-1} in water.

Iodine Complexes of α-Glucans. Lugol solution (1 mg mL^{-1} I$_2$ and 0.1 mg mL^{-1} KI; typically 0.5 vol) was added to samples of isolated polysaccharides, with further dilution in water as appropriate, and absorbance was measured either using a PerkinElmer Lambda 25 spectrophotometer (Coventry, United Kingdom) or a BMG Labtech CLARIOstar microtiter plate spectrophotometer (Ortenberg, Germany).

Reduction and Debranching of α-Glucans. Saccharides were reduced by adding aqueous NaBH$_4$ (typically 1 mg in 5 μL) to aqueous samples (in 25 μL). The mixtures were incubated at 21 °C for 1 h, acidified with dilute acetic acid, and treated with absolute ethanol (5 vol). The resulting precipitated material was isolated by centrifugation and dried.

Polymer samples (typically in 25 μL) were debranched using *Pseudomonas* sp. isoamylase (with typically 2 μL of a 1000 U mL^{-1} stock solution) from Megazyme (Bray, Ireland) in 100 mM sodium acetate buffer, pH 4.0, at 37 °C for 3 h (as illustrated in Figure 3A). NMR spectroscopy confirmed the loss of α-1,6 linkages. Reactions were terminated by heating at 95 °C for 5 min, and the products were analyzed either directly or after precipitation with ethanol (5 vol). Precipitated samples were dissolved in water (50 μL) and freeze-dried to yield an amorphous powder.

Mass Spectrometry. Matrix-assisted laser desorption/ionization mass spectra were acquired as described previously[3] on Bruker Ultraflex or Autoflex Speed MALDI-TOF/TOF mass spectrometers (Coventry, United Kingdom), equipped with a pulsed nitrogen laser emitting at 337 nm or Smartbeam-II 2 kHz laser, respectively. Samples were typically mixed 1:2 v/v with a 20 mg mL^{-1} solution of 2,5-dihydroxybenzoic acid in 50% aqueous acetonitrile and spotted on a target plate (Bruker MTP 384 Polished Steel TF Target). Data were processed using Bruker flexAnalysis software. The relative abundance of malto-oligosaccharides of DP $\geq$ 4 was determined at least in duplicate by measuring the area under the peaks of all isotopes associated with each species of a given degree of polymerization using the Sophisticated Numerical Annotation Procedure with sugar-sodium adduct average composition. When experiments involved quantifying both reduced and nonreduced malto-oligosaccharides in a sample, the intensity of the main isotopologue was used to quantify each species, taking into account the contribution of any nonreduced isotopologues with +2 Da when quantifying reduced species of the same mass. The percentage contribution of such nonreduced isotopologues for each DP of malto-oligosaccharide was determined from experimental spectra of authentic compounds. Any difference between the ionization efficiencies of reduced and nonreduced malto-octaose was <15% in control spectra (data not shown), showing that quantitative comparison between such species is possible. Ion mobility mass spectrometry was carried out as described previously.[25] Samples were analyzed using a Synapt G2 HDMS mass spectrometer (Waters Corp., Manchester, UK) using electrospray ionization in traveling-wave mobility mode with nitrogen and helium gas flows in negative mode.

Capillary Electrophoresis. Reducing saccharides were labeled with 8-aminopyrene-1,3,6-trisulfonic acid to facilitate their separation on the basis of charge and their detection by fluorescence spectrophotometry. Samples (typically 5 μL) were treated with 8-aminopyrene-1,3,6-trisulfonic acid (1 μL of 50 mM in 15% aqueous glacial acetic acid) and freshly prepared NaCNBH$_3$ in tetrahydrofuran (1 μL of 1 M). Reaction
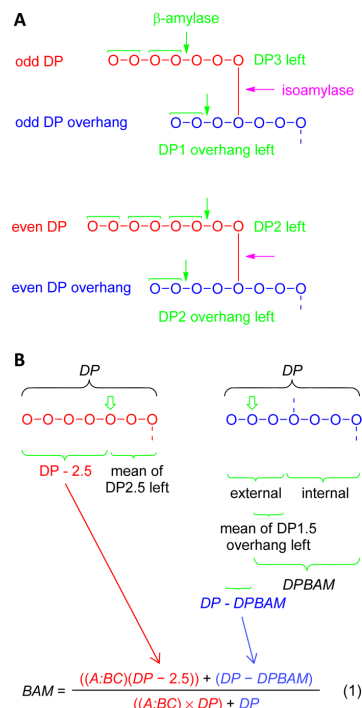
Figure 3. Cartoon showing the degradation of α-glucan by β-amylase and isoamylase. (A) Isoamylase hydrolyzes the α-1,6-linked branch points (indicated in purple). β-Amylase hydrolyses maltose units iteratively from the nonreducing ends of linear chains (indicated in green). A chains (shown in red) are left with a DP of either 3 or 2, depending on whether the original DP was odd or even, respectively. B chains (shown in blue) are left with an overhang of either DP 1 or 2 depending on whether the original overhang was odd or even, respectively. C chains are degraded in the same way as for B chains (not shown). (B) A chains are therefore extensively hydrolyzed by β-amylase leaving a mean of 2.5 glucosyl residues (i.e., DP2.5). B and C chains are hydrolyzed leaving a mean overhang of 1.5 glucosyl residues (indicated by the broad green arrows). The contribution of the hydrolysis of A and B/C chains to the extent to which β-amylase degrades α-glucan (BAM) is indicated in the context of eq 1, which can be rearranged to give eq 2 that defines A:BC. Thus, the right-hand side of eq 1, in both the numerator and the denominator, relates to the extent of hydrolysis of B plus C chains. The left-hand side relates to the extent of hydrolysis of the A chains, where A:BC is the ratio of the numbers of A to B plus C chains normalized to the number of B plus C chains.

mixtures were incubated at 37 °C for 4 h, diluted 10-fold with water. Samples were subjected to capillary electrophoresis using a Beckman Coulter PA800 ProteomeLab Protein Characterization System (High Wycombe, United Kingdom) using laser-induced fluorescence detection with an excitation wavelength of 488 nm and an emission band-pass filter of 520 nm ±10 nm. An SCIEX N-CHO capillary (50 μm internal diameter × 50.2 cm; 40 cm effective length to detector) was used with a constant voltage of −30 kV, the anode at the detector, and a temperature of 25 °C. The capillary was rinsed with SCIEX N-CHO buffer, and the samples were introduced at a pressure of 0.5 psi for 5 s.

Determination of the A:BC Chain Ratio. α-Glucan and glycogen were each exhaustively treated for at least 1 h at 37 °C with β-amylase from barley Megazyme (Bray, Ireland) in 10 mM MES pH 6.0. The only malto-oligosaccharide produced was maltose according to mass spectrometry as expected. The amount of maltose liberated by β-amylase was determined using the 3,5-dinitrosalicylic acid method. A solution of 3,5-dinitrosalicylic acid (43.8 mM 3,5-dinitrosalicylic acid containing 1.06 M potassium sodium L-(+)-tartrate and 0.4 M NaOH) was added to a sample (1:1 v/v) with water and incubated at 95 °C for 5 min. The mixtures were cooled on ice for 10 min, diluted 2-fold, and the absorbance of the mixtures was measured at 540 nm and compared with a glucose calibration curve. The amount of glucose residues liberated would have been 2 times the amount of maltose detected. In separate experiments, all glucosyl residues within a polymer were liberated with *Bacillus licheniformis* α-amylase (EC 3.2.1.1 in 50 mM MOPS buffer, pH 7.0, at 99 °C for 6 min to generate shorter dextrins) followed by *Aspergillus niger* amyloglucosidase (EC 3.2.1.3 in 200 mM sodium acetate buffer, pH 4.5, at 50 °C for 30 min to hydrolyze all remaining α-1,4 and α-1,6 linkages) from the Megazyme Total Starch Assay kit K-TSTA 05/2008 (Bray, Ireland),[26] allowing the total glucose content to be determined using 3,5-dinitrosalicylic acid. All measurements were carried out at least in duplicate. For example, three independent replicate experiments were carried out with *S. venezuelae* α-glucan giving 21.0 ± 4.7 (SD of replicate means).

The proportion of glucose residues liberated as maltose from A and B/C chains by β-amylase (BAM) is defined by eq 1 as illustrated in Figure 3.

$$\text{BAM} = \frac{((\text{A:BC})(\text{DP} - 2.5)) + (\text{DP} - \text{DPBAM})}{((\text{A:BC}) \times \text{DP}) + \text{DP}} \qquad (1)$$

where A:BC is the ratio of the numbers of A to B plus C chains normalized to the number of B plus C chains (where, in this context, the sole C chain per molecule is equivalent to a B chain because it bears at least one branch also).

DP is the mean degree of polymerization of all chains (it is assumed that all types of chain have a similar mean DP; nevertheless, the lower the A:BC chain ratio, the less sensitive eq 2 below becomes to differences in the mean DP of A chains); e.g., five independent replicate experiments were carried out with *S. venezuelae* α-glucan giving 7.3 ± 0.4 (SD of replicate means).

DPBAM is the mean degree of polymerization of linear B chains after digestion with β-amylase determined by mass spectrometry (where maltose and malto-triose were ignored because they were derived from A chains); e.g., four independent replicate experiments were carried out with *S. venezuelae* α-glucan giving 6.5 ± 0.2 (SD of replicate means).

Equation 1 can be rearranged to define the A:BC ratio to give eq 2

$$\text{A:BC} = \frac{\text{DP} - \text{DPBAM} - (\text{BAM} \times \text{DP})}{(\text{BAM} \times \text{DP}) - \text{DP} + 2.5} \qquad (2)$$

Production of Recombinant Enzymes. The DNA sequences of the *glgB* genes from *M. tuberculosis* H37Rv, *Mycobacterium smegmatis* mc^2 155, *Streptomyces coelicolor* A3(2) isoform 1, and *Streptomyces venezuelae* ATCC 10712 were synthesized with optimized codon usage for expression in *Escherichia coli* (Genscript Corporation, Piscataway, NJ, U.S.A.), allowing the production of each enzyme with an N-terminal His$_6$ tag and TEV cleavage site. Each gene was flanked at the 5′ end by an *Nde*I restriction site and a sequence encoding a His$_6$ tag and a TEV protease cleavage site, and at the 3′ end by a *Bam*HI restriction site. Each gene, with exceptions described below, was subcloned in to a pET-21a(+) (Novagen, Watford, United Kingdom) vector using *Nde*I and *Bam*HI restriction sites and heterologously expressed in *E. coli* BL21(DE3). The *M. tuberculosis* H37Rv *glgB* gene sequence was mutated into that of *M. tuberculosis* CCDC5180 by changing the codon for Ser470 into one for Pro using QuikChange (Agilent Technologies, Santa Clara, CA, U.S.A.). Cells containing the resulting expression plasmid were cultured in Lysogeny Broth containing carbenicillin (100 μg mL^{-1}) at 37 °C. When the cells reached an OD600 nm of 0.6, expression was induced by the addition of 0.5 mM isopropyl β-D-thiogalactopyranoside. After incubation at 30 °C for 2 h and then 16 °C for 17 h, the cells were harvested by centrifugation and resuspended in 50 mM Tris-HCl, pH 8.0, containing 15 mM imidazole, 300 mM NaCl, DNase 1, and a Complete protease inhibitor cocktail tablet (Roche, Burgess Hill, United Kingdom). A similar strategy was used for the production of the other recombinant proteins. The cells were disrupted with a TS Series Benchtop 1.1 kW cell disruptor (Constant Systems Ltd.) at 25 kPSI. The resulting cell lysate was separated from the cell debris by centrifugation at 20000*g* for 15 min at 4 °C. Enzyme were purified from each cell lysate by application onto a 1 mL FF HisTrap column (GE Healthcare, Amersham, United Kingdom). Protein was eluted with 15−300 mM imidazole gradient. Enzymes were further purified to remove any traces of *E. coli* amylase activities using a HiLoad 26/60 Superdex 200 gel filtration column (GE healthcare) equilibrated with 20 mM sodium phosphate buffer, pH 7.4, containing 20 mM NaCl. The purified proteins were homogeneous according to SDS-PAGE with Coomassie staining.

The production of *S. venezuelae* GlgB was not successful with pET21a vector. The *glgB* gene was ligated into a lower copy-number pBAD43 vector[27] using *Nco*I and *Hind*III restriction sites, placing the gene under the more stringent control of an arabinose-inducible promoter. GlgB was produced in *E. coli* TOP10 cells (Novogen) grown at 37 °C to an OD at 600 nm of 0.6 in Lysogeny Broth before induction with 0.02% w/v L-arabinose followed by a further 5 h of incubation. The enzyme was purified using both Ni-affinity and size exclusion chromatographies as described above.

Enzyme-Catalyzed Reactions. To generate synthetic α-glucan and to determine the branch length specificity of GlgB enzymes, 1 mM malto-hexaose and 10.1 mM α-maltose-1-phosphate were incubated with 7 μg each of recombinant GlgE and GlgB (giving <2 μM final concentrations) in 100 mM bis-Tris propane, pH 7.0, containing 50 mM sodium chloride in a total volume of 50 μL. After 18 h at 22 °C, a second aliquot of α-maltose-1-phosphate was added. After a further 90 h, the sample was heated at 95 °C for 5 min, cooled, and washed 5 times using a 10 000 Da cutoff membrane with 100 mM sodium acetate buffer, pH 4.0, to remove small malto-oligosaccharides. The sample was analyzed by electron microscopy, mass spectrometry, or capillary electrophoresis with or without debranching as appropriate.

In other experiments, malto-oligosaccharides (typically 5 mg) were extended with GlgE (14 μg) and 6.6 mM α-maltose 1-phosphate in a total volume of 122 μL at 37 °C for 3 h. The enzyme was denatured by heating to 96 °C for 5 min. As appropriate, GlgB (70 μg) was then added, for example, and the

Table 1. Properties of α-Glucans and Glycogen

	M. tuberculosis		*S. venezuelae*	rabbit liver	*E. coli*
	capsular	cytosolic	cytosolic	glycogen	glycogen
Size					
TEM β particle (nm)[a]	63 ± 7	58 ± 2	42 ± 3	38 ± 2	34 ± 5
after β-amylase[a,b]	nd[c]	nd[c]	42 ± 3	29 ± 2	nd[c]
DLS β particle (nm)[d]	64 ± 8	64 ± 5	51 ± 11	59 ± 10	nd[c]
after β-amylase[b,d]	nd[c]	nd[c]	54 ± 1	40 ± 1	nd[c]
NMR diffusion coefficient (m^2 s^{-1})[e]	nd[c]	nd[c]	2.5×10^{-12}	2.9×10^{-12}	nd[c]
NMR α particle (nm)[e]	nd[c]	nd[c]	~280	~240	nd[c]
NMR α particle (Da)[e]	nd[c]	nd[c]	$\sim 5 \times 10^{6}$	$\sim 4 \times 10^{6}$	nd[c]
Linear Chain DP[f]					
overall mean (DP)	7.0 ± 2.0	7.2 ± 1.9	7.3 ± 1.8	10.8 ± 4.6	11.0 ± 2.9
C mean[g]	nd[c]	nd[c]	9.1 ± 2.5	nd[c]	nd[c]
Extent of Branching from					
overall mean DP[h]	0.14	0.14	0.14	0.09	0.09
NMR spectroscopy[i]	nd[c]	nd[c]	0.14	0.09	nd[c]
β-Amylase Treatment					
mean DP (DPBAM)[f]	6.1 ± 1.4	6.2 ± 1.8	6.5 ± 1.6	8.6 ± 3.5	9.8 ± 3.3
residues liberated % (BAM)	31.1	33.0	21.0 ± 4.7	47.5	40.5
internal chain length[j]	4.6 ± 1.4	4.7 ± 1.8	5.0 ± 1.6	7.1 ± 3.5	8.3 ± 3.3
A:BC chain ratio[k]	0.6	0.6	0.2 ± 0.2	0.9	0.8
mean branches per B/C chain[l]	~1.6	~1.6	~1.2 ± 0.2	~1.9	~1.8

[a]The diameters of individual particles were determined by transmission electron microscopy (TEM) from at least 10 measurements (see Figure 2). [b]The size of glycogen particles after treatment with β-amylase. [c]Not determined. [d]The hydrodynamic diameter of the dominant species of <100 nm in each sample was determined using dynamic light scattering where at least five measurements were made. [e]The diffusion coefficient, hydrodynamic diameter, and molecular mass of the dominant large species assigned as α particles were determined using pulse field gradient NMR spectroscopy (see Figure S2 for the calibration curve). [f]The DP mean (mass average) was determined for products with a DP $\geq$ 4 using mass spectrometry (see Figure 4 for representative spectra). Chain length distributions are expressed as SD. [g]The DP of C chains was determined as described in Figure 5. [h]Extent of branching (α-1,6 linkages divided by the sum of α-1,6 and α-1,4 linkages) equates to the reciprocal mean branch length. [i]Extent of branching was also determined by integrating the appropriate anomeric resonances in ^{1}H NMR spectra without solvent suppression of samples subjected to repeated freeze-drying and resolubilization in D$_2$O. [j]Calculated from DPBAM minus 1.5. [k]The ratio of the numbers of A to B plus C chains normalized to the number of B plus C chains, from eq 2. [l]Essentially A:BC plus one.

mixture was incubated at 30 °C for 2 h. The enzyme was then denatured by heating and analyzed as described above.

The GlgE-dependent extension of acceptors with α-maltose 1-phosphate was monitored using malachite green to detect the production of inorganic phosphate as described previously.[3,19,28]

RESULTS

Isolation of α-Glucan Particles from Actinomycetes and Classical Glycogen Particles from *E. coli*. Before characterizing GlgB branching enzymes from actinomycetes, we characterized the α-glucans produced by these bacteria together with classical glycogens to allow direct comparisons to be made. To this end, α-glucans/glycogen were first isolated from the cytosols of *S. venezuelae*, *M. tuberculosis* (the default strain in this study is H37Rv), and *E. coli*, from the capsule of *M. tuberculosis*, and from whole cells of *M. smegmatis*. We included *S. venezuelae* because we have previously shown that this organism generates α-glucan solely by the GlgE pathway.[8] This contrasts with *E. coli*, which generates its glycogen only via the classical GlgA glycogen pathway.[2,29] NMR spectra of the isolated materials gave characteristic resonances of α-1,4- and α-1,6-linked glucose polymers (examples of solution- and solid-state NMR spectra of *S. venezuelae* α-glucan are shown in Figure S1).[17,30−32] Transmission electron microscopy showed that each of the isolated polymers comprised β particles that sometimes aggregated into larger α particles (Figure 2 and Table 1).[33] According to both transmission electron micros-

copy and dynamic light scattering, the β particles had diameters ranging from ~30 to ~60 nm, consistent with previous reports.[15−17,30,34] Sectioned cells of *S. venezuelae* and *M. smegmatis* that were stained for α-glucan showed the presence of cytosolic β particles of a similar diameter (~30−45 nm) to that of isolated material (Figure 2). Only faint speckled staining was observed on the exterior of *M. smegmatis* cells, suggesting the presence of either little or diffuse capsular α-glucan in these growth conditions. Pulse field gradient NMR spectroscopy showed that isolated *S. venezuelae* α-glucan and rabbit liver glycogen contained slowly diffusing α particles with diameters of ~240−280 nm, equating to ~5 × 10^6 Da assemblies comprising ~30 000 glucosyl residues (Table 1 and Figure S2). Similar values have been reported for material isolated from *M. smegmatis*.[35]

Actinomycete α-Glucans Are Composed of Short Linear Chains of Length ~7−8. It has previously been reported that the α-glucans of mycobacteria contain shorter linear chains[10,15−17,30,34,36] than those of classical glycogen.[29] In order to establish whether this was the case with *S. venezuelae* α-glucan, the isolated material was treated with isoamylase to hydrolyze the α-1,6 linkages (Figure 3A). Such a sample remained in solution and did not give strongly colored complexes with iodine according to absorbance spectroscopy (Figure S3), as would be expected with liberated linear chains having a degree of polymerization (DP) of <15.[37] Indeed the absorbance of the iodine complexes was only slightly greater than with malto-heptaose. By contrast, debranched rabbit liver glycogen yielded a little precipitate in the absence of iodine,
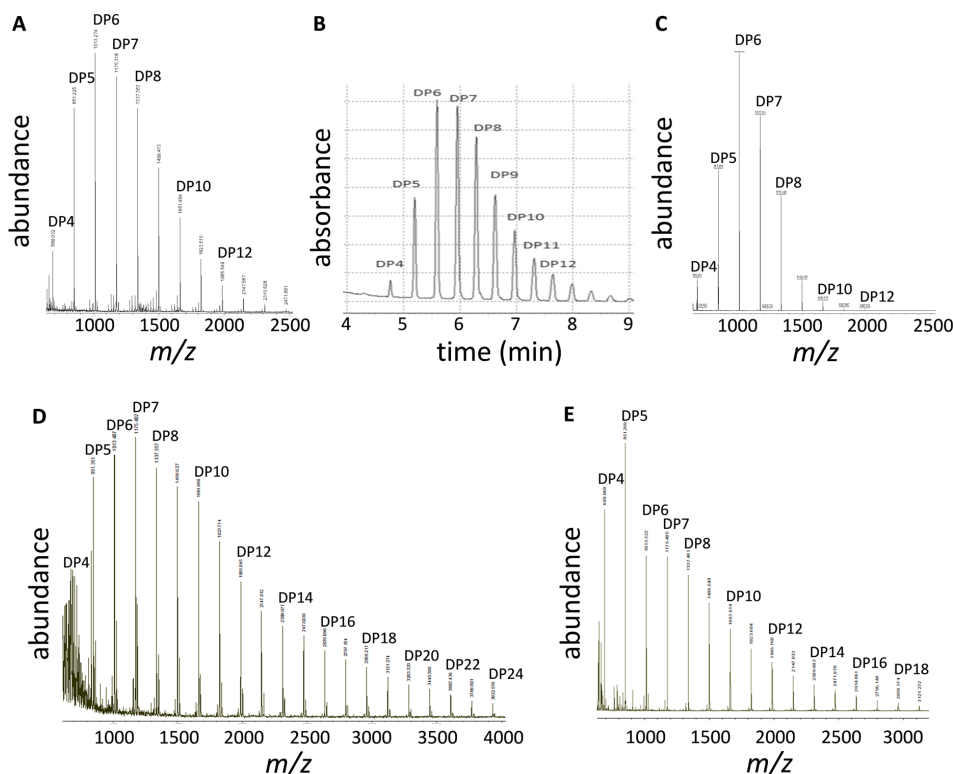
Figure 4. Branch length distribution of *S. venezuelae* α-glucan and rabbit liver glycogen. The α-1,6 linkages in the α-glucan were hydrolyzed using isoamylase to <1% according to ^{1}H NMR spectroscopy and subjected to either (A) MALDI MS or (B) capillary electrophoresis after treatment with 8-aminopyrene-1,3,6-trisulfonic acid and NaCNBH$_3$, showing similar chain length distributions. Only after treatment with isoamylase were malto-oligosaccharides observed (data not shown). (C) A sample treated with β-amylase before debranching was analyzed by MALDI-MS and showed a loss of longer DPs and a relative gain of medium DPs of ~6. Mass spectrometry of a sample that was treated solely with β-amylase showed the presence of only maltose as expected (data not shown). Debranched rabbit liver glycogen was subjected to MALDI MS either (D) without or (E) with pretreatment with β-amylase, leading to the substantial net reduction in DP. Only after treatment with isoamylase were malto-oligosaccharides observed, and a sample that was treated solely with β-amylase showed the presence of only maltose as expected (data not shown). Additional samples were analyzed as summarized in Table 1.

indicative of the liberation of linear malto-oligosaccharides that were too long to remain in solution (>DP20). In the presence of iodine, such a sample exhibited significant absorbance with a λ_{max} of ~537 nm, revealing the liberation of linear malto-oligosaccharides of DP up to ~30.[37]

Matrix assisted laser desorption ionization mass spectrometry showed that that debranched *S. venezuelae* α-glucan and rabbit liver glycogen had mean DPs (mass average) of 7.3 and 10.8, respectively (Figure 4 and Table 1), the latter consistent with published values.[29] The narrower distribution of DPs in α-glucan compared with glycogen was also reflected in the standard deviations of their means of 1.8 and 4.6, respectively. The reliability of this analysis was independently verified using capillary electrophoresis (Figure 4). For example, the *S. venezuelae* material gave a value of 7.7 ± 2.2, which was similar to the value of 7.3 ± 1.8 obtained with mass spectrometry. Analysis of the materials from *M. tuberculosis* capsule and cytosol showed that they comprised DPs of ~7 that were similar not only to that of *S. venezuelae*, but also each other (Table 1). A similar value was obtained with material isolated

from whole *M. smegmatis* colonies (7.1 ± 2.0). By contrast, the *E. coli* material had a longer mean DP of 11.0 ± 2.9 (Table 1), as expected for a classical glycogen.[29]

Polymer samples treated with pullulanase from either *Bacillus acidopullulyticus* or *Klebsiella pneumonia* also led to the cleavage of α-1,6 linkages. However, the hydrolysis of α-1,4 linkages also occurred such that glucose and maltose were the sole products after 24 h. This side reaction and/or contamination of pullulanase preparations might explain the very narrow distribution of very short chain lengths reported in one study.[16]

The degree of branching (the ratio between the number of α-1,6 linkages and the total number of linkages) can be calculated from the reciprocal of the mean chain DP. This gave 0.14 and 0.09 for actinomycete α-glucans and glycogens, respectively (Table 1). These values were independently verified by integrating the anomeric resonances observed using ^{1}H NMR spectroscopy (Table 1). Furthermore, the values were consistent with previous reports for materials isolated from mycobacteria based on either methylation analysis, which gave a degree of branching of 0.14,[10,17,36] or enzymatic analyses.[34]

The C Chain of *S. venezuelae* α-Glucan has a DP of ~9.
Classical glycogen is an arboreal (tree-like) polymer. From the sole C chain with its reducing end stem one or more α-1,6-linked B chains that, in turn, bear further α-1,6 linked branches. The extremities of the nonreducing ends of the polymer comprise A chains that do not themselves bear branches. The C chain is uniquely addressable through reduction with sodium borohydride leading to an increase in mass by 2 Da. The debranching of reduced *S. venezuelae* α-glucan allowed the detection of small amounts of reduced linear malto-oligosaccharides using mass spectrometry. For example, chains with a DP of 8 in nonreduced control samples gave the natural isotopic distribution of isotopologues (Figure 5). However, a
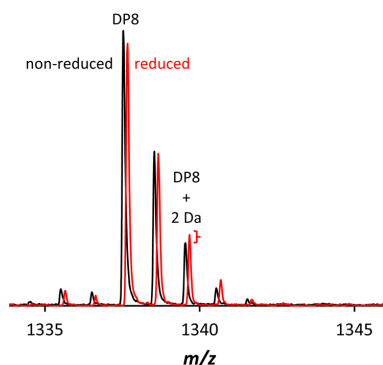

Figure 5. Detection of short C chains in *S. venezuelae* α-glucan. The mass spectra show the isotopic distributions of the DP8 product of debranched α-glucan with or without prior reduction. Without prior reduction (black), the natural abundance isotopic distribution of isotopologues was observable. With prior reduction with NaBH$_4$ (red; displaced slightly along the horizontal axis for clarity), a small but significant relative enhancement of the isotopologue with two addition mass units (1339.4 *m/z* for DP8) was observed (highlighted by the bracket), indicative of the presence of small quantities of C chains of DP8 that had been reduced at their reducing ends. Spectra of products of DP4 to 14 were similarly recorded, and the absolute differences in the signals with two additional mass units for each DP allowed the chain length distribution of C chains to be calculated.

reduced sample showed a relative increase in the abundance of the isotopologue with an additional two mass units. Such reduced species must have been liberated C chains because malto-oligosaccharides are detected by mass spectrometry in samples of α-glucan only when they are debranched. The reduced products had a mean DP ± SD of 9.1 ± 2.5 (Table 1). Therefore, the C chains were slightly longer than the average of ~7 of all chains within the polymer.

Internal Chain Length of Actinomycete α-Glucan B Chains Is 4.6−5.0. To define the internal chain length of B chains (i.e., between the branch point of B chains and the furthest branch they bear as illustrated in Figure 3), α-glucan samples were treated with β-amylase. This exoacting enzyme hydrolyzes maltose units from the nonreducing ends of B chains to within one or two glucosyl residues of the first branch point it encounters, depending on whether the external chain has an odd or even number of sugar residues.[29,38] Given that there is no detectable odd/even bias within the polymer chains, the average overhang beyond the outermost branch after digestion is 1.5 glucosyl residues in length. By contrast, A

chains are almost completely digested to give maltose and maltotriose (Figure 3). After treatment, the mean DPs of actinomycete α-glucans of DP ≥ 4 were 6.1−6.5 (Table 1 and Figure 4), showing that a typical B chain of seven residues bears a branch 4.6−5.0 residues from its own branch point (i.e., roughly half way along their length on their fourth or fifth residue). This contrasts with classical glycogens from rabbit liver and *E. coli* that have internal chain lengths of ~7−8 (Table 1).

α-Glucan from *S. venezuelae* Has a Smaller A:BC Chain Ratio than Classical Glycogen. The ratio of the number of A vs B plus C chains (A:BC) of classical glycogen is typically close to 1.[29] The A:BC chain ratio can be determined from the mean DP before and after treatment with β-amylase together with the extent of hydrolysis by this enzyme (where eq 2 is derived from eq 1, which is also illustrated in Figure 3). The mean number of branches per B/C chain in a polymer of finite size is essentially the A:BC chain ratio plus one. For example, consider a 17 kDa polymer consisting of 105 glucosyl residues in 15 chains with a mean DP of 7 (Figure S4). If each B and C chain bore only one branch (Figure S4A), the polymer would comprise 1 A chain, 13 B chains one C chain, giving an A:BC chain ratio of close to zero (0.07). If the polymer comprised B and C chains that bore two chains each (Figure S4B), it would have eight A chains, six B chains, and 1 C chain, giving an A:BC chain ratio of close to 1 (1.14).

α-Glucans were conspicuously less prone to hydrolysis by β-amylase (21−33% glucose liberated) than *E. coli* classical glycogen (~40−47%) (Table 1). This was also evident from electron microscopy and dynamic light scattering (Table 1) that showed that only glycogen particles were significantly diminished in size. The A:BC chain ratio with *S. venezuelae* α-glucan was 0.2 ± 0.2, giving ~1.2 ± 0.2 branches per B/C chain (Table 1). The values for both types of *M. tuberculosis* α-glucan were a little higher at 0.6 and ~1.6, respectively. These values contrast with those of 0.8−0.9 and 1.8−1.9, respectively, for classical glycogens (Table 1).[29]

GlgB Enzymes from Actinomycetes Generate Relatively Short Branches of DP ∼ 7−8. It might be expected that the branch length specificity of GlgB branching enzyme from actinomycetes reflects the shorter lengths present in the α-glucans from these bacteria. The treatment of α-1,4-linked amylose with a short or medium DP with *M. tuberculosis* GlgB branching enzyme led to a loss of absorbance in the presence of iodine (Figure S5), as previously described.[39] The chain length specificity of this enzyme was then determined by exposing α-maltose 1-phosphate to GlgE and GlgB to generate a branched α-glucan product that could be debranched and subjected to mass spectrometry (Figure 6). The chains generated by *M. tuberculosis* GlgB in the presence of *M. tuberculosis* GlgE had a mean DP of 8.6 ± 2.5 (Table 2). This value was verified using capillary electrophoresis, which gave 8.0 ± 3.1 (Table 2). No branched products were observed when malto-octaose was exposed to GlgB suggesting this malto-oligosaccharide was not long enough to be branched. Mean values of between 7.2 and 8.5 were obtained when using *S. coelicolor* GlgE with GlgB from either *M. tuberculosis* or a number of other *Mycobacterium* and *Streptomyces* species (Table 2). This contrasts with chain DPs of >10 observed with branching enzymes from organisms that produce classical glycogen, such as that from *E. coli*.[18,40] The actinomycete GlgB branching enzymes therefore generate relatively short branches, as predicted.
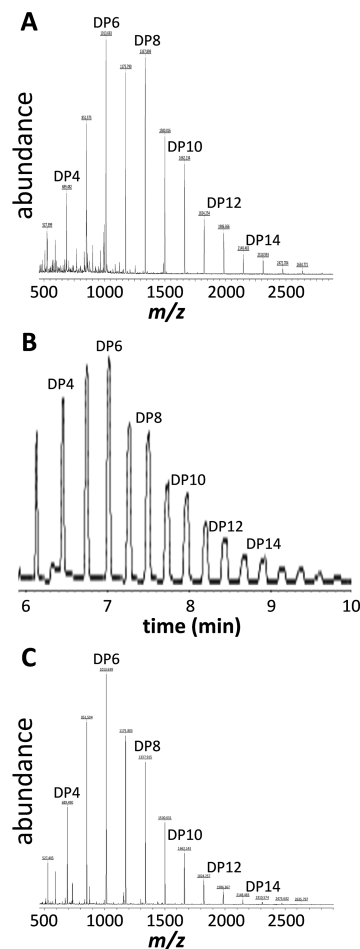
Figure 6. Branch length specificity of GlgB branching enzymes. The polymer, generated using *M. tuberculosis* GlgB, *S. coelicolor* GlgE, and α-maltose 1-phosphate, was filtered to remove small malto-oligosaccharides and then treated with isoamylase. The resulting debranched sample was then analyzed using either (A) MALDI mass spectrometry or (B) capillary electrophoresis. (C) A mass spectrum is also shown of a sample generated with GlgB enzyme from *S. venezuelae*. The mean branch lengths are shown in Table 1.

Table 2. Branch Length Distributions Generated by GlgB Branching Enzymes in the Presence of GlgE

| | | branch length[a] | |
source of GlgB	source of GlgE	MS[b]	CE[c]
M. tuberculosis	*M. tuberculosis*	8.6 ± 2.5	8.0 ± 3.1
	S. coelicolor	8.1 ± 2.7	7.5 ± 3.5
A chains	na[d]	8.3 ± 1.9[e]	nd[f]
odd length C chains	na[d]	9.0 ± 2.9[e]	nd[f]
M. tuberculosis CCDC5180	*S. coelicolor*	8.4 ± 2.5	nd[f]
M. smegmatis mc[2] 155	*S. coelicolor*	8.5 ± 2.5	nd[f]
S. venezuelae	*S. coelicolor*	7.9 ± 2.4	nd[f]
S. coelicolor isoform 1	*S. coelicolor*	7.2 ± 2.2	7.1 ± 3.1

[a]Length distributions shown are mean ± SD of products of DP ≥ 3 of samples of malto-hexaose treated with GlgB, GlgE, and α-maltose 1-phosphate followed by isoamylase. [b]Determined using mass spectrometry. [c]Determined using capillary electrophoresis. [d]Not applicable. [e]The A and C chain length distributions were determined from a single branching event as shown in Figure 7. [f]Not determined.

with NaBH$_4$, and treated with isoamylase to cleave the α-1,6 linkages. This would liberate reduced C chains and nonreduced A chains. All nonreduced species would be A chains. By contrast, the reduced C chains would be significantly contaminated with starting materials with an even number of residues that would also be reduced. GlgB does not discriminate between odd and even chain transfers. Therefore, reduced species that had an odd DP would be liberated C chains, with only a low level of contamination from hydrolytic side-products, as described below.

Mass spectrometry was used to quantify the abundance of reduced C chains that had an odd DP, giving a mean DP of 9.0 ± 2.9 (Figure 7B and Table 2). For A chains, the distribution of nonreduced species with an even DP (mean DP 8.1 ± 2.7; data not shown) was similar to that of the odd species (mean DP 8.5 ± 2.6 shown in Figure 7B) giving an overall mean DP of 8.3 ± 2.7 (Table 2). This value was consistent with the DP of 8.1−8.6 generated in the presence of GlgE as described above (Table 2). The overall higher abundance of reduced over nonreduced chains was consistent with GlgB generating some hydrolytic products. From this, the hydrolysis to branching ratio was estimated to be ∼0.2:1 for such single branching events with short-chain substrates.

Branching by *M. tuberculosis* GlgB Is Strictly Intra-chain. Branching enzymes act by transferring a segment from the nonreducing end of a malto-oligosaccharide to a hydroxyl group at the six position of a glucosyl residue along a malto-oligosaccharide, via cleavage of an α-1,4 linkage and a glycosyl-enzyme intermediate.[41] Both intra- and interchain branching has been reported with GlgB enzymes.[18,42,43] In order to establish whether *M. tuberculosis* GlgB carries out intra- and/or interchain transfers, we first generated ^{13}C-labeled and unlabeled populations of malto-oligosaccharides (Figure 8 and Figure 9). This was possible by exposing unlabeled malto-tetraose (DP4) to GlgE in the presence of labeled or unlabeled α-maltose 1-phosphate. These two populations of molecules were then boiled to denature the GlgE, mixed, boiled again, and then cooled, to allow any intermalto-oligosaccharide associations that might potentially occur to be formed at random. On exposure to GlgB, there would be three possible outcomes (Figure 7). If the branching reaction were intrachain, the masses of the products would be identical to those of the substrates. If it were interchain, the label would be redistributed

***M. tuberculosis* GlgB Transfers a Nascent A Chain of DP ∼ 8 to a Recipient Nascent C Chain of DP ∼ 9.** The mean chain DPs of the polymers generated in the presence of GlgB and GlgE would include all A, B, and C chains, where the B chains would dominate, and the C chains would be the least represented. We therefore devised an experiment to define the DP of nascent A chains that are transferred onto nascent C chains (Figure 7A). Malto-octaose (DP8) was first treated with GlgE and α-maltose 1-phosphate to extend it by maltose units to give a series of products with an even number of glucosyl residues. The sample was then boiled, to denature the GlgE, followed by treatment with GlgB to introduce a single branch. The sample was boiled again, to denature the GlgB, reduced
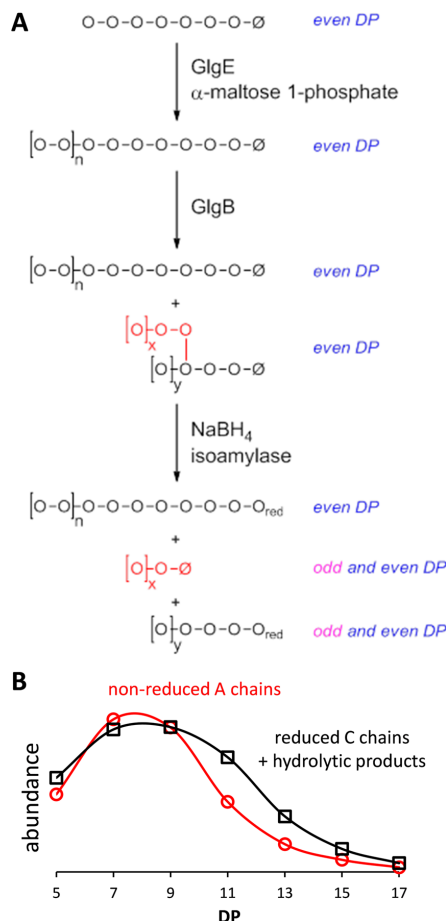
Figure 7. Branch length distributions of A and C chains generated by *M. tuberculosis* GlgB are similar. (A) The scheme shows the experimental design. Linear molecules and C chains are shown in black, and A chains are shown in red, where Ø and O_{red} refer to reducing end and reduced glucose residues, respectively. A population of malto-oligosaccharides with even DPs was generated with GlgE, malto-octaose, and α-maltose 1-phosphate. After heat-denaturation of GlgE, the sample was treated with GlgB. After heat-denaturation of GlgB, the sample was reduced and treated with isoamylase to cleave the α-1,6 linkages. (B) The final products from the experiment were subjected to mass spectrometry. The abundance of malto-oligosaccharide ions are plotted as a function of their degree of polymerization (DP). The quantification of reduced species took into account the natural abundance of isotopes from experimental mass spectra of debranched *S. venezuelae* α-glucan. This was necessary because the larger the malto-oligosaccharide, the more naturally abundant species with $(m + 2)/z$ were in control samples. Nonreduced species with an odd DP would have originated from A chains created by GlgB (plotted in red), while reduced species would have originated from C chains and hydrolytic products (plotted in black).

in the products. If both types of reaction were to occur, there would be a partial redistribution of the label in the products.

It was possible to generate labeled and unlabeled starting materials with a DP of at least 20 (Figure 9). Closer inspection
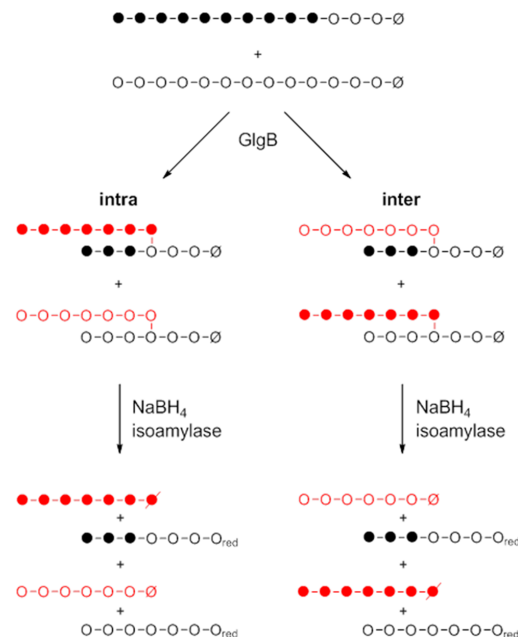
Figure 8. Experimental scheme to distinguish between intra- and interchain mechanisms of GlgB. Linear molecules and C chains are shown in black, and A chains are shown in red. Linear malto-oligosaccharides were generated without or with labels by extending malto-tetraose with *M. tuberculosis* GlgE using either $[U^{-12}C]$ or $[U^{-13}C]\alpha$-maltose 1-phosphate, respectively (see Figure 9). A mixture of these materials was then exposed to *M. tuberculosis* GlgB to establish whether the distribution of the label was retained in the products or not (see Figure 10). The filled circles, Ø, and O_{red} refer to labeled, reducing end, and reduced glucosyl residues, respectively.

of the unlabeled starting material showed that it contained the expected natural abundance isotopic distribution (Figure 9B). By contrast, the labeled material was dominated by a species with the expected number of ^{13}C labels, together with some higher natural abundance isotopes associated largely with the four unlabeled glucosyl residues (Figure 9D). In addition, there was a low level of some species with one or two labels missing, reflecting the fact that the α-maltose 1-phosphate was not quite 100% labeled.

Exposure of the mixed starting material to *M. tuberculosis* GlgB led to little obvious change in the overall distribution of chain lengths (Figure 10A). However, some hydrolysis had occurred because of a slight increase in the presence of species with an odd DP (Figure 10A) compared with the starting materials, which contained relatively little (Figure 9). For example, the unlabeled DP7 hydrolysis product had an abundance 8% that of the unlabeled DP6 species, showing that hydrolysis was a minor side reaction.

Most importantly, closer inspection of the main species with an even DP showed that the products contained the same distribution of label as the starting materials. For example, the DP16 species, which would be expected to include a product with a single branch point, had masses identical to those of the starting materials (Figure 10B) by being either fully unlabeled or with all but four glucosyl residues labeled. This is consistent
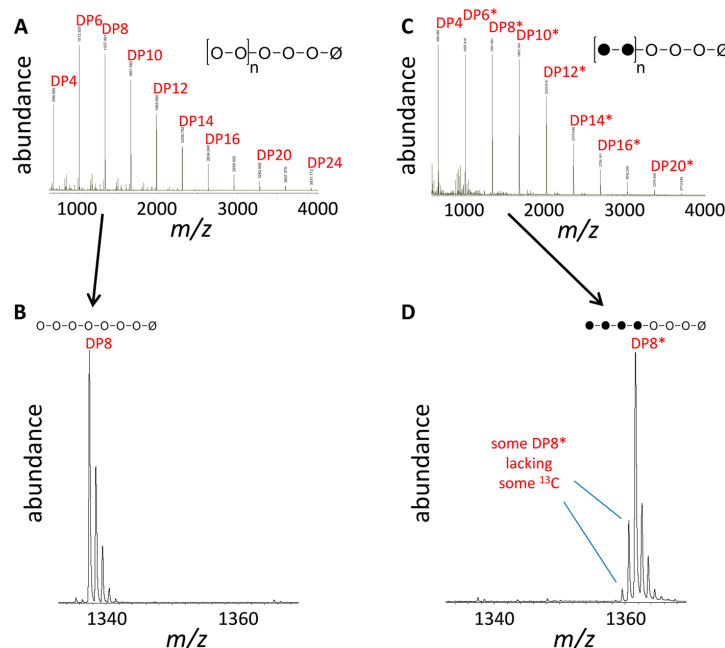
Figure 9. Enzymatic synthesis of labeled and unlabeled malto-oligosaccharides. Malto-tetraose (DP4) was extended by *M. tuberculosis* GlgE with either [U-^{12}C] or [U-^{13}C]α-maltose 1-phosphate as depicted in Figure 8. The filled circles and Ø in the structures refer to labeled and reducing end glucosyl residues, respectively. To stop the reaction, the samples were boiled to denature the enzyme. (A) A mass spectrum of unlabeled compounds showed the presence of a range of products. (B) The ions associated with each unlabeled product gave the expected natural abundance isotopic distribution, as illustrated by the DP8 product shown. (C) A mass spectrum of the labeled compounds also showed the presence of a range of products. (D) The ions associated with each labeled product showed the presence of the label, as illustrated by the DP8* product. The asterisk refers to the presence of ^{13}C-labeled glucosyl residues observed in all but the original DP4 acceptor.

with strictly intrachain branching because interchain branching would have given additional species with intermediate amounts of label.

In order to confirm that branching had indeed occurred, the sample was reduced and then debranched (Figure 8). It was immediately apparent that the overall distribution of species was shifted to shorter DPs (Figure 10C) consistent with the cleavage of branch points. Closer inspection of the products confirmed that this was indeed the case. For example, the reduced, labeled DP8 species had the identical isotopic distribution (Figure 10D) as the starting material (Figure 9D), except for the addition of 2 Da to all ions. This is consistent with starting material and liberated C chains (and likely some hydrolytic products). Furthermore, this shows that reduction was complete because no unreduced labeled species were detected. By contrast, the unlabeled DP8 product contained not only species with the same isotopic distribution as starting material plus 2 Da, but also nonreduced species amounting to over a quarter of the total unlabeled DP8 species (Figure 10D). The unreduced species can only have been derived from A chains formed by the branching reaction of GlgB. Such nonreduced species, with and without labels, were detected with a DP of between 5 and 10 giving a mean DP of 8.0 ± 1.4 for all liberated A chains, again consistent with the branching specificity experiments described above. These observations are therefore strongly supportive of a strictly intrachain mechanism.

Re-Evaluation of the Ability To Detect Malto-Oligosaccharides Containing an α-1,6 Branch Using Ion Mobility Mass Spectrometry. We previously interpreted the presence of two isomers of a malto-oligosaccharide with a DP of 11 in samples of amylose treated with *M. tuberculosis* GlgB as evidence for the ability to resolve linear and branched materials using ion-mobility mass spectrometry.[25] Careful re-evaluation of reaction mixtures with and without enzyme revealed that the relative abundance of the two observable isomers was in fact somewhat variable and, more critically, independent of GlgB (Figure S6). Reduction of a sample of linear malto-oligosaccharide (amylose 2800) with NaBH$_4$ clearly showed the collapse of the two isomers of DP11 into one. This phenomenon resembles what we reported with malto-hexaose (DP6), which was interpreted as the presence of two conformers, where one is extended and the other forms the first complete turn of a left-handed helix. Each turn of the helix involves the formation of O(6)−O(2) hydrogen bonds between glucosyl residues along the helical axis. Reduction of the reducing end would open up the terminal glucopyranose ring, disrupting the potential for such hydrogen bonding involving this residue. We therefore now interpret the presence of two isomers in the material with a DP of 11 in a similar way; it could adopt either a conformation with two complete turns of the helix or another more extended conformation. The variability in the relative abundance of the two isomers likely reflects very small differences in the stabilities of the two
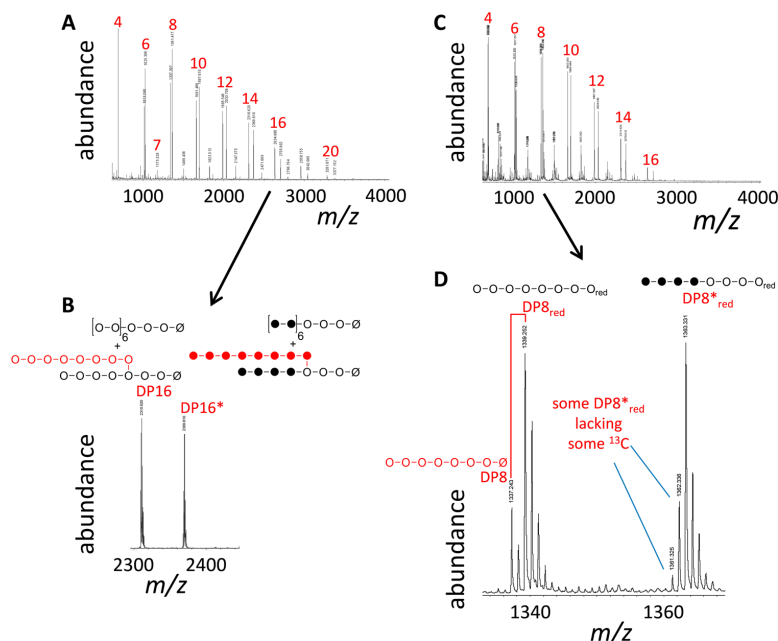
Figure 10. *M. tuberculosis* GlgB catalyzes intrachain branching. Labeled and nonlabeled compounds (see Figure 9) were mixed in equal proportions, boiled, cooled and treated with GlgB (see Figure 8 for experimental scheme). (A) A mass spectrum showed only minor changes in the profile of DP compared with starting materials (see Figure 9). (B) None of the products with an even DP showed any redistribution of ^{13}C labels, as illustrated with the DP16 product. The filled circles, Ø, and O$_{red}$ in the structures refer to labeled, reducing end, and reduced glucosyl residues, respectively. Structures in black were derived from linear molecules and C chains, while those in the color red were derived from A chains. (C) The sample was reduced with NaBH$_4$ and treated with isoamylase to hydrolyze all of the α-1,6 linkages in order to confirm the presence of branches. The mass spectrum showed the loss of compounds with the highest DP, consistent with the cleavage of branches. (D) The reduced products illustrate how the reduced products (DP8$_{red}$ and DP8*$_{red}$), derived from C chains and linear molecules, had the same isotopic distribution as the starting material, while the presence of nonlabeled, nonreduced product (DP8) showed that sample must have contained A chains generated by GlgB (see Figure 8 for the experimental scheme and the main text for further details).

conformations in the gas phase, and reduction would disrupt the formation of two complete helical turns.

GlgE Preferentially Extends A Chains over B and C Chains. GlgE must extend chains to attain the requisite length for GlgB to act. The B and C chains of an α-glucan cannot be frequently extended by GlgB because the majority of B and C chains bear only one branch, and their lengths clearly remain unchanged after being formed by GlgB. To test this hypothesis, the activity of *M. tuberculosis* GlgE was determined with rabbit liver glycogen and *S. venezuelae* α-glucan, which have A:BC chain ratios of 0.9 and 0.2, respectively (Table 1). This was done by monitoring inorganic phosphate release from the donor substrate of GlgE, α-maltose 1-phosphate. The k_{cat}/K_m of the enzyme was 6.4-fold higher with rabbit liver glycogen (Figure S7), consistent with GlgE exhibiting a preference for A chains. We then treated an equal mass of either oligosaccharides that contained a single branch, glycogen or α-glucan with GlgE and ^{13}C-labeled α-maltose 1-phosphate (for example, see Figure S8). Mass spectrometry of debranched samples showed the chains were extended by 94, 38, and 18%, respectively, again showing a correlation between the number of available A chains and the degree of extension.

Particles of α-Glucan Generated by GlgE and GlgB Resemble Those Isolated from Cells. Polymers generated by *M. tuberculosis* GlgE and GlgB from α-maltose 1-phosphate

and malto-hexaose (DP6) resembled those isolated from cells according to electron microscopy (Figure 2). The mean diameter $\pm$ SE of the synthetic material was about half (32 $\pm$ 2 nm) that of the biological polymer isolated from *M. tuberculosis*, but their morphologies were similar. Similar sized particles were generated when the reaction was carried out with GlgE from *S. coelicolor* (30 $\pm$ 2 nm, n = 40). Comparable diameters were also obtained when both *S. coelicolor* enzymes were used either with malto-hexaose (29 $\pm$ 5 nm, n = 10) or without (25 $\pm$ 4 nm, n = 10), showing that polymer initiation is possible without a primer.

■ **DISCUSSION**

GlgB enzymes from actinomycetes generate polymers in the presence of GlgE with DPs of typically 8.1 $\pm$ 2.5, which is reasonably similar to the typical DP of 7.2 $\pm$ 1.9 observed in polymer material isolated from cells. These are, to the best of our knowledge, among the shortest branch lengths reported. The shortest reported thus far for GH13 branching enzymes appear to be 7.9 $\pm$ 3.4 from *Vibrio vulnificus*[44] and 8.0 $\pm$ 2.6 from *Deinococcus geothermalis*.[45] We and others[18] have attempted to correlate the specificity of branching enzymes with their amino acid sequence but have failed to identify any obvious motif or domain responsible for any particular property. Nevertheless, the two N-terminal domains of

actinomycete GlgB enzymes, which include a family 48 carbohydrate binding module, do appear to be important because chimeric enzymes have altered chain length specificities.[44,45]

Interestingly, branching by *M. tuberculosis* GlgB was shown to be strictly intrachain. This implies that the nascent recipient chain, and by definition the remainder of the polymer, must bind very tightly and noncovalently to the enzyme until the nascent A chain is fully transferred to the recipient chain via a glycosyl-enzyme intermediate. We and others[41] have observed that branching enzymes can generate some hydrolytic by-products. This was more prevalent with short substrates, implying fewer potential contacts with the protein lead to weaker binding. With the *M. tuberculosis* enzyme at least, it would appear that water, rather than another polymer chain, acts as the acceptor after premature dissociation of the original recipient chain.

We have shown that the internal chain length of α-glucan is shorter than for classical glycogen. Many previous studies on glycogens have estimated internal chain lengths based solely on the degree of hydrolysis by β-amylase. This is inappropriate given that glycogen contains almost as many A chains and B chains, and A chains are almost completely digested.[38] This artificially decreases the calculated internal chain lengths to such an extent that they would often not be long enough to support the degree of branching that is known to occur. In order to avoid this problem, we directly determined the length of B chains with mass spectrometry after treatment with β-amylase.

It is now possible to conceive of how α-glucans are generated in organisms such as *S. venezuelae* (Figure 11), which possesses the GlgE pathway and no other α-glucan pathway.[8] It is known that GlgE can initiate polymer synthesis without any primer, such as glycogenin or a malto-oligosaccharide. It does this by slowly hydrolyzing α-maltose 1-phosphate to give maltose, which in turn is slowly extended to give malto-tetraose, one of its preferred acceptor substrates.[3,19] As soon as the malto-oligosaccharide attains a length of ~16 glucosyl residues, GlgB will generate the first branch giving an A and a C chain of DP ~ 8−9. GlgE will then preferentially extend the A chain until it is long enough to undergo branching to generate a new A chain and the first B chain. The minimum length that seems to be transferred can occasionally be three but is more typically four or more *in vivo* and *in vitro* (Figure 4 and Figure 6). Each new A chain forms a branch as close as three residues from the branch point of the recipient chain, but more often on the fourth or fifth. By contrast, the *E. coli* enzyme has been reported to prefer the third residue for the first branch to be formed.[18] GlgE and GlgB will continue to extend and create new A chains, respectively. Normally, a newly formed B chain will not be long enough to allow a second branch to be introduced by GlgB. However, GlgE appears to occasionally extend a B chain and, when it attains sufficient length, GlgB introduces a second branch into this chain. These additional events ultimately generate the arboreal structure of the α-glucan. However, the introduction of a second branch in a B chain is not a frequent event because the mean number of branches per B chain is 1.2−1.6, considerably less than the value of 1.8−1.9 routinely observed in classical glycogens. Indeed the A:BC chain ratios for α-glucan are the smallest that have, to our knowledge, been reported.[29]

Importantly, the majority of chains in an α-glucan are B chains. Therefore, the overall mean linear chain DP of an α-
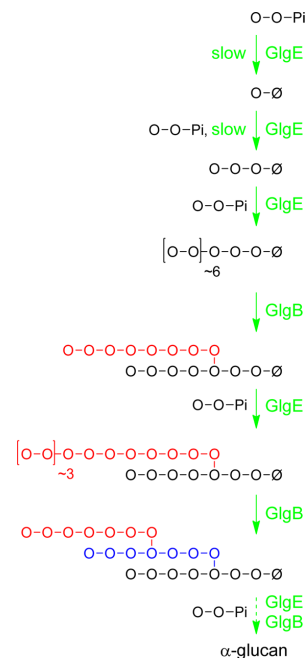
Figure 11. Scheme of α-glucan assembly by GlgE and GlgB. Linear molecules and C chains are shown in black, A chains in red, and B chains in blue. The Ø and Pi in the structures refer to the reducing end glucose residue and phosphate, respectively.

glucan must be principally defined by GlgB. This is because the length of each B chain is initially defined by the recipient chain length specificity of GlgB. We have established that *M. tuberculosis* GlgB creates C chains of DP ~ 9 Table 1 and Figure 5), which matches that observed in the polymer isolated from cells. It is reasonable to assume that it also generates B chains of a length that matches those observed in the polymer of ~7−8. If GlgE readily extended B chains, one would expect that the mean linear chain DP in the polymer would be much greater, but this is clearly not the case. Rather, it would appear that the occasional extension of individual B chains by GlgE simply leads to GlgB acting again, significantly restricting any potential to extend B chains within the polymer. Consistent with this, the branch lengths generated by GlgB *in vitro* are not obviously influenced by the presence or absence of GlgE in our *in vitro* assays.

GlgE is capable of not only extending malto-oligosaccharides with α-maltose 1-phosphate but also moving maltosyl units between them in a disproportionation reaction.[3,19] Perhaps this second activity is important to prevent any chains from being excessively extended because linear malto-oligosaccharides with a DP of >20 have limited solubility in aqueous conditions.

Having previously used a genetic approach to show that the GlgE pathway is both necessary and sufficient for the biosynthesis of α-glucan in *S. venezuelae*,[8] we have now shown that GlgE and GlgB are sufficient to generate the polymer from solely α-maltose 1-phosphate *in vitro*. Importantly, the α-glucans isolated from the cytosol and capsule of *M. tuberculosis* were very similar not only to each other, but also to that from *S. venezuelae*, which is derived exclusively from the

GlgE pathway. Although others have reported differences between the two mycobacterial materials,[15,16] such as in optical rotation, particle size, and hydrolytic susceptibility, these differences are subtle and could reflect differences in the way in which the materials are isolated. The similarity in the structures of the cytosolic and capsular materials implies that they could be synthesized solely by the GlgE pathway. It therefore remains to be established what role GlgA and Rv3032 have in the biosynthesis of α-glucan polymers in mycobacteria, despite genetic evidence suggesting they could have a direct role.[15] Furthermore, how a kDa-MDa dendrimer with a diameter of tens of nanometer is exported from a mycobacterial cell remains unclear, since secretion via vesicles appears to have been ruled out.[46]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.bio-chem.6b00209.

> NMR spectra of α-glucan, calibration curve for pulse field gradient NMR spectroscopy, UV/visible absorbance spectra of iodine complexes, cartoons of theoretical structures, ion mobility mass spectra, and kinetics and mass spectrometric analysis of polymer extension by GlgE (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: +44 1603 450741. E-mail: stephen.bornemann@jic.ac.uk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

DP, degree of polymerization of linear α-1,4-linked chains

■ REFERENCES

(1) Preiss, J. (2009) Glycogen biosynthesis, in *The Encyclopedia of Microbiology Vol. 5* (Schaechter, M., Ed.) pp 145−158, Elsevier, Oxford, U.K.

(2) Chandra, G., Chater, K. F., and Bornemann, S. (2011) Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology 157*, 1565−1572.

(3) Kalscheuer, R., Syson, K., Veeraraghavan, U., Weinrick, B., Biermann, K. E., Liu, Z., Sacchettini, J. C., Besra, G., Bornemann, S., and Jacobs, W. R. (2010) Self-poisoning of *Mycobacterium tuberculosis*

by targeting GlgE in an α-glucan pathway. *Nat. Chem. Biol. 6*, 376−384.

(4) Elbein, A. D., Pastuszak, I., Tackett, A. J., Wilson, T., and Pan, Y. T. (2010) Last step in the conversion of trehalose to glycogen: a mycobacterial enzyme that transfers maltose from maltose 1-phosphate to glycogen. *J. Biol. Chem. 285*, 9803−9812.

(5) Bornemann, S. (2016) α-Glucan biosynthesis and the GlgE pathway in *Mycobacterium tuberculosis*. *Biochem. Soc. Trans. 44*, 68−73.

(6) Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M., and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Eng., Des. Sel. 19*, 555−562.

(7) Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res. 42*, D490−495.

(8) Miah, F., Bibb, M. J., Barclay, J. E., Findlay, K. C., and Bornemann, S. (2016) Developmental delay in a *Streptomyces venezuelae glgE* null mutant is associated with the accumulation of α-maltose 1-phosphate, *Microbiology 162*, DOI: 10.1099/mic.0.000296.

(9) Jackson, M., and Brennan, P. J. (2009) Polymethylated polysaccharides from *Mycobacterium* species revisited. *J. Biol. Chem. 284*, 1949−1953.

(10) Lemassu, A., and Daffé, M. (1994) Structural features of the exocellular polysaccharides of *Mycobacterium tuberculosis*. *Biochem. J. 297*, 351−357.

(11) Schwebach, J. R., Glatman-Freedman, A., Gunther-Cummins, L., Dai, Z. D., Robbins, J. B., Schneerson, R., and Casadevall, A. (2002) Glucan is a component of the *Mycobacterium tuberculosis* surface that is expressed *in vitro* and *in vivo*. *Infect. Immun. 70*, 2566−2575.

(12) Cywes, C., Hoppe, H. C., Daffé, M., and Ehlers, M. R. W. (1997) Nonopsonic binding of *Mycobacterium tuberculosis* to complement receptor type 3 is mediated by capsular polysaccharides and is strain dependent. *Infect. Immun. 65*, 4258−4266.

(13) Gagliardi, M. C., Lemassu, A., Teloni, R., Mariotti, S., Sargentini, V., Pardini, M., Daffé, M., and Nisini, R. (2007) Cell wall-associated α-glucan is instrumental for *Mycobacterium tuberculosis* to block CD1 molecule expression and disable the function of dendritic cell derived from infected monocyte. *Cell. Microbiol. 9*, 2081−2092.

(14) Geurtsen, J., Chedammi, S., Mesters, J., Cot, M., Driessen, N. N., Sambou, T., Kakutani, R., Ummels, R., Maaskant, J., Takata, H., Baba, O., Terashima, T., Bovin, N., Vandenbroucke-Grauls, C. M. J. E., Nigou, J., Puzo, G., Lemassu, A., Daffé, M., and Appelmelk, B. J. (2009) Identification of mycobacterial α-glucan as a novel ligand for DC-SIGN: involvement of mycobacterial capsular polysaccharides in host imune modulation. *J. Immunol. 183*, 5221−5231.

(15) Sambou, T., Dinadayala, P., Stadthagen, G., Barilone, N., Bordat, Y., Constant, P., Levillain, F., Neyrolles, O., Gicquel, B., Lemassu, A., Daffé, M., and Jackson, M. (2008) Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: biosynthesis and impact on the persistence in mice. *Mol. Microbiol. 70*, 762−774.

(16) Dinadayala, P., Sambou, T., Daffé, M., and Lemassu, A. (2008) Comparative structural analyses of the α-glucan and glycogen from *Mycobacterium bovis*. *Glycobiology 18*, 502−508.

(17) Ortalo-Magné, A., Dupont, M. A., Lemassu, A., Andersen, A. B., Gounon, P., and Mamadou, D. (1995) Molecular composition of the outermost capsular material of the tubercle bacillus. *Microbiology 141*, 1609−1620.

(18) Sawada, T., Nakamura, Y., Ohdan, T., Saitoh, A., Francisco, P. B., Jr., Suzuki, E., Fujita, N., Shimonaga, T., Fujiwara, S., Tsuzuki, M., Colleoni, C., and Ball, S. (2014) Diversity of reaction characteristics of glucan branching enzymes and the fine structure of α-glucan from various sources. *Arch. Biochem. Biophys. 562*, 9−21.

(19) Syson, K., Stevenson, C. E. M., Rejzek, M., Fairhurst, S. A., Nair, A., Bruton, C. J., Field, R. A., Chater, K. F., Lawson, D. M., and Bornemann, S. (2011) Structure of a *Streptomyces* maltosyltransferase GlgE: a homologue of a genetically validated anti-tuberculosis target. *J. Biol. Chem. 286*, 38298−38310.

(20) Yam, K. C., D'Angelo, I., Kalscheuer, R., Zhu, H., Wang, J.-X., Snieckus, V., Ly, L. H., Converse, P. J., Jacobs, W. R., Jr., Strynadka, N.,

and Eltis, L. D. (2009) Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*. *PLoS Pathog. 5*, e1000344.

(21) Sani, M., Houben, E. N. G., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S. R., Jiménez, C. R., Daffé, M., Appelmelk, B. J., Bitter, W., van der Wel, N., and Peters, P. J. (2010) Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog. 6*, e1000794.

(22) Miah, F., Koliwer-Brandl, H., Rejzek, M., Field, R. A., Kalscheuer, R., and Bornemann, S. (2013) Flux through trehalose synthase flows from trehalose to the alpha anomer of maltose in mycobacteria. *Chem. Biol. 20*, 487−492.

(23) Miller, M. C., Klyosov, A., Platt, D., and Mayo, K. H. (2009) Using pulse field gradient NMR diffusion measurements to define molecular size distributions in glycan preparations. *Carbohydr. Res. 344*, 1205−1212.

(24) Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000) *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich, United Kingdom.

(25) Rashid, A. M., Saalbach, G., and Bornemann, S. (2014) Discrimination of large maltooligosaccharides from isobaric dextran and pullulan using ion mobility mass spectrometry. *Rapid Commun. Mass Spectrom. 28*, 191−199.

(26) McCleary, B. V., Gibson, T. S., and Mugford, D. C. (1997) Measurement of total starch in cereal products by amyloglucosidase-α-amylase method: collaborative study. *J. AOAC Int. 80*, 571−579.

(27) Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation and high-level expression by vectors containing the arabinose P_{BAD} promotor. *J. Bacteriol. 177*, 4121−4130.

(28) Syson, K., Stevenson, C. E. M., Rashid, A. M., Saalbach, G., Tang, M., Tuukkanen, A., Svergun, D. I., Withers, S. G., Lawson, D. M., and Bornemann, S. (2014) Structural insight into how *Streptomyces coelicolor* maltosyl transferase GlgE binds α-maltose 1-phosphate and forms a maltosyl-enzyme intermediate. *Biochemistry 53*, 2494−2504.

(29) Manners, D. J. (1991) Recent developments in our understanding of glycogen structure. *Carbohydr. Polym. 16*, 37−82.

(30) Dinadayala, P., Lemassu, A., Granovski, P., Cérantola, S., Winter, N., and Daffé, M. (2004) Revisiting the structure of the anti-neoplastic glucans of *Mycobacterium bovis* Bacille Calmette-Guérin - Structural analysis of the extracellular and boiling water extract-derived glucans of the vaccine substrains. *J. Biol. Chem. 279*, 12369−12378.

(31) Bittencourt, V. C. B., Figueiredo, R. T., da Silva, R. B., Mourao-Sa, D. S., Fernandez, P. L., Sassaki, G. L., Mulloy, B., Bozza, M. T., and Barreto-Bergter, E. (2006) An α-glucan of *Pseudallescheria boydii* is involved in fungal phagocytosis and toll-like receptor activation. *J. Biol. Chem. 281*, 22614−22623.

(32) Tan, I., Flanagan, B. M., Halley, P. J., Whittaker, A. K., and Gidley, M. J. (2007) A method for estimating the nature and relative proportions of amorphous, single, and double-helical components in starch granules by C^{-13} CP/MAS NMR. *Biomacromolecules 8*, 885−891.

(33) Ryu, J.-H., Drain, J., Kim, J. H., McGee, S., Gray-Weale, A., Waddington, L., Parker, G. J., Hargreaves, M., Yoo, S.-H., and Stapleton, D. (2009) Comparative structural analyses of purified glycogen particles from rat liver, human skeletal muscle and commercial preparations. *Int. J. Biol. Macromol. 45*, 478−482.

(34) Antoine, A. D., and Tepper, B. S. (1969) Characterization of glycogens from mycobacteria. *Arch. Biochem. Biophys. 134*, 207−213.

(35) van de Weerd, R., Berbís, M. A., Sparrius, M., Maaskant, J. J., Boot, M., Paauw, N. J., de Vries, N., Boon, L., Baba, O., Cañada, F. J., Geurtsen, J., Jiménez-Barbero, J., and Appelmelk, B. J. (2015) A murine monoclonal antibody to glycogen: characterization of epitope-fine specificity by saturation transfer difference (STD) NMR spectroscopy and its use in mycobacterial capsular α-glucan research. *ChemBioChem 16*, 977−989.

(36) Misaki, A., and Yukawa, S. (1966) Studies on cell walls of *Mycobacteria*. II. Constitution of polysaccharides from BCG cell walls. *J. Biochem.(Tokyo) 59*, 511−520.

(37) Bailey, J. M., and Whelan, W. J. (1961) Physical properties of starch. 1. Relationship between iodine stain and chain length. *J. Biol. Chem. 236*, 969−973.

(38) Summer, R., and French, D. (1956) Action of β-amylase on branched oligosaccharides. *J. Biol. Chem. 222*, 469−477.

(39) Garg, S. K., Alam, M. S., Kishan, K. V. R., and Agrawal, P. (2007) Expression and characterization of α-(1,4)-glucan branching enzyme Rv1326c of *Mycobacterium tuberculosis* H37Rv. *Protein Expression Purif. 51*, 198−208.

(40) Guan, H. P., Li, P., ImparlRadosevich, J., Preiss, J., and Keeling, P. (1997) Comparing the properties of *Escherichia coli* branching enzyme and maize branching enzyme. *Arch. Biochem. Biophys. 342*, 92−98.

(41) Pal, K., Kumar, S., Sharma, S., Garg, S. K., Alam, M. S., Xu, H. E., Agrawal, P., and Swaminathan, K. (2010) Crystal structure of full length *Mycobacterium tuberculosis* H37Rv glycogen branching enzyme: insights of N-terminal β-sandwich in substrate specificity and enzymatic activity. *J. Biol. Chem. 285*, 20897−20903.

(42) Hernández, J. M., Gaborieau, M., Castignolles, P., Gidley, M. J., Myers, A. M., and Gilbert, R. G. (2008) Mechanistic investigation of a starch-branching enzyme using hydrodynamic volume SEC analysis. *Biomacromolecules 9*, 954−965.

(43) Roussel, X., Lancelon-Pin, C., Vikso-Nielsen, A., Rolland-Sabaté, A., Grimaud, F., Potocki-Véronèse, G., Buléon, A., Putaux, J.-L., and D'Hulst, C. (2013) Characterization of substrate and product specificity of the purified recombinant glycogen branching enzyme of *Rhodothermus obamensis*. *Biochim. Biophys. Acta, Gen. Subj. 1830*, 2167−2177.

(44) Jo, H.-J., Park, S., Jeong, H.-G., Kim, J.-W., and Park, J.-T. (2015) *Vibrio vulnificus* glycogen branching enzyme preferentially transfers very short chains: N1 domain determines the chain length transferred. *FEBS Lett. 589*, 1089−1094.

(45) Palomo, M., Kralj, S., van der Maarel, M. J. E. C., and Dijkhuizen, L. (2009) The unique branching patterns of *Deinococcus* glycogen branching enzymes are determined by their N-terminal domains. *Appl. Environ. Microbiol. 75*, 1355−1362.

(46) Prados-Rosales, R., Baena, A., Martinez, L. R., Luque-Garcia, J., Kalscheuer, R., Veeraraghavan, U., Camara, C., Nosanchuk, J. D., Besra, G. S., Chen, B., Jimenez, J., Glatman-Freedman, A., Jacobs, W. R., Jr., Porcelli, S. A., and Casadevall, A. (2011) Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice. *J. Clin. Invest. 121*, 1471−1483.

7.2 Appendix 2: Supplementary X-ray crystallography data

MsGlgB structures presented in Chapter 2

Dataset ID	Apo-WT-MsGlgB	DP4-WT-MsGlgB	DP8-D416A-MsGlgB	ACR-E469A-MsGlgB	DP3-ACR-MsGlgB
Data Collection					
Beamline	I04-1	I03	I04-1	I03	I04
Wavelength (Å)	0.9282	0.9763	0.9282	0.9762	0.9795
Detector	Pilatus 6M	Pilatus3 6 M	Pilatus 6M	Pilatus3 6 M	Pilatus 6M
Resolution range (Å)	84.49-1.72 (1.75-1.72)	54.10-1.46 (1.48-1.46)	84.7-1.88 (1.92-1.88)	48.50-1.60 (1.63-1.60)	62.28-1.45 (1.47-1.45)
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
a, b, c (Å)	98.49, 164.41, 54.52	97.36, 162.31, 53.75	98.79, 164.57, 54.42	97.00, 162.04, 53.64	97.67, 161.69, 53.79
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Total observations	912,423 (46,131)	1,949,958 (86,737)	972,120 (60,979)	729,137 (37,025)	2,005,551 (97,080)
Unique reflections	94,955 (4,676)	148,349 (7219)	73,128 (4,408)	111,249 (5,452)	151,467 (7,404)
Multiplicity	9.6 (9.9)	13.1 (12.0)	13.3 (13.8)	6.6 (6.8)	13.2 (13.1)
Mean $I/\sigma(I)$	9.6 (1.2)	16.3 (1.2)	11.3 (1.5)	9.7 (1.3)	16.8 (1.4)
Completeness (%)	100.0 (100.00)	100.0 (99.9)	100.0 (100.0)	99.2 (98.8)	100.0 (99.9)
R_{merge}[a]	0.152 (1.846)	0.073 (2.015)	0.184 (1.851)	0.095 (1.688)	0.082 (1.779)
R_{meas}[b]	0.161 (1.949)	0.076 (2.105)	0.192 (1.922)	0.104 (1.828)	0.086 (1.874)
$CC_{1/2}$[c]	0.998 (0.606)	1.000 (0.738)	0.998 (0.753)	0.998 (0.463)	1.000 (0.685)
Wilson B value (Å^2)	20.7	21.0	25.0	22.0	25.0
Refinement					
Resolution range (Å)	84.49-1.72 (1.75-1.72)	54.10-1.46 (1.48-1.46)	84.7-1.88 (1.92-1.88)	48.50-1.60 (1.63-1.60)	62.28-1.45 (1.47-1.45)
Reflections: working/free[d]	90,195/4,684	140,775/7,483	69,343/3,715	105,686/5501	143,858/7,518
Final R_{work}	0.164	0.125	0.173	0.176	0.124
Final R_{free}[e]	0.186	0.160	0.206	0.192	0.161
Estimated coordinate error (Å)[f]	0.089	0.052	0.125	0.078	0.054
r.m.s bond distance deviation (Å)	0.0152	0.0103	0.0102	0.0102	0.0099
r.m.s. bond angle deviation (°)	1.511	1.411	1.393	1.394	1.484
No. of protein residues per chain (ranges)	723 (9-375; 380-736)	723 (9-375; 380-736)	722 (9-94; 99-736)	722 (9-94; 101-736)	722 (9-94; 100-736)
No. of heterogen residues: sugar/water/other[g]	0/759/9	4/737/9	35/636/6	13/411/15	17/901/18
Mean B factors: protein/ligands/water/overall (Å^2)	25/34/37/25	27/36/41/27	13/42/40/30	32/41/40/32	23/35/39/23
Ramachandran: favoured/allowed/disallowed[h] (%)	97.6/2.3/0.1	98.2/1.7/0.1	98.5/1.4/0.1	97.8/2.1/0.1	98.0/1.9/0.1
Protein Data Bank accession code	N/A	N/A	N/A	N/A	N/A

[a] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

[b] $R_{meas} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl, $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl.

[c] $CC_{1/2}$ is the correlation coefficient between symmetry equivalent intensities from random halves of the data set.

[d] The data set was split into "working" and "free" sets consisting of 95 and 5% of the data, respectively. The free set was not used for refinement.

[e] The R factors R_{work} and R_{free} were calculated as follows: $R = \sum(| F_{obs} - F_{calc} |)/\sum| F_{obs} |$ where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[f] Based on R_{free} as calculated by REFMAC5.[1]

[g] "Other" includes imidazole, glycerol, sodium ions, nickel and chloride ions.

[h] As calculated using MolProbity.[2]

Summary of validation reports

Apo-WT-MsGlgB
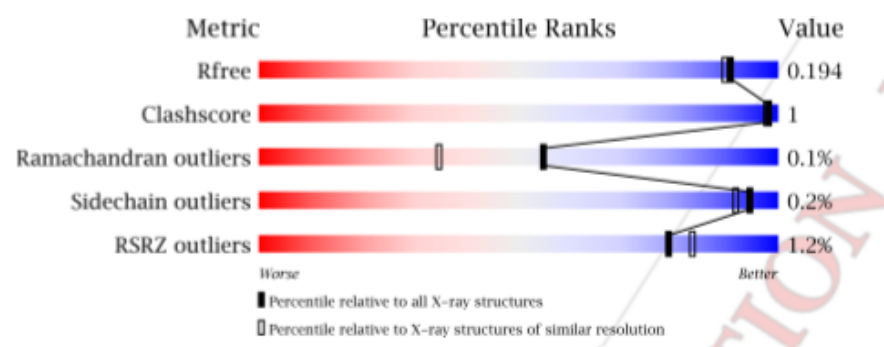
DP4-WT-MsGlgB
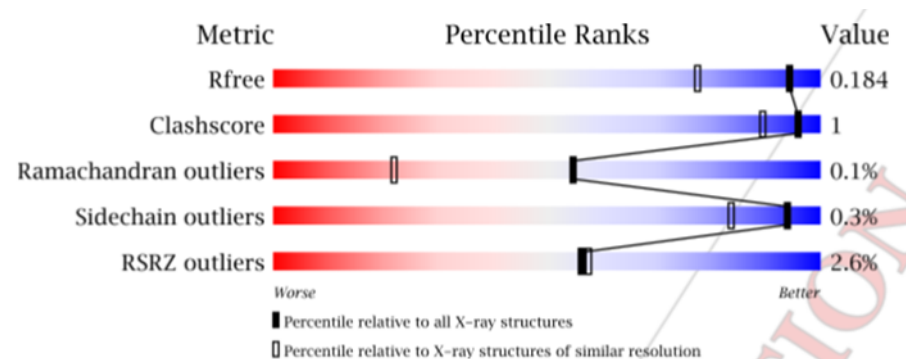
PRIVATEER

Name	Chain	Q^1	Phi	Theta	Anomer	D/L^2	Conformation	$RSCC^3$	<Bfactor>	Diagnostic
GLC	A	0.511	219.68	8.61735	alpha	D	$^4C_1$	0.80	35.075	Ok
GLC	A	0.587	350.509	10.8085	alpha	D	$^4C_1$	0.78	45.1464	Ok
GLC	A	0.604	27.5162	7.86233	alpha	D	$^4C_1$	0.73	36.9482	Ok
GLC	A	0.570	154.594	3.29564	alpha	D	$^4C_1$	0.77	33.7645	Ok

DP8-D416A-MsGlgB
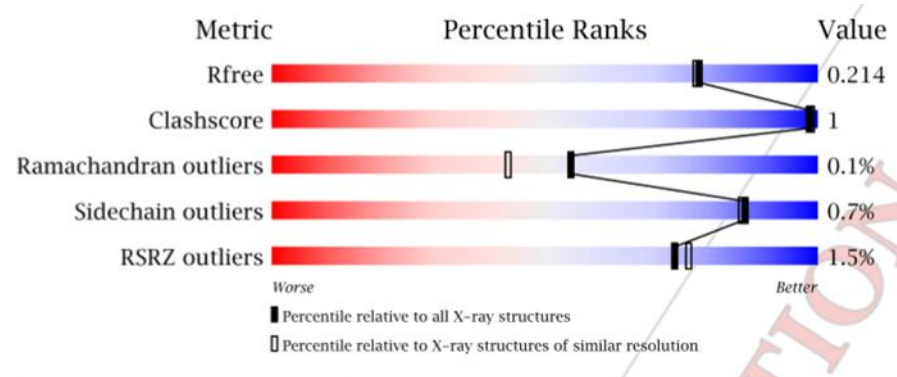
PRIVATEER

Name	Chain	Q^1	Phi	Theta	Anomer	D/L^2	Conformation	$RSCC^3$	<Bfactor>	Diagnostic
BGC	A	0.595	86.8336	9.67867	beta	D	$^4C_1$	0.75	18.565	Ok
GLC	A	0.577	288.799	5.03079	alpha	D	$^4C_1$	0.85	17.8491	Ok
GLC	A	0.565	172.664	3.55629	alpha	D	$^4C_1$	0.86	18.9373	Ok
GLC	A	0.571	345.327	0.604983	alpha	D	$^4C_1$	0.82	20.1836	Ok
GLC	A	0.559	37.067	5.87266	alpha	D	$^4C_1$	0.87	20.3118	Ok
GLC	A	0.582	29.6918	8.86196	alpha	D	$^4C_1$	0.90	18.7455	Ok
GLC	A	0.581	5.51538	11.7087	alpha	D	$^4C_1$	0.73	19.2773	Ok
GLC	A	0.581	95.0614	11.6734	alpha	D	$^4C_1$	0.72	20.1064	Ok
GLC	A	0.559	243.453	7.18061	alpha	D	$^4C_1$	0.70	65.6225	Ok
GLC	A	0.573	42.9558	5.21885	alpha	D	$^4C_1$	0.85	49.8527	Ok
GLC	A	0.556	158.581	2.90187	alpha	D	$^4C_1$	0.88	41.69	Ok
GLC	A	0.574	87.9007	6.83329	alpha	D	$^4C_1$	0.87	34.24	Ok

GLC	A	0.548	43.0559	4.58215	alpha	D	$^4C_1$	0.86	33.9445	Ok
GLC	A	0.546	280.702	2.91125	alpha	D	$^4C_1$	0.83	33.8973	Ok
GLC	A	0.530	278.637	6.8716	alpha	D	$^4C_1$	0.69	56.9383	Ok
GLC	A	0.569	193.402	3.0405	alpha	D	$^4C_1$	0.82	49.7664	Ok
GLC	A	0.574	274.484	3.83747	alpha	D	$^4C_1$	0.76	49.6091	Ok
GLC	A	0.545	76.6182	1.14197	alpha	D	$^4C_1$	0.79	46.6155	Ok
GLC	A	0.555	91.4727	0.83588	alpha	D	$^4C_1$	0.82	43.8091	Ok
GLC	A	0.568	104.849	1.68537	alpha	D	$^4C_1$	0.85	39.6327	Ok
GLC	A	0.558	351.358	5.60606	alpha	D	$^4C_1$	0.87	33.2055	Ok
GLC	A	0.548	260.225	6.13912	alpha	D	$^4C_1$	0.82	46.14	Ok
GLC	A	0.543	79.6826	5.80724	alpha	D	$^4C_1$	0.83	37.6573	Ok
GLC	A	0.533	43.4975	3.52916	alpha	D	$^4C_1$	0.85	45.2236	Ok
GLC	A	0.519	287.537	8.01385	alpha	D	$^4C_1$	0.69	59.5067	Ok
GLC	A	0.550	231.544	2.51996	alpha	D	$^4C_1$	0.66	57.8264	Ok
GLC	A	0.547	200.239	3.03789	alpha	D	$^4C_1$	0.73	78.6817	Ok
GLC	A	0.542	358.445	6.75598	alpha	D	$^4C_1$	0.85	55.0145	Ok
GLC	A	0.547	34.2687	7.86902	alpha	D	$^4C_1$	0.90	39.2318	Ok
GLC	A	0.541	46.4177	6.40408	alpha	D	$^4C_1$	0.91	29.5955	Ok
GLC	A	0.549	90.9705	6.11333	alpha	D	$^4C_1$	0.89	30.9273	Ok
GLC	A	0.532	20.7939	8.77359	alpha	D	$^4C_1$	0.81	46.4873	Ok
BGC	A	0.532	36.4979	9.7292	beta	D	$^4C_1$	0.59	69.5133	Ok
BGC	A	0.559	42.3375	4.5499	beta	D	$^4C_1$	0.55	51.755	Ok
GLC	A	0.556	220.545	14.5887	alpha	D	$^4C_1$	0.71	58.0975	Ok

ACR-E469A-MsGlgB

Metric	Percentile Ranks	Value
Rfree		0.201
Clashscore		1
Ramachandran outliers		0.1%
Sidechain outliers		0.9%
RSRZ outliers		6.9%

Worse — Better

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution
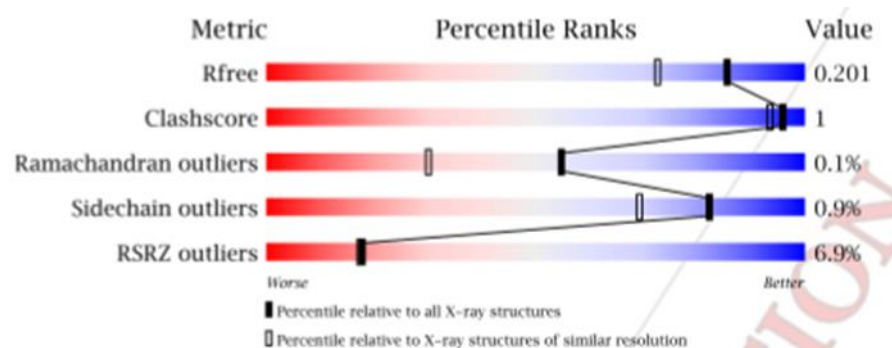
DP3-ACR-MsGlgB
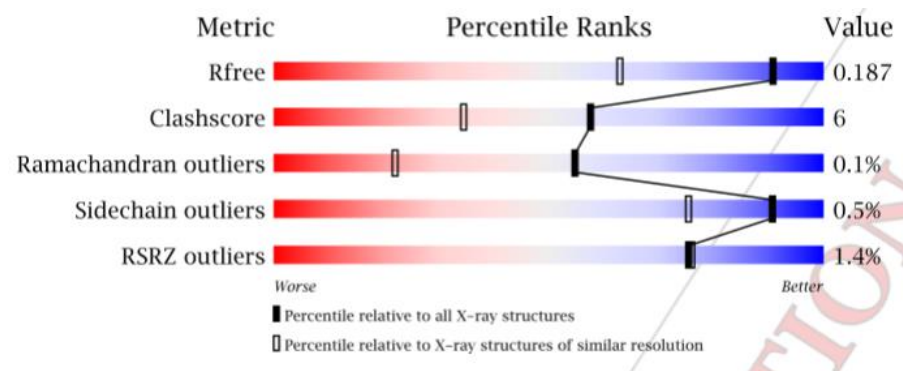
PRIVATEER

Name	Chain	Q[1]	Phi	Theta	Anomer	D/L[2]	Conformation	RSCC[3]	<Bfactor>	Diagnostic
GLC	A	0.532	108.486	8.39016	alpha	D	$^4C_1$	0.93	22.9991	Ok
GLC	A	0.567	87.365	5.27104	alpha	D	$^4C_1$	0.89	35.0673	Ok
GLC	A	0.558	189.226	9.81099	alpha	D	$^4C_1$	0.75	70.8918	Ok
GLC	A	0.533	333.195	4.05168	alpha	D	$^4C_1$	0.69	50.8375	Ok
GLC	A	0.573	13.354	5.6516	alpha	D	$^4C_1$	0.75	40.2191	Ok
GLC	A	0.558	39.973	10.6909	alpha	D	$^4C_1$	0.90	40.9055	Ok
GLC	A	0.539	33.6354	5.43116	alpha	D	$^4C_1$	0.91	29.4827	Ok
GLC	A	0.526	105.007	2.77795	alpha	D	$^4C_1$	0.92	28.2291	Ok
GLC	A	0.497	89.4325	22.4068	alpha	D	$^4C_1$	0.73	40.6618	Ok
GLC	A	0.573	304.065	5.91823	alpha	D	$^4C_1$	0.49	50.9383	Ok
GLC	A	0.589	151.759	9.08439	alpha	D	$^4C_1$	0.73	40.1382	Ok
GLC	A	0.550	137.474	7.77946	alpha	D	$^4C_1$	0.82	25.9327	Ok
GLC	A	0.627	36.7132	6.8425	alpha	D	$^4C_1$	0.88	20.67	Ok
GLC	A	0.561	80.0154	6.16169	alpha	D	$^4C_1$	0.86	19.2382	Ok
GLC	A	0.591	321.059	7.66566	alpha	D	$^4C_1$	0.85	22.2482	Ok
GLC	A	0.537	278.175	21.1158	alpha	D	$^4C_1$	0.75	29.3183	Ok

MsGlgB structures that require further refinement

Dataset ID	Apo-D416A-MsGlgB	Apo-E469A-MsGlgB	Apo-D416A-E469A-MsGlgB	DP8-D416A-E469A-MsGlgB
Data Collection				
Beamline	I04-1	I04-1	I04-1	I04-1
Wavelength (Å)	0.9282	0.9282	0.9282	0.9282
Detector	Pilatus 6M	Pilatus 6 M	Pilatus 6 M	Pilatus 6 M
Resolution range (Å)	84.67-1.68 (1.71-1.68)	98.41-1.67 (1.70-1.67)	98.21-1.47 (1.50-1.47)	84.43-1.42 (1.44-1.42)
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
a, b, c (Å)	98.72, 164.70, 54.55	98.41, 163.63, 54.26	98.23, 164.18, 54.46	98.48, 163.97, 54.23
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Total observations	1,358,284 (69,014)	1,362,173 (67,063)	1,991,857 (91,830)	1846138 (90948)
Unique reflections	102,252 (5,037)	102,557 (5,053)	150,312 (7,375)	166069 (8084)
Multiplicity	13.3 (13.7)	13.3 (13.3)	13.3 (12.5)	11.1 (11.3)
Mean $I/\sigma(I)$	11.0 (1.4)	15.1 (1.4)	15.6 (1.4)	13.6 (1.2)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (99.9)
R_{merge}[a]	0.160 (1.932)	0.111 (1.849)	0.089 (1.876)	0.091 (1.944)
R_{meas}[b]	0.167 (2.007)	0.116 (1.923)	0.092 (1.956)	0.095 (2.038)
$CC_{1/2}$[c]	0.998 (0.742)	0.999 (0.725)	0.999 (0.727)	0.999 (0.720)
Wilson B value (Å²)	17.1	21.0	17.3	15.7
Refinement				
Resolution range (Å)	84.67-1.68 (1.71-1.68)	98.41-1.67 (1.70-1.67)	98.21-1.47 (1.50-1.47)	84.43-1.42 (1.44-1.42)
Reflections: working/free[d]	97,047/5,129	97,374 (5,106)	142,772 (7,451)	157,667 (8,310)
Final R_{work}	0.169	0.161	0.132	0.127
Final R_{free}[e]	0.190	0.194	0.174	0.171
Estimated coordinate error (Å)[f]	0.085	0.089	0.058	0.052
r.m.s bond distance deviation (Å)	0.0147	0.0197	0.0274	0.0287
r.m.s. bond angle deviation (°)	1.475	1.782	2.172	2.361
No. of protein residues per chain (ranges)	722 (9-375; 381-736)			
No. of heterogen residues: sugar/water/other[g]	0/748/8			
Mean B factors: protein/ligands/water/overall (Å²)	23/34/34/23			
Ramachandran: favoured/allowed/disallowed[h] (%)	98/1.9/0.1			
Protein Data Bank accession code	N/A	N/A	N/A	N/A

[a] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

[b] $R_{meas} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl, $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl.

[c] $CC_{1/2}$ is the correlation coefficient between symmetry equivalent intensities from random halves of the data set.

[d] The data set was split into "working" and "free" sets consisting of 95 and 5% of the data, respectively. The free set was not used for refinement.

[e] The R factors R_{work} and R_{free} were calculated as follows: $R = \sum(|F_{obs} - F_{calc}|)/\sum|F_{obs}|$ where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[f] Based on R_{free} as calculated by REFMAC5.[1]

[g] "Other" includes glycerol and sodium ions

[h] As calculated using MolProbity.[2]