D-Arabinan hydrolysis and its roles in Mycobacteriales biology

By

Samuel Benedict

A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences
College of Life and Environmental Sciences
University of Birmingham

November 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

1 Acknowledgements

I would like to thank Dr. Patrick Moynihan for his constant support, advice, and belief in me during my PhD. I do not think I would have completed without his guidance.

Furthermore, I would like to thank all members of the Moynihan laboratory for their friendship and advice, being surrounded by such brilliant scientists and wonderful people has made the experience of doing a PhD during a pandemic so much more enjoyable.

I would also like Dr. Elisabeth Lowe for both her amazing glycobiology knowledge, mentorship, and friendship. Furthermore, I would like to thank Prof. Spencer Williams and Prof Todd Lowary for their provision of the weird sugars, glycosides, and other synthetic analogues needed to make this project possible.

I would like to thank my family, in particular both of my parents, both for supporting me through my education, and for finally making me believe in myself enough. Finally, I want to thank my partner, Ryan, for dealing with my late-night structural biology-induced panics, and for his gentle, constant support.

I also want to thank my late friend, Olly Beswetherick, for instilling in me a love of science that has carried me through this whole process. Even though you're not physically here to see me finally finish this PhD, having you in my memories has been such a source of strength, even though I miss you like mad.

Table of Contents

1 Acknowledgements 2

2 Glossary 9

3 Introduction 11

3.1 Microbial glycobiology 11

3.2 CMNR group organisms 11

3.3 Mycobacteria 12

3.4 Mycobacterial disease 13

3.5 Structure of the mycobacterial cell wall 16

3.6 Lipids of mycobacteria 19

3.7 Biogenesis of the mycobacterial cell wall 21

3.7.1 Synthesis of Peptidoglycan 21

3.7.2 Synthesis of arabinogalactan 23

3.7.3 Synthesis of mycolic acids 26

3.7.4 Synthesis of lipoarabinomannan 28

3.8 Cell wall turnover in mycobacteria 30

3.8.1 Glycoside hydrolases 34

3.9 CAZymes in the mycobacterial genome 38

3.10 Bacteriodetes spp. 38

4 Summary of findings 41

5 Materials and Methods 43

5.1 Strains and Growth conditions 43

5.2 Sauton's media preparation 44

5.3 Plasmids and primers used in this study 44

5.4 SDS-PAGE gel electrophoresis 45

5.5 Agarose Gel electrophoresis 46

5.6 Restriction-free cloning 46

5.7 Q5 site-directed mutagenesis 47

5.8 Induction and protein production 48

5.9 Protein buffer preparation 48

5.10 Protein purification by Immobilised Metal Affinity Chromatography 49

5.11 Preparative size exclusion Chromatography 49

5.12 Protein Crystallography 50

5.13 X-ray diffraction data collection 50

5.13.1 Crystallography of Rv3707c 50

5.13.2 Crystallography of *N. brasiliensis* GH172 50

5.13.3 Crystallography of MAB_{GH183} 51

5.13.4 Crystallography of $MAP_{complex}$ 51

5.13.5 Phasing and structural refinement 52

5.14 Mycobacterium smegmatis cell wall preparation 52

5.15 Mycobacterial glycolipid extraction 53

5.16 *Pseudomonas aeruginosa* PA1596 pilin glycan preparation 53

5.17 Porous graphitic carbon chromatography 54

5.18 Polysaccharide enzymatic digestion assays 54

5.19 Thin layer chromatography of carbohydrates 54

5.20 High performance anion chromatography with pulsed amperometric detection (IC-PAD) 55

5.21 Isothermal titration calorimetry 55

5.22 Analytical size exclusion chromatography of protein complexes.....56

5.23 Nocardia species carbon source assay 56

5.24 Spot dilution assays.....57

5.25 Proteomics 57

6 Recombinant expression and purification of endogenous GH183 enzymes from Mycobacteriales 59

6.1 Contribution statement.....59

***6.2* Identification of putative GH183 enzymes within the genome of *M. tuberculosis* 59**

6.3 Recombinant expression of Rv3707c 63

6.4 Identification of full length Rv3707c 65

6.5 Removal of proline-rich region does not stabilise Rv3707c in solution 67

6.6 Rv3707c can be produced recombinantly using a 30 amino acid N-terminal truncation 69

6.7 Large-scale production of Rv3707c_P30M 71

6.8 Identification of MSMEG_2107.....73

6.9 Recombinant expression and purification of MSMEG_2107.....75

6.10 Rv3707c_P30M is hydrolytic against D-arabinan-containing polysaccharides 77

.....78

6.11 PGC-SPE allows removal of endogenous arabinanase reaction products.....79

6.12 Optimisation of PGC-SPE to allow IC-PAD analysis of reaction products 81

6.13 MSMEG_2107 possesses a restricted substrate preference compared to Rv3707c 83

6.14 Rv3707c is a five-bladed beta propeller glycoside hydrolase 85

6.15 Structural comparison of Rv3707c reveals a conserved active site across DUF4185 enzymes 89

6.16 Structural analysis successfully identifies catalytic residues in Rv3707c 92

6.17 Discussion 95

7 Biochemical characterisation of a putative arabinogalactan pathway in Nocardia ... 99

7.1 Contribution statement 99

7.2 Introduction 99

7.2.1 Central carbon metabolism in CMNR group bacteria 99

7.2.2 Pentose Phosphate Pathway 100

7.3 Results 103

7.3.1 Identification of homologs of HMPREF9455_02467 in *Nocardia* species 103

7.3.2 Syntenic analysis of the *N. brasiliensis* DUF2961 locus suggests a route for D-arabinan recycling. 107

7.3.3 Purification of DUF2961-containing protein O3I_017420 from *N. brasiliensis* 110

7.3.4 $NBras_{Gh172}$ is an α-1,5 exo-D-arabinofuranosidase 113

7.3.5 $NBras_{Gh172}$ is a retaining glycoside hydrolase. 116

7.3.6 $NBras_{GH172}$ is trimeric 118

7.3.7 Multimerisation state of GH_{172} differs across family members 124

7.3.8 Active site architecture of $NBRAS_{Gh172}$ 126

7.3.9 Michaelis-Menten kinetics of $NBras_{GH172}$ 129

7.3.10 *Nocardia* species are capable of growth on arabinogalactan as the sole carbon source 131

7.3.11 Purification of *N. brasiliensis* metabolic enzymes. 133

7.3.12 FGGY Kinases 1 and 2 have differing substrate specificities 135
7.3.13 Identification of a putative arabinose epimerase 137
7.3.14 Characterisation of *N. brasiliensis* putative GH43 139
7.3.15 Mass spectrometry proteomics of *N. brasiliensis* 141
7.3.16 Additional lysis step is required for representative proteomic extraction. 146
7.3.17 Improved proteomic analysis of *N. brasiliensis* 148
7.4 Discussion 151
8 Identification of a putative GH183 Type 7 effector-immunity pair in Mycobacteriales 155
8.1 Introduction 155
8.1.1 Interbacterial competition 155
8.1.2 Type VII secretion in Mycobacteriales species 157
8.1.3 Mechanism of T7 secretion in mycobacteria 158
8.1.4 Identified T7SS toxins 159
8.2 Results 161
8.2.1 Count of GH183 enzymes across Mycobacteriales indicates gene duplication. 161
8.2.2 Alphafold analysis indicates possible type 7 secretion of MAP_{GH183} and MAB_{GH183} 162
8.2.3 Purification of MAP_{GH183} and MAB_{GH183} 166
8.2.4 MAB_{GH183} is a retaining glycoside hydrolase 168
8.2.5 Structure of MAB_{GH183} 170
8.2.6 Identification of small ORFs upstream of MAB_{GH183} and MAP_{GH183} suggest formation of a complex 174
8.2.7 Purification of $MAB_{Inhibitor}$ and $MAP_{inhibitor}$ proteins 176
8.2.8 MAB_{GH183} and MAP_{GH183} form complexes with their cognate inhibitors 178
8.2.9 Complexes formed between MAB_{GH183},MAP_{GH183}, and upstream proteins are inhibitory 180
8.2.10 MAP_{GH183} forms an extremely high affinity complex with its inhibitor protein 182
8.2.11 Structure of MAP_{GH183} in complex with its inhibitor protein 184

8.2.12 Structure of $MAP_{inhibitor}$ 189

8.2.13 $MAP_{complex}$ is a very closely association of the two proteins 191

8.2.14 Mechanism of inhibition 193

8.2.15 Cross species inhibition is prevented by subtle conformational changes in the GH183 protein 195

8.2.16 Exogenous expression of MAB_{GH183} in *M. smegmatis* leads to toxicity 197

8.3 Discussion 199

9 Discussion 204

10 References 209

2 Glossary

Arabinogalactan- a complex polysaccharide consisting of glycosidically-linked arabinose and galactose subunits. Two classes of arabinogalactan exist, derived from either plants or microbes, with differing structural characteristics.

Arabinose- A 5-carbon monosaccharide, widespread in plant cell walls as the L-isomer, but also found more rarely in the D-isomerisation state in the cell wall of Mycobacteriales and a small number of other bacteria.

Cytosine- One of four nucleobases found in DNA. Forms 3 hydrogen bonds with guanine

Furanose- a conformation adopted by aldose monosaccharides composed of a 5-membered ring system and at least one hydroxyl outside of the ring.

Galactose- a hexose monosaccharide found throughout nature in hemicellulose of plants and as a ubiquitous component of protein glcyosylations in Eukaryotes and prokaryotes.

GlcNAc- an acetylated derivative of glucose, and one of the two monosaccharide components of the peptidoglycan backbone.

Glycan- a compound consisting of glycosidically-linked monosaccharide subunits.

Guanine- One of four nucleobases found in DNA. Forms 3 hydrogen bonds with cytosine

Hexose- A monosaccharide consisting of 6 carbon atoms .

Leprosy- The disease caused by the bacterium *Mycobacterium leprae*, it is characterised by thickening of skin in reddish-pink lesions, loss of sensation in the hands and feet, and cartilage damage.

Lipoarabinomannan- a lipid-linked glycan found in the cell wall of mycobacteria, responsible for providing cell wall rigidity and immune interactions with the host.

Mannose- hexose monosaccharide

MurNAc- an n-acetylated derivative of muramic acid, is one of the two monosaccharides found in peptidoglycan. Is the attachment point for peptide stems which crosslink the glycan strands of peptidoglycan.

Mycobacteriales- A phylogenetic order of bacteria consisting of *Corynebacteriaceae, Dietziaceae, Gordoniaceae, Lawsonellaceae, Mycobacteriaceae, Segniliparaceae, Tsukamarellaceae.*

Pentose- A monosaccharide consisting of 5 carbon atoms

Peptidoglycan- A rigid polysaccharide conjugated with crosslinked peptide strands that provides mechanical rigidity and support against osmotic lysis to the bacterial cell wall.

Pyranose- A conformation adopted by aldose monosaccharides composed of a 6-membered ring system. The ring system consists of the anomeric oxygen and 5 carbon atoms.

Tuberculosis- The disease caused by the bacterium *Mycobacterium tuberculosis*, active disease is indicated by chest pains, bloody sputum, fever, chills and weight loss.

3 Introduction

3.1 Microbial glycobiology

Glycobiology is the study of the contributions of sugar-containing molecules to the lifestyles of living organisms. Glycans are ubiquitous in nature and have been found to play pivotal roles in diverse fields. They act as virulence factors in bacteria, viruses, and eukaryotic pathogens such as *Plasmodium spp.* (Pearce, 2018; Behrens *et al.*, 2016; Ben Ami Pilo *et al.*, 2022; Ocaña-Ortega *et al.*, 2023). Furthermore, glycans function as storage molecules for multicellular lifeforms, and as structural motifs, providing rigidity and support to a huge range of organisms (Bisson-Filho *et al.*, 2017; Mills *et al.*, 2004; Mistretta *et al.*, 2019). Until recently, studying sugars *in vivo* has been difficult. However, with advances in chemical biology techniques such as click chemistry, the mechanisms by which these sugars exert their phenotypic effects, both on the host and the pathogen have begun to be elucidated (Kamariza *et al.*, 2018; Swarts *et al.*, 2012). Studying the glycobiology of mycobacterial pathogens represents a particularly exciting avenue of research due to the diverse contributions of glycans to their lifestyle, governing their cell wall ultrastructure, cellular metabolism, and immune interactions. In this project we initially aimed to explore the importance of novel D-arabinan hydrolases in mycobacteria, but rapidly their importance across Mycobacteriales species became clear.

3.2 CMNR group organisms

The *Corynebacterium-Mycobacterium-Nocardia-Rhodococcus* group of bacteria are a group of organisms found within the order Mycobacteriales (Tindall, 2019; Arahal *et al.*, 2022). This order contains a wide variety of important soil-dwelling bacteria, as well as a significant

human pathogens including *Corynebacterium diphtherieae, Nocardia asterioides,* and *Mycobacterium tuberculosis* (Hoskisson, 2018; Beaman and Beaman, 1994). These organisms are differentiated from other bacteria through their unique cell wall ultrastructure, making them resistant to acid destaining in Gram-staining, hence them sometimes being referred to as acid-fast organisms (Minnikin and Goodfellow, 1980). This phenotype is caused by exceptional retention of lipophilic molecules due to the hydrophobic mycolic acids in the mycobacterial outer membrane which are discussed later.

3.3 Mycobacteria

The genus *Mycobacterium* represents a huge global health burden. In 2021, *Mycobacterium tuberculosis,* the most prominent pathogen from the genus, caused an estimated 1.78 million deaths (WHO, 2022). In addition to this, non-tuberculous mycobacteria (NTM), such as *Mycobacterium avium* complex and *Mycobacterium abscessus* are of increasing concern worldwide (Gopalaswamy *et al.*, 2020). Whilst those with cystic fibrosis are particularly vulnerable to infection by NTMs, those without pre-existing health conditions are often infected, and between 1997 to 2003, incidence of NTM infection increased from 9.1 to 14.1 cases/ 100,000 (Gopalaswamy *et al.*, 2020).

Mycobacteria, especially *Mycobacterium tuberculosis* are some of humanity's oldest pathogens. Initial bioinformatic analysis based on the lack of single nucleotide polymorphisms in modern lineages of extant mycobacteria indicated an origin of around 35,000 years ago (Brosch *et al.*, 2002). But more recent analysis suggests modern mycobacterial pathogens represent a recent clonal expansion from a significantly older mycobacterial progenitor, between 2.6-2.8 million years ago (Gutierrez *et al.*, 2005). Regardless, *M. tuberculosis* co-

expanded with early humans during the Neolithic expansion out of Africa, and has been causing disease for all of known human history (Comas *et al.*, 2013).

Over the course of this ancient evolutionary relationship between *Homo sapiens* and *M. tuberculosis,* the organism has evolved to efficiently evade the human immune system, hijacking our own cellular processes to benefit it in the course of infection (Jankute, Nataraj, *et al.*, 2017).

3.4 Mycobacterial disease

As mentioned before, mycobacteria cause diverse pathology in a wide variety of organisms, as summarised in Table 1. Mycobacteria, excluding *M. tuberculosis* and *Mycobacterium leprae,* are ubiquitous environmental organisms that can be found in water and soil (Honda *et al.*, 2018).

Table 1. Table outlining the host range and disease states of mycobacterial pathogens

Species	Hosts	Disease caused (organs affected)	Citations
Mycobacterium tuberculosis	Humans	Tuberculosis (lungs)	(Keane *et al.*, 1997)
Mycobacterium abscessus	Humans	Chronic lung infection (lungs) Other (central nervous system, soft tissue, eye)	(Lee *et al.*, 2015; Wiersma *et al.*, 2020)
Mycobacterium ulcerans	Humans	Buruli Ulcer (skin)	(Marsollier *et al.*, 2007)
Mycobacterium bovis	Cows Humans	Tuberculosis (lungs)	(Malone *et al.*, 2018; O'Reilly and Daborn, 1995)
Mycobacterium marinum	Fish Humans	Fish tuberculosis (Gills) Skin infections	(Aronson, 1926; Aubry *et al.*, 2002)
Mycobacterium avium	Humans	Chronic lung infection (lungs)	(Guerrant *et al.*, 2011)
Mycobacterium leprae	Humans	Leprosy (Peripheral nervous system, skin)	(Chavarro-Portillo *et al.*, 2019)

The pathogenesis of *Mycobacterium tuberculosis* is complex. The organism is spread primarily through droplets produced by an infected patient coughing, which can remain in the air for several hours (Escombe *et al.*, 2008). The organism is highly infectious, with an estimated 30% of household contacts of an infected person going on to become infected themselves (Chang *et al.*, 1992). Symptomatic tuberculosis only occurs in 10% of infections, with 90% not becoming acutely symptomatic (Sáenz *et al.*, 2013). In the 10% of symptomatic cases, *M. tuberculosis* bacilli go on to infiltrate alveolar macrophages, which are the principal cell type infected by the bacilli (Keane *et al.*, 1997).

Macrophages normally envelop pathogens into phagosomes, which then undergo a series of fusions with other intracellular compartments, delivering antibacterial agents, reactive oxygen species, and ATPases which acidify the compartment (Stewart *et al.*, 2005; Buter *et al.*, 2019; Simeone *et al.*, 2012). However, through a host of factors secreted by the type 7 secretion system ESX-1 and direct interactions with the glycolipid lipoarabinomannan (LAM), tuberculosis bacilli prevent this maturation of the phagosome (Houben *et al.*, 2014; Dinadayala *et al.*, 2006; Kang *et al.*, 2005). This allows the bacterium to occupy a privileged niche, protected from the adaptive immune response by the body's own cells, where it subsists on host-derived long chain fatty acids (Gouzy *et al.*, 2021). At this point, most people can sequester this infection into lesions referred to as granuloma. These are aggregations of multinucleated fused phagocytes encased in a hard exterior, with dormant bacteria within (Davis and Ramakrishnan, 2009). In most cases, bacteria within granuloma remain dormant for life. However, they can be can become resuscitated upon immunosuppression, causing rupture of the granuloma (Kana and Mizrahi, 2010). This leads to active infection. Interaction of mycobacterial sulfolipid-I with nociceptive (damage-sensing) neurons, leads to tuberculosis

cough via an inflammation-independent route, thereby distributing active bacteria to new hosts (Ruhl *et al.*, 2020). Without treatment, the rate of mortality for tuberculosis is 50%. However, using a combination of isoniazid, ethambutol, and rifampicin, mortality is relatively low (World Health Organisation, 2018). Unfortunately, as is increasingly common in bacterial disease, antibacterial resistance is proliferating rapidly, and the drug development pipeline remains sparse (Sterling *et al.*, 2003; Tahlan *et al.*, 2012).

3.5 Structure of the mycobacterial cell wall

Whilst it is important to frame mycobacteria in the context of their pathogenicity, this focus can lose sight of the commonalities between the members of the genus *Mycobacterium*. As mentioned previously, the core cell wall ultrastructure is conserved not only in mycobacterial species, but also in other acid-fast pathogens. Often, only minor modifications in the cell wall cause the differences seen in their pathogenicity and host range seen in Table 1 (Angala *et al.*, 2018; Bacon *et al.*, 2014). The cell wall core of CMNR group organisms is broadly constituted of 4 major components, this ultrastructure is shown in the below figure.

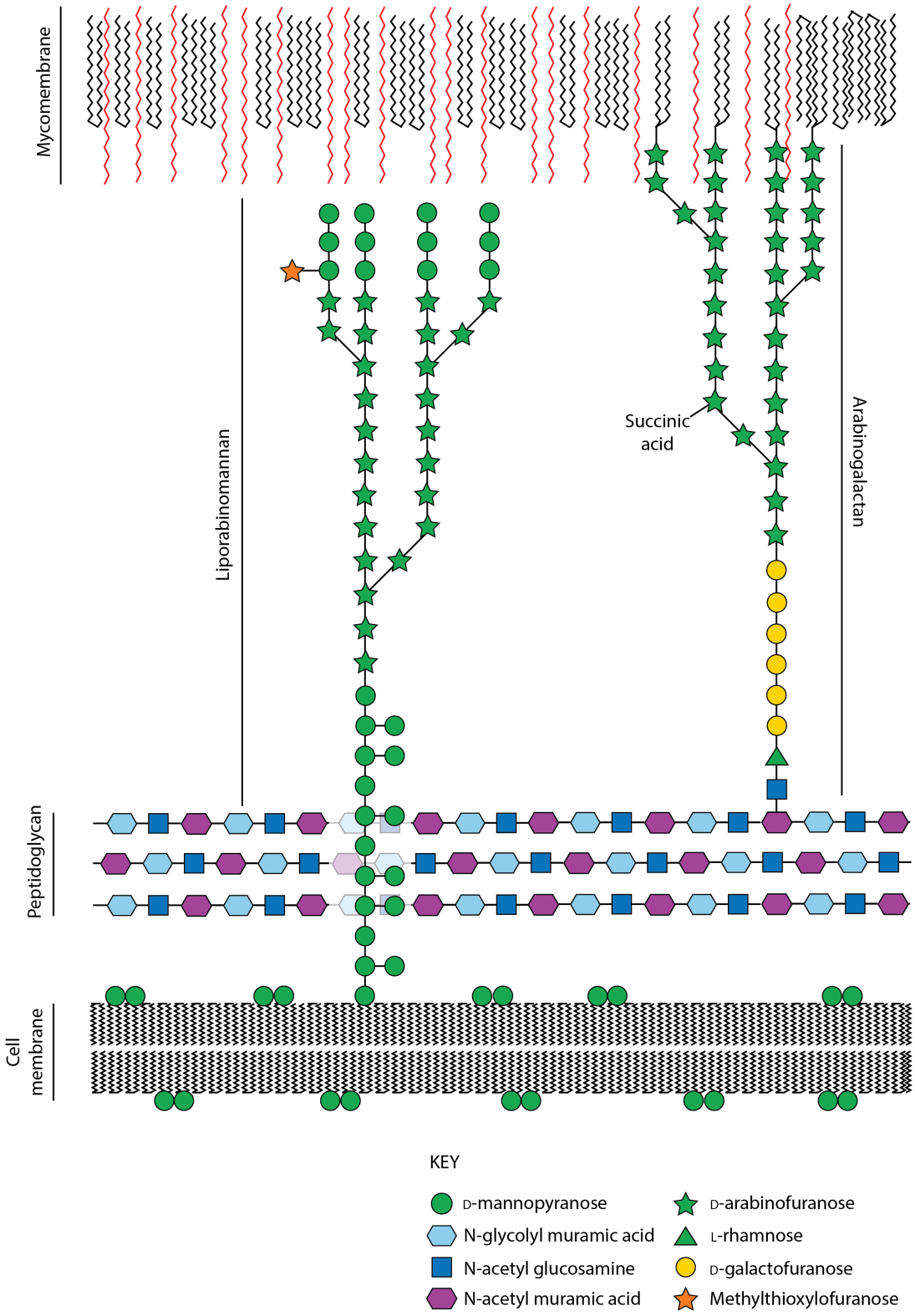


Figure 1. Schematic of the cell wall ultrastructure of *M. tuberculosis* (adapted from Dulberger *et al.*, 2020). The innermost leaflet is a phospholipid bilayer composed primarily of PIMs. Exterior to this are LAM and PG. Covalently attached to the PG is AG, and the outermost leaflet is the mycomembrane. Note that the site of arabinose attachment in AG is not the terminal galactose, the structure has been shown in this way for ease of presentation. Additionally, mycolic acids external to LAM are also covalently attached to AG, but for presentation purposes this has not been shown under the LAM molecule. Mycolic acids in black represent trehalose mycolate, those in red represent other species such as PDIM and phenolic glycolipids.

The innermost leaflet of the cell envelope is the cell membrane composed of phosphatidyl-*myo*-inositol mannosides (PIMs), cardiolipin (CL) and phosphatidylethanolamine (PE). PIMs are unique to Mycobacteriales and are the majority species in the membrane, outnumbering CL and PE at a ratio of 5:1 based on dry mass (Bansal-Mutalik and Nikaido, 2014; Bansal-Mutalik and Nikaido, 2011; Brown *et al.*, 2023). Unlike other bacteria, the remainder of the core cell wall ultrastructure is a single covalent unit named the mycolyl-arabinogalactan-peptidoglycan (mAGP) complex. Exterior to the cell membrane is the peptidoglycan sacculus, which provides support to maintain turgor pressure of the cell. Mycobacterial peptidoglycan is unusual amongst bacteria due to the fact it contains *N*-glycolylmuramic acid (MurNGlyc) as well as *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) (Hansen *et al.*, 2014; Mahapatra *et al.*, 2005). The β-1,4-linked sugar backbone is then extensively crosslinked by the lactoyl-linked peptide crosslinks of muramic acid (Baranowski *et al.*, 2018). Covalently linked to this PG layer is a branching heteropolysaccharide called arabinogalactan, which is composed of two distinct moieties of furanoses. The first of these domains is D-galactan, which is linked to peptidoglycan by a rhamnose-GlcNAc-phosphate linker unit (Harrison *et al.*, 2016; McNeil *et al.*, 1990). D-galactan is a linear chain of alternating β-1,5 and β-1,6 galactofuranose (Gal*f*) units, which at the 8th, 10th, and 12th monosaccharides are substituted with β-1,3-linked arabinofuranose (Ara*f*) (Justen *et al.*, 2020). These Ara*f* substituents are then polymerised in a branching manner with α-1,5 Ara*f* monomers, and then further branched with β-1,3 linkages (Larrouy-Maumus *et al.*, 2012).

Arabinan chains are terminated with β-1,2 caps of D-Ara*f* (Alderwick *et al.*, 2015). These caps are then lipidated with mycolic acids, a class of long-chain lipid unique to Mycobacteriales organisms (Tahlan *et al.*, 2012). The structure of these molecules is discussed further below.

These covalently attached lipids then serve as an anchor for a wide variety of other lipids to intercalate, forming an exterior lipid membrane (Dover *et al.*, 2004), which similar to diderm Gram-negative organisms is asymmetric (Bansal-Mutalik and Nikaido, 2011).

In addition to this core cell wall unit is another glycoconjugate named lipoarabinomannan (LAM) and its biosynthetic precursor, lipomannan (LM). LAM is a hyperglycosylated PIM which has a α-1,6 mannopyranose (Man*p*) backbone, with β-1,2 Man*p* sidechains ligated to an arabinan core similar to that of arabinogalactan (Choudhary *et al.*, 2018; Guerardel *et al.*, 2002). The outermost layer of the cell wall ultrastructure is the capsule. This component is arguably the most cryptic and poorly understood, but is important in pathogenesis (Kalscheuer *et al.*, 2019). Similar to other cell wall components, its composition varies with species and growth phase, being described as an electron-transparent zone which resisted staining by established pathology methodologies, leading to a halo-like occlusion under both light and electron microscopy (Brieger and Glauert, 1956; Yamamoto *et al.*, 1958). More recent analyses revealed that it is primarily composed of α-glucan, mannan, arabinomannan, and lipids. The ratio of these three major components seems to widely vary between mycobacterial species and growth phase (Dinadayala *et al.*, 2008).

3.6 Lipids of mycobacteria

The thick layer of complex lipids forming the mycomembrane distinguishes Mycobacteriales bacteria, causing them to have a uniquely hydrophobic cell surface. This gives them an appearance similar to candle wax on agar plates (Jankute, Nataraj, *et al.*, 2017). As mentioned previously, a critical element of the mycomembrane are the mycolic acids, a family of lipids

which constitute 50% of the mycomembrane (Tahlan *et al.*, 2012; Dai *et al.*, 2020; Bansal-Mutalik and Nikaido, 2014). These are a class of long, two chain, 3' hydroxylated (C_{60-90}) fatty acids (Daffé *et al.*, 1991). Within the mycomembrane, mycolic acids can either exist in their fatty acid forms, or covalently attached to trehalose to form an amphipathic glycolipid. Following their synthesis, detailed later, trehalose monomycolate is incorporated into the mycomembrane in a variety of ways through the Antigen 85 enzyme complex (Armitige *et al.*, 2000). Firstly, the acyl chain can be transferred to arabinan termini to produce covalent mycolate (Alderwick, Dover, *et al.*, 2006). Secondly, one acyl chain can be transferred to another trehalose monomycolate to produce trehalose dimycolate (TDM) also known as cord factor which is a principle virulence factor in *M. tuberculosis* (Indrigo *et al.*, 2003; Silva *et al.*, 1985; Ojha *et al.*, 2010). Finally, the fatty acid can be transferred to outer membrane proteins (Huc *et al.*, 2010). A proportion of trehalose dimycolate are modified with sulfate groups on the 3' position of each glucose monomer of the trehalose to form sulfolipid-1, which is the primary driver of tuberculosis transmission through induction of cough (Ruhl *et al.*, 2020; Gilmore *et al.*, 2012)

Interspersed between these covalent anchors are an assortment of other lipids. The most important of which are pthiocerol dimycoceroserate (PDIM). These lipids are abundant in the mycomembrane and are implicated in the pathogenesis of *M. tuberculosis*, where they are responsible for the rupture of the phagolysosome in concert with ESAT-6 (Cambier *et al.*, 2020; Simeone *et al.*, 2012). Additionally, due to their lack of polarity, they provide an efficient means of exclusion of water-soluble chemical aggressors such as drugs and damaging small molecules (Camacho *et al.*, 2001).

3.7 Biogenesis of the mycobacterial cell wall

3.7.1 Synthesis of Peptidoglycan

Peptidoglycan is a near-ubiquitous bacterial cell wall polysaccharide usually consisting of linear chains of β-1,4 linked disaccharides of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) crosslinked by pentapeptide crosslinks (Healy *et al.*, 2020; Boneca *et al.*, 2000; Glauner *et al.*, 1988). As in many bacterial species, mycobacterial peptidoglycan synthesis relies on lipid-borne sugar precursors which are polymerised by glycosyltransferases and transpeptidases (Gupta *et al.*, 2010). As mentioned previously, a proportion of MurNAc in mycobacterial peptidoglycan is replaced by MurNgly. This monosaccharide is believed to be one reason why the mycobacterial cell wall possesses such extraordinary immunostimulatory effects, with a preparation of it named Freund's Complete Adjuvant and was previously widely used in vaccines in order to increase their potency (Dubé *et al.*, 2019). Mycobacterial peptidoglycan synthesis is initiated by formation of uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) by GlmU (Rv1018c)(Zhang *et al.*, 2009). This is then further processed in two subsequent pathways to generate the three monosaccharides used in peptidoglycan synthesis in mycobacteria. MurA (Rv1315) and MurB (Rv0842) catalyse the conversion of a proportion of UDP-GlcNAc to UDP-MurNAc by addition of a enolpyruvate moiety from phosphoenolpyruvate and subsequent reduction into UDP-MurNAc using NADPH (De Smet *et al.*, n.d.; Eniyan *et al.*, 2018). UDP- MurNglyM is then generated by NamH from UDP-MurNAc via hydroxylation (Raymond *et al.*, 2005). The UDP-MurNglyM are then decorated with the pentapeptide L-alanine-D-*iso*glutamine-*meso*-diaminopimelic acid-D-alanine-D-alanine by the ligases MurC, MurD, MurE and MurF (Eniyan

et al., 2018; Basavannacharya *et al.*, 2010). This UDP MurNglyM-pentapeptide unit is then ligated onto decaprenyl-phosphate (C50-P) on the inner membrane leaflet by MraY (Rv2156c) to produce Lipid I (Chung *et al.*, 2013). The MurG glycosyltransferase then catalyses the formation of lipid II through generation of a β-1,4 linkage of UDP-GlcNAc to the UDP-NGlyM of Lipid I, thus generating the complete disaccharide on its lipid carrier, termed Lipid II (Mengin-Lecreulx *et al.*, 1991).

Subsequently, the entire lipid is flipped to present the GlcNAc-MurNglyM-pentapeptide to the periplasm by MurJ (García-Heredia *et al.*, 2021; Zheng *et al.*, 2018). Once on the extracellular face, PBPs and Shape, Elongation, Division, and Sporulation (SEDS) proteins catalyse the condensation of disaccharides into the long β-1,4 chains characteristic of peptidoglycan (Sher *et al.*, 2020; Meeske *et al.*, 2016; Nygaard *et al.*, 2023). Bifunctional PBPs represent a subclass of these enzymes (Rv0050 and Rv3682) which also possess transpeptidase activity, catalysing crosslinks between *meso*-diaminopimelic acid and D-Ala between peptide stems in neighbouring glycan chains (Baranowski *et al.*, 2018). This nature of crosslinking is quickly replaced in 80% of peptide linkages by the activity of L,D-transpeptidases which cleaves the next D-ala, forming the final 3-3 linkage (mDAP-mDAP) crosslinked peptidoglycan characteristic of mycobacteria (Baranowski *et al.*, 2018; Gupta *et al.*, 2010; Lavollay *et al.*, 2008).

3.7.2 Synthesis of arabinogalactan

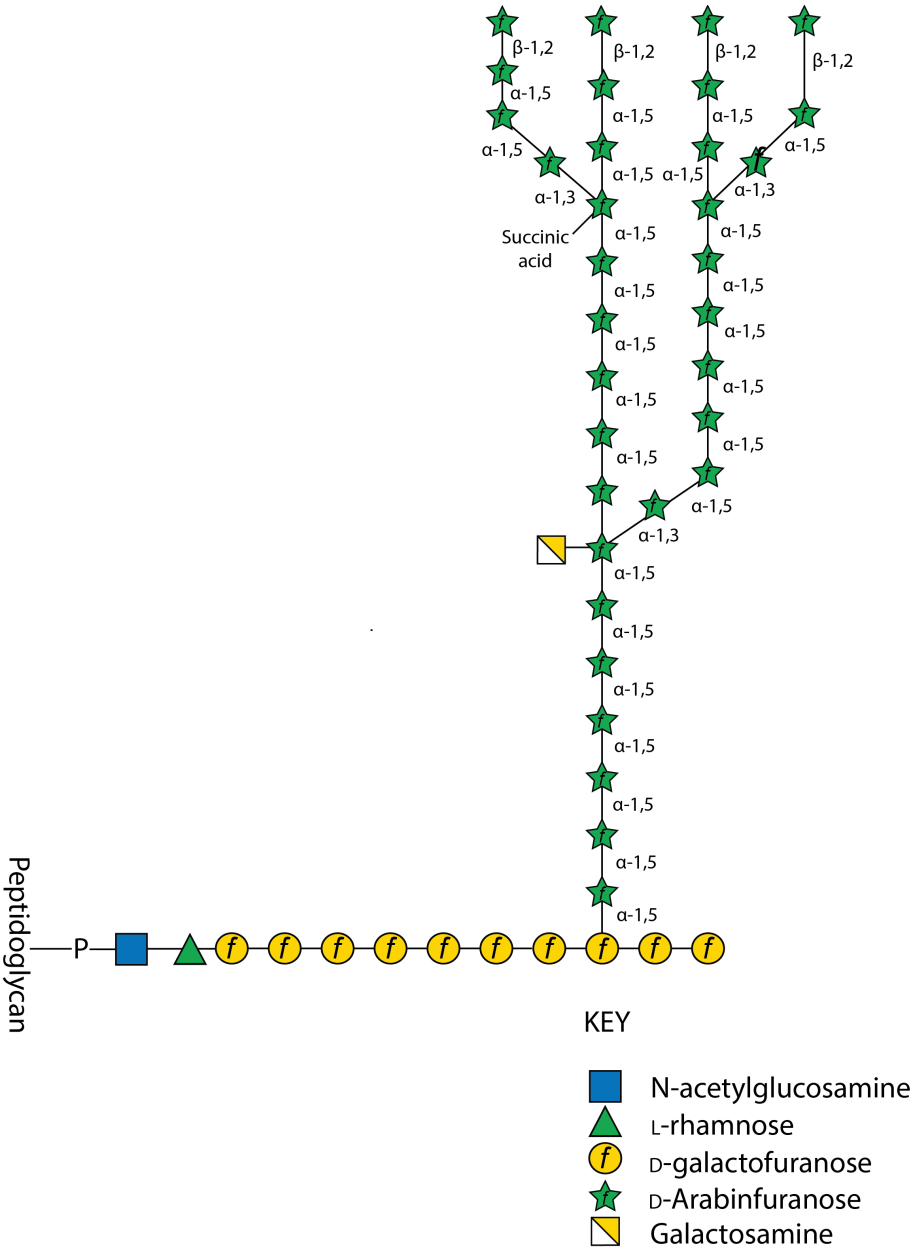


Figure 2. Schematic of the final polymerised structure of arabinogalactan, the structure of which has been elucidated through the work of many groups in collaboration, as reviewed by Alderwick et al (2015).

Arabinogalactan is the heteropolysaccharide which serves to anchor the external mycolic acids to the peptidoglycan sacculus in the cell wall, acting as a scaffold between the layers of the mAGP complex. Like peptidoglycan, its synthesis is dependent on long chain polyprenyl (C_{50}) borne monosaccharide components which are transferred to the periplasm and processively polymerised. Arabinogalactan is comprised of 3 key moieties. The first of these is the linker unit binding it to peptidoglycan which is GlcNAc linked to α-L-rhamnose by an α-1,3 linkage (Mills *et al.*, 2004; McNeil *et al.*, 1990). Attached to this is the D-galactan moiety, consisting of 28-30 residues of D-galactofuranose (Gal*f*) joined by alternating β-1,5 and β-1,6 linkages (Beláňová *et al.*, 2008; Poulin and Lowary, 2016). From this extends the D-arabinan domain, consisting of α-1,5 linear chains of D-arabinose, with branch points beginning with an α-1,3 linkage and then terminating in β-1,2 linked residues (Jankute *et al.*, 2018).

Arabinogalactan synthesis begins with the condensation of the rhamnose-GlcNAc linker unit formed by the phosphotransferase enzyme Lcp1 onto GlcNAc residues on PG (Harrison *et al.*, 2016). The reducing end of the rhamnose then serves as the acceptor for the bifunctional galactosyltransferase Glft1. This enzyme first primes the rhamnose by addition of a β-1,4-linked Gal*f* residue from an undecaprenol-D-Gal*f* donor, followed by another D-Gal*f* residue with a β-1,5-linkage (Beláňová *et al.*, 2008). The processive glycosyltransferase Glft2 then polymerises the D-galactan backbone through successive addition of around 28 alternating β-1,5 and β-1,6-linked D-Gal*f* residues (Poulin and Lowary, 2016). After polymerisation of the D-galactan backbone, the D-arabinan domain is polymerised using the lipid donor decaprenylphosphoryl-β-D-arabinofuranose (DPA). This substrate donor is synthesised from the pool of ribose-5-phosphate from the pentose phosphate pathway in central carbon metabolism (Huang *et al.*, 2005). The first stage of this is addition of a pyrophosphate group

to the ribose by PrsA. This phosphate group is then displaced with a decaprenyl phosphate (C_{50}-P), which is the lipid carrier for the molecule. The additional phosphate group is then cleaved from the molecule, thereby creating decaprenylphosphoryl-β-D-phosphoribose. This then undergoes two stages of epimerisation by DprE1 and DprE2 (Rv3790 and Rv3791, respectively), to produce DPA (Brecik *et al.*, 2015). The DPA is then flipped to the periplasmic face of the cell membrane by an unidentified transport protein.

The D-arabinan of AG is then polymerised successively by a suite of glycosyltransferases in the periplasm, using DPA as the donor. First, the 5' carbon on the 8th, 10th and 12th β-1,6 Galf residues is primed with three Ara*f* residues with α-1,5 linkages (Alderwick, Seidel, *et al.*, 2006). The D-arabinan is then polymerised by the Emb proteins, so named due to their inhibition by the drug ethambutol (Zhang *et al.*, 2020). Studies using the closely-related *Corynebacterium glutamicum*, knocking out its singular Emb arabinosyltransferase, resulted in complete loss of all arabinan other than the 3 priming monosaccharides (Jankute *et al.*, 2018). Knockdown of the EmbA and EmbB genes in *M. smegmatis* clearly implicates the gene products in polymerisation of the α-1,5 arabinan, but each of their precise enzyme functions remain unclear (Escuyer *et al.*, 2001; Bakuła *et al.*, 2013; Tan *et al.*, 2020). The AftC and AftD proteins introduce single α-1,3-linked residues which act as branch points to be further elongated by the EmbA and EmbB proteins (Jian Zhang *et al.*, 2011). This processive addition results in a heterogenous, branched final structure. This branching network of D-arabinan is terminated by the AftB protein introducing a β-1,2 hexa-Ara*f* cap, thus completing the final structure of arabinogalactan (Escuyer *et al.*, 2001; Birch *et al.*, 2008).

3.7.3 Synthesis of mycolic acids

The outermost layer of the mycobacterial cell wall is analogous to the outer membrane of Gram-negative bacteria, albeit consisting of a highly unusual mix of covalently bound and non-covalently-associated long chain lipids named mycolic acids (Daffé *et al.*, 2017). Seminal work of Prof. David Minnikin and collaborators identified them as C_{16}-C_{24} in size, constituting a monolayer at the outermost point of the mycobacterial cell wall (Minnikin and Goodfellow, 1980; Minnikin *et al.*, 1984; Villeneuve *et al.*, 2005)

Fatty Acid Synthase I (FasI- Rv2524c) mediates synthesis of mycolic acids (Zimhony *et al.*, 2000). The enzyme consists of seven domains, each catalysing a different step in the sequential synthesis of a long chain fatty acid. These enzymes are often likened to molecular assembly lines, with each domain acting sequentially (Elad *et al.*, 2018). Synthesis begins in the acyltransferase domain of FasI, with the transfer of an acetate moiety from acyl-CoA to the Acyl carrier protein (ACP) domain (Kremer *et al.*, 2002). The lipid is then elongated processively using malonate (C_3) moieties donated by malonyl-CoA (Kremer *et al.*, 2001). The malonate moieties are then dehydrated and saturated by auxiliary modules of the FasI complex to produce mycolic acids (Dong *et al.*, 2015)

By an as yet unknown mechanism, the final products of FASI follow a bimodal distribution, with equal proportions of C16-C18 (short) and C24-C26 (long) chain lengths being produced by the same enzyme (Zimhony *et al.*, 2000). The nascent lipid chain is then shuttled to the FASII complex by malonyl acyl carrier protein M (AcpM), whereupon the lipid undergoes a series of cis/trans cyclopropanations, methylations and methoxy and keto group ligations, to

produce a mature mycolic acid (Bhatt *et al.*, 2007). Polyketide synthase 13 will then catalyse the condensation between a short and long chain mycolic acid (57). The mycolic acid will then be ligated to its carrier sugar, trehalose, to produce trehalose monomycolate (TMM) (Gavalda *et al.*, 2014; Belisle *et al.*, 1997). TMM is then transported across the cytoplasmic membrane by Major Membrane Protein Large 3 (Mmpl3) where it acts as the substrate for the tripartite antigen 85 complex for its incorporation into the cell wall (Zhang *et al.*, 2019).

3.7.4 Synthesis of lipoarabinomannan

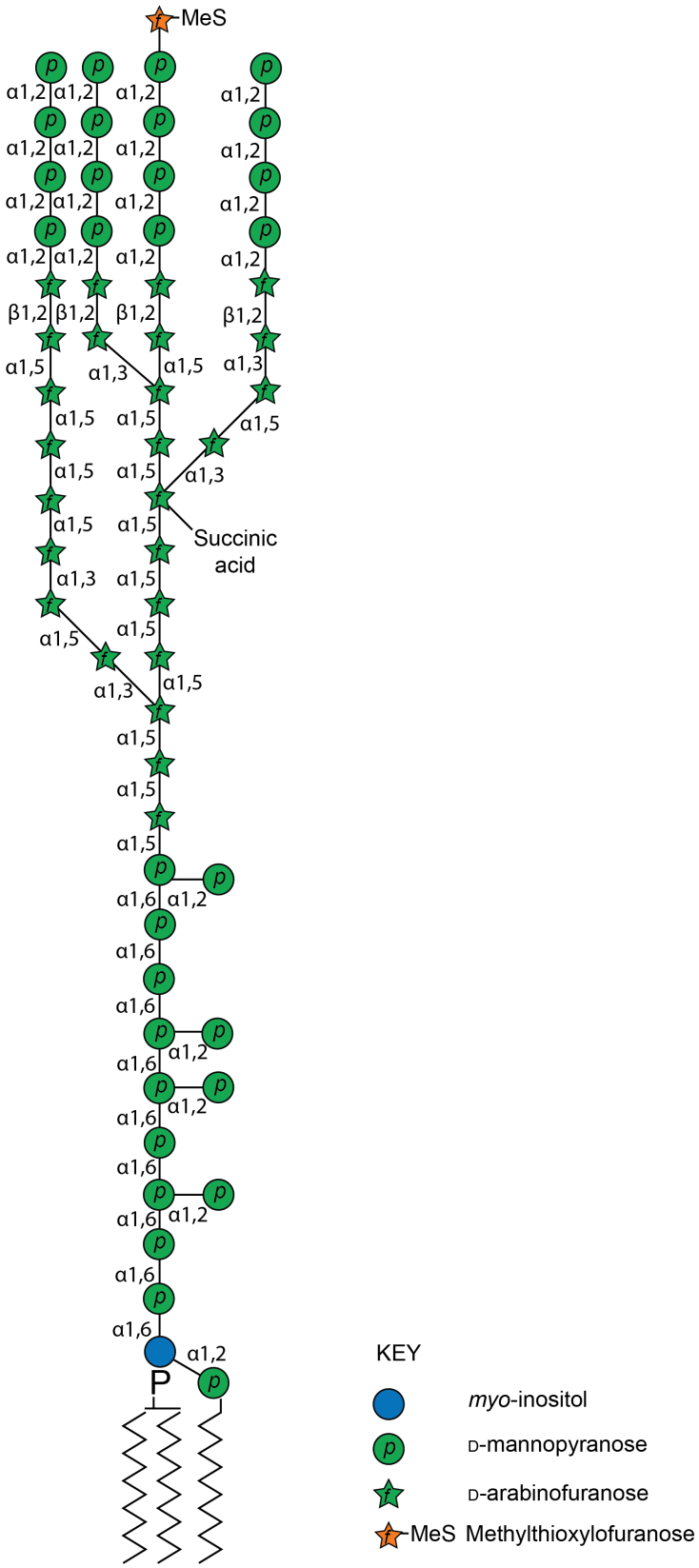


Figure 3. Schematic of lipoarabinomannan structure adapted from Kaur *et al.*, 2014. Monosaccharides are annotated using symbolic nomenclature for glycans (SNFG) linkages are indicated by numbers between residues.

The final polymerised structure of LAM is summarised in Figure 3. LAM is a cell wall glycolipid that begins as a PIM. Synthesis of LAM is initiated by the glycosyltransferase family A protein PimA (Korduláková *et al.*, 2002). This enzyme catalyses the formation of a glycosidic bond between the 2' hydroxyl of inositol of PI and an α-D-mannopyranose (Man*p*) from GDP-mannose (Korduláková *et al.*, 2002), thus forming PIM_1. This initiating mannose is then acylated by Rv2611 (Korduláková *et al.*, 2003), forming $AcPIM_1$. A further Man*p* residue is linked to the 6' hydroxyl of the first mannose by PimB (Guerin *et al.*, 2009) to form AC_1PIM_2. A further Man*p* monomer is then added to the PIM by PimC to form Ac_1PIM_3 and then a final Man*p* residue is added by PimD to produce $Ac_{1/2}PIM_4$ which is the branching point for glycolipid synthesis (Morita *et al.*, 2006). A proportion of these mature PIMs will be hyperglycosylated to produce LAM (Mishra *et al.*, 2008) This process is believed to be initiated by flipping of the $Ac_{1/2}PIM_4$ to the periplasmic face of the membrane by an unknown translocase. PimE then adds an α-1,2 Man*p* residue, the α-1,6 backbone is extended with further Man*p* residues by MptA and MptB to form the mannan core of LAM (Morita *et al.*, 2006; Mishra *et al.*, 2007; Mishra *et al.*, 2008). MptC then catalyses the intermittent addition of α-1,2-linked Man*p* monomer sidechains (Khoo *et al.*, 1996). EmbC then polymerises the linear α-1,5 arabinan chains, 12-16 residues in length, on the terminus of the mannose backbone (Alderwick *et al.*, 2011). As in arabinan from AG, this is intermittently branched with α-1,3-linked Ara*f* residues by AftC (Jian Zhang *et al.*, 2011). Finally the arabinan domain is terminated with β-1,2-arabinofuranoside caps to form LAM (Jankute, Alderwick, *et al.*, 2017; Seidel *et al.*, 2007).

In slow-growing mycobacteria such as *M. tuberculosis* and *Mycobacterium marinum*, this glycan acquires further structural complexity from α-1,2-linked mannose caps on the

arabinan termini by the CapA enzyme (Dinadayala *et al.*, 2006). Significant evidence for the immunomodulatory capabilities of these mannose caps on LAM exists within the literature (Kaur *et al.*, 2008; Maeda *et al.*, 2003; Torrelles *et al.*, 2012).

Following complete synthesis of the LAM and AG, they can then be further modified in some species to change the biophysical characteristics of the polysaccharide. The most thoroughly investigated of these modification is succinylation, which can be found on the C2 position of the α-1,3-D-arabinofuranose branch points of arabinan in LAM and AG (Xie *et al.*, 2015). This modification is found in many mycobacteria, including *M. smegmatis, M. tuberculosis*, and *M. abscessus* (Bhamidi *et al.*, 2008a). The modification occurs in a stoichiometry of around 1-4 residues per arabinogalactan molecule (Bhamidi *et al.*, 2008a). The transferase SucT was shown to be responsible for this modification in *M. smegmatis* (Palčeková *et al.*, 2019). Loss of succinylation has diverse effects on cell surface hydrophobicity, fitness in infection, and mannose capping of LAM in *M. tuberculosis* (Palčeková *et al.*, 2023; Palcekova *et al.*, 2020).

3.8 Cell wall turnover in mycobacteria

Cell wall turnover is a dynamic process carried out in bacteria whereby soluble fragments of the cell wall ultrastructure are released in response to changes in environment. The process was first observed in 1962 in *Bacillus megaterium*, whereby gradual release of radioactive material of chemostat-grown cultures of the bacteria which had been previously labelled with ^{14}C-mDAP (Chaloupka, Křečková, *et al.*, 1962; Chaloupka, Kreckova, *et al.*, 1962). It was also noted that release of this radioactive material was concomitant with changes in tolerance to

treatment with lysozyme, implying the process on some level changed the biochemical properties of the bacteria.

Today, cell wall turnover is understood to be the cornerstone of binary fission. The process of cell division in bacteria is a tightly regulated process, coordinated by FtsZ, a tubulin homolog which polymerises in a ring at the site of division(Bi and Lutkenhaus, 1991; Bisson-Filho *et al.*, 2017). FtsZ causes the concerted recruitment of enzymes such as lytic transglycosylases which cleave peptidoglycan, as well as penicillin binding proteins and SEDS proteins which synthesise and crosslink nascent peptidoglycan, preventing bacterial lysis during replication, resulting in daughter cells (Bisson-Filho *et al.*, 2017; Meeske *et al.*, 2016).

However, in recent years, it has been shown that cell wall turnover is critical in multiple other processes, particularly in immune interactions. Nucleotide-binding Oligomerisation Domain-containing (NOD) proteins one and two are critical in the detection and response coordination against peptidoglycan-containing pathogens (Deshmukh *et al.*, 2009). The minimal ligand for NOD1 is γ-D-glutamyl-*meso*-diaminopimelic acid (muramyl dipeptide), whilst NOD2 detects muramyl dipeptide (*N*-acetylmuramyl-L-alanyl-L-*iso*glutamine) (Girardin *et al.*, 2003; Inohara *et al.*, 2003). Formation of substrate-NOD protein complexes results in induction of expression of NF-κB, which in turn triggers production of proinflammatory cytokines. This makes NOD proteins master regulators of the innate immune response in humans (Deshmukh *et al.*, 2009). Futhermore, errors in NOD signalling have been associated with the incidence of Crohn's disease (Inohara *et al.*, 2003).

In a similar manner, the lectin Galectin-9 is implicated in detection of mycobacterial infections by binding to arabinogalactan (Wu *et al.*, 2021). However, mechanisms regarding how soluble

fragments of arabinogalactan could be detected by the immune system remain unclear. Lipoarabinomannan can, however, be detected in the urine of HIV-TB coinfected patients, and represents a promising avenue for tuberculosis diagnosis in this cohort (Broger *et al.*, 2019). Taken together, the significance of the interplay between enzymatic products of bacterial cell wall homeostasis and the human immune system is difficult to overstate.

Cell wall turnover in mycobacteria is a growing field, with trehalose catabolism being one of the earliest explorations into the question of whether and how mycobacteria remodel their unusual cell wall components. The process is mediated by the high-affinity trehalose LpqY-SugA-SugB-SugC reuptake system (Kalscheuer *et al.*, 2010). The system is critical in virulence, with knockout of SugC causing a 3-log drop in bacterial burden in a mouse infection model (Kalscheuer *et al.*, 2010). Further study found that the recycling is not essential but enables energy-efficient cell envelope remodelling in response to redox stress and carbon starvation (Pohane *et al.*, 2021). Given the importance of trehalose as a carrier molecule in the synthesis of mycolic acids, it is obvious why this recycling pathway is so important in *M. tuberculosis* Furthermore, whilst no evidence exists to show mycobacteria directly recycle the glycan portion of their PG, they can survive on the lactate liberated from muramic acid by LpqI (Moynihan *et al.*, 2019). Whilst no transporter has yet been identified, it is likely that there exists a transporter for the lactate to allow it to return to central carbon metabolism within the cell (Moynihan *et al.*, 2019).

The presence of soluble cell wall components liberated from the bulk cell wall ultrastructure of *M. tuberculosis* is reminiscent of the 1962 chemostat experiments in *Bacillus megaterium.* Therefore, it is likely that the genome of *Mycobacterium tuberculosis* encodes enzymes capable of the depolymerisation of other cell wall components than peptidoglycan, likely

essential in viability due to the need to allow division as mediated by FtsZ. Therefore, the discovery of arabinogalactan remodelling enzymes in Mycobacteriales species could be informative for understanding disease progression of mycobacterial pathogens.

3.8.1 Glycoside hydrolases

The two principal families of enzymes that are capable of breaking glycosidic bonds are polysaccharide lyases and glycoside hydrolases. Polysaccharide lyases catalyse cleavage of a polysaccharide by elimination or other mechanisms, whilst glycoside hydrolases are so named for their inclusion of water in their cleavage mechanism (Sutherland, 1995; Ardèvol and Rovira, 2015).

Glycoside hydrolases play many pivotal roles in biology. Within humans, saliva contains large amounts of the starch-specific glycoside hydrolase amylase in order to liberate glucose from food (Pajic *et al.*, 2019). Furthermore, GHs function biogeochemically in saprophytic organisms, allowing release of carbon-containing monosaccharides from dead and rotting organic matter, thus are critical in carbon cycling (Attia *et al.*, 2018). Additionally they are critical in global cycling of sulfur through turnover of the sugar sulfoquinovose, with 10 billion tons of it produced and turned over yearly (Sharma *et al.*, 2022; Harwood and Nicholls, 1979). They also play diverse roles in microorganisms. LacZ, the regulator of the *lac* operon in *E. coli,* is a GH and neuraminidases are important virulence factors in influenza viruses (Huber *et al.*, 1976; Taylor and von Itzstein, 1994). Given the global importance of glycoside hydrolases, both at the micro- and macroscopic scales, their study is widespread, and in response to this the Carbohydrate-Active enZyme DataBase (CAZYDB) was established in order to curate the wealth of information on enzymes active on carbohydrates (Henrissat, 1991). Glycoside hydrolases within CAZYDB are primarily organised by sequence identity. As sequence leads to structure this classification broadly groups enzymes by mechanism, substrate, and product profile.

3.8.1.1 Enzyme mechanism

The two major catalytic mechanisms for glycoside hydrolases are defined by the resultant anomeric configuration, either retaining this configuration or inverting it. For example, a retaining glycoside hydrolase will catalyse the breakdown of an α-glycoside and produce an α-glycoside. Whereas an inverting glycoside hydrolase produces a β-glycoside from an α-glycoside. This difference of final anomeric configuration is due to the steps taken in enzymatic catalysis.

As summarised in Figure 4, retaining glycoside hydrolases proceed by a double displacement reaction via a covalent glycoside-enzyme intermediate (Tanaka *et al.*, 1994). The first step of this catalysis is glycosylation of the enzyme. In this stage, the nucleophilic catalytic residue, typically a glutamate or aspartate, attacks the anomeric centre forming an oxocarbonium ion transition state to displace the aglycon and form the glycoside-enzyme intermediate (McCarter and Withers, 1994). The second step, deglycosylation, occurs when a water molecule is deprotonated by a catalytic base residue, also usually aspartate or glutamate. This deprotonated water then hydrolyses the intermediate, liberating the final saccharide products (McIntosh *et al.*, 1996). Structural information indicates that retaining glycoside hydrolases typically position their active site residue side chains ~5.5 Å from one another (Davies and Henrissat, 1995).

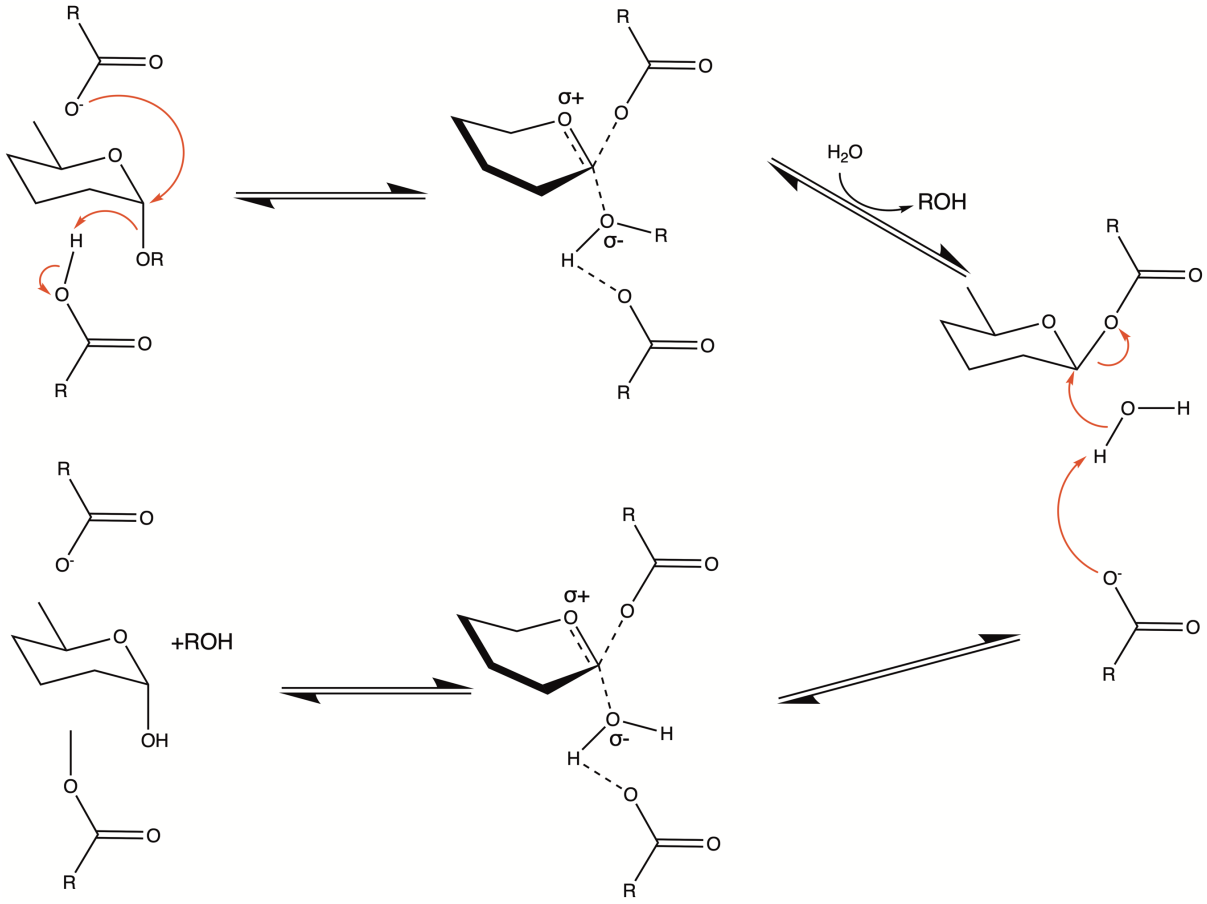


Figure 4. Mechanism of catalysis of a hydrolysis of a glycosidic bond by a glycoside hydrolase with retention of anomeric configuration, adapted from CazyDB.

Inverting glycoside hydrolases catalyse the glycosidic cleavage through a single step reaction whereby the basic residue deprotonates a water which then attacks the anomeric carbon forming a non-covalent oxocarbonium transition state, with the acidic side chain then protonating the aglycon to liberate product saccharides, this mechanism is summarised in Figure 5 below. Catalytic sidechains of inverting glycoside hydrolases will tend to be 6-11 Å away from each other (McCarter and Withers, 1994). These distinctions matter if working on a glycoside hydrolase therapeutically, as in the case of retaining catalysis, the process involving a covalent enzyme-substrate intermediate is critical in the synthesis of small molecule inhibitors.

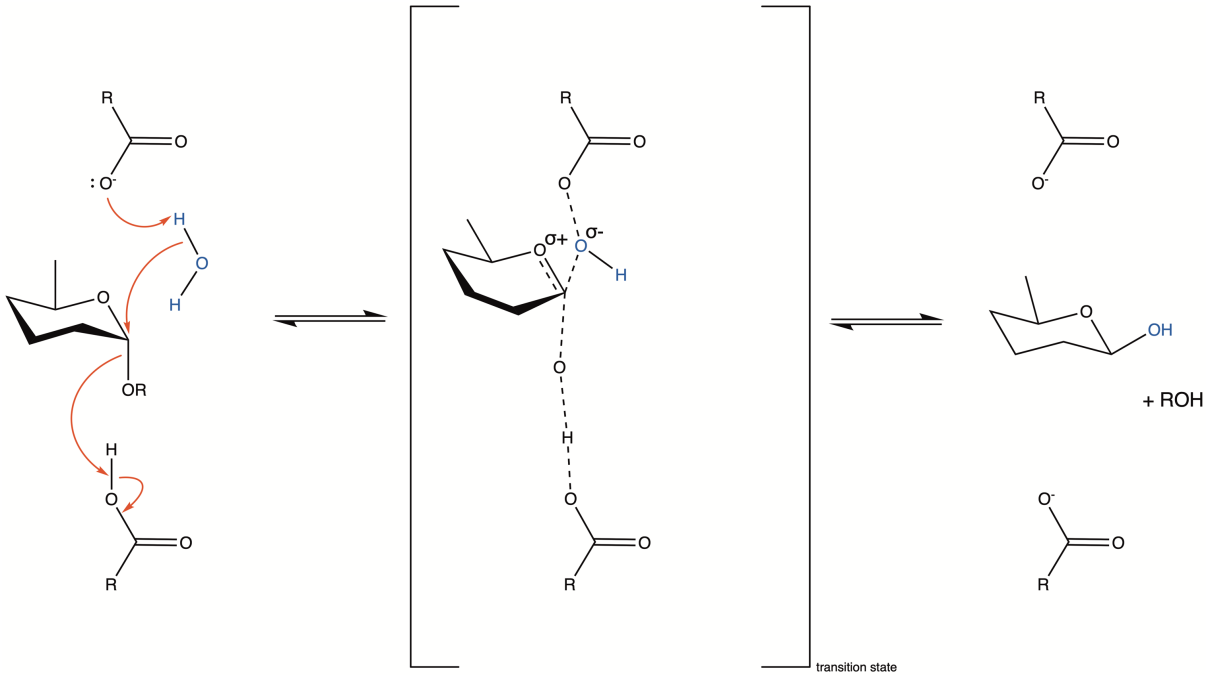


Figure 5. Mechanism of catalysis of hydrolysis of a glycosidic bond by a glycoside hydrolase with inversion of the anomeric configuration, adapted from CazyDB.

Glycoside hydrolases can also be categorised by their products, as summarised in Figure 6. Endo-acting GHs take a longer oligosaccharide and cleave it into two shorter oligosaccharides. Exo-acting glycoside hydrolases successively depolymerise an oligosaccharide by releasing monosaccharides from a longer chain. With this incredible diversity of substrate, glycoside hydrolases are capable of degradation of almost every known polysaccharide and are pivotal in carbon acquisition in all known life forms (Wardman *et al.*, 2022).

3.9 CAZymes in the mycobacterial genome

Carbohydrates play a uniquely important role in the physiology of mycobacteria. Within the cell wall they play multiple functions as structural, immunological and storage molecules (Bhamidi *et al.*, 2008b; Lavollay *et al.*, 2008; Torrelles *et al.*, 2012). This means that the genome of *M. tuberculosis* is rich in CAZymes responsible for the synthesis, remodelling, and maintenance of glycan-containing structures. The full repertoire of CAZymes is underexplored yet due to the lack of functional understanding of the mycobacterial genome. As of 2017, roughly 40% of mycobacterial gene products lacked functional characterisation, experimentally or bioinformatically (van Wyk *et al.*, 2017). However, recent advances in large language models such as eCAMI, AlphaFold, and accompanying search tools such as FoldSeek have rapidly accelerated the rate at which functions can be predicted before experimental methodologies (Xu *et al.*, 2020; Jumper *et al.*, 2021; van Kempen *et al.*, 2023).

3.10 Bacteriodetes spp.

Members of the genus Bacteroidota are a prominent component of the gut microbiota, representing approximately 20% of the species found commonly in the healthy human gut (King *et al.*, 2019). A reason for their ubiquity within this niche is due to their capacity for breakdown of dietary polysaccharides, allowing them to utilise a greater variety of carbon sources than their competitors (El Kaoutari *et al.*, 2013). Additionally, this extensive glycolytic capacity makes them important symbionts with their human hosts. The advantages Bacteroidota members endow to their hosts are twofold. First, the human capacity for carbon utilisation is limited, with our own genomes encoding the breakdown of only three

carbohydrates, namely starch, lactose, and sucrose. It is thought that the breakdown of all other carbohydrates we gain is due to processing by gut-resident organisms. Furthermore, short-chain fatty acids produced by these organisms during carbohydrate fermentation are extensively implicated with improved human health outcomes (Arumugam *et al.*, 2011; Wu *et al.*, 2011).

A key genomic feature of carbohydrate degraders are Polysaccharide Utilisation Loci (PULs) (Bjursell *et al.*, 2006). Initially thought to be unique to Bacteroidota, but found in other genera, these are discrete, co-expressed, co-regulated genomic loci which comprise the set of proteins required for the recognition, uptake, and processing of a polysaccharide into its monomeric components. PULs therefore consist of surface glycan binding proteins, carbohydrate transporters, glycolytic enzymes such as glycoside hydrolases and polysaccharide lyases, as well as auxiliary enzymes such as sulfatases, lytic polysaccharide monooxygenases (LPMO), and proteases (Grondin *et al.*, 2017). Therefore, the complexity of a polysaccharide is proportional to the size of the PUL required for its catabolism. This elegant genomic feature therefore allows for screening to be carried out using *Bacteroides* as a search tool.

As mentioned previously, to date no enzymes had been isolated which were capable of the breakdown of mycobacterial arabinogalactan. To address this gap in the literature, we grew *Bacteroides* species on mycobacterial arabinogalactan and used proteomics to identify GHs active on AG, leading to the identification of DUF4185 (GH183) and DUF2961 (GH172) enzymes (Al-Jourani *et al.*, 2023). BLAST analysis, and later analyses making use of the recently released program Alphafold allowed identification of conserved homologs of these

enzymes across the order Mycobacteriales, this work has been published and serves as the basis of this thesis.

4 Summary of findings

The work described in this thesis is the identification, biochemically characterisation, and biological interpretation of the functions of D-arabinofuranases and D-arabinofuranosidases produced by members of the Mycobacteriales. Therefore, the work is split into three key chapters.

1. **Characterisation of D-arabinofuranases discovered through high throughput screening**

Mycobacterium tuberculosis appears to encode homologs of arabinogalactan-active glycoside hydrolases discovered in the Bacteroides species *Dysgonomonas gadei* within its genome. Within this chapter we set out to recombinantly express, purify and characterise mycobacterial homologs of the *D. gadei* proteins.

2. **Dissect the role of of D-arabinofuranosidases in cell wall recycling in species from the genus *Nocardia*.**

Nocardia are an unusual, neglected human and animal pathogen which bear many similarities to *Mycobacterium*, this chapter identifies and explores a complete pathway of arabinogalactan recycling in key pathogenic species of the genus to understand their contribution to microbial fitness.

3. **Explore secreted D-arabinofuranosidases as contributors to interspecies competition between mycobacteria and their microbial peers.**

Whilst D-arabinofuranosidases appear to play housekeeping roles in mycobacteria, many mycobacteria possess further homologues whose function could not be ascertained, within this chapter we study related enzymes which we hypothesise to be Type 7 effectors.

5 Materials and Methods

5.1 Strains and Growth conditions

During this thesis a variety of strains were used. Table 2 summarises the relevant strains, media, and growth temperatures.

Table 2. Table summarising strains grown, and media used for their cultivation along with growth temperatures and selectable markers.

Media	**Strains grown**	**Selections used**	**Temperatures used**
Lysogeny Broth	*Pseudomonas aeruginosa* PA7, *Escherichia coli* 10-beta, *Escherichia coli* T7 Shuffle, *Escherichia coli* BL21-DE3	Kanamycin 50 µg/mL, ampicillin 100 µg/mL, hygromycin B 50 µg/mL	30 °C, 37 °C
Tryptic soy broth	Mycobacterium smegmatis $MC^2$2155	kanamycin 50 µg/mL, hygromycin B 50 µg/mL	37 °C
Brain heart infusion Broth	*Nocardia brasiliensis* (ATCC 19296), *Nocardia farcinica* (ATCC 3318), *Nocardia asterioides* (ATCC 19247)	None used	30 °C
Middlebrook 7h9 Broth	*Mycobacterium smegmatis* $MC^2$2155, *Mycobacterium bovis* BCG Danish, *Mycobacterium marinum* M, *Mycobacterium abscessus* ATCC 19977	Kanamycin 50 µg/mL, hygromycin B 50 µg/mL	30 °C, 37 °C
7H10 Agar	*Mycobacterium smegmatis* $MC^2$2155, *Mycobacterium bovis* BCG Danish, *Mycobacterium marinum* M, *Mycobacterium abscessus* ATCC 19977	Kanamycin 50 µg/mL, hygromycin B 50 µg/mL, anhydrotetracycline 100 ng/mL	30 °C, 37 °C

5.2 Sauton's media preparation

Sauton's media was prepared following Sauton's recipe (Trémolières, 1952). Briefly, for one litre of media 0.5 g potassium phosphate monobasic, 0.5g magnesium sulfate, 4.0g L-asparagine, 0.05 g ferric ammonium citrate, 2.0 g citric acid were dissolved in 800 mL of deionised water. One-hundred µL of a 1% zinc sulfate solution was then added to this solution. The pH was then adjusted to 7.0 with sodium hydroxide and final volume adjusted to 1 L before sterilisation using a 0.22 µm membrane.

5.3 Plasmids and primers used in this study

During this thesis, a variety of plasmids were used for protein overexpression for protein biochemistry and toxicity assays. Table 3 summarises the plasmids used in this project.

Table 3. Plasmids used in this study

Plasmid name	**Function**
pET28a	IPTG-inducible expression of hexa-histidine tagged proteins of interest in *E. coli*. Kanamycin resistance cassette.
pMYBADc-KAN	Mycobacterial shuffle plasmid for L-arabinose inducible overexpression of a target protein. Kanamycin resistance cassette.

Table 4. Primers used in this study

Primer name	Function	Sequence (5'-3')
Rv3707cE270ARev70	Site directed mutagenesis of pET28a_Rv3707c_P30M	CACCCAGCTGGTCTGGCC
Rv3707cE270AFor70	Site directed mutagenesis of pET28a_Rv3707c_P30M	GCGGATGTCCATCCGTCAGATGGACG
Rv3707cD86ARev69	Site directed mutagenesis of pET28a_Rv3707c_P30M	CGCCACAQACCTGCAGGA
Rv3707cD86AFor69	Site directed mutagenesis of pET28a_Rv3707c_P30M	GCGTTCTTTCGCAGGCCAGGG
Rv3707cD69ARev71	Site directed mutagenesis of pET28a_Rv3707c_P30M	CGGTCGCGCCAATGCCGT
Rv3707c D69AFor71	Site directed mutagenesis of pET28a_Rv3707c_P30M	GCGCCTGTGCGAATTTCGTAGAATTCCCG

5.4 SDS-PAGE gel electrophoresis

Gels were cast according to the procedure of Laemmli and colleagues (Laemmli, 1970). Briefly, a solution of 12-15% bis-acrylamide (Bio-Rad) was prepared in 375 mM Tris pH 6.8 and 0.001% SDS. This was then polymerised using 25 µL 10% ammonium persulfate and 12.5 % trimethylamine. Before polymerisation began, the solution was transferred into a Mini-Protean glass gel cassette. This was then covered with a layer of propan-2-ol and allowed to cure for 30 minutes. Following removal of the propan-2-ol, the loading gel was then poured over the top and wells generated using combs. These were then stored in damp tissue until used. For SDS-PAGE, protein samples were prepared in 5x SDS-page loading dye (30 % Glycerol; 250 mM Tris ph 6.8; 10% SDS; 5% β-mercaptoethanol; 0.02% bromophenol blue) and boiled for 10 minutes before loading into the gel. The gel was then separated using a

current of 50 mA and 200 V for 50 minutes in 1x premade TGX running buffer (Thermo Fisher Scientific). Proteins were visualised using Ready Blue (Thermo Fisher Scientific) protein stain.

5.5 Agarose Gel electrophoresis

For a 1% (w/v) agarose gels, agarose (Merck) was dissolved in tris acetate-EDTA buffer (Thermo Fisher Scientific). The agarose solution was then melted in a microwave for 1 min and allowed to cool slightly before addition of 5 μL midori green (Nippon Biochemicals) and then cast in a gel form. Gels were then loaded with samples and electrophoresed at 140 V, 400 mA for 40 mins using tris acetate-EDTA as a running buffer.

5.6 Restriction-free cloning

Restriction free cloning was carried out using established protocols outlined by Löwe and van den Ent (van den Ent and Löwe, 2006). Briefly, 50 bp primers were designed with 25 base pairs complementary to the insert and 25 base pairs complementary to the destination. These were then used to amplify the insert from template (IDT DNA synthesised geneblock or genomic DNA) using polymerase chain reaction (PCR) with Q5 2x Mastermix (New England Biolabs) using the following steps:

This fragment was then separated using agarose gel electrophoresis. The fragment of interest was excised from the gel using a clean scalpel and purified from the gel using a GenElute Gel Extraction kit (ThermoFisher Scientific). 4 μL of this product was then used as primers for a 100 μL second PCR using 1 ng of the destination plasmid as template. This reaction was then split across three tubes and subjected to gradient PCR with the following steps:

Temperature	Time	No. cycles
98 °C	1 minute	1x
98 °C	30 seconds	35x
Tm	30 seconds	
72 °C	1 minute/ kilobase	
72 °C	3 minutes	1x

The 3 reactions were then pooled and digested with 1 μL of DpnI to remove methylated template DNA, and then cleaned using a Qiagen PCR clean kit. 20 μL of the final product was then transformed into heat-competent *Escherichia coli* 10-β cells (NEB) and then plated on LB plates supplemented with kanamycin 50 μg/mL. Successful transformants were screened using colony PCR with appropriate primers and Sanger sequencing (Source Biosciences).

5.7 Q5 site-directed mutagenesis

Primers to create the mutation of interest were designed using the NEBchanger web tool (http://nebasechanger.neb.com/). PCR was carried out using Q5 2x mastermix (NEB) using plasmid of interest as template at a concentration of 1 ng of DNA as template. PCR cycling conditions were:

Temperature	Time	No. cycles
98 °C	1 minute	1x
98 °C	30 seconds	35x
Tm	30 seconds	
72 °C	1 minute/ kilobase	
72 °C	3 minutes	1x

PCR products were analysed using 0.9% agarose gel electrophoresis, and the fragment of interest was excised and purified using the GenElute Gel Extraction Kit (ThermoFisher). This was then treated with T4 DNA ligase, DpnI, and polynucleotide kinase (NEB) for 1 h at room temperature and then transformed into DH5α *E. coli* and plated on LB agar supplemented with appropriate antibiotic.

5.8 Induction and protein production

One litre of terrific broth containing appropriate antibiotic was inoculated with the bacteria from 1 plate of freshly transformed T7 SHuffle cells transformed with an appropriate plasmid. Flasks were incubated at 37 °C until the OD reached 0.5-0.8. Flasks were cooled on ice and induced with an appropriate concentration of isopropyl-thio-galactopyranoside (IPTG). Cells were allowed to grow overnight at 18 °C and then harvested by centrifugation at 5000 x *g*; 4 °C.

5.9 Protein buffer preparation

Buffers used for protein purification in this thesis:

Table 5. Composition of buffers used for the purification of recombinantly expressed proteins in this thesis.

Name	Components	Proteins purified with
Buffer 1	400 mM NaCl; 25 mM HEPES pH 8	MAP_{4185}; $MAP_{Inhibitor}$; $Ga0_{4185}$; $Ga0_{Inhibitor}$; $Ga0_{truncation}$; $NBras_{GH172}$
Buffer 2	400 mM NaCl; 25 mM HEPES pH 8; 500 mM imidazole pH 8	MAP_{4185}; $MAP_{Inhibitor}$; $Ga0_{4185}$; $Ga0_{Inhibitor}$; $Ga0_{truncation}$; $NBras_{GH172}$
Buffer 3	400 mM NaCl; 25 mM HEPES pH 8; 5% glycerol; 50mM l-arginine/ l-glutamic acid; 2 mM β-mercaptoethanol	Rv3707c_P30M; MSMEG_2107
Buffer 4	400 mM NaCl; 25 mM HEPES pH 8; 5% glycerol; 50mM l-arginine/ l-glutamic acid; 2 mM β-mercaptoethanol; 500 mM imidazole	Rv3707c_P30M; MSMEG_2107
Storage Buffer 1	200 mM NaCl; 25 mM HEPES pH 8	MAP_{4185}; $MAP_{Inhibitor}$; $Ga0_{4185}$; $Ga0_{Inhibitor}$; $Ga0_{truncation}$; $NBras_{GH172}$
Storage Buffer 2	400 mM NaCl; 25 mM HEPES pH 8; 5% glycerol; 50mM l-arginine/ l-glutamic acid; 2 mM dithiothreitol	Rv3707c_P30M; MSMEG_2107

5.10 Protein purification by Immobilised Metal Affinity Chromatography

Following harvest, cell pellets were resuspended in lysis buffer (as outlined above) to a final volume of 37 mL. DNAse A and lysozyme were then added to the cell suspension to a final concentration of 20 μg/mL and 50 μg/mL, respectively. After incubation on ice for 30 min, suspensions were lysed using either three passages through a French pressure cell, or by sonication (30 sec on; 45 sec off; 10 cycles). Following lysis, lysates were clarified by centrifugation at 40,000 x *g*. The resulting lysates were then loaded onto a column containing metal affinity chromatography resin and flowthrough collected. Columns were then washed with three column volumes (~ 60 mL) lysis buffer before elution using defined mixtures of the above buffers. Collected fractions were then assessed for purity using SDS-PAGE. Protein-containing fractions were pooled and dialysed against 3 L of storage buffer as above at 4 °C. Following dialysis, proteins were concentrated using Pierce™ protein concentrators (ThermoFisher Scientific). Proteins were then flash frozen in liquid nitrogen in aliquots for later use.

5.11 Preparative size exclusion Chromatography

Following initial purification, size exclusion chromatography was carried out on crude protein samples. Size exclusion was carried out on an Äkta Pure 25 system using a Superdex 26/600 200 pg column at a flow rate of 3 mL/min. Prior to chromatographic separation, the column was equilibrated with 1.5 column volumes of buffer. Fractions of 2 mL were collected using a fraction collector and pure fractions determined by SDS-PAGE.

5.12 Protein Crystallography

After polishing, fractions containing purified protein were pooled and concentrated to 10 mg/mL. The resultant protein solution was crystallised using sitting drop vapour diffusion. Briefly 2 µL of the protein solution was mixed with 2 µL of the buffer to produce the mother liquor. The tray was then allowed to equilibrate and upon successful crystallisation, the crystal was transferred a buffer identical to the mother liquor but for the addition of 30% glycerol or ethylene glycol.

5.13 X-ray diffraction data collection

5.13.1 Crystallography of Rv3707c

Rv3707c was concentrated to 7 mg/mL and loaded into a sitting drop crystallisation tray (Clover) using JCSG-plus crystallisation screen (Molecular Dimensions) at a ratio of 2 µL protein with 2 µL reservoir solution. Crystals were formed in a buffer containing 0.8 M succinic acid (pH 7) over a period of 1 month. The crystals were cryoprotected in a solution of 0.8 M succinic acid pH 7 with 30% glycerol and flash frozen in liquid nitrogen. X-ray diffraction data were collected at beamline I04 of Diamond light source, Oxford.

5.13.2 Crystallography of *N. brasiliensis* GH172

Protein was concentrated to a 10 mg/ml before loading into a sitting drop vapour diffusion tray (Clover) using the JCSG-plus crystallisation screen in a 4 µL drop containing a 1:1 mix of protein solution and mother liquor. Crystals formed in a condition containing 0.1 M Phosphate/citrate pH 4.2 40 % v/v PEG 300 in three days. Following crystallisation, crystals

were cryoprotected using the above buffer supplemented with 30% ethylene glycol. X-ray diffraction data were collected at beamline I04 of Diamond light source, Oxford.

5.13.3 Crystallography of MAB_{GH183}

Protein was concentrated to a 10 mg/ml before loading into a sitting drop vapour diffusion tray (Clover) using the JCSG-plus crystallisation screen in a 4 µL drop containing a 1:1 mix of protein solution and mother liquor. Crystals formed in a condition containing 0.15 M Potassium bromide 30 % w/v PEG 2000 MME. Following crystallisation, crystals were cryoprotected using a final concentration of 30% glycerol before collection of diffraction data at beamline I04 of Diamond light source, Oxford.

5.13.4 Crystallography of $MAP_{complex}$

Following analytical size exclusion chromatography on an AKTA Pure 25 system equipped with a Sephacryl 26/600 200 pg column, fractions from the centre of the complex peak were pooled and concentrated to 6 mg/ml and loaded into a sitting drop vapour diffusion tray (Clover) using the BCS crystallisation screen in a 4 µL drop containing a 1:1 mix of protein solution and mother liquor. Crystals formed in a buffer condition containing 0.05 M Ammonium acetate, 0.15 M Magnesium sulphate heptahydrate, and 0.1 M HEPES 7.0; 12% v/v PEG Smear Medium. Crystals were cryoprotected in a 70:30 mix of this buffer with ethylene glycol before collection of diffraction data at beamline I04 of Diamond light source, Oxford.

5.13.5 Phasing and structural refinement

Data were auto-processed using Xia2 and general file manipulations were performed using the CCP4 suite of programs (Winn *et al.*, 2011). Data were phased by molecular replacement using the Alphafold structural prediction of Rv3707c as a search model using PHASER (TFZ of 21.1) (Liebschner *et al.*, 2019). The structure was then auto built in PHENIX with rounds of refinement carried out by PHENIX-refine and manually in COOT (Emsley and Cowtan, 2004) . Final models were then submitted to the PDB-REDO web server to carry out structural adjustments to optimise for PDB deposition (Joosten *et al.*, 2014).

5.14 *Mycobacterium smegmatis* cell wall preparation

Large scale purification of mycobacterial arabinogalactan was achieved by established methodologies. Briefly, 8 L cultures of appropriate mycobacteria were grown to mid logarithmic phase then pelleted and washed once in PBS. The resulting cell suspension was lysed using an Emulsiflex. Following lysis, lysate was heated to 100 ℃ following the addition of SDS to a final concentration of 1% and refluxed for one hour. The insoluble material was then lyophilised and washed 10-15 times to remove SDS. Mycolic acids were subsequently removed from the material by saponification using 1% potassium hydroxide in methanol at 37 ℃ for three days. Free fatty and mycolic acids were then washed by three washes with diethyl ether. The phosphopdiester linker unit between PG and AG was then hydrolysed using 0.1 M H_2SO_4 and heating the mixture to 95 ℃. Following cooling, the mixture was neutralised using anhydrous sodium carbonate. The free arabinogalactan was then isolated through collection of the supernatant through centrifugation. The solution was desalted using 3.5 kDa molecular weight cutoff dialysis tubing against 5 litres of deionised water overnight. The arabinogalactan was then lyophilised and yielded around 22.5 mg/ litre of cells collected.

5.15 Mycobacterial glycolipid extraction

Cultures of mycobacterial species were grown in appropriate media to an OD_{600} of 0.5-0.8 and harvested by centrifugation at 5000 x *g*. Cells were then washed 3x in sterile PBS supplemented with 0.05% Tween-80 and sonicated using 5 min total process time consisting of 30 s ON; 1 mins OFF. The lysate was then transferred to a Teflon-capped glass centrifuge tube and incubated at 65 °C with an equal volume of 90% phenol (Merck) for 3 h, with the mixture vortexed at hourly intervals. The suspension was then centrifugated at 3000 x *g* for 15 min and the aqueous upper layer was transferred to a fresh glass centrifuge tube and phenol washes were repeated two further times. The purified glycolipids were then dialysed exhaustively against tap water overnight and then into deionised water for 3 hours using 3.5 kDa molecular weight cutoff dialysis membrane (Servapor). The resulting glycolipids were then frozen and lyophilised for 3 h.

5.16 *Pseudomonas aeruginosa* PA1596 pilin glycan preparation

Plates of LB agar were inoculated with liquid culture of *Pseudomonas aeruginosa* at an OD of 0.1-0.5. The plates were grown at 30 °C overnight. Plates were then scraped into 2 mL/plate of ice-cold PBS containing COmplete protease inhibitor tablets (Roche) and 0.02% sodium azide. The resulting suspension was transferred into 50 mL falcon tubes and vortexed at max speed for 3 bursts of 1 min, cooling on ice between mixing. Cells were pelleted at 4000 x *g* and then the supernatant transferred into Oak Ridge centrifuge tubes and insoluble debris removed by centrifugation at 16,000 x *g* for 40 min at 4 °C. The clarified solution containing sheared pili was filter sterilised and pilins were precipitated using 30% v/v acetone. The

resultant cloudy suspension was centrifuged as previously, and pilins collected as a smudge in the bottom of the tube. Pilin glycans were liberated by β-elimination using 1 M sodium hydroxide at 37 °C overnight. Pilins were then isolated using porous graphitic carbon chromatography as below.

5.17 Porous graphitic carbon chromatography

Hypercarb PGC columns (ThermoFisher Scientific) were washed in 50% tetrahydrofuran in water followed by 80% acetonitrile in water, using a syringe and a vacuum manifold to accelerate flow. The column was then exhaustively equilibrated using water (10 column volumes minimum). Following equilibration, sample was loaded onto the equilibrated column and flowthrough collected. The column was then washed using 5 column volumes water and then sample eluted using 50% acetonitrile in water with 0.1% trifluoroacetic acid.

5.18 Polysaccharide enzymatic digestion assays

Purified protein solutions at a known concentration (1-5 μM) were incubated with substrates at a final concentration of 1 mg/mL for a defined time at 37 °C. At the end of the assay, the reaction was terminated either by addition of sodium hydroxide to a final concentration of 5 mM or by boiling for 10 min.

5.19 Thin layer chromatography of carbohydrates

For the analysis of reaction products 10 μL of sample was spotted 1 cm from the base of a silica gel sheet60 F_{254} (Merck). The TLC was then developed twice in a system of 2:1:1 (butan-1-ol: acetic acid: water). After drying, TLCs were stained with a solution of 500 mg 4-

methylresorcinol dissolved in 100 mL 70% ethanol in water, acidified with 10 mL concentrated sulfuric acid. TLCs were charred using a heat gun.

5.20 High performance anion chromatography with pulsed amperometric detection (IC-PAD)

Oligosaccharides produced by enzymatic treatment of polysaccharide digestions were analysed using an ICS-6000 ion chromatography system fitted with a CARBOPAC PA-300 analytical colum (Thermo Fisher Scientific). Oligosaccharides were detected using pulsed amperometric detection using the standard Carbo Quad waveform. The following program was used for the separation of oligosaccharides. 0-10 mins 100% Buffer A (100 mM NaOH) followed by 60 minutes of a linear gradient increasing to 60% Buffer B (100 mM NaOH; 500 mM $NaCOOH_3$) 70-80 minutes was 100% Buffer B. the final 20 minutes of the prpgram, the column was washed with 500 mM NaOH before a 10 minute recalibration into buffer A. L-arabino-oligosaccharides (DP = 2-8) obtained commercially (Megazyme) were used as standards at a concentration of 25 µM. Data were processed using Chromeleon™ Chromatography Management System V.6.8. Final graphs were created using GraphPad Prism 10.0.1

5.21 Isothermal titration calorimetry

All ITC experiments were carried out on a Malvern Panalytical MicroCal PEAQ-ITC instrument. The reference cell temperature was set to 25 °C with a reference power of 0.1 µcal/min and a needle rotation of 250 RPM. Unless otherwise stated, concentration of protein in the chamber was 20 µM and in the needle was 200 µM. Each experiment comprised 20x 10

microlitre injections with a 1.5-min delay between injections carried out in technical triplicate. All analysis was carried out on Malvern Panalytical ITC software.

5.22 Analytical size exclusion chromatography of protein complexes

Concentrations of protein solutions were assayed using the ThermoFisher Scientific BCA assay following manufacturer's instructions. Samples were then diluted to 150 μM and then mixed 1:1 with inhibitor proteins or buffer (25 mM HEPES; 400 mM NaCl), to a final concentration of 75 μM. Samples were then injected using a 100 μL loop and subjected to size exclusion chromatography using an Akta Pure 25 system equipped with a Superdex 75 increase 10/300 GL column (Cytiva), at a flow rate of 0.8 mL/min, equilibrated with with Storage Buffer 1. Abs_{280} readings were taken at a rate of 5 Hz.

5.23 Nocardia species carbon source assay

Starter cultures of *N. brasiliensis* ATCC 19296*, N. farcinica* ATCC 3308*, and N. asterioides* ATCC 19247 were first grown in Brain-Heart Infusion Medium (Merck) supplemented with 0.05% Tween-80. These were diluted into Sauton's media supplemented with 1% glycerol and 0.05% Tyloxapol and allowed to grow overnight to an OD_{600} 0.6-0.8. Cultures were pelleted and washed 3x in sterile PBS + 0.05 % tyloxapol and OD_{600} measured. Cultures were then diluted to an OD_{600} of 0.05 in Sauton's media containing no carbon source and 0.05% tyloxapol and 99 μL/well inoculated into a sterile 96 well plate, the wells had been primed with 1 μL of a 50% solution of appropriate carbon source. The plate was wrapped in parafilm and incubated at 37 °C for 21 hours. At this point 30 μL of a 0.15 mg/mL solution of Resazurin in 0.05% tyloxapol was added to each well and the plate was incubated for a further 3 hours at 37 °C.

Absorbance was measured on a Clariostar plate reader (BMG Lab Tech). Data were plotted using GraphPad Prism 10, each value is representative of 8 technical replicates, and each experiment was carried out in biological duplicate.

5.24 Spot dilution assays

Mid to late log bacterial suspensions were prepared to an OD_{600} of 0.8 and then diluted serially 1 in 10 in a sterile Greiner 96-well plate. Dilutions 3 to 9 were spotted onto 1.5% agar plates supplemented with appropriate antibiotics and stressors and allowed to grow.

5.25 Proteomics

N. brasiliensis was grown to an OD_{600} of 1 and then washed twice in PBS with 0.05% tyloxapol, the bacteria were then disaggregated using a fine gauge needle and luer syringe and sonication for 30 seconds. A suspension of washed bacteria was then prepared in 2x Sautons media to an OD_{600} of 0.1. This was then mixed 1:1 into filter sterilised stock solutions of carbon source prepared in ddH_2O (arabinogalactan 20 mg/mL; glucose 20 mg/mL; glycerol 1%) to yield a bacterial suspension with an OD_{600} of 0.05 and final concentrations of arabinogalactan, glucose and glycerol of 10 mg/mL, 10 mg/mL, and 0.5%, respectively. These were then incubated at 30 °C until late stationary phase. The bacteria were then pelleted by centrifugation (30 °C; 4000 x *g*; 10 mins) and washed once in 500 μL of PBS. The bacteria were then pelleted once again and resuspended in 600 μL 100 mM Tris-HCL and bead beaten twice at a velocity of 6 m/s for 45 seconds, cooling on ice for 2 minutes between cycles. 500 μL of this lysate was transferred into a protein lobind tube (Eppendorf). SDS was added to this at a final concentration of 4% in 720 μL and the samples were boiled with shaking for 20 minutes

to fully denature the proteins. The proteins were then transferred into 15 mL Falcon tubes containing 80 μL 1 M NaCl, and proteins were precipitated by addition of 3.2 mL frozen acetone and tubes were incubated for 2 days at -20 °C. Precipitated proteins were then pelleted by centrifugation at 3 °C and were washed with ice cold acetone once. Following this the proteins were dried using a Speedvac at room temperature.

6 Recombinant expression and purification of endogenous GH183 enzymes from Mycobacteriales

6.1 Contribution statement

The pET28a-based constructs used in results section 5.3 were cloned by Abigail Layton. Portions of this chapter have been published in a recent article from our laboratory (Al-Jourani *et al.*, 2023)

6.2 Identification of putative GH183 enzymes within the genome of *M. tuberculosis*

After the successful identification of D-arabinan active glycoside hydrolases encoded in the genome of *Dysgonomonas gadei*, we reasoned that it is likely that these genes may play functions within the lifestyles of bacteria that possess arabinogalactan. To identify potential homologues of HMPREF9455_02480 and HMPREF9455_02481, we carried out BLAST analysis for homologs in mycobacteria. Later in the project, with the release of the Alphafold database (AFDB) and FoldSeek, we revisited this analysis to confirm homologs were not missed in our initial screen. Using the 3-dimensional structural prediction of HMPREF9455_02480 as a search model, we restricted searches to *M. tuberculosis* H37Rv.

Table 6. Table summarising the top 6 structural homologues of HMPREF8455_02480 encoded in the genome of *M. tuberculosis H37Rv*. Structural homologues were identified using Foldseek using the Alphafold structural prediction of HMPREF9455_02480 as the search model. Probability was calculated by FoldSeek using the expected bit score of a coincidental match compared to the bit score distribution of the experimental structure, as outlined in Van Kempen et al, 2023*(van Kempen* et al*., 2023)*

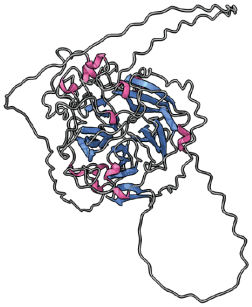
Target	Description	Prob.	Seq. Id.	E-Value	Score	Query Pos.
Rv1754c	Conserved protein	1	27.2	2.56E-28	996	15-374
Rv3707c	DUF4185 domain-containing protein	1	20.3	3.45E-15	511	47-370
Rv1502	Uncharacterized protein Rv1502	1	11.5	2.04E-08	239	45-372
Rv3748	Dioxygenase	1	9.8	2.40E-04	90	72-372
LppL	Probable conserved lipoprotein	0.9	8.8	5.09E-03	67	15-370
LppZ	Probable conserved lipoprotein	0.9	10.8	6.63E-03	67	2-357

As can be seen in Table 6, the genome of *M. tuberculosis* encodes 2 distinct structural homologues of HMPREF9455_02480 with a probability of 1.00 and e values higher than $1e^{-10}$. The accessions for these genes are Rv1754c and Rv3707c. Excitingly, both these proteins contain the domain of unknown function 4185 (DUF4185), which is also found in *D. gadei* homologues. Therefore, they are likely to encode gene products which catalyse similar reactions. Lower confidence hits in the results encode proteins with five-bladed beta-propeller structures, these are likely involved in synthesis, turnover, or binding of other carbohydrate moieties found within mycobacteria. Rv1502 is one such example, whose *M. marinum* homologue has been shown to catalyse the synthesis of lipooligosaccharide (Chen *et al.*, 2016). These were therefore not pursued in more depth.

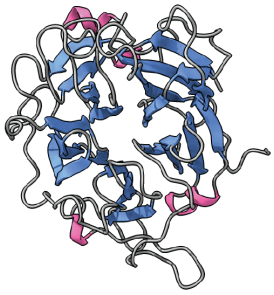
A major benefit of this methodology over sequence identity methodologies such as BLAST is that it identifies conserved protein structural motifs independent of their primary structure. As can be seen in Figure 7c and d, whilst both the amino acid and nucleotide identity of all

proteins identified is low, the structural homologues identified by Foldseek possess a very high degree of similarity in their tertiary structures, all possessing RMSD values lower than 1.2Å. Therefore, we set out to biochemically characterise these mycobacterial homologues to verify that they possess D-arabinanase activity.

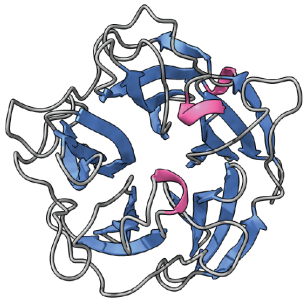
A.



Rv1754c
RMSD: 0.793 Å

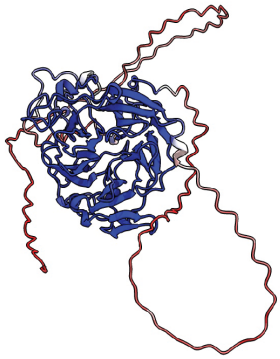


Rv3707c
RMSD: 1.168 Å

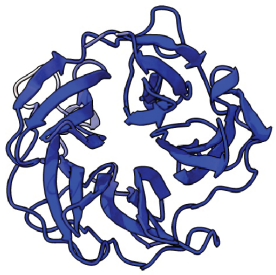


Rv1502
RMSD: 1.160 Å

B.









C.

	Rv0782	Rv1502	Rv1754c	Rv3435	Rv3707c
Rv0782		22.046	31.778	20.071	22.373
Rv1502	22.046		19.955	15.697	21.586
Rv1754c	31.778	19.955		17.057	21.165
Rv3435	20.071	15.697	17.057		28.402
Rv3707c	22.373	21.586	21.165	28.402	

D.

	Rv0782	Rv1502	Rv1754c	Rv3435	Rv3707c
Rv0782		8.289	10.165	7.820	5.760
Rv1502	8.289		6.958	9.135	5.155
Rv1754c	10.165	6.958		5.793	10.968
Rv3435	7.820	9.135	5.793		4.928
Rv3707c	5.760	5.155	10.968	4.928	

Figure 6. Comparison of outputs generated by bioinformatic searches for homologues of HMPREF9455_02480. A. Side by side comparison of top hits generated by Foldseek structural homology search. RMSD have been calculated using Chimera X matchmaker using Rv1754 as the reference scaffold. Values displayed are the paired RMSD of the regions where the proteins align, not full length. B. Alphafold structural predictions coloured by pLDDT (Jumper et al., 2021). C/D. Nucleotide and amino acid sequence identity comparison of selected hits. Alignment grid produced using MUSCLE alignment. (Edgar, 2004).

6.3 Recombinant expression of Rv3707c

A pET28a vector containing the C-terminal His-tagged annotated genomic locus of Rv3707c (annotated amino acids 1-336) was generated using restriction-free cloning. Cell pellets resulting from a 3-hour induction using 250 μM IPTG were first lysed by sonication in Buffer 1 supplemented with a COmplete protease inhibitor tablet (Roche). The fractions were purified using an imidazole gradient generated by mixing of Buffers 1 and 2 in increasing proportions on a gravity column containing a bed of Ni-NTA resin (Thermo Fisher Scientific). The expected molecular weight of the protein was 36.482 kDa. The protein appeared to express very poorly, as can be seen in Figure 8, with no protein being visibly over-expressed in the whole cell lysate lane. As highlighted by the red box over the 100 and 250 mM Imidazole elutions, there was a very small amount of protein of roughly 37 kDa eluting in these fractions. Regardless, the protein was left to dialyse overnight against 3 litres of 25 mM Tris-HCl pH 8 with 400 mM NaCl at 4 °C, but the protein was insoluble and had precipitated. For the purposes of structural biology and efficient biochemistry we wanted to increase the solubility and therefore yield of the recombinant protein, therefore we set about identifying potential truncations of the protein we could generate to produce protein with more favourable characteristics. To ensure that the correct start site had been identified, we carried out thorough bioinformatic analysis of the genetic locus to clarify its characteristics.

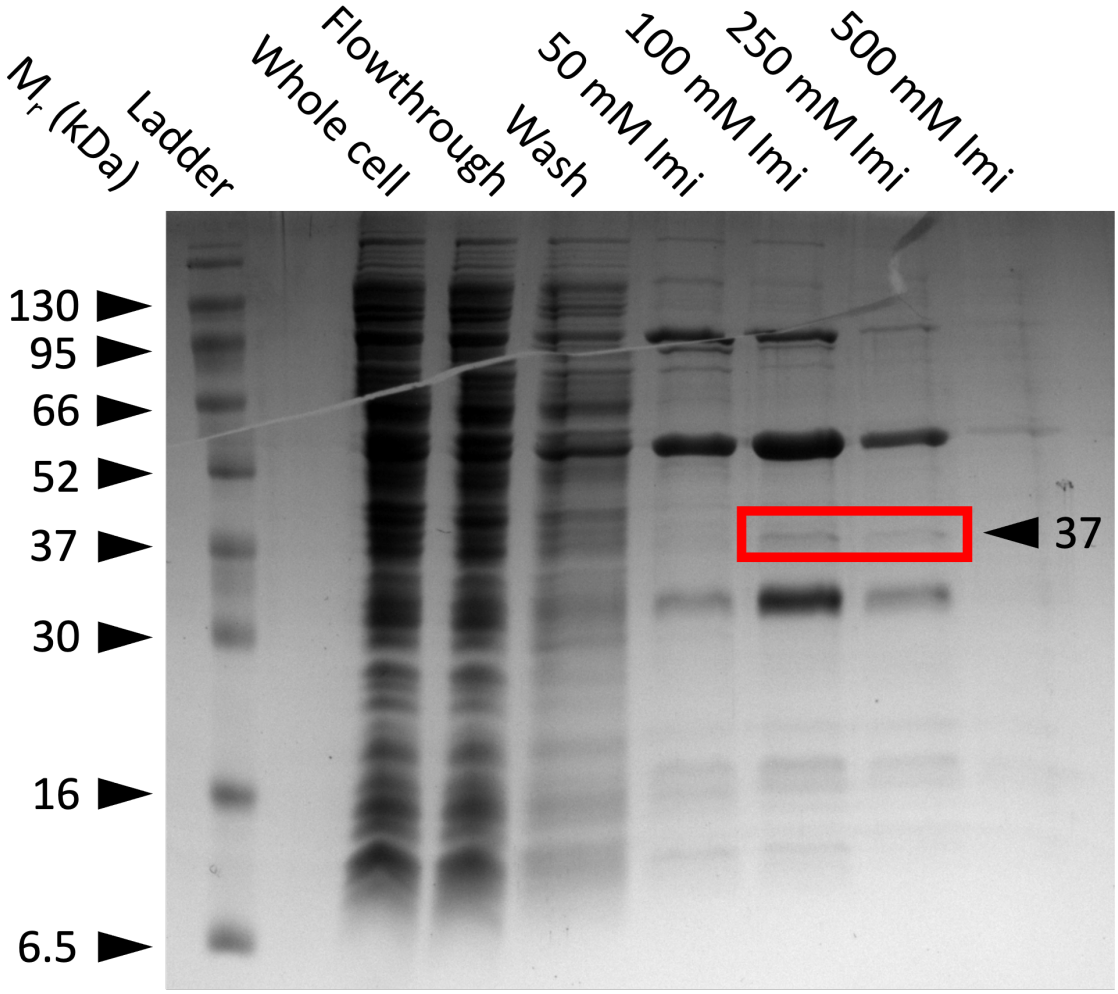
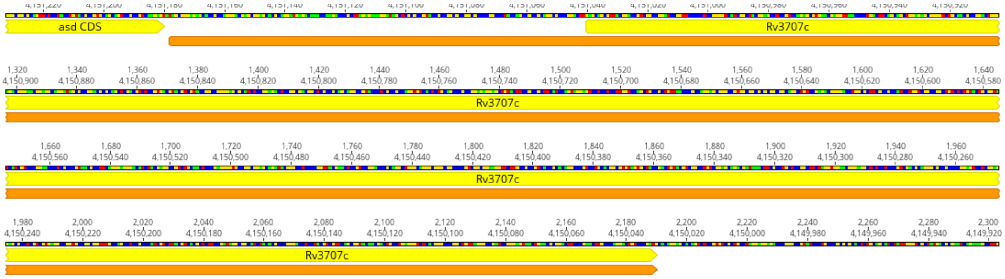


Figure 7. SDS-PAGE electrophoresis gel of a protein purification of Rv3707c expressed from a pET28a vector. Cell pellets were harvested from one litre cultures induced overnight at 18 °C with 250 μM IPTG. Cells were lysed using a French pressure cell for three passages before clarification of the lysate through centrifugation. Processing of lysate was carried out on a drip column with Ni-NTA resin using Buffer 3, and an imidazole gradient generated through mixture of increasing proportions of buffer 4 diluted into buffer 3.

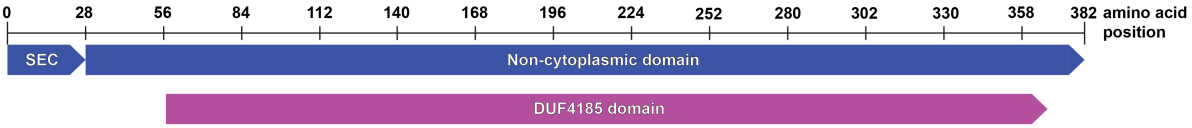
6.4 Identification of full length Rv3707c

To exert D-arabinanase activity within the cell, we hypothesised that Rv3707c would require a secretion signal to be exported into the periplasm where it can then catalyse the breakdown of arabinogalactan. Based on the current annotation of the *M. tuberculosis* Rv3707c locus, there was no canonical secretion signal encoded, despite experimental proof of its secretion *in vivo* (Perkowski *et al.*, 2017). Therefore, we searched for alternative start sites within the region to find a correct start site. As shown in in Figure 9a, ORFinder identified multiple different start sites within the region. Interestingly, upon closer analysis of the ORF outlined in red in Figure 9a using Interproscan, the new ORF appeared to contain both a canonical SEC signal peptide, as well as a proline-rich linker region between the signal peptide and the DUF4185 region (Figure 9b) (Jones *et al.*, 2014). This was further supported by SignalP 6.0 analysis of the sequence in Figure 9c (Teufel *et al.*, 2022). We therefore hypothesised that the protein we had previously cloned had a part of this proline-rich linker region which would contribute to poor solubility due to translational stalling towards the start of translation (Brewer and Wagner, 2022; Krafczyk *et al.*, 2021). To further improve our purification efforts we had the DUF4185 portion of Rv3707c codon-optimised for *E. coli* and synthesised in a pCOLD backbone by Twist Biosciences, which we designated Rv3707$_{\Delta 59}$(Sugiki *et al.*, 2017).

A.



B.



C.

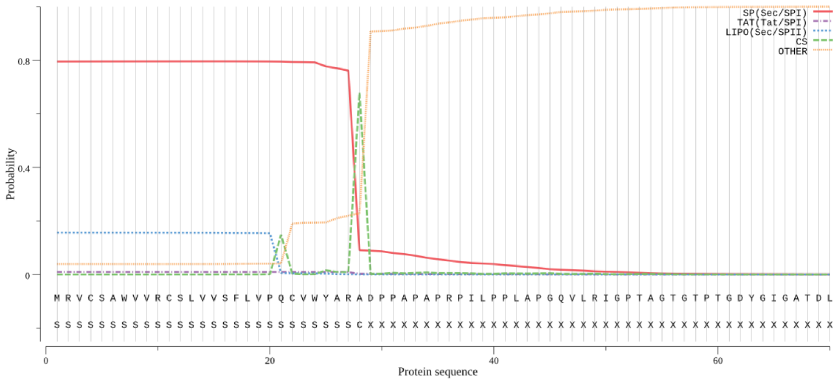


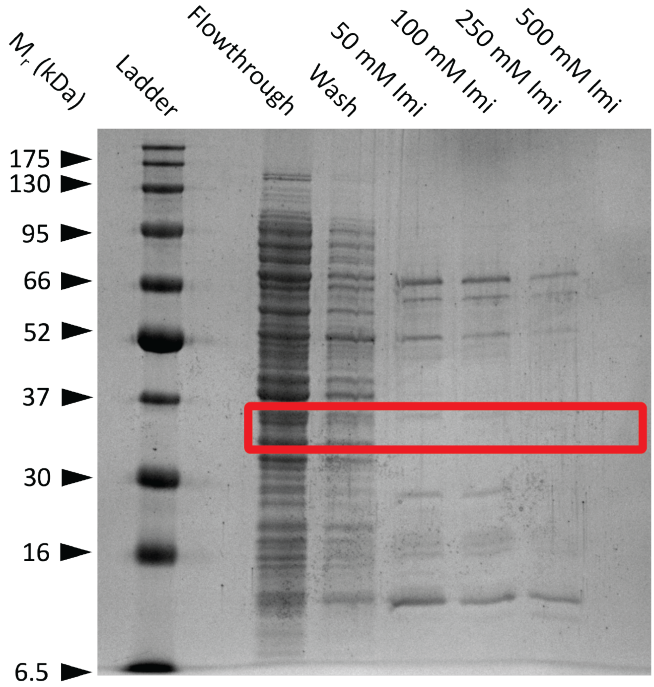
Figure 8. Bioinformatic analysis of the locus encoding Rv3707c. A. comparison of annotated ORF encoding Rv3707c, with proposed different start site in orange. B. output resulting from InterproScan using the full-length protein sequence from A, including signal peptide and linker region. C. SignalP-5.0 output for Rv3707c, identifying it as a Sec signal peptide.

6.5 Removal of proline-rich region does not stabilise Rv3707c in solution

Protein purification was attempted using cell pellets from 1 L cultures of BL21-DE3 *E. coli* transformed with pCOLD-Rv3707c$_{\Delta 59}$ which had been induced overnight at 4 °C. This construct was selected as it is reported to produce poorly folding proteins in good yield (Sugiki *et al.*, 2017). The pellets were lysed in Buffer 1 and then processed by IMAC on a gravity column containing Ni-NTA resin, then eluted using imidazole gradient generated by mixing buffers 1 and 2 together in increasing proportions. Figure 10 displays the SDS-PAGE gel generated from the IMAC purification of Rv3707$_{\Delta 59}$. The expected M_R of the truncated protein was 34.188 kDa. Unfortunately, no soluble protein was produced.

Following discussions with Prof. Andrew Lovering, we hypothesised that the proline-rich region of Rv3707c could be important structurally, as such regions can sometimes act as 'zipper' motifs which maintain structural integrity of proteins (Gachon *et al.*, 2011). Therefore, using TBLASTN against the Mycobrowser plus database, an alignment of the N terminal region of Rv3707c and its homologues was generated. Despite the protein often being truncated, it could be seen from Figure 10b that the prolines were deeply conserved, particularly prolines 52 and 59, both of which had been removed in the Rv3707c$_{\Delta 59}$ construct. Due to the deep structural conservation of these residues within the protein we reasoned that the protein produced must be destabilised due to loss of this N-terminal proline-rich region that stabilises the protein by some mechanism. We therefore had further codon-optimised pET28a plasmids synthesised by Twist Biosciences, this time with 3 different start points, at M1, L13, and P30. These three start points represent full-length, signal peptide truncation, and a small N-terminal truncation, respectively.

A.



B.

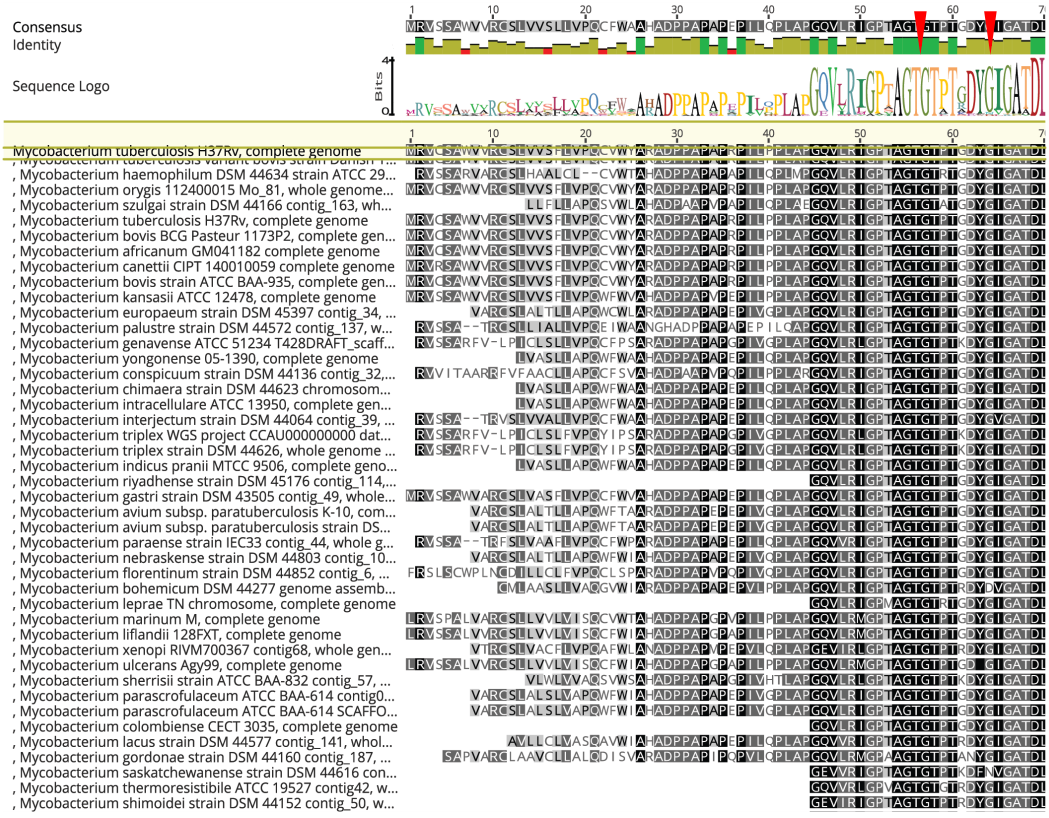


Figure 9. A. SDS-PAGE gel analysis of an attempt to purify Rv3707$_{\Delta 59}$ From cell pellets after cold shock induction overnight, protein was purified in Buffer 3. B. MSA of the N terminal proline rich region using BLAST against the Mycobrowser Plus database, figure generated using Geneious. Proline region highly conserved.

6.6 Rv3707c can be produced recombinantly using a 30 amino acid N-terminal truncation

An aspect of protein purification that we had not yet optimised in purification of Rv3707c was buffering conditions. As shown in Figure 11a, we carried out protein solubility trials from 10 mL cell culture pellets lysed in Tris-HCl (pH 8.0) with 300 mM NaCl as well as the supplements below. Rv3707c_M1 has an expected M_R of 38.058 kDa, and as shown it is expressed reasonably well during induction. However, the intact Sec secretion signal means that the protein is likely being exported from the cell and is therefore not present in the lysate samples due to pelleting of the cells and subsequent removal of the protein-containing supernatant. Rv3707c_L13M appears to not be as well expressed as M1 version, with no notable band in the post induction lane. However, a notable band of protein can be observed in the post-induction sample of the P30M construct. This is of the correct molecular weight and is not present in the pre-induction sample, meaning its presence is likely due to induction and not another cellular process. Additionally, the presence of 5% glycerol in the lysis buffer appears to efficiently solubilise the protein with a significant proportion of it in the glycerol lane, whilst the other buffer components do not as efficiently extract the protein into the soluble fraction.

To optimise induction conditions, we also carried out a trial expression using autoinduction media. The lysates were this time prepared in Tris-HCl (pH 8.0) with 300 mM NaCl and 5% glycerol. All sugar mixtures appeared to produce the protein in good yield, and therefore we proceeded to purify Rv3707c using BL21-DE3 cells transformed with pET28a_Rv3707c_P30M.

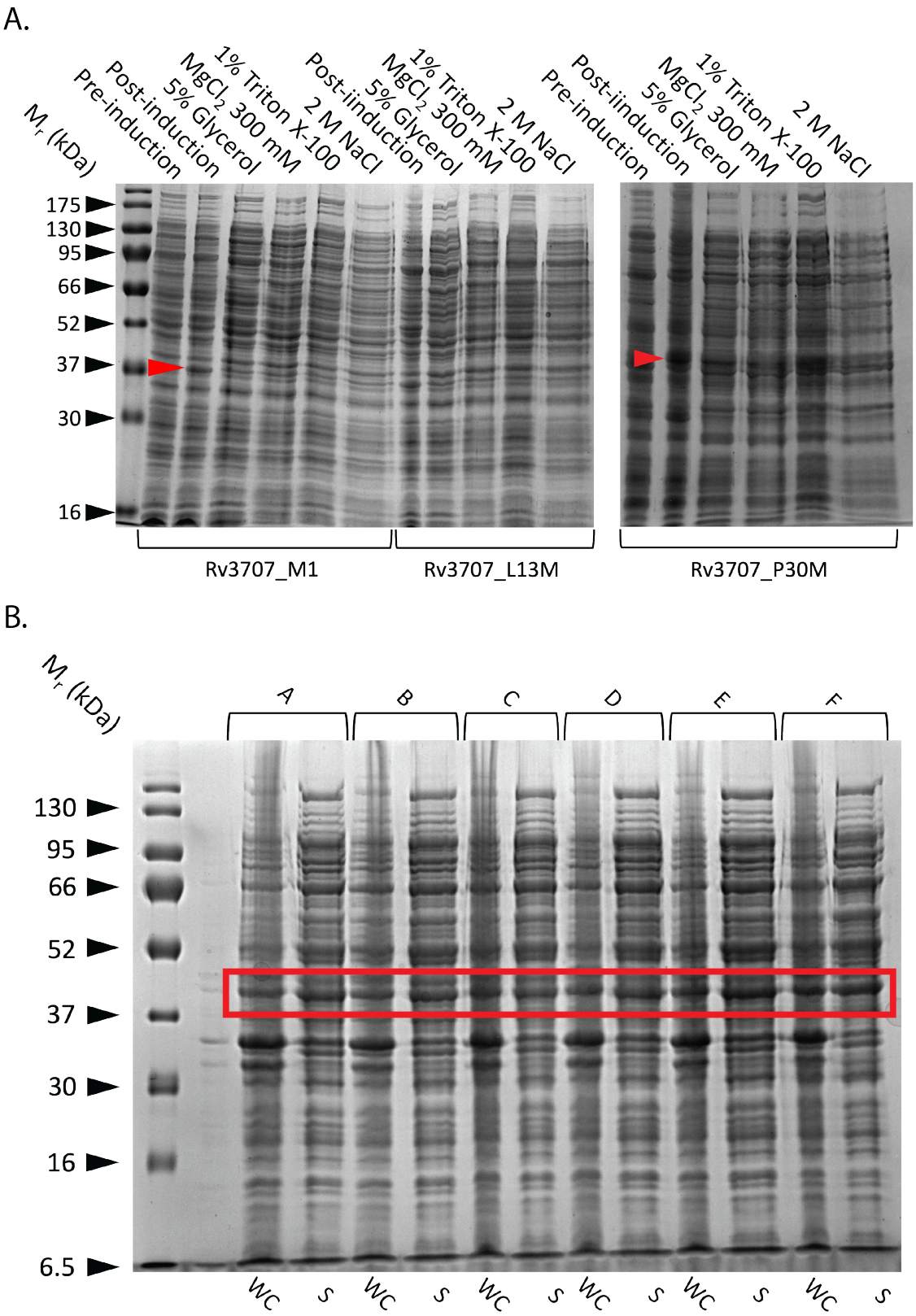


Figure 10. Induction trials of RV3707c_M1, Rv3707c_L13M, and Rv3707c_P30M. A. SDS-PAGE analysis of trial induction of protein expression from T7 Shuffle E. coli transformed with respective pET28a construct and induced overnight at 18 °C overnight with 250 µM IPTG. B. SDS-PAGE analysis of an induction trial using autoinduction media from BL21-DE3 cells transformed with respective plasmids A-E refers to different sugar solutions used for autoinduction.

6.7 Large-scale production of Rv3707c_P30M

Following the successful expression of the 30 amino acid truncation of Rv3707c_P30M in small-scale cultures, we grew one litre cultures of the protein using similar conditions. Figure 12 summarises SDS-PAGE analysis of fractions obtained following IMAC. After lysis and IMAC in the optimised buffer we analysed yield and found that the protein expressed with reasonable yield. However, the protein bound poorly, with a significant amount of protein in the flowthrough and wash fractions. Additionally, most of the protein eluted in the 50 mM imidazole fraction and was therefore extensively contaminated with non-target proteins. Therefore, we pooled the 100- and 250 mM fractions (purity of 51 and 75%, respectively) and dialysed them overnight. Unfortunately, the protein was not soluble for long periods of time, we therefore supplemented all buffers with 2 mM β-mercaptoethanol and 50 mM L-arginine and L-glutamic acid, to produce Buffers 3 and 4. These buffer components were chosen due to the presence of an unpaired cysteine in RV3707c_P30M, which likely needed to be reduced to carry out its function. Additionally, L-arginine and L-glutamic acid have been reported to stabilise proteins in solution (Shukla and Trout, 2011). The inclusion of these components allowed the protein to be stably purified without precipitation during overnight dialysis. Protein aliquots were then flash frozen in liquid nitrogen and then stored at -80 °C until use.

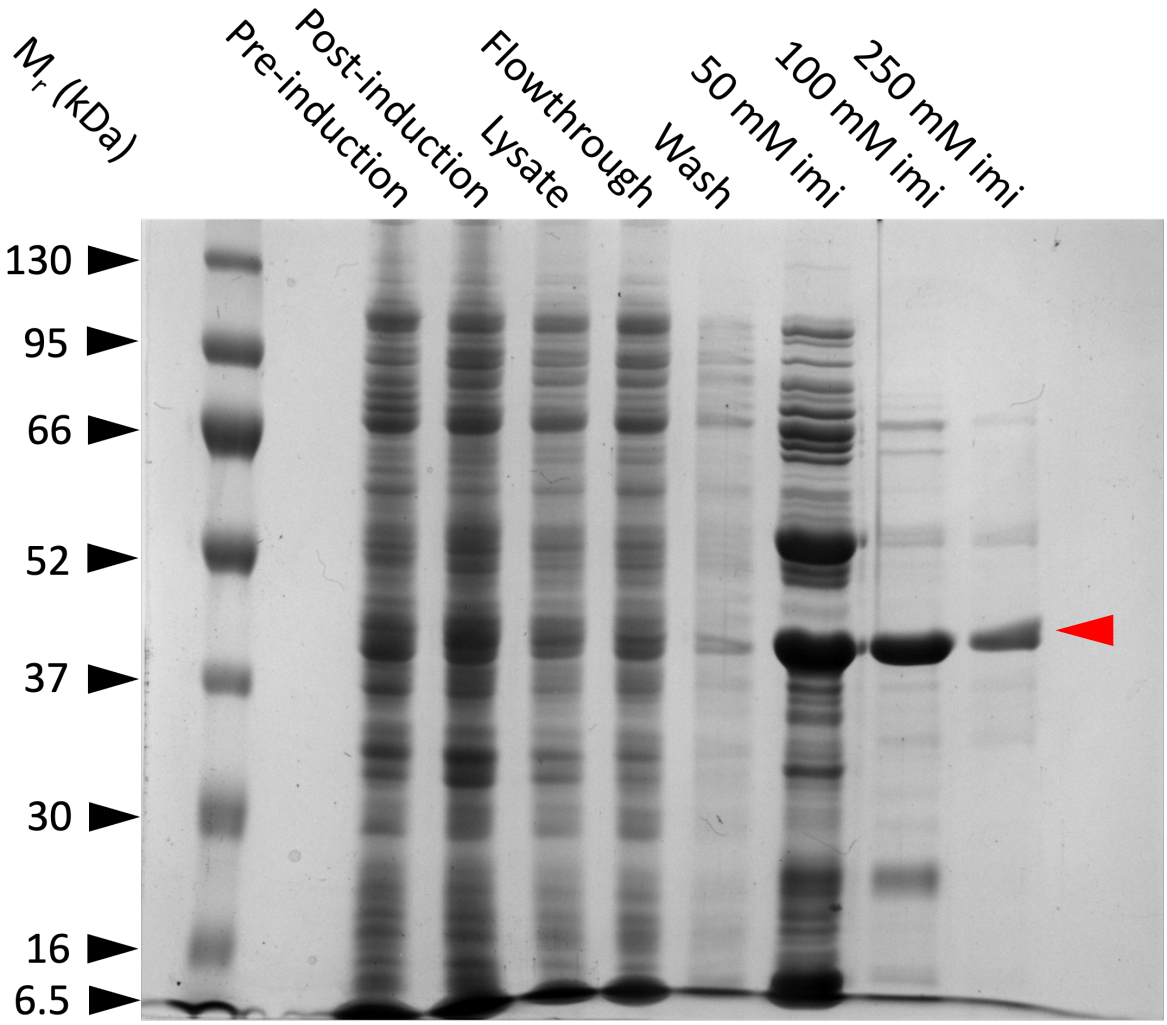
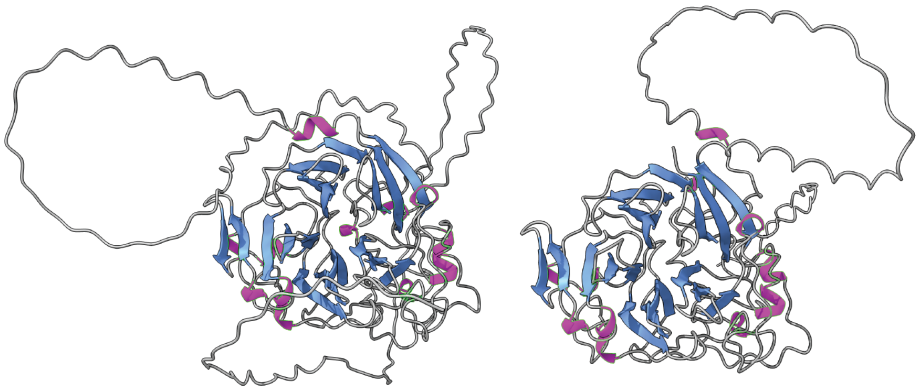


Figure 11. SDS-PAGE analysis of fractions resulting from IMAC purification of Rv3707_P30M from 1 litre cultures. Lysate was prepared by three passages through a French Pressure cell before centrifugation and fractionation by IMAC on nickel resin. Protein elutes reasonably cleanly in 100 and 250mM fractions, shown using the red indicator.

6.8 Identification of MSMEG_2107

Attempts to purify Rv1754c from any construct were unsuccessful. Therefore, we sought homologues in other mycobacteria which may have better solubility. In Figure 13a, the very long disordered N-terminal loop domain can be seen looping around the predicted structure. This is common to all homologues of Rv1754c we could identify. Initial attempts at purifying the protein made use of a construct without the N-terminal domain of the H37Rv homologue, but seemingly the N-terminal domain plays a structural role, but its presence during recombinant expression caused irreversible association with the membrane (data not shown). Therefore, we ordered a synthetic pET28a plasmid containing MSMEG_2107 excluding it's signal peptide (amino acids 28-441), as the 2107 N-terminal disordered domain was shortest of any mycobacterial species and would likely be most favourable in protein purification.

A.



Rv1754c MSMEG_2107 RMSD= 0.54 Å

	MSMEG_21...	Rv1754c
MSMEG_2107		68.546%
Rv1754c	68.546%	

Figure 12. Structural and sequence comparison of Rv1754c and MSMEG_2107. A. Side-by-side comparison of the Alphafold structural predictions of Rv1754c and MSMEG_2107, coloured by secondary structure (blue= strand; pink= helix; grey= coil) RMSD calculated using Chimera X MatchMaker and RMSD calculated based on paired atoms.

6.9 Recombinant expression and purification of MSMEG_2107

Following induction, MSMEG_2107 was purified using IMAC and then SEC to further purify the protein. The predicted M_R of the expressed protein was 49 kDa and the protein was expressed well upon induction, as shown in the band in the post-induction lane in Figure 14a. As seen in Rv3707c_P30M, the protein was poorly retained by the nickel resin, with equal amounts of the target protein seen in the lysate and flowthrough, this significantly decreased yield. Relatively clean protein (purity of 47.7%) was obtained in 50-500 mM imidazole elutions. However, a double band was observed for the target protein, which may be due to a disulfide bridge in the protein not being completely reduced during sample preparation. To obtain protein of a higher purity, the dialysed protein containing fractions were concentrated and polished using SEC. Protein containing peaks were analysed for purity by a second SDS-PAGE as shown in Figure 14b. The protein obtained was more than 95% pure but was contaminated with a very small amount of a protein of roughly 20 kDa, which was likely dimerising in solution to cause it to co-elute with the 39 kDA target protein. Despite this, the protein was frozen and stored for biochemical analysis later.

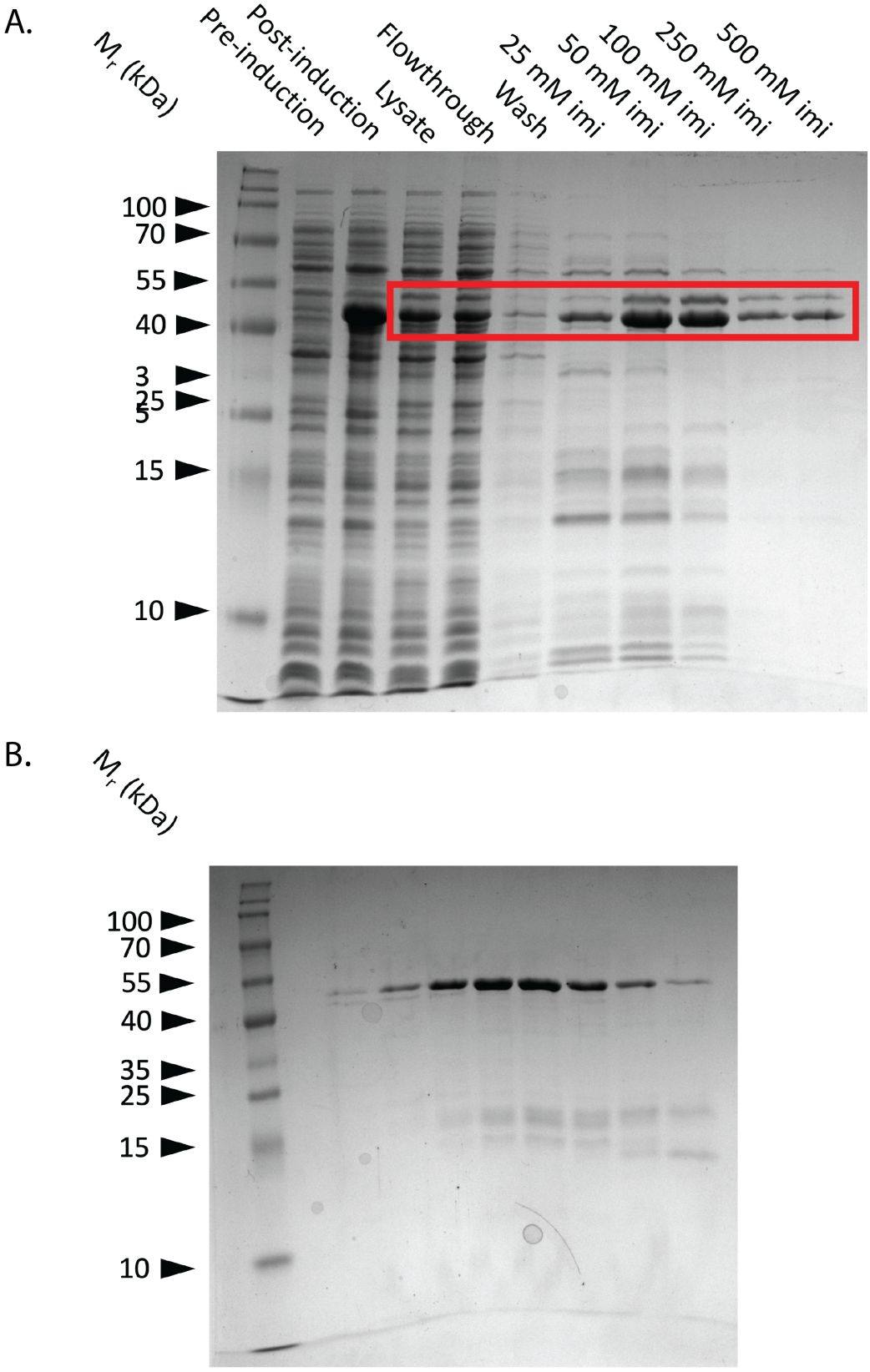


Figure 13. Purification of MSMEG_2107. A. IMAC of MSMEG_2107 following autoinduction. Protein purified in buffer 3. B. SDS-PAGE analysis of flowthrough from size exclusion chromatography of MSMEG_2107

6.10 Rv3707c_P30M is hydrolytic against D-arabinan-containing polysaccharides

Rv3707c_P30M was incubated with lipoarabinomannan and arabinogalactan purified from *M. smegmatis* overnight. Reaction products were visualised using orcinol stained TLCs and compared to commercially available L-arabinofuranose oligosaccharide standards. As shown in Figure 15, activity overall was weak, but notable activity was observed against LAM. Rv3707c produced two distinct groups of oligomers either 3 or 4 units in length (Rf of 0.6 and 0.53, respectively). Against arabinogalactan, activity was not readily discernible. However, oligosaccharides could be observed in the region of 6 and greater monosaccharides in length in both the enzyme treated and substrate only controls. We hypothesised that these contaminating oligomers are likely the products of the endogenous DUF4185 enzymes found within *M. smegmatis*. Therefore, these products likely co-migrate with reaction products from the recombinant enzymes. To remove this contaminating substrate, we used porous graphitic carbon solid phase extraction (PGC-SPE) to bind small oligosaccharides, whilst likely not retaining the larger, highly branched whole arabinogalactan molecules, improving our ability to discern reaction products of Rv3707c more easily.

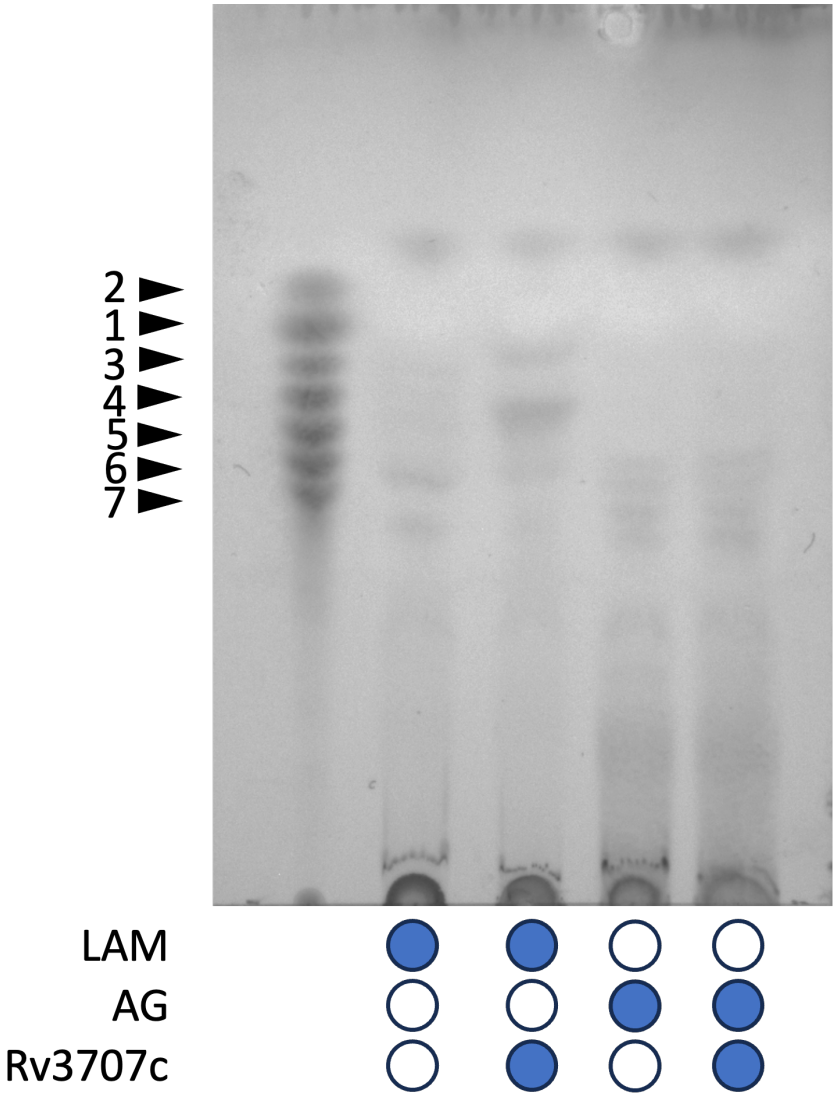


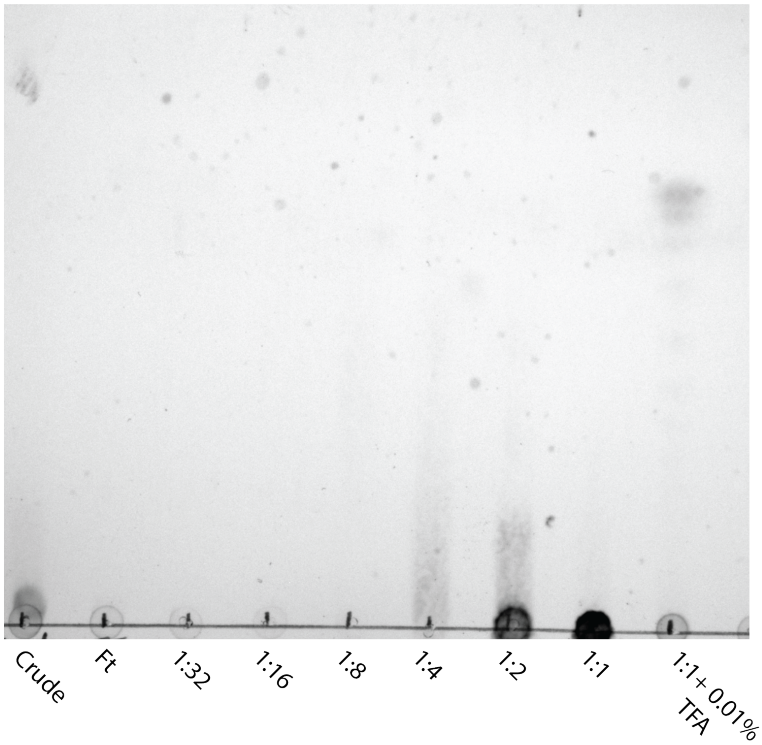
Figure 14. Thin layer chromatogam (TLC) analysis of Rv3707c_P30M (5µM) overnight digestion products using arabinogalactan purified from *M. smegmatis* (1 mg/ml) as a substrate. Following hydrolysis, reaction mixtures were separated using a system of 2:1:1 butan-1-ol:acetic acid:water

6.11 PGC-SPE allows removal of endogenous arabinanase reaction products

Porous graphitic carbon is a reversed phase chromatography substrate which specifically binds phenolic and cyclic solutes, including mutarotationally fixed monosaccharides, such as those in oligosaccharides and glycosides through π-bonding. A 1 g bed weight column was used to remove endogenous oligosaccharides from our arabinogalactan preparations. Figure 16a summarises TLC analysis of fractions produced from the PGC-SPE clean-up of arabinogalactan, the flowthrough to 1:16 fractions, as well as the 1:1 elution have a circular mark at the indentation which was concluded to be whole arabinogalactan. These fractions were pooled and dried using a speed-vac. In elutions 1:4 and 1:2, a smear of undefined length high-order oligosaccharides elute from the cartridge. However, elution of the native DUF4185 products from the cartridge required the addition of 0.01 % trifluoracetic acid as a counterion to cause elution, meaning the oligosaccharides were strongly retained by the matrix.

Use of PGC cleaned arabinogalactan in a hydrolysis assay where the pH was varied through use of phosphate/citrate buffers shows that the endogenous oligosaccharides were no longer present in the substrate, allowing clear visualisation of two groups of oligosaccharides produced from arabinogalactan during a one-hour digest in the presence of exogenous Rv3707c_P30M in Figure 16b. The arabinan species produced by Rv3707c migrate similarly to the 6-8mer L-arabinose standards, with a similar retention to the endogenous D-arabinan oligosaccharides found in arabinogalactan preparations from *M. smegmatis*. This supports the designation of these enzymes as endogenous arabinan hydrolases from the Mycobacteriales. These data also indicate that the pH optimum of the enzyme is in the region of 7-9, with enzyme activity completely lost at pH 5 and below, likely due to protonation of the catalytic carboxylic acid side chains.

A.



B.

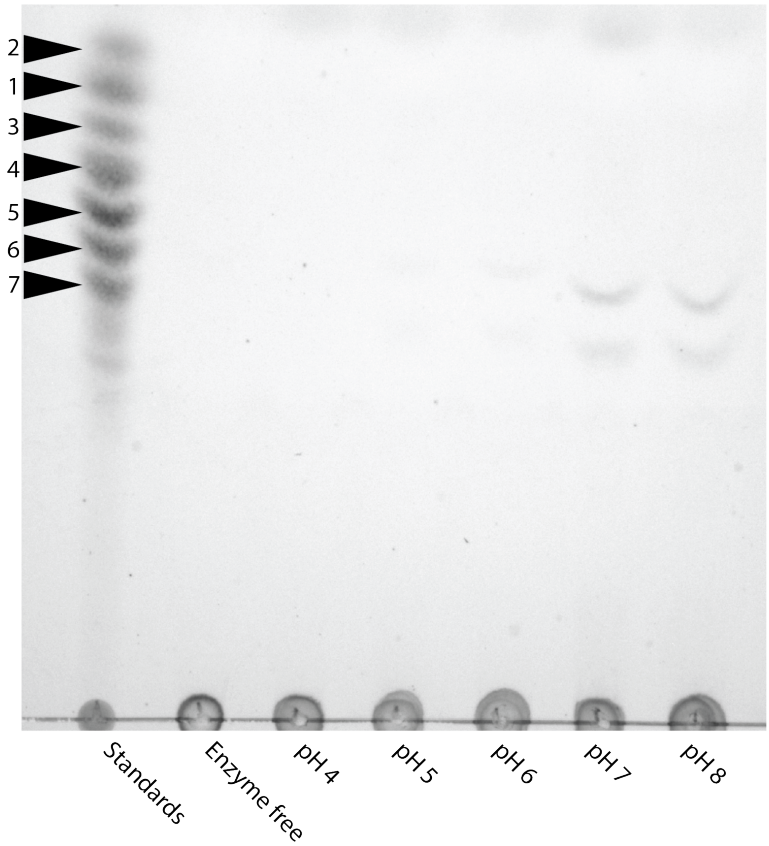


Figure 15. PGC-SPE of endogenous arabinanase oligosaccharides contaminating AG preparations. Following loading of crude AG sample onto a PGC column, the column was washed with dilutions of butan-1-ol in water, tlc analysis of resulting fractions. B. Use of PGC-SPE AG (1 mg/ml) using Rv3707c (5 µM) following overnight digestion allows visualisation of reaction products due to removal of endogenous arabinan oligosaccharides. Analysis by TLC.

6.12 Optimisation of PGC-SPE to allow IC-PAD analysis of reaction products

Ion chromatography with pulsed amperometric detection (IC-PAD) allows for the exceptionally detailed analysis of sugars and other negatively charged analytes. However, prior to analysis of the reaction products of enzymes prepared in Buffer 3, the amino acid contamination needed to be removed, as arginine and glutamic acid are both charged and easily reduced, meaning they would overload the PAD detector. To remove amino acid contamination from the reaction mixtures, PGC-SPE was once again used. Mock reactions mixtures were produced by mixing 10 μL of 1 mM L-arabinose standards with 90 μL of Buffer 3. This mixture was then loaded into a 50 mg bed-weight PGC-SPE cartridge. The cartridge was washed with 5 mL water and then oligos were eluted with 80% acetonitrile in water. Resulting fractions were then analysed in 2:1:1 butan-1-ol: acetic acid: water and TLC were stained with either ninhydrin for amino acids or orcinol for carbohydrates. As shown in Figure 17, the amino acids were poorly retained by the stationary phase and eluted in the flowthrough or wash fractions, a very small amount of one of the amino acids remained, but the overall quantity of amino acids was orders of magnitude higher in these samples compared to enzyme reactions using 5 μM enzyme. In Figure 17, we can see that most oligos in the samples were very well retained, eluting in 80% acetonitrile. Some of the oligosaccharides were not retained by the cartridge, but this may be non-specific staining of the amino acids eluting in the wash fraction. This methodology could therefore be used for reaction clean-up before IC-PAD analysis of the reaction products.

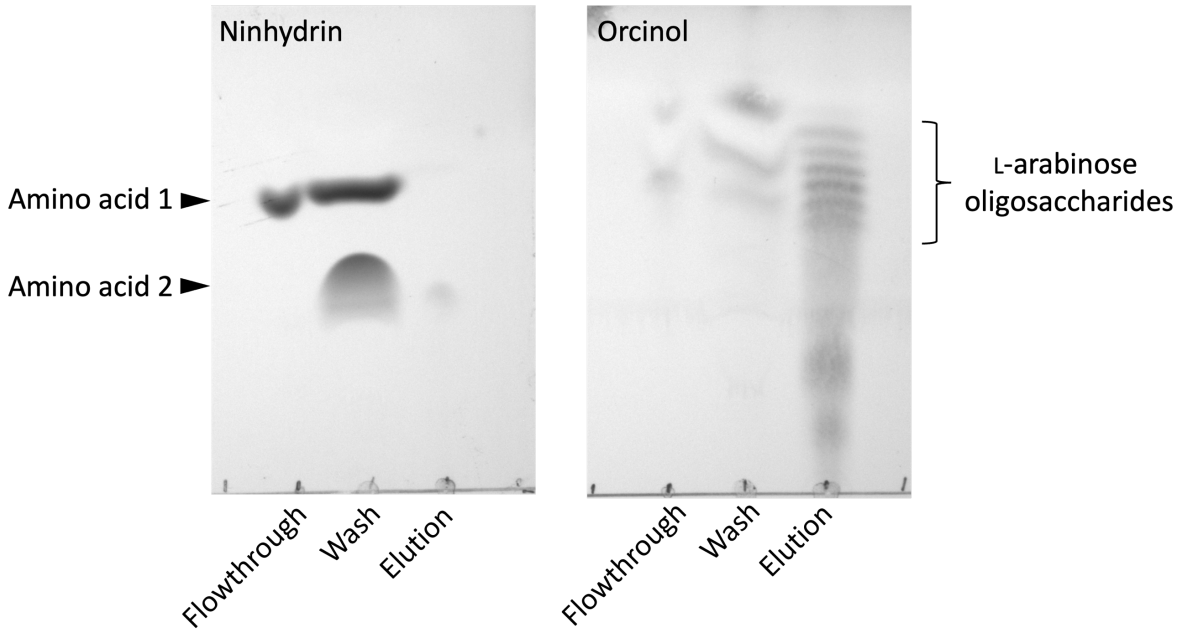


Figure 16. Trial solid phase extraction 1 mM L-arabinose oligosaccharides in protein purification buffer 3. Following elution with 80% acetonitrile in water with 0.01% trifluoroacetic acid, fractions were developed teice using thin layer chromatography in a system of 2:1:1 butan-1-ol:acetic acid:water. Oligosaccharides were well retained by the matrix whilst contaminating amio acids were efficiently washed from the sample.

6.13 MSMEG_2107 possesses a restricted substrate preference compared to Rv3707c

Overnight reactions using purified enzymes against LAM and AG from *M. smegmatis* were solid phase extracted as previously described and resulting oligosaccharides were resuspended in 100 μL water prior to analysis using IC-PAD. Figure 18a is a chromatogram of reaction products liberated from LAM by each enzyme. MSMEG_2107 is highly active on the substrate, producing short oligosaccharides eluting with a similar retention time to L-arabinose oligosaccharides 2-3 monosaccharides in length. In contrast, Rv3707c is active, albeit weakly against LAM, producing a small amount of oligosaccharide between 3 and 4 monosaccharides in length.

The chromatogram shown in Figure 18b summarises activity against arabinogalactan. In this case, MSMEG_2107 possesses no visible activity against AG, with no difference between the native oligosaccharides present and the enzyme treated sample. Rv3707c is once again weakly hydrolytic against the substrate, but this time producing oligosaccharides with a chain length of 6. Based on present knowledge of the structures of the arabinan domains of AG and LAM, they are highly structurally similar, but these data suggest that the domains differ significantly in structure due to the differences in product profiles produced by each enzyme. In addition, the two proteins appear to possess different substrate preferences, with MSMEG_2107 highly active against LAM and Rv3707c weakly active on both. Though these low activities may be due to lower solubility of Rv3707c in the assay conditions.

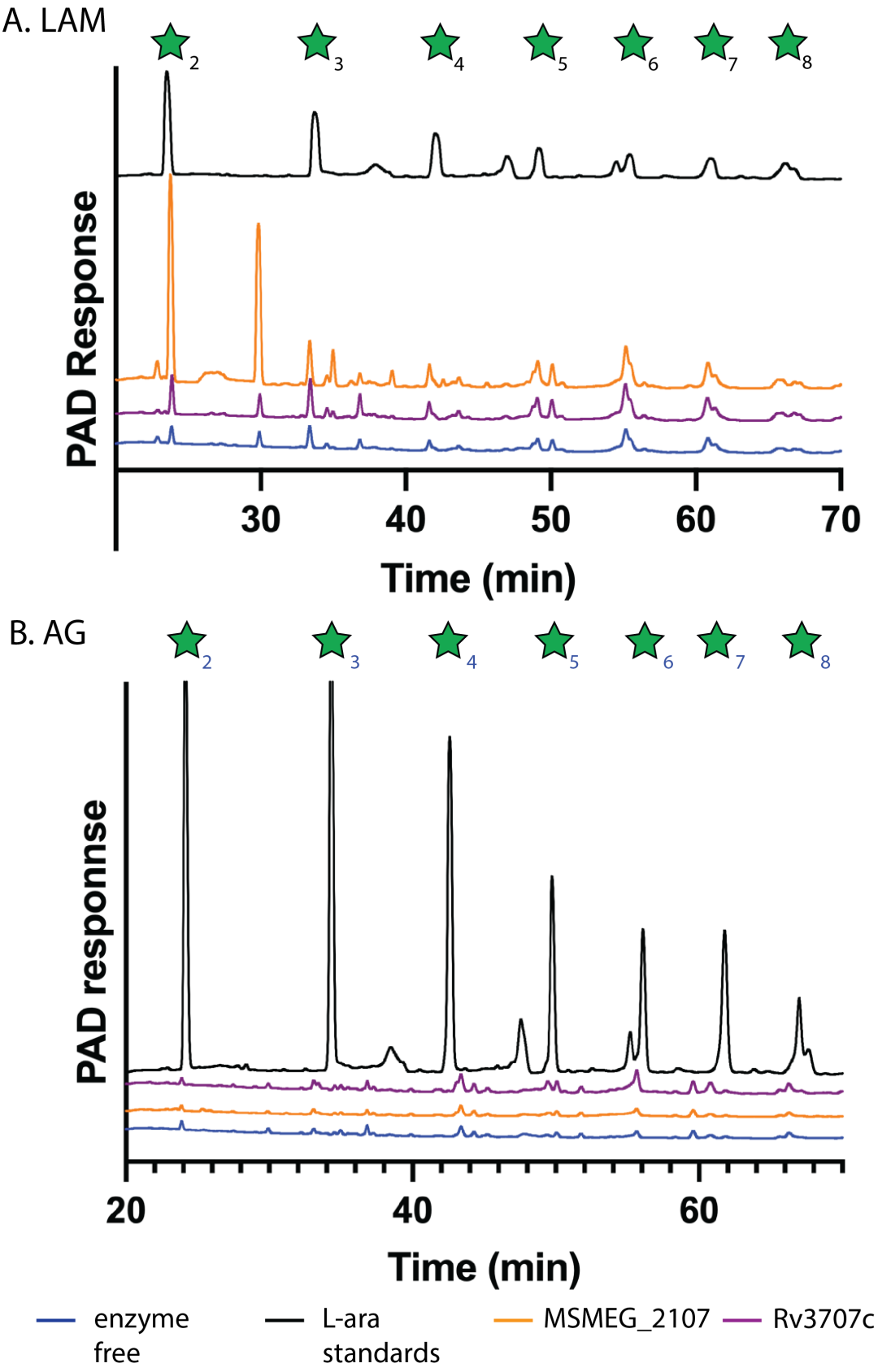


Figure 17. IC-PAD analysis of PGC-SPE cleaned hydrolysis reactions. Stadards are linear α-1,5 oligos of L-arabinose of defined length. Reactions were carried out overnight. A. Analysis of reaction products resulting from treatment of LAM (1 mg/ml) with MSMEG_2107 (5 µM) and Rv3707c (5 µM). B. Analysis of reaction products resulting from treatment of AG (1 mg/ml) with MSMEG_2107 (5 µM) and Rv3707c (5 µM).

6.14 Rv3707c is a five-bladed beta propeller glycoside hydrolase

Following successful purification of recombinant Rv3707c, attempts to crystallise the proteins were made to provide structural data on this new family of glycoside hydrolases. Small crystals were obtained from the JCSG-plus screen (Molecular Dimensions) in a condition containing 0.8 M Succinic acid pH 8 after roughly two months. The crystal structure of Rv3707c_P30M was determined at a resolution of 2.4 Å in the apo form using the Alphafold structural prediction (average pLDDT= 96.5) as a search model. The protein crystallised coordinating a succinic acid molecule from the crystallisation condition. As shown in Figure 19D, the crystalline lattice contained large amounts of water, causing the region where N and C termini meet to have a significant amount of mobility within the asymmetric unit, causing B-factor to increase in this region, as shown in Figure 19c. The protein also possesses a high degree of flexibility in residues 1-10, which we hypothesise may allow conformational change to occur during substrate binding, thus allowing Rv3707c to accommodate the branched arabinan domains of arabinogalactan and LAM.

The Five-bladed beta-propeller structure of Rv3707c is shown in Figure 19a. Also depicted is the N-terminal proline-rich region which contributes the final strand of the 5th beta sheet blade of the propeller, thus explaining its essentiality for protein solubility, and why expression of a truncated version of the protein lacking the N-terminal 59 amino acids was insoluble. The five-bladed beta propeller structure is notably deformed in the region of amino acids 108-159, highlighted in light green in figures 19C and 19D.

Table 7. Refinement statistics for Rv3707c

	Rv3707c
Wavelength	0.9795
Resolution range	43.5 - 2.41 (2.497 - 2.41)
Space group	R 3 2: H
Unit cell	145.674 145.674 92.6806 90 90 120
Total reflections	153262
Unique reflections	14695 (1458)
Completeness (%)	99.88 (99.73)
Wilson B-factor	47.35
$CC_{1/2}$(%)	1.0 (0.382)
Reflections used in refinement	14680 (1454)
Reflections used for R-free	714 (67)
R-work	0.2046 (0.2942)
R-free	0.2495 (0.3172)
Number of non-hydrogen atoms	2417
macromolecules	2295
ligands	16
solvent	106
Protein residues	297
RMS(bonds)	0.024
RMS(angles)	1.95
Ramachandran favored (%)	95.47

Ramachandran allowed (%)	3.48
Ramachandran outliers (%)	1.05
Rotamer outliers (%)	4.45
Clashscore	11.93
Average B-factor	53.92
macromolecules	54.34
ligands	53.57
solvent	44.85

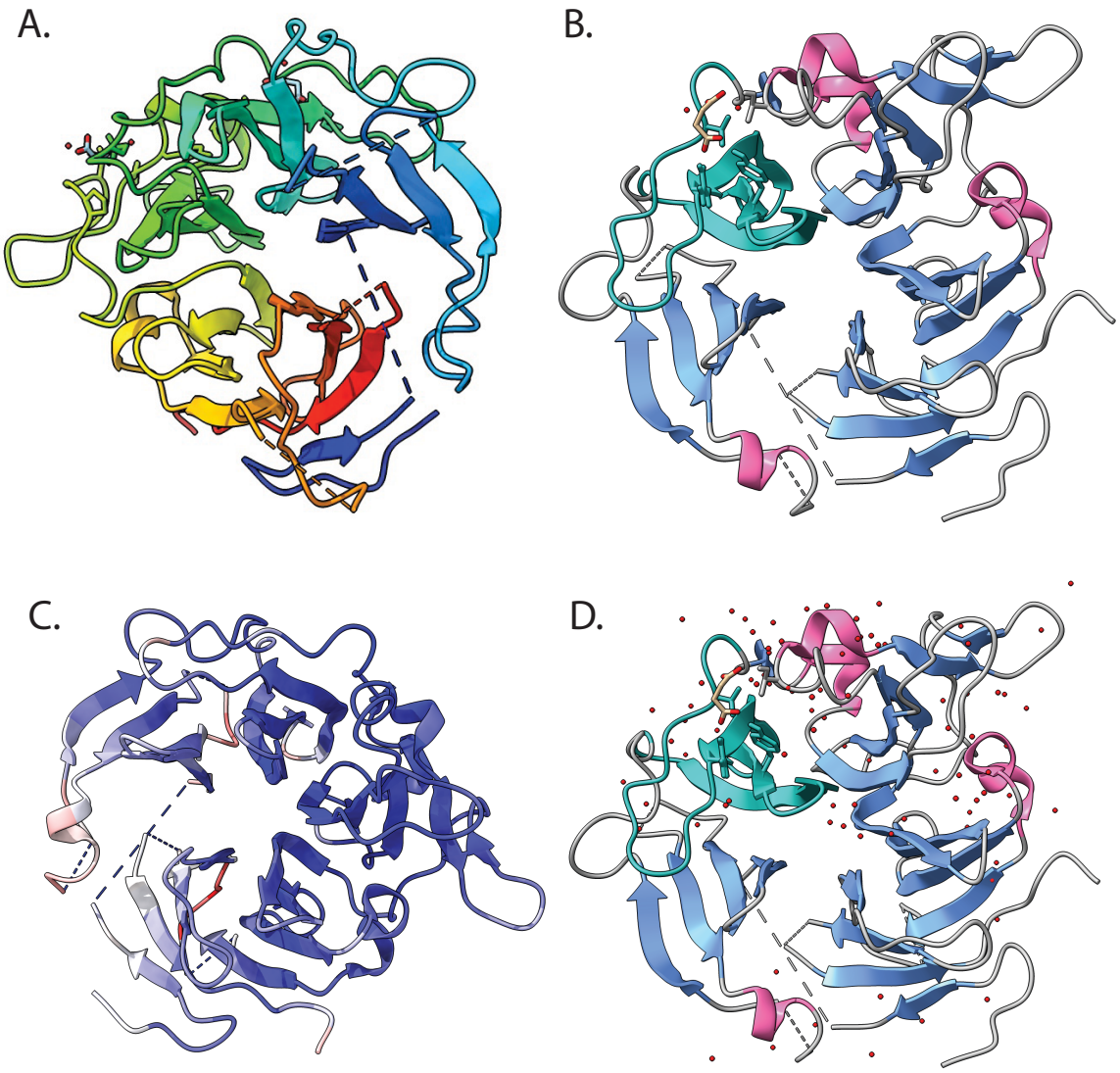


Figure 18. Analysis of the crystal structure of Rv3707c A. top view of the protein, colo chchred by secondary structure progression (blue= N-terminus; Red= C-terminus) B. degenerate blade of propeller highlighted in green. C. structure colourd by b factor, which is roughly uniform but for a highly mobile loop containing an alpha helix in the left of the diagram (Blue = 26.14 $Å^2$, White= 74.255 $Å^2$, Red= 122.37 $Å^2$) D. Water content of the asymmetric unit.

6.15 Structural comparison of Rv3707c reveals a conserved active site across DUF4185 enzymes

Superpositioning of the experimental and Alphafold-predicted structures of Rv3707c using MatchMaker in ChimeraX achieved an RMSD of 0.832 Å, with variance mostly driven by absence of coils found in the Alphafold model (Meng *et al.*, 2023). Given that the structures recapitulated each other well, we carried out BLAST analysis within ChimeraX using the Alphafold structural model to search against the Uniprot100 database, aligning 150 sequences. Using the resulting alignment, we coloured the ribbon representation by conservation with red indicating a high level of conservation, as shown in Figure 20a. Figure 20b is a visualisation of this colour scheme mapped onto the protein surface, highlighting a highly conserved cleft within the protein which could accommodate a glycan molecule.

During our work, another group reported the structure of a distantly related enzyme (Shimokawa *et al.*, 2023b). Including this enzyme in our analysis allowed for improved analysis of our structure. Overlaying the experimental, Alphafold, and a ligand-bound structure of a homolog from *Microbacterium arabinogalactanolyticum* (PDB ID= 8IC1) allowed visualisation of a highly conserved active site architecture. The *M. arabinogalactanolyticum* active site is shown in Figure 20c. To generate a Michaelis complex, the authors mutagenized aspartic acid 51 to asparagine, therefore removing the catalytic acid/base for the enzyme and rendering the enzyme inactive. The distance between the catalytic residues in the *M. arabinogalactanolyticum* structure is 6.31 Å, whilst the Asp33 nucleophile is positioned 2.87 Å from the anomeric carbon of the +1 D-arabinofuranose. Therefore, the Asp33 residue is posed to carry out nucleophilic attack, initiating hydrolysis of the bond. The distance between the two catalytic residues makes it difficult to deduce the mechanism by which catalysis is

carried out. 5.5 Å implies a retaining mechanism, whereas a distance of 6-11 Å implies an inverting mechanism. Due to the clear structural variability of these enzymes, as well as flexibility in the enzyme, the specific mechanism of catalysis is unclear and further experimentation is required.

Figure 20c shows the active site from the experimental structure of Rv3707c. Whilst Asp33 is in the same location, the coordinating tyrosine and catalytic base Asp aren't located anywhere near the active site. This therefore indicates that the experimental structure has likely crystallised in an enzymatically inactive conformation. However, this could be indicative of adaptive conformational change in response to substrate binding. But without a ligand form of Rv3707c, this is difficult to prove.

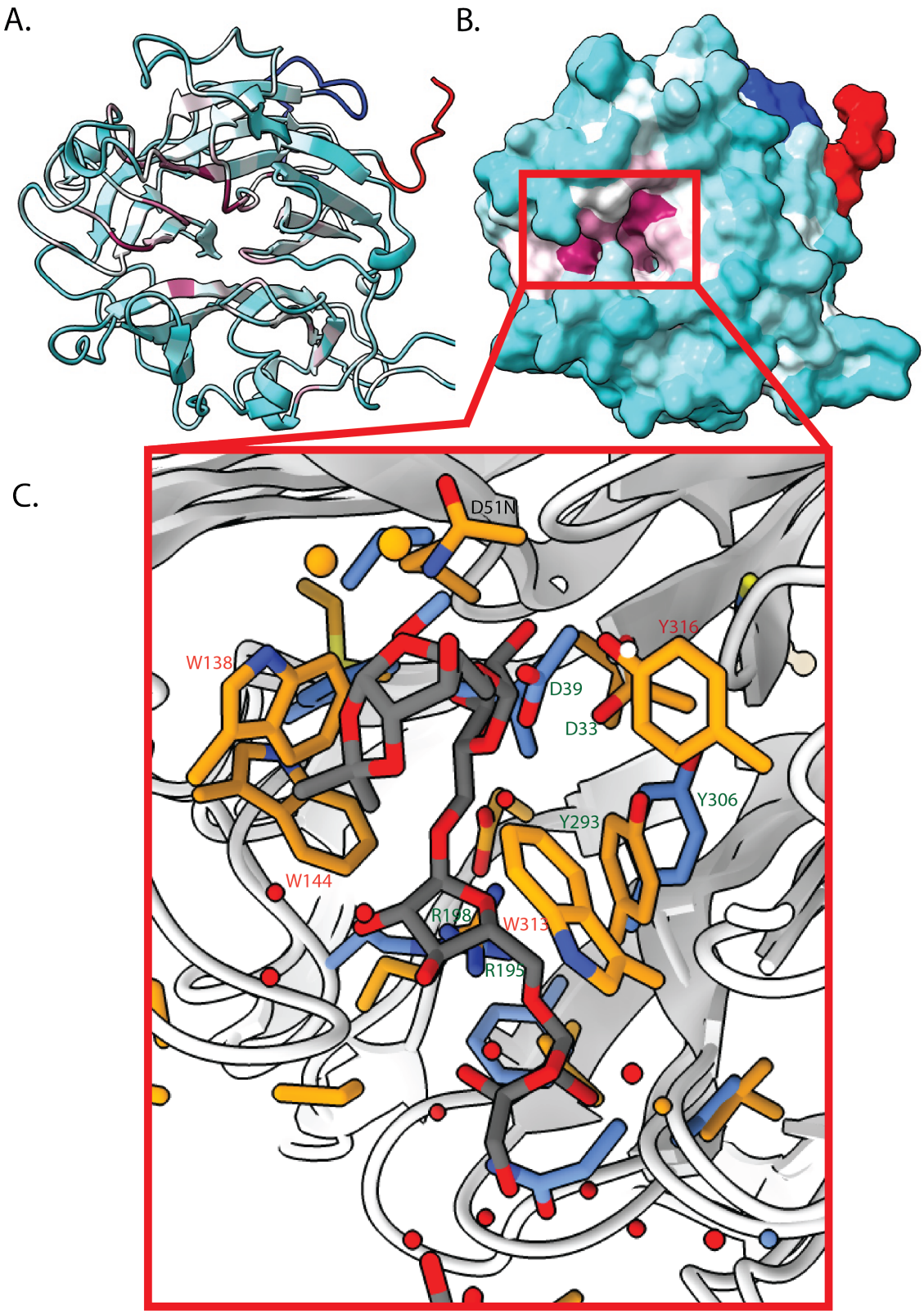


Figure 19. In depth structural analysis of Rv3707c compared to other DUF4185 enzymes. A. structure of Rv3707c coloured by sequence conservation (Maroon= highly conserved, Cyan= poorly conserved) using a multi-sequence alignment resulting from BLAST against the Unireff100 database. N-terminus coloured in Blue, C terminus coloured in red due to poor conservation. B. surface analysis of the protein sequence conservation identified highly conserved cleft on surface of the protein. C. Overlay of active sites from DUF4815 enzymes (orange= 8IC1, Blue= Rv3707c experimental), Residue numbers are green where residues from each structure are colocalised in each structure, and red for those that are not.

6.16 Structural analysis successfully identifies catalytic residues in Rv3707c

Following comparative structural analysis, putative active site residues were mutagenized to alanine using the Q5 DNA polymerase site-directed mutagenesis protocol. Residues targeted were Asp39 Asp59, and Asp86. Following IMAC purification summarised in Figure 21, using Ni-NTA resin in Buffer 3, the purified proteins were used in hydrolysis reactions of LAM. The wildtype enzyme, as expected, produced oligomers. However, the D86A mutant remained catalytically active, showing that the residue is not involved in catalysis, whereas the D39A and D56A mutants, whilst not being fully catalytically null were notably less active, with none of the same products produced as in WT Rv3707c. We therefore propose that D39 is the nucleophile and D56 acts as the catalytic acid/base, these data are summarised in Figure 22.

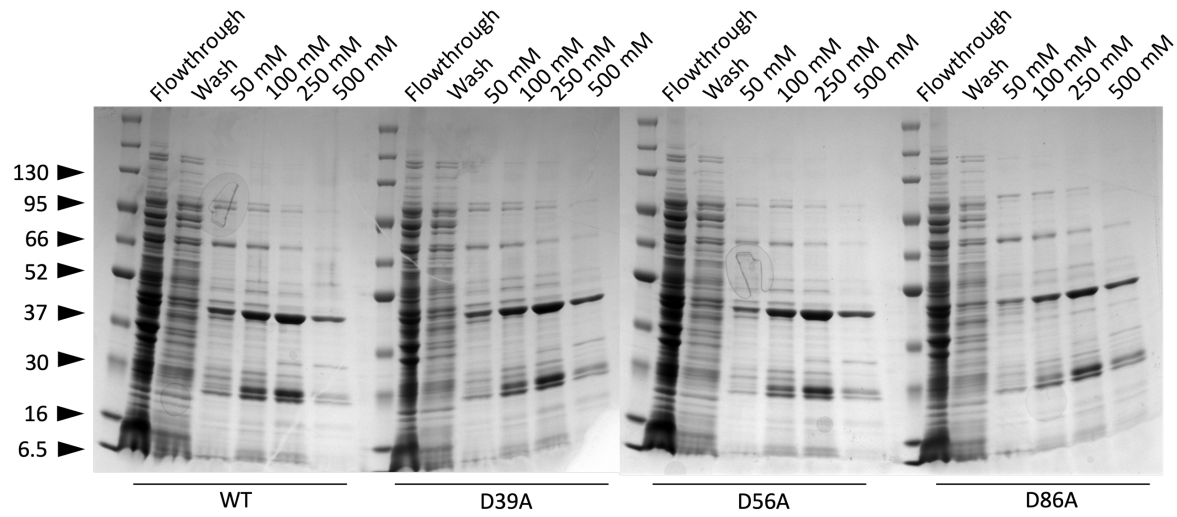


Figure 20. SDS-PAGE analysis of fractions resulting from IMAC of site-directed mutants of Rv3707c-P30M in putative catalytic residues. Pellets were lysed in Buffer 3 using sonication and then processed on a gravity column using NI-NTA resin

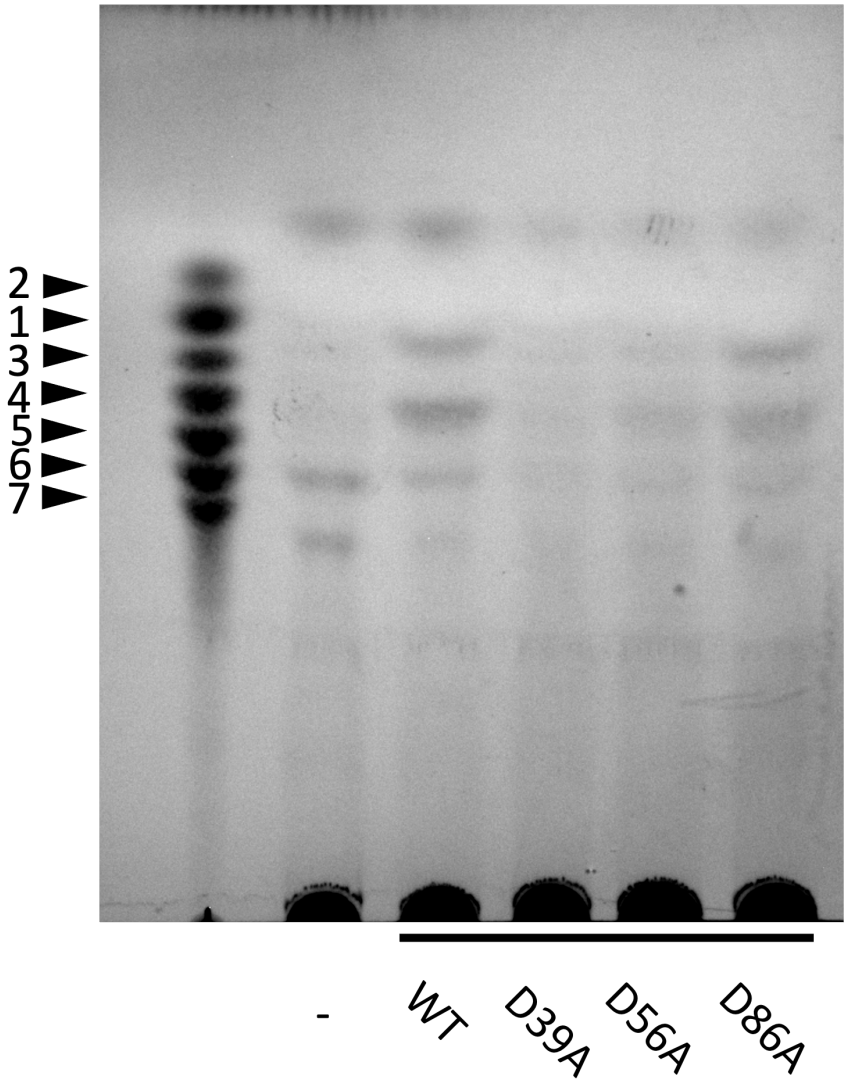


Figure 21. Thin layer chromatographic analysis of reaction products resulting from the overnight hydrolysis of LAM purified from *M. smegmatis* (1mg/ml) in an *in vitro* reaction against catalytic mutants of Rv3707c (5µM). TLCs were developed twice in a system of 2:1:1 butan-1-ol: acetic acid: water.

6.17 Discussion

In this chapter, we have purified, characterised, and structurally analysed two members of the DUF4185 family. The enzymes presented have endo α-1,5-D-arabinofuranosidase activity. This activity has been known to exist within the field of mycobacteriology for more than 30 years, both within mycobacteria (Xin *et al.*, 1999), and within saprophytic organisms that likely coexist with environmental Mycobacteriales, such as *Cellulomonas* spp. (Chatterjee *et al.*, 1993). Further reinforcing this is that there are multiple DUF4185 homologs encoded in the genomes of *Cellulomonas spp.* Purified lysates of these species have been used as a source of nonspecific endo-D-arabinanase activity for many years (Chatterjee *et al.*, 1993). We have proposed that the family DUF4185 is reannotated as GH183, and this has been updated in the CAZY database (Henrissat, 1991). The biological functions of these GHs within the lifestyles of mycobacteria remain elusive.

Whilst homologues of these enzymes from gut-resident bacteria were relatively easily purified by our collaborators at Newcastle University, the mycobacterial homologues were challenging to work with (Al-Jourani *et al.*, 2022). The proteins required complex buffering conditions, and extensive construct design consideration. Additionally, we have developed methodologies to characterise the enzyme activity of the endogenous GH183 enzymes *in vitro* using PGC-SPE. We believe that PGC is underutilised as a stationary phase in the analysis of carbohydrates, and that it has proven to be instrumental in the analysis of the presented work. Finally, until recently, we were the first group to crystallise and solve the structure of any GH183 enzyme. However, during the preparation of this thesis, Fujita and colleagues produced high quality apo- and ligand structures of GH183 paralogs found in *Microbacterium arabinogalactanolyticum.* This structure proved instrumental in understanding the dataset

we produced of Rv3707c, allowing us to better parse the structure we had produced. Biochemical characterisation of the two enzymes indicates that they possess differing substrate preferences, with Rv3707c-like enzymes seemingly possessing hydrolytic activity against both AG and LAM, whilst Rv1754c-like enzymes are restricted to LAM. We therefore propose that enzymes with homology to Rv3707c be annotated as AraH1, and Rv1754c homologs as AraH2. Further work must be carried out to determine the mechanism by which this substrate preference is determined.

The mAGP complex presents a unique engineering challenge to the bacteria that possess it in coordinating cell growth, due to the fact the entire ultrastructure is covalently bound- therefore meaning the division of the cell wall must differ from the elongasome and divisome organisation found in bacteria possessing lateral growth (*Escherichia coli, Bacillus subtilis*) (de Pedro *et al.*, 1997; Spratt, 1975). This has been a well-established fact for many years, with Mycobacteriales growing in a polar manner. Present understanding of cell division in bacteria requires the presence of autolysins, enzymes which cleave cell wall components to release daughter cells. To date, autolysins active on arabinan have not been identified. We therefore propose that these enzymes may play a significant role in the division of mycobacteria and may constitute another part of the divisome.

There has been a recent research interest in the mycobacterial divisome. In two studies using co-immunoprecipitation of Rv0954 and LamA in mycobacteria both proteins have been shown to participate in the mycobacterial divisome, but like many proteins involved in cell division in other organisms, their deletion leads to a subtle phenotype, leading to uniform polar growth instead of polar growth with more active growth localised at the old pole (Rego

et al., 2017; Wang and Ehrt, 2021). An immediate avenue of work will be to conduct co-immunoprecipitation of tagged AraH1/2 to identify interaction partners and regulators using the pRIBO plasmid series (Ahamed *et al.*, 2022). We can be confident that these genes play important roles in the lifestyle of pathogenic organisms due to the fact they have appeared in multiple high throughput screens of mycobacteria as described below.

The *Mycobacterium bovis* BCG homologue of Rv3707c was identified in a transposon sequencing of macrophage infections as insertion mutants of this gene could no longer control phagosome maturation (Stewart *et al.*, 2005). Additionally, knockout of Rv1754c caused a significant defect in pathogenicity in mice infection models (Perkowski *et al.*, 2017). Furthermore, the amoeba predator *Mycobacterium llaterenze* significantly upregulates expression of its Rv3707c homolog during infection of *Acanthamoeba castellanii* when it is resisting phagocytosis (Delafont *et al.*, 2017). We therefore hypothesise that Rv3707c improves fitness of a pathogen by three possible mechanisms. The first is that the enzyme reaction products have a conserved binding partner in eukaryotes, whose interaction causes immunosuppression by some unknown mechanism. Arabinogalactan-binding lectins such as Galectin-9 have been associated with disease progression in *M. tuberculosis* (Wu *et al.*, 2021). However, in the a more recent paper, Galectin-9 signalling results in worse disease outcomes, complicating Galectin-9's involvement in tuberculosis disease progression (Morrison *et al.*, 2023). Recently published work in a triple galectin (galectins 3, 8, and 9) knockout mouse model shows that none of these are required for the acute suppression of *M. tuberculosis*, but are less able to contain the bacteria in granuloma (Morrison *et al.*, 2023).

Alternatively, mycobacteria must undergo significant cell wall remodelling during the transition from extra- to intracellular, and AraH enzymes may enable this. Finally, the AraH enzymes could play a role in the extrusion of secretion systems used in delivering pathogenicity factors such as ESAT-6 and PE-PPE proteins into the host and causing illness. However, to better dissect these hypotheses, microbiological work must be carried out on gene knockout strains of these organisms, which is a next avenue of work for the project.

7 Biochemical characterisation of a putative arabinogalactan pathway in *Nocardia*

7.1 Contribution statement

The protein biochemistry sections of this chapter involving the FGGY kinases and isomerases were carried out in collaboration with MSci student, Amar Gudka, whom I supervised. I also thank Dr. Hannah Doherty for producing diagrams of the proteomics data presented. The expression construct for catalytically null $NBras_{GH172}$ was produced using Q5 SDM by Dr. Jennifer Ross at the University of Newcastle.

7.2 Introduction

7.2.1 Central carbon metabolism in CMNR group bacteria

Central carbon metabolism is a fundamental cellular process responsible for the generation of intracellular energy sources such as ATP. The pathway comprises of glycolysis, the pentose phosphate pathway, the citric acid cycle, glyoxylate shunt, and gluconeogenesis (Sassetti and Rubin, 2003; Li *et al.*, 2019). Pathogenic mycobacteria require a highly adaptive CCM to enable survival in restrictive host environments such as the phagolysosome, whilst also needing to grow in competitive environments such as soil, bodies of water, and the lung microbiome. Selection in biologically diverse communities often favours generalist bacteria able to utilise a wide variety of carbohydrate sources (Briliūtė *et al.*, 2019). Therefore, extensive work has been carried out to characterise the exceptional capability of mycobacteria to switch their preferred carbon source throughout infection (Gouzy *et al.*, 2021; Nandakumar *et al.*, 2014; Baughn *et al.*, 2009; Tian *et al.*, 2005). An early study from the Youmans laboratory showed that only 3 carbon sources could support the growth of *M. tuberculosis* in laboratory media.

These were glucose, glycerol and maltose (α-1,4-glucobiose) (Youmans and Youmans, 1953). Recent studies by Gouzy and colleagues have shown that carbohydrate utilisation in *M. tuberculosis* is pH-dependent, with glycerol preferred at neutral pH, whilst host-derived long chain fatty acids such as oleic and arachidonic acid support growth of *M. tuberculosis* at pH as low as 4.5 (Gouzy *et al.*, 2021). This ability is unique amongst intracellular pathogens, and likely amongst bacteria.

Another question remains over how CMNR organisms balance demand for ribose pyrophosphate (PRPP) between purine, histidine, and D-arabinose synthesis. Additionally, a pathway allowing recycling of D-arabinose into DNA and other cellular currencies could be advantageous for slowly replicating intracellular populations of pathogenic CMNR group bacteria, where carbon can be a growth limiting nutrient due to the exclusion of phagocytosed pathogens from the rest of the cell through membrane partitioning. However, no evidence in the literature exists for such a pathway, but our discovery of arabinan active glycoside hydrolases is a pivotal first step in uncovering such a mechanism.

7.2.2 Pentose Phosphate Pathway

The pentose phosphate pathway is ultimately responsible for the conversion of non-typical sugars into glucose for glycolysis. The pathway is deeply conserved across all kingdoms of life, and it consists of two arms, the oxidative and the nonoxidative. The oxidative arm is responsible for producing ribose-5-phosphate (R5P), a precursor for many fatty acids, nucleotides and other critical molecules in mycobacteria, such as D-arabinan (Gouzy *et al.*, 2021). Whilst the nonoxidative arm produces NADPH, which is the principle cofactor

responsible for the reduction of biomolecules within the cell (Hostetler and Landau, 1967; Zhong *et al.*, 2019).

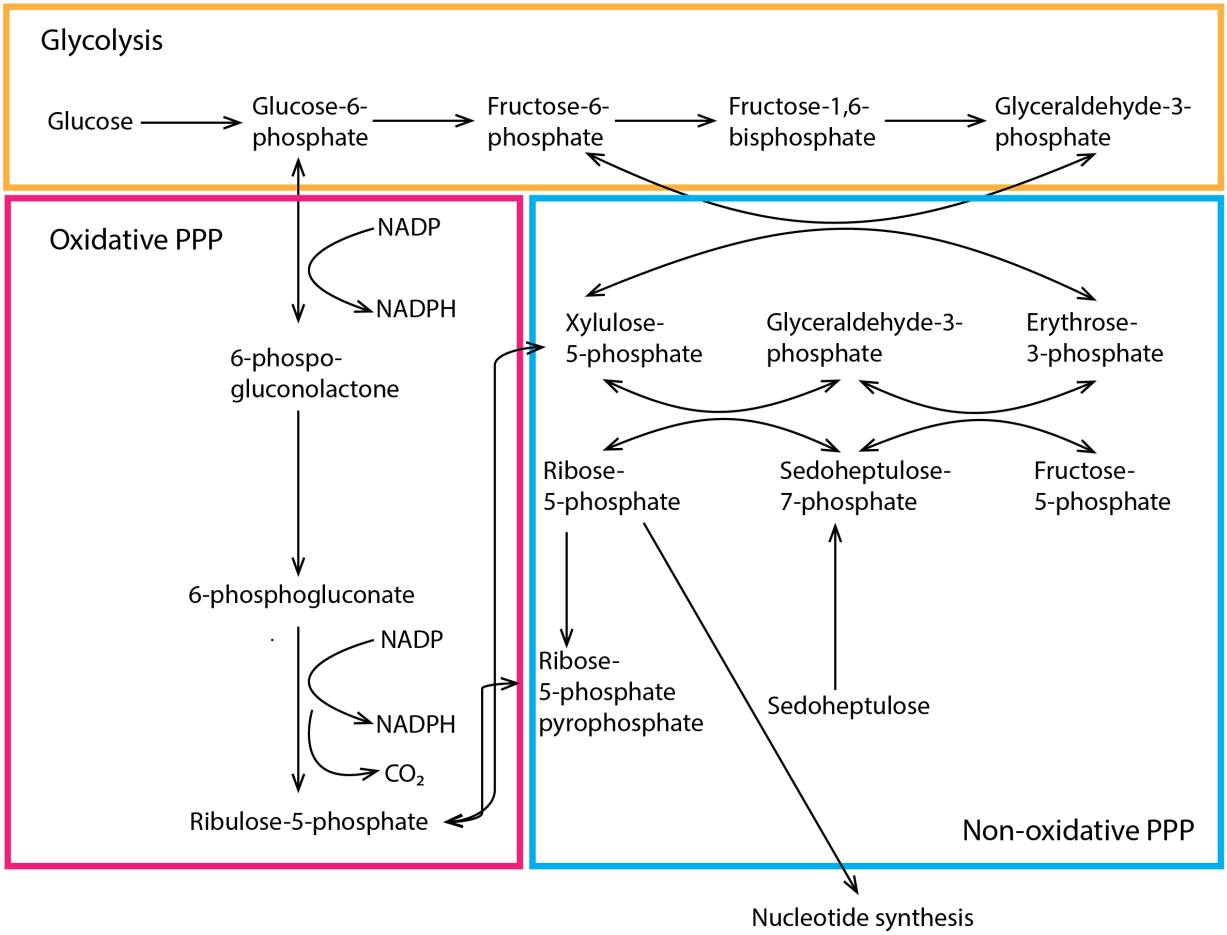


Figure 22. Schematic of the pentose phosphate pathway, this diagram highlights the central role of the ketopentoses ribulose and xylulose as transition states between inputs and outputs, and ultimately are what allow for the conversion of glucose to ribose-5-phosphate for DNA synthesis, cell wall metabolism and glycolysis in mycobacteria. Figure adapted from TeSlaa et al., 2023.

The two arms of the pentose phosphate pathway are summarised in Figure 23. As can be seen, glucose-6-phosphate is redirected from glycolysis into 6-phosphogluconolactone by glucose-6-phosphate dehydrogenase to liberate a proton which can be used to reduce one NADP molecule to NADPH. This is then further processed to 6-phosphogluconate and then dehydrogenated once further to ribulose-5-phosphate (Ru5P), liberating a further proton to produce another NADPH molecule. This Ru5P is then processed in the nonoxidative arm of the pentose phosphate pathway into R5P. In mycobacteria, the fate of this molecule can either be in nucleotide synthesis or in the eventual synthesis of D-arabinan through the phosphoribosylpyrophosphate synthase PrsA producing 5-phospho-α-D-ribose-1-diphosphate (PRPP)(Breda *et al.*, 2012).

Due to the intensive study of central carbon metabolism across organisms in the last 60 years, the biochemical steps governing synthesis of PRPP is well known, which have been recently reviewed by TeSlaa and colleagues (TeSlaa *et al.*, 2023). Nevertheless, given that we have shown that mycobacteria and other related organisms readily turn-over their AG, we wondered if AG-derived D-arabinose could be directly recycled in any Mycobacteriales species.

7.3 Results

7.3.1 Identification of homologs of HMPREF9455_02467 in *Nocardia* species

As described earlier, there are three DUF2961-containing proteins encoded in *D. gadei* PUL42 (Al-Jourani *et al.*, 2023). In the previous chapter we characterised endo-acting homologues of the GH183 enzymes within the same PUL. To use complex carbohydrates as a carbon source, the bacteria require exo-acting enzymes like the GH172 family, to catalyse the complete depolymerisation of oligosaccharides. We therefore hypothesised that the presence of GH172 family enzymes in a Mycobacteriales species could be a marker of D-arabinose recycling.

We wanted to see if the GH172 enzymes were distributed similarly to the GH183 family and so we carried out BLAST analysis against the Refseq-protein database using BlastP, restricting search results to the order Mycobacteriales (Taxonomic ID: 85007). When compared to the distribution of GH183 enzymes, DUF2961 encoding genes appear to be found in a different set of organisms, with none being found in *M. tuberculosis* complex, or other common model mycobacteria, including *M. smegmatis.* Instead, they are concentrated in *Nocardia, Gordonia,* and *Rhodococcus* and environmental or non-human pathogenic species of *Mycobacterium*, as can be seen in Tables 8-10. These bacteria are widespread in the environment, suggesting that DUF2961 enzymes play a different function in comparison to mycobacterial GH183 enzymes. To further verify this, we carried out syntenic analysis of the DUF2961 protein-containing locus in *Nocardia brasiliensis* (Uniprot accession= O3I_017420). This bacterium was selected due to its clinical relevance as a rare but severe pathogen of the central nervous system. In addition to this, SignalP 6.0 analysis indicated that this protein does not possess a signal peptide, and therefore is likely cytosolic (Teufel *et al.*, 2022). This localisation is consistent with selfish-mechanisms of carbohydrate breakdown, exo-acting GHs typically act

as the final degradative step within the cytosol or periplasm to release monosaccharides for use in central carbon metabolism (Cuskin *et al.*, 2015).

Table 8. Top 30 hits resulting from BLAST analysis using HMPFREF9455_02467 against the REFSEQ-protein database. Taxonomy was restricted to Mycobacteriales (TaxID: 85007). All proteins were annotated as DUF2961-containing proteins so protein ID column has been removed

02467-Descriptions

Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Fodinicola feengrottensis	406	406	97%	3E-139	55.41	362	WP_163567793.1
Fodinicola acaciae	405	405	96%	1E-138	54.79	357	WP_246281905.1
Fodinicola acaciae	389	389	97%	2E-132	52.16	362	WP_163509140.1
Mycobacterium heckeshornense	348	348	95%	1E-116	50.41	357	WP_048892883.1
Mycobacterium heckeshornense	346	346	95%	1E-115	50.14	357	WP_203800010.1
Mycobacterium xenopi	346	346	95%	1E-115	49.03	353	WP_003919415.1
Mycobacterium xenopi	344	344	95%	5E-115	48.48	353	WP_085195037.1
Nocardia	341	341	90%	5E-114	48.39	335	WP_063026085.1
Nocardia sp. BSTN01	339	339	90%	2E-113	48.68	335	WP_194808931.1
Antrihabitans stalactiti	337	337	90%	1E-112	48.24	338	WP_169592358.1
Mycobacterium talmoniae	336	336	95%	7E-112	49.03	356	WP_071028892.1
Nocardia sp. MDA0666	335	335	90%	7E-112	48.25	335	WP_106953469.1
Nocardia kruczakiae	333	333	90%	4E-111	47.51	335	WP_310398417.1
Nocardia nova	333	333	90%	5E-111	48.10	335	WP_104377283.1
Nocardia tenerifensis	333	333	90%	7E-111	49.12	339	WP_040741814.1
Nocardia cerradoensis	333	333	90%	1E-110	47.51	335	WP_039784108.1
Nocardia cerradoensis	333	333	90%	1E-110	47.51	335	WP_094024998.1
Nocardia veterana	332	332	90%	2E-110	47.51	335	WP_040717842.1
Nocardia violaceofusca	331	331	90%	6E-110	47.66	335	WP_063063036.1
Nocardia africana	330	330	90%	8E-110	47.21	335	WP_062967632.1
Nocardia aobensis	330	330	90%	9E-110	47.66	335	WP_036502673.1
Nocardia kruczakiae	330	330	90%	9E-110	46.92	335	WP_063008563.1
Nocardia bhagyanarayanae	330	330	90%	1E-109	47.51	338	WP_185757181.1
Nocardia spumae	330	330	90%	1E-109	48.68	335	WP_227983593.1
Nocardia vermiculata	329	329	90%	2E-109	47.09	338	WP_067880654.1

Table 9. Top 30 hits resulting from BLAST analysis using HMPFREF9455_02471 against the REFSEQ-protein database. Taxonomy was restricted to Mycobacteriales (TaxID: 85007)

02471-Descriptions

Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Gordonia rhizosphera	158	158	75%	1E-38	26.65	677	WP_040530813.1
Nocardia takedensis	139	139	76%	7E-32	25.93	1021	WP_040713023.1
Fodinicola acaciae	139	139	76%	8E-32	25.54	1074	WP_246281702.1
Fodinicola acaciae	131	131	72%	4E-29	26.11	1097	WP_163506300.1
Nocardia albiluteola	122	122	73%	2E-26	23.98	690	WP_245400914.1
Mycobacterium colombiense	119	119	70%	1E-25	25.21	680	WP_065027544.1
Mycolicibacterium vulneris	118	118	70%	4E-25	25.21	680	WP_085292570.1
Mycobacterium colombiense	117	117	70%	4E-25	25.98	680	WP_044487753.1
Nocardia vaccinii	114	114	72%	3E-24	25.44	525	WP_084525821.1
Mycobacterium mantenii	115	115	70%	4E-24	25.83	680	WP_067910335.1
Mycobacterium colombiense	114	114	70%	4E-24	25.77	680	WP_248607048.1
Mycobacterium colombiense	114	114	70%	6E-24	25.82	670	WP_040630722.1
Nocardia sp. CS682	112	112	72%	7E-24	26.08	520	WP_240772408.1
Mycobacterium colombiense	114	114	71%	8E-24	26.07	680	WP_076102655.1
Nocardia miyunensis	112	112	72%	1E-23	25.45	525	WP_084532308.1
Nocardia sp. NRRL S-836	113	113	70%	1E-23	26.78	1085	WP_063798867.1
Mycobacterium sp. 1465703.0	112	112	70%	2E-23	24.79	680	WP_066989618.1
Mycobacterium malmoense	112	112	76%	3E-23	24.52	676	WP_065443052.1
Mycobacterium malmoense	112	112	76%	3E-23	24.52	653	WP_255309351.1
Mycobacterium malmoense	111	111	76%	3E-23	24.52	598	WP_065495265.1
Mycobacterium parascrofulaceum	111	111	74%	3E-23	23.53	653	WP_244268820.1
Mycobacterium colombiense	112	112	70%	3E-23	26.03	680	WP_076055814.1
Mycobacterium fragae	111	111	72%	4E-23	25.65	679	WP_085199937.1
Mycobacterium mantenii	111	111	70%	5E-23	26.13	680	WP_067834322.1

Table 10. Top 30 hits resulting from BLAST analysis using HMPFREF9455_02479 against the REFSEQ-protein database. Taxonomy was restricted to Mycobacteriales (TaxID: 85007)

02479-Descriptions

Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Mycobacterium sp. KBS0706	221	221	89%	3E-66	36.08	391	WP_138890758.1
Rhodococcus sp. D2-41	218	218	87%	2E-65	39.09	360	WP_277831809.1
Fodinicola acaciae	153	153	78%	1E-40	32.81	362	WP_163509140.1
Fodinicola feengrottensis	147	147	77%	2E-38	32.80	362	WP_163567793.1
Nocardia brasiliensis	135	135	76%	5E-34	29.08	342	WP_174817079.1
Fodinicola acaciae	135	135	79%	5E-34	31.03	357	WP_246281905.1
Nocardia anaemiae	135	135	85%	6E-34	29.95	343	WP_062985556.1
Nocardia vaccinii	134	134	76%	1E-33	29.45	344	WP_067890850.1
Nocardia bhagyanarayanae	134	134	75%	2E-33	29.28	338	WP_185757181.1
Nocardia amikacinitolerans	133	133	75%	2E-33	28.95	338	WP_253809698.1
Nocardia coffeae	133	133	76%	3E-33	29.77	343	WP_235747360.1
Nocardia ninae	133	133	81%	3E-33	29.06	341	WP_147140347.1
Nocardia blacklockiae	133	133	76%	3E-33	29.87	346	WP_195099557.1
Nocardia suismassiliense	132	132	81%	4E-33	28.77	341	WP_107656310.1
Nocardia amikacinitolerans	132	132	75%	4E-33	29.04	338	WP_253798129.1
Nocardia beijingensis	132	132	81%	4E-33	29.34	329	WP_228828270.1
Nocardia vinacea	132	132	85%	5E-33	29.48	343	WP_040701404.1
Nocardia miyunensis	132	132	75%	6E-33	30.49	343	WP_067665207.1
Nocardia acididurans	132	132	75%	8E-33	32.46	335	WP_232852094.1
Nocardia macrotermitis	131	131	76%	1E-32	30.55	343	WP_153411466.1
Nocardia amikacinitolerans	131	131	75%	1E-32	28.62	338	WP_245909808.1
Nocardia araoensis	131	131	81%	1E-32	29.01	336	WP_051020692.1
Nocardia amikacinitolerans	131	131	75%	2E-32	28.62	338	WP_253781653.1
Nocardia sp. CS682	130	130	75%	2E-32	29.80	341	WP_135236999.1

7.3.2 Syntenic analysis of the *N. brasiliensis* DUF2961 locus suggests a route for D-arabinan recycling.

To understand if the *N. brasiliensis* genes were organised into a common syntenic locus, the amino acid sequence of O3I_017420 was submitted to the WebFLAGS server (Saha *et al.*, 2021). The result of this search is summarised in Figure 24, demonstrating strong synteny across *Nocardia* species. The black triangle in the middle is O3I_017420 from *N. brasiliensis*, which is common to all loci. Towards the bottom of the data is the operon for *N. brasiliensis*, which contains genes (left to right) 4, 1, 3, 8 and 7. WebFLAGS annotates gene 6 as a pseudogene, but manual analysis indicates that the coding sequence possesses a start and stop codon, therefore this annotation is possibly incorrect. Full protein information for each protein and its corresponding number is summarised in Table 11.

Common to all the loci is the presence of at least one copy of an FGGY kinase. These are a large family of carbohydrate active kinases active on ketopentoses such as ribulose and xylulose, especially enriched in the pentose phosphate pathway (Saha *et al.*, 2021). However, most species (79%) possess two FGGY kinases in the locus. Additionally, the *N. brasiliensis* operon possesses hypothetical protein 3, which is common to a bottom cluster of species highlighted in the red box. Finally, *N. brasiliensis* also includes genes 7 and 8, which encode a putative maleylpyruvate isomerase and an amidohydrolase family protein respectively. Gene 6 encodes a putative thioredoxin. Given the importance of xylulose and ribulose kinases in metabolism through the pentose phosphate pathway we speculated whether the DUF2961 enzyme could then work with the encoded kinases.

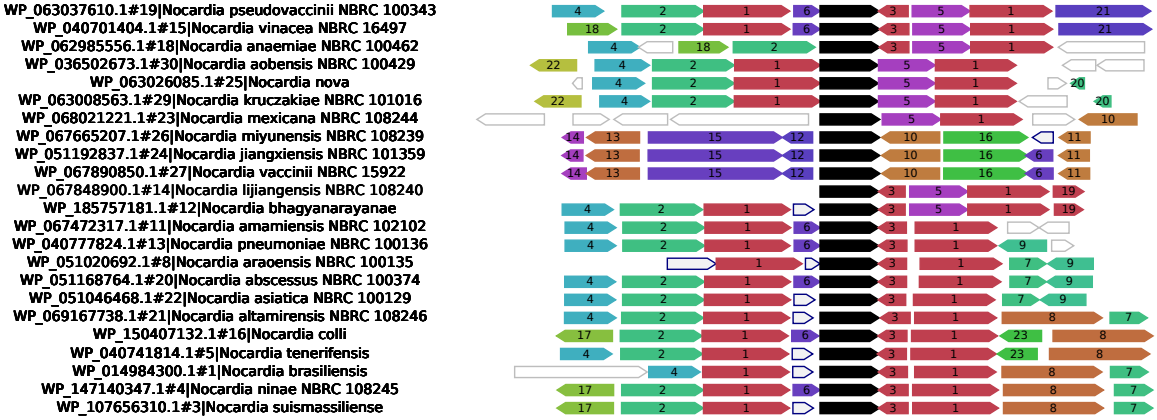


Figure 23. WebFLAGS analysis using O3I_017420 as a search model. The left panel is a maximum likelihood phylogeny with genetic distances based on local gene neighbourhood highlighted on each branch of the tree, scale is shown in the bottom left. On the right is a schematic of the gene content of each locus including directionality of genes relative to the search model, with numbered annotations in Table 8 below. Highlighted by the red box is the cluster containing *N. brasiliensis.*

Table 11. Gene products labelled from Figure 23

Number	Gene product	Number	Gene product
1	FGGY Kinase	**13**	DinB protein
2	L-fucose isomerase	**14**	Hypothetical protein
3	Hypothetical protein	**15**	LuxR C- transcriptional regulator
4	GH32/43	**16**	MFS transporter
5	Zn-dependent alcohol dehydrogenase	**17**	carbohydrate kinase
6	Thioredoxin	**18**	NAD(P)-dependent oxidoreductase
7	maleylpyruvate isomerase	**19**	N-acetyltransferase
8	amidohydrolase family protein	**20**	hypothetical protein
9	helix-turn-helix transcriptional regulator	**21**	α/β hydrolase
10	helix-turn-helix domain-containing protein	**22**	α/β hydrolase
11	hypothetical protein	**23**	GH25 lysozyme

7.3.3 Purification of DUF2961-containing protein O3I_017420 from *N. brasiliensis*

Due to the prior difficulty we experienced expressing protein recombinantly from high-GC genomic DNA, we had the full-length O3I_017420 synthesised into a pET28a plasmid, which was transformed into T7 Shuffle Express cells (Lobstein *et al.*, 2012). Cell pellets obtained following induction were lysed by three passages through a French pressure cell before IMAC.

Fortunately, upon induction, a notable band appeared with a molecular weight slightly below the 37 kDa standard (expected M_r= 38.984 kDa), as shown in Figure 25. The protein bound to IMAC resin with high affinity, with none eluting in the wash fraction. The protein was then eluted with an imidazole gradient, with a large amounts of protein eluting at 80% homogeneity in the 100-, 250-, and 500 mM fractions based on ratiometric analysis using Image Lab. The protein was then further purified using size exclusion chromatography. As shown in the chromatogram and gel in Figure 26, the protein eluted earlier than expected based on its predicted molecular weight, indicating it was multimerising. It eluted at 148 mL with a purity greater than 95%, which based on comparison to known standards is consistent with a molecular weight of approximately 120 kDa. This molecular weight is consistent with the protein being trimeric in nature. This hypothesis was supported with SEC-light scattering by our collaborators (data published in Al-Jourani *et al.,* 2023).

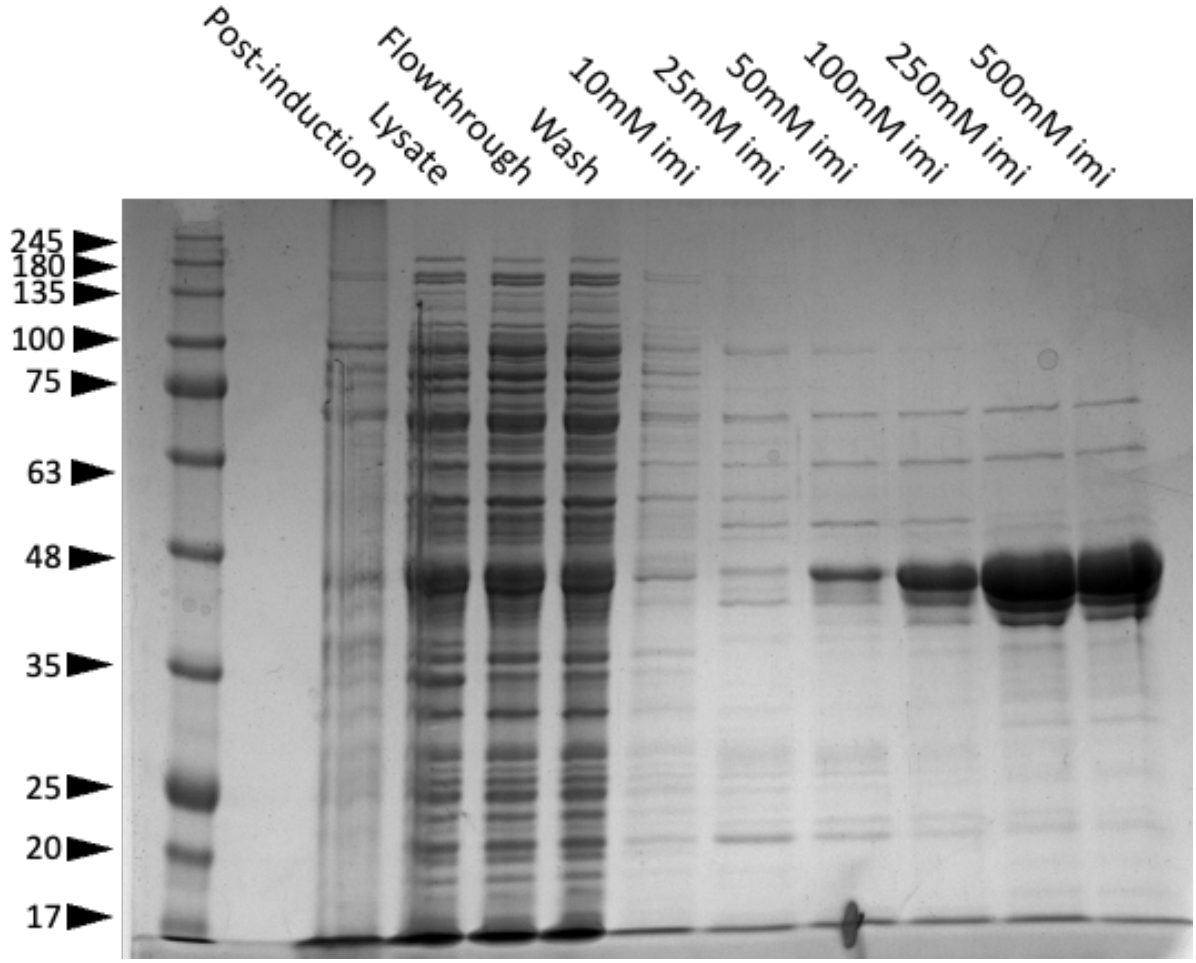


Figure 24. SDS-PAGE analysis following IMAC purification of O3I_017420 from T7 SHuffle pellets harvested from overnight 1L cultures induced with 250 μM IPTG. Cells were lysed by sonication and then processed on a gravity column using Ni-NTA resin. Protein was purified in a buffer containing 25 mM HEPES pH 8 containing 400 mM NaCl.

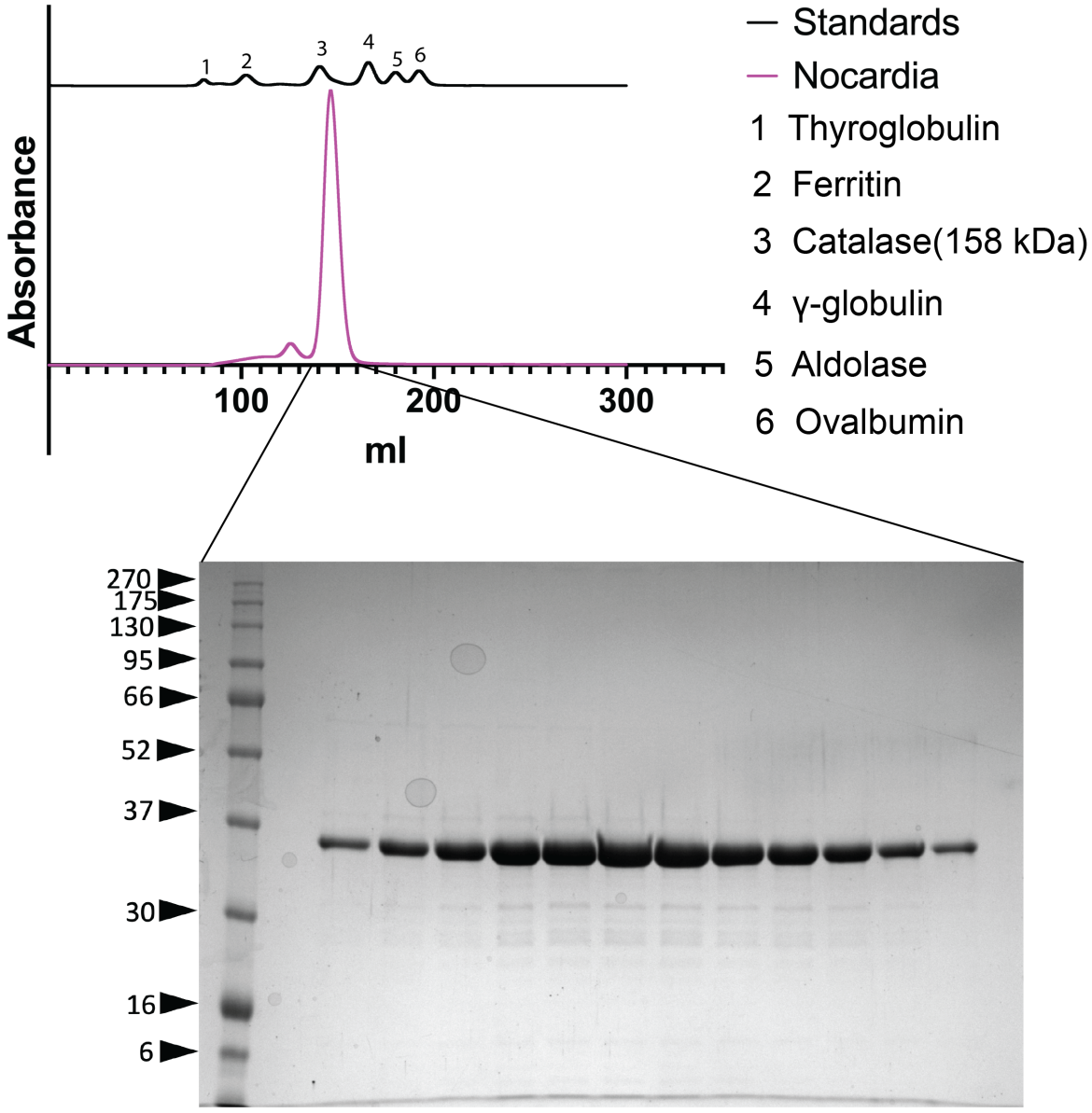


Figure 25. Size exclusion chromatography of O3I_017420 protein elutes just after the catalase standard which has a molecular weight of 158 kDa, indicating its biological assembly is slightly smaller than that, confirming a multimeric assembly in solution. SDS-PAGE analysis of the major peak indicate protein is a monomer around 37 kDa in mass. Analysis of band intensity in ImageLab (Biorad) indicated that the protein was approximately 95% pure.

7.3.4 NBras$_{Gh172}$ is an α-1,5 exo-D-arabinofuranosidase

After successful purification of the O3I_017420 gene product, we hypothesised that the protein may be an exo-D-arabinofuranosidase, as other DUF2961 family members purified by colleagues had been shown to possess this activity. Therefore, we carried out an *in vitro* analysis of reaction products formed from *M. smegmatis* arabinogalactan in the presence of O3I_017420.

As shown in Figure 27, the protein rapidly produces a species which migrates with a retention factor of 0.68, which is consistent with being D-arabinose, in a time-dependent manner. This supports our hypothesis that the protein is an exo-acting D-arabinofuranosidase. Despite the qualitative nature of TLC analysis, we can see that the product is present after 30 min but does not appear to change in quantity after 90 min. The D-arabinose present in the reaction versus the analytical standard appears to migrate slightly differently, this is due to different concentrations of phosphate buffer in the reaction conditions causing retardation in the reaction samples.

Due to the varied nature of D-arabinose linkages present in the whole arabinogalactan, use of it as a substrate did not allow us to define the linkage specificity of the enzyme. We hypothesised that the protein would likely cleave α-1,5 linkages between arabinose monomers, given that is the dominant linkage in arabinogalactan. We therefore used the pilin glycans purified from *Pseudomonas aeruginosa* PA7, which are linear chains of α-1,5-D-arabinofuranose (Kus *et al.*, 2008). As shown in Figure 27b, the enzyme liberates D-arabinofuranose from the glycan, therefore confirming that the enzyme is an α-1,5 exo-D-arabinofuranosidase.

At the time of the work, this specificity of any DUF2961 enzyme had not yet been published. However, during our work a paper was published on a DUF2961 protein from *Bifidobacterium dentium* (Kashima *et al.*, 2021). This publication established DUF2961s as the CAZy family GH172. Therefore, for the remainder of this thesis, O3I_017420 will be referred to as $NBras_{GH172}$.

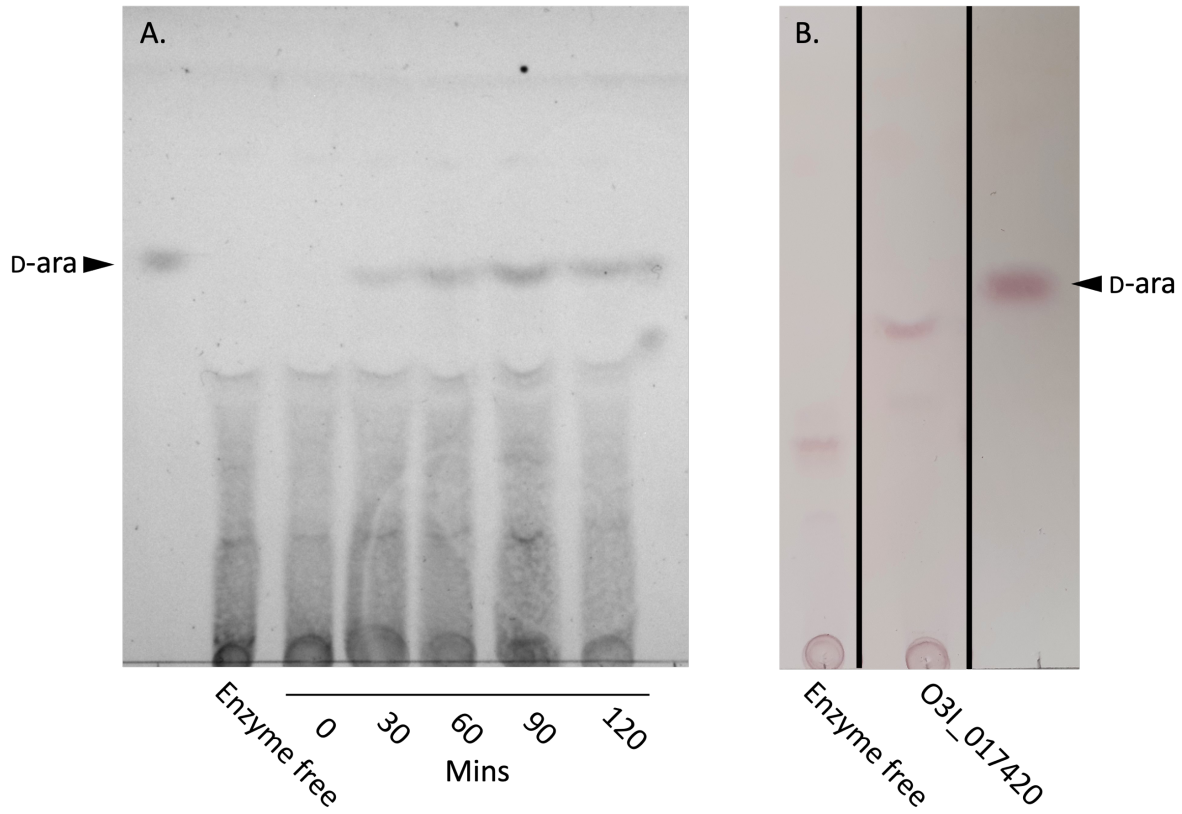


Figure 26. TLC analysis of Products formed by O3I_017420. A. Time course of an in vitro hydrolysis reaction using 5mM purified enzyme with 1 mg/ml M. smegmatis arabinogalactan. B. In vitro hydrolysis reaction using 5mM purified enzyme against 1 mg/ml PA7 pilin glycans. TLC has been cropped from a larger TLC for ease of understanding.

7.3.5 $NBras_{Gh172}$ is a retaining glycoside hydrolase.

Methanolysis is a method for experimentally verifying whether a glycoside hydrolase is retaining or inverting. It involves carrying out an *in vitro* hydrolysis reaction in the presence of 10% methanol within the reaction mixture. A retaining glycoside hydrolase typically requires a water molecule to resolve the oxazolinium intermediate produced in the first inversion step of retaining catalysis. In 10% methanol, a proportion of product will thereby be methylated due to replacement of this water by a methanol in the active site.

Figure 28 summarises TLC analysis of the reaction products produced from PA7 pilins by $NBras_{GH172}$. As can be observed in the methyl-D-arabinose standard, methylation causes a decrease in polarity, thereby reducing retention factor and causing the product to migrate higher on the TLC. In the absence of methanol, $NBras_{GH172}$ produces only D-arabinose, but with the inclusion of methanol, a second species, migrating with a similar retention factor (0.85) to the methyl-D-arabinose standard (0.9) is observed, this subtle difference is likely due to the calculation of retention factors between TLCs. This therefore supports the designation of $NBras_{GH172}$ as a retaining glycoside hydrolase which proceeds through a covalent enzyme-substrate complex.

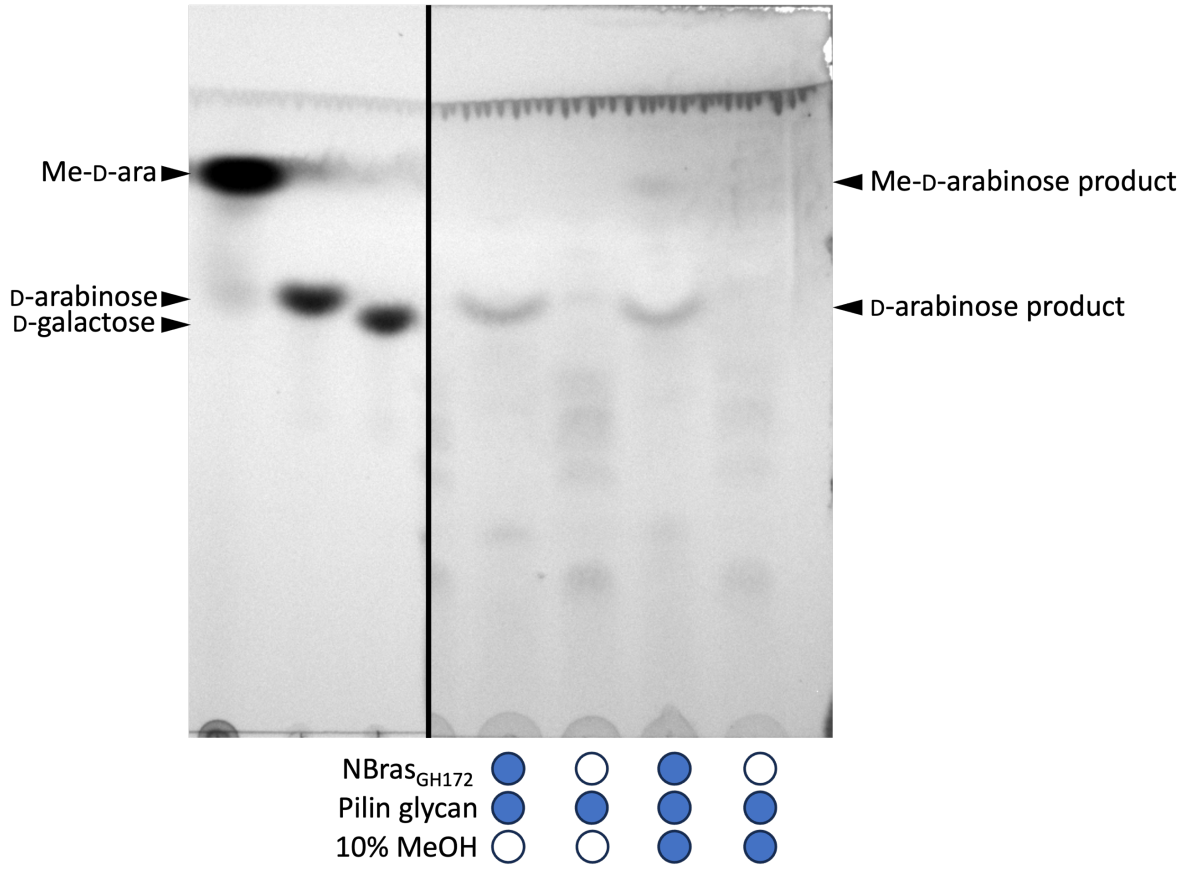


Figure 27. TLC analysis of methanolysis assay using Nbras$_{GH172}$ (5μM) and PA7 pilin glycans (1 mg/ml). TLC has been cropped from a larger, single TLC for illustrative purposes. TLC has been developed twice in 2:1:1 butan-1-ol:acetic acid: water and visualised using orcinol.

7.3.6 $NBras_{GH172}$ is trimeric

Given the trimerisation of $NBras_{Gh172}$, we sought to structurally characterise the protein using X-ray crystallography. We generated highly purified protein by SEC in Buffer B and crystallised it using sitting drop vapour diffusion. The JCSG-plus screen was used as a sparse matrix crystallisation screen and combined 1:1 with protein solution at 10 mg/mL. After approximately 4 days, large crystals were observed in a condition containing 0.1 M phosphate/citrate pH 4.2 with 40% PEG-2000 MME. These crystals were then cryoprotected using the above buffer supplemented with 30% ethylene glycol as a cryoprotectant and flash frozen in liquid nitrogen. Datasets were then collected on Beamline I04 at the Diamond light source.

The crystals diffracted to a resolution of 1.54 Å and the structure was solved by molecular replacement using the structure of a hypothetical protein crystalised by the Joint Center for Structural Genomics from *Bacteroides uniformis* (PDB ID: 4KQ7) as a search model. The asymmetric unit contained a trimeric structure possessing C_3 group symmetry. Superposition of each monomer onto the others revealed no significant structural deviations with an RMSD of 0.2 for each pair, as shown in Figure 29a. This supports the observation from size exclusion chromatography that the biological assembly is likely trimeric.

The monomers are composed of two β-jelly fold motifs, which is common to all DUF2961 family proteins, which can be seen in Figures 29b and c. Use of Foldseek against the PDB100 database provided low confidence hits (E-value greater than or equal to 9.98e-5) against several phage capsid proteins, as shown in Table 13. This indicates a potential evolutionary ancestor, which was recently reported by Krupovic and colleagues(Krupovic *et al.*, 2022). This

fold is unusual and so far has only been observed in DUF2961 proteins and these phage capsid assemblies.

Using the BLAST tool in ChimeraX, we analysed the sequence of $NBras_{GH172}$ against the Uniprot100 database and then used the resulting sequence logo to colour by sequence conservation. This analysis indicates there are multiple regions of high conservation in the protein (maroon regions on protein surface in Figure 29d). We hypothesised that these regions are involved in catalysis and multimerisation of the protein.

Table 12. Collection and refinement statistics for $_{NBrasGH172}$.

	***Nocardia brasiliensis* GH172**
Wavelength	0.9794
Resolution range	59.39 - 1.54 (1.56 - 1.54)
Space group	P 1 21 1
Unit cell	69.19 115.94 69.29 90 91.72 90
Total reflections	1318184 (55926)
Unique reflections	160730 (5278)
Multiplicity	6.8 (5.8)
Completeness (%)	99.93 (99.77)
Mean I/sigma(I)	15.5(0.6)
Wilson B-factor	28.61
R-merge	0.039
R-meas	0.042
R-pim	0.016
CC1/2	1 (0.4)
Reflections used in refinement	160730 (5278)
Reflections used for R-free	7676 (285)
R-work	0.1837 (0.4144)
R-free	0.2058 (0.4159)
Number of non-hydrogen atoms	8259
macromolecules	7848
ligands	18

solvent	393
Protein residues	998
RMS(bonds)	0.006
RMS(angles)	0.85
Ramachandran favored (%)	95.33
Ramachandran allowed (%)	4.36
Ramachandran outliers (%)	0.30
Rotamer outliers (%)	0.76
Clashscore	3.80
Average B-factor	44.10
macromolecules	44.00
ligands	41.21
solvent	46.19

Statistics for the highest-resolution shell are shown in parentheses.

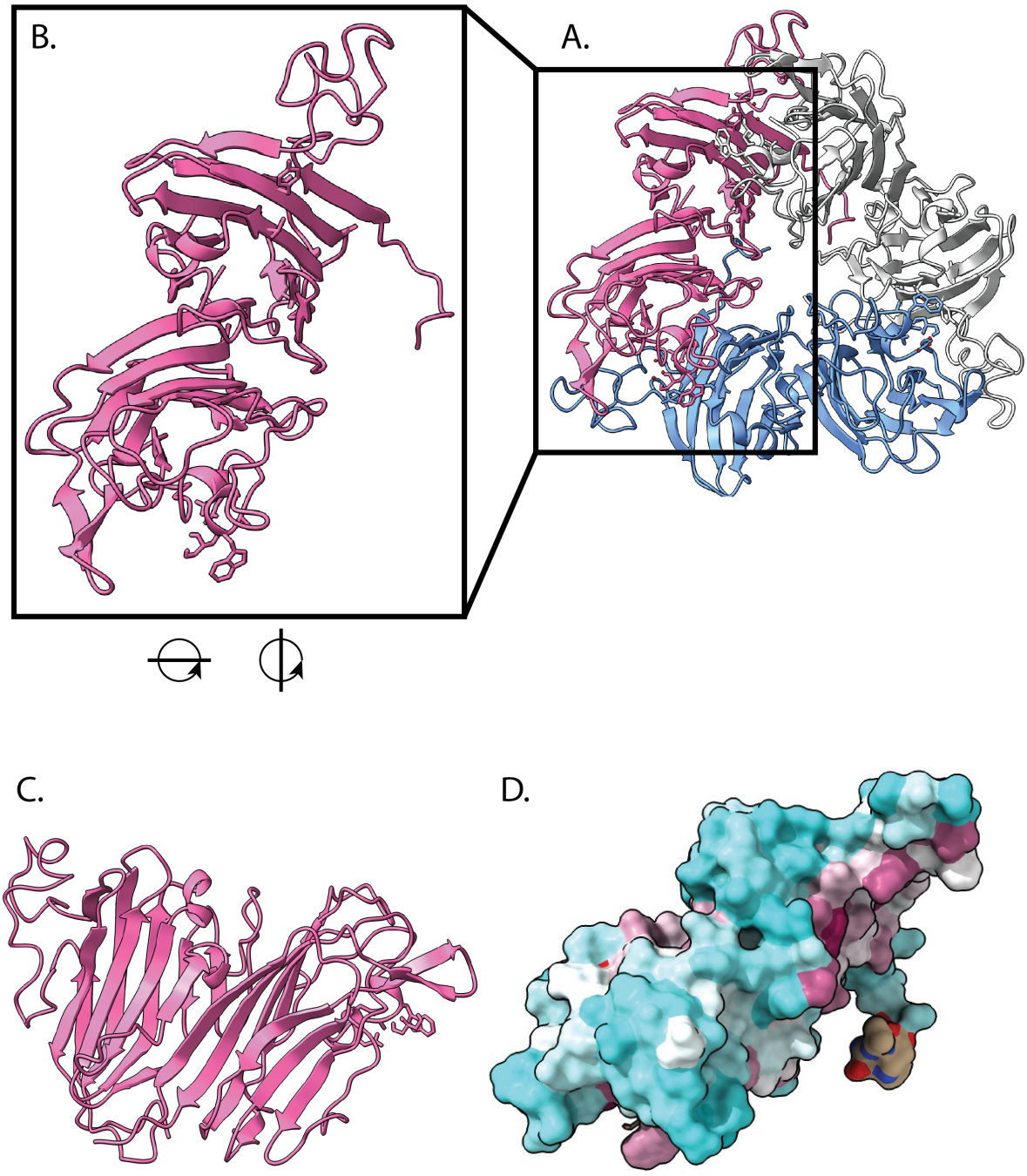


Figure 28. Structural analysis of $NBras_{GH172}$. A. Top view of the trimeric biological assembly of $NBras_{GH172}$, with each protomer coloured in a different colour. B and C. monomeric form of $NBras_{GH172,}$ highlighting the double β-jelly roll fold of the monomer, C has been rotated horizontally and vertically by 180 degrees. D. protein surface coloured by sequence conservation based on an alignment produced from BLAST search against the UniRef100 database. Maroon= conserved Cyan= nonconserved

Table 13. Hits arising from Foldseek analysis of the solved crystal structure of $NBras_{Gh172}$. RMSD calculated in ChimeraX using MatchMaker with $NBras_{GH172}$ as reference. Note that for many of the below structures, RMSD was calculated from as few as 18 paired atom pairs

PDB ID	Description	Probability	Sequence ID (%)	E-value	Pruned RMSD (Å)
2vvf_F	Crystal structure of the major capsid protein P2 from Bacteriophage PM2	1	9.6	9.98E-05	0.914
6q5u_G	High resolution electron cryo-microscopy structure of the bacteriophage PR772	1	10.4	7.83E-06	1.2
8a01_C	Bacteriophage phiCjT23 major capsid protein trimer type 1	1	9.9	6.06E-05	0.7
1w8x_F	Structural analysis of PRD1	1	10.4	3.49E-05	1.2
6oj0_Z	Cryo-EM reconstruction of Sulfolobus polyhedral virus 1 (SPV1)	1	9.7	3.64E-03	1.191
8h2i_bC	Near-atomic structure of five-fold averaged PBCV-1 capsid	0.91	9.1	3.85E-03	0.89

7.3.7 Multimerisation state of GH_{172} differs across family members

The *B. dentium* homolog found in the paper published during thesis preparation, A0A6L9SN29_9BIFI, which we will refer to as $BDEN_{GH172}$ (PDB: 7V1W) for the remainder of the chapter. Figure 30 summarises the structural comparison of the *B. dentium* homolog to the experimental structure we obtained of $NBras_{GH172}$. As shown in Figure 29a, the two proteins possess the core double β-jelly roll motif which typifies DUF2961 proteins. The proteins are highly structurally similar, with an RMSD for paired atoms of 0.8 Å. However, $BDEN_{GH172}$ has a C-terminal extension which forms an extended alpha helix, shown in in figures 30b and c, consisting of amino acids 396-449. Through these alpha helices, the $BDEN_{GH172}$ forms a dimer of trimers, whereas $NBras_{GH172}$ forms one trimer due to the lack of this encapsulating α-helix.

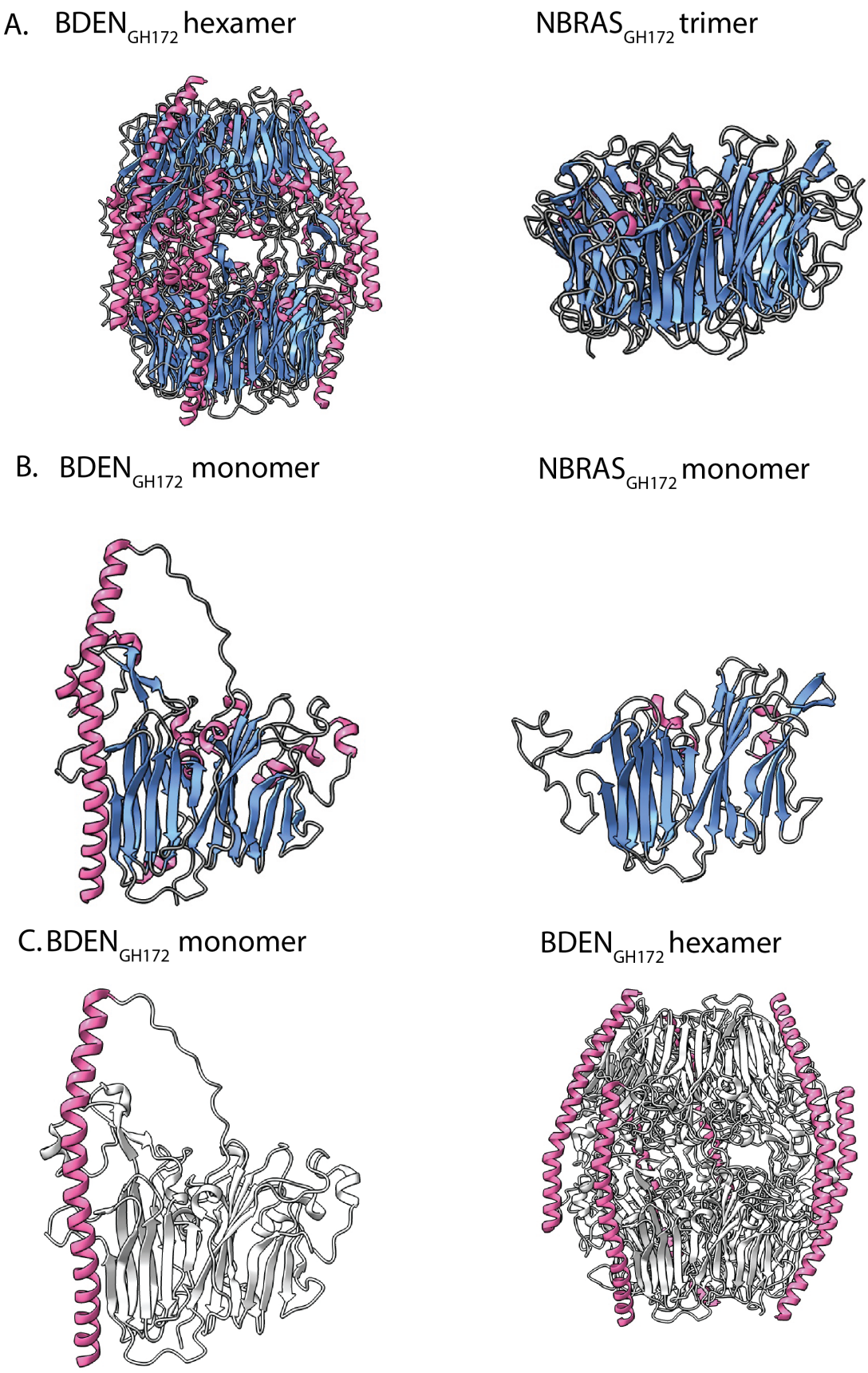


Figure 29. Structural comparison of BDen$_{GH172}$ (PDB ID: 71VW) with NBras$_{GH172}$. A. multimeric assembly of each protein coloured by secondary structure (Blue= sheet; pink= helix; grey= coil). Monomeric forms of each protein coloured as in A. C. Highlighting of encapsulating helices causing associating of two trimers.

7.3.8 Active site architecture of NBRAS$_{Gh172}$

GH172 enzymes are highly unusual as the active site is found at the interface between two protomers, with residues from both contributing to active site architecture. As stated above, it has been hypothesised that these proteins are ancestrally related to phage capsids, perhaps explaining this unusual architecture (Krupovic *et al.*, 2022). NBras$_{GH172}$ crystallised with a molecule of glycerol and five water molecules in each active site, as shown in Figure 31a. In contrast, BDen$_{GH172}$ was crystallised as a product complex, therefore the active site contains a D-arabinose molecule as well as three water molecules and a calcium ion. As shown in Figure 31c, this calcium ion was absent in the structure of NBras$_{GH172}$, instead K72 from the adjacent protomer replaces the positive charge provided by the ion, thus stabilising two adjacent water molecules which likely play a role in H-bonding and stabilising the substrate within the active site. Y183 from Chain B of NBras$_{GH172}$ also protrudes into the active site, perpendicular to the two hydrogens from the C2 of glycerol, likely producing π-stacking interactions and stabilising the substrate during binding and catalysis, as well as engaging the protomer face to stabilise the protein ultrastructure.

Identification of the catalytic residues of the enzymes is difficult due to the presence of 3 carboxylate side chains in appropriate poses for catalysis within the active site. In NBras$_{GH172,}$ the likely residues are Glu245, Glu267, and Asp268. Their counterparts in BDen$_{GH172}$ are Glu270, Glu291, and Asp292 with E270 and E292 having been confirmed as catalytic. To better understand the catalytic mechanism of NBras$_{GH172}$, we drew insight from BDen$_{GH172}$. As shown in Figure 31b, E270 and E291 are positioned 6.191 Å from each other, perpendicular to the C1 hydroxyl of D-arabinose which is the linkage they would be breaking in nature. Additionally, replacement of these two residues by Koshani and colleagues in BDen$_{GH172}$ abolished catalytic

activity of that enzyme. Informed by knowledge of the structural features of $NBras_{GH172}$, we hypothesised that the conserved glutamic acid residues within the active site were involved in the hydrolysis of the glycosidic linkage within the substrate. To test this hypothesis, we carried out Michaelis-Menten kinetic analysis of the WT enzyme and compared it to the putative catalytic-null mutant E245A.

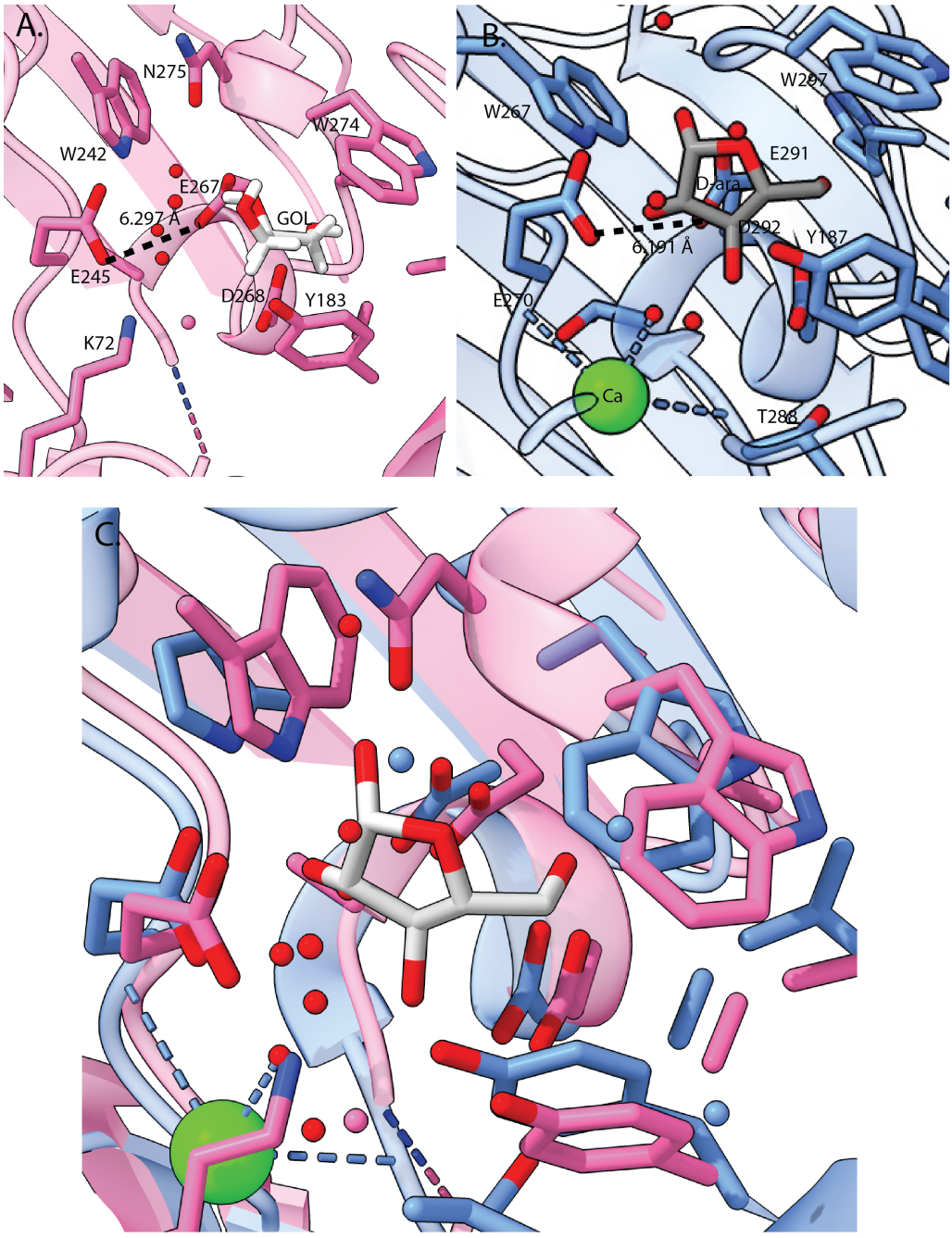
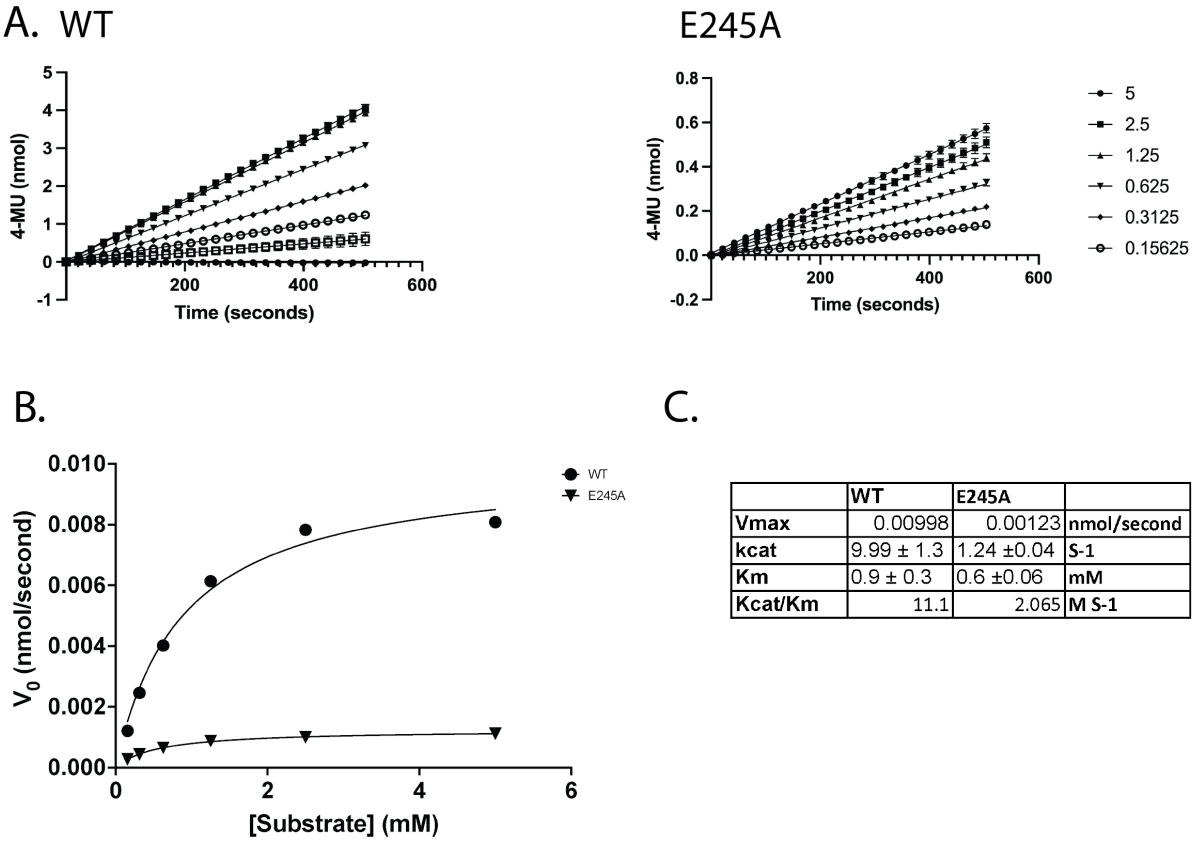


Figure 30. Superpositioning of the $BDen_{GH172}$ with $Nbras_{GH172}$ carried out using Matchmaker in ChimeraX reveals conserved active site architecture (Meng et al., 2023). Active site site of $NBras_{GH172}$ (A.). $BDen_{GH172}$ (B.) C. Superposition of $BDen_{GH172}$ (blue) with $NBras_{GH172}$ (pink) (RMSD for paired atoms = 0.23Å).

7.3.9 Michaelis-Menten kinetics of $NBras_{GH172}$

Michaelis-Menten analysis was carried out using a D-arabinofuranose glycoside, which is conjugated to 4- methylumbelliferone (4-MU) at the 1' position with an alpha linkage, thereby allowing it to act as a substrate analogue. This was synthesised in the laboratory of Prof. Spencer Williams. This fluorogenic substrate enabled facile kinetic analysis of the protein. These data are summarised in Figure 32, with the turnover rate at each substrate concentration summarised in section A, whilst the velocity plot is in Figure 32b. Section C summarises the Michaelis-Menten parameters for the two protein forms. Replacement of the putative catalytic residue, E245, with alanine interestingly does not completely abolish catalytic activity, instead reducing V_{max} approximately 10-fold. Additionally, loss of E245 decreases the K_M, from 0.88 mM to 0.56 mM, possibly through the loss of a negative charge in the active site allowing the slightly acidic sugar into the active site more easily.

Loss of catalytic efficiency can also be observed in the reduction of K_{cat}, decreasing K_{cat} from 9.987 s^{-1} in the WT to 1.239 s^{-1} in the putative catalytically null protein. However, replacement of E245 does not completely abolish activity, which may be due to the presence of the proximal aspartate residue E245 to facilitate suboptimal catalysis.



	WT	E245A	
Vmax	0.00998	0.00123	nmol/second
kcat	9.99 ± 1.3	1.24 ±0.04	S-1
Km	0.9 ± 0.3	0.6 ±0.06	mM
Kcat/Km	11.1	2.065	M S-1

Figure 31. Michaelis-Menten kinetic analysis of $NBras_{GH172}$ using 4-methylumbelliferone-D-arabinofuranoside as a substrate. Methylumbelliferone was used at a maximum concentration of 5 mM, and $NBras_{GH172}$ at a fixed concentration of 10 nM. Error bars calculated but are smaller than symbols denoting sample reference. Variance plotted in C as the 95% confidence interval.

7.3.10 *Nocardia* species are capable of growth on arabinogalactan as the sole carbon source

Given the robust data indicating that *N. brasiliensis* encodes at least one gene capable of returning arabinan into D-arabinose, we hypothesised that *Nocardia* species might be able to grow on arabinogalactan as a sole-carbon source. To verify this, we carried out a survival assay of representative *Nocardia* species against a panel of carbon sources, including arabinogalactan purified from *M. smegmatis*. As can be seen in Figure 33, sole carbon source assays in *Nocardia* were limited by two significant challenges. First, the minimal media we use for mycobacterial growth sustained growth of *Nocardia* in the absence of additional carbon. Secondly, the filamentous bacteria were difficult to disaggregate, especially *N. asteroides*, causing inoculation discrepancies in the data. With this in mind, we designed our assays to provide end-point data based on cellular respiration using resazurin, which is significantly more robust than optical density measurements (Chen *et al.*, 2018). Consequently, sufficient dynamic range and repeatability exists for us to analyse statistical significance.

Nocardia farcinica was able to demonstrate growth above background on glycerol and arabinogalactan, whilst *N. brasiliensis* seemed to grow, albeit weakly in some cases, using all carbon sources. Finally, *N. asteroides* displayed the most robust carbon utilisation capability, growing significantly above background in all conditions. Importantly, significant utilisation of arabinogalactan as a carbon source was observed for all species (1-way ANOVA; CI= 95%; P value = 0.02-0.0001). *N. brasiliensis* appears to have the smallest increase in growth in the presence of arabinogalactan. To better understand which proteins enable the growth of *N. brasiliensis* on arabinogalactan, we carried out biochemical analysis of further genes within the locus.

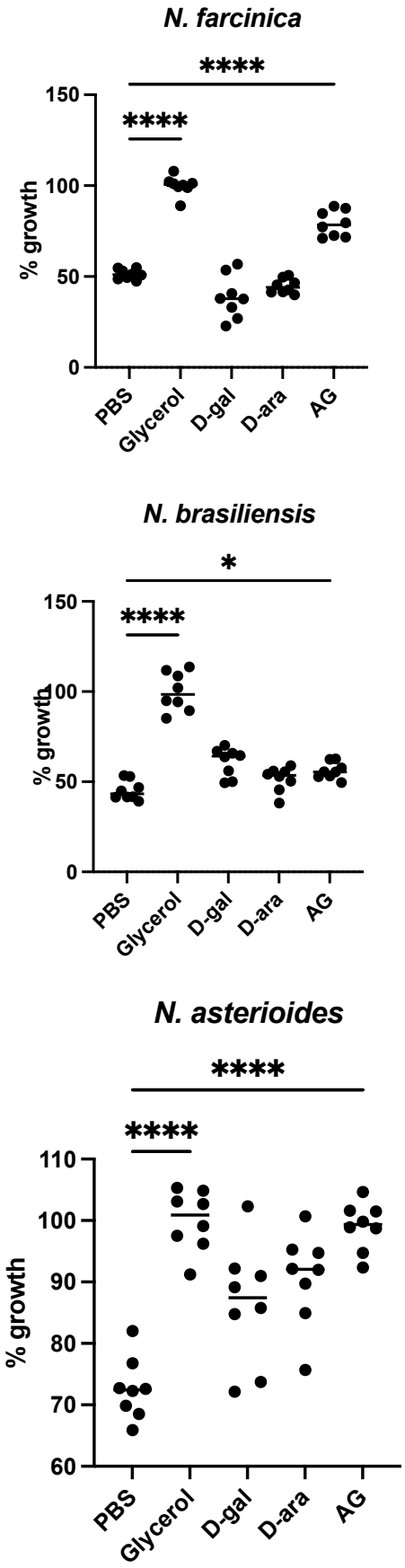


Figure 32. Growth of Nocardia *species against a panel of carbon sources in Sauton's media.* Resazurin fluorescence was used to quantify relative survival. Data are plotted as % survival compared to growth on glycerol. Glycerol was included at 0.5%, all other carbon sources were used at 5 mg/ml. Data were collected 24 hours after inoculation at 0.05. Each condition was carried out with 8 technical replicates. Statistical analysis carried out using a one-way ANOVA against PBS only control. ****= p= 0.00001, *= p= 0.01

7.3.11 Purification of *N. brasiliensis* metabolic enzymes.

To reconstitute the metabolic pathway we hypothesise is responsible for arabinan recycling in *N. brasiliensis*, the proteins first needed to be purified. As previously, pET28a plasmids containing synthetic genes of interest were ordered from Twist Biosciences. Table 14 summarises the plasmid and protein inserts studied in this section, and Figure 34 summarises.

Table 14. Summary of metabolic proteins purified from the N. brasiliensis locus. Recombinant expression plasmids were synthesised by Twist Biosciences in a pET28a backbone. The expected molecular weight of the expressed protein is in the final column.

Locus tag	Protein	Species	M_r (kDa)
O3I_017410	FGGY kinase 1	*N. brasiliensis*	53.4
O3I_017430	FGGY kinase 2	*N. brasiliensis*	48.3
O3I_017435	GH43	*N. brasiliensis*	32.7
HMPREF9455_02472	Fucose isomerase	*D. gadei*	67
HMPREF9455_02482	Fucose kinase	*D. gadei*	62.8

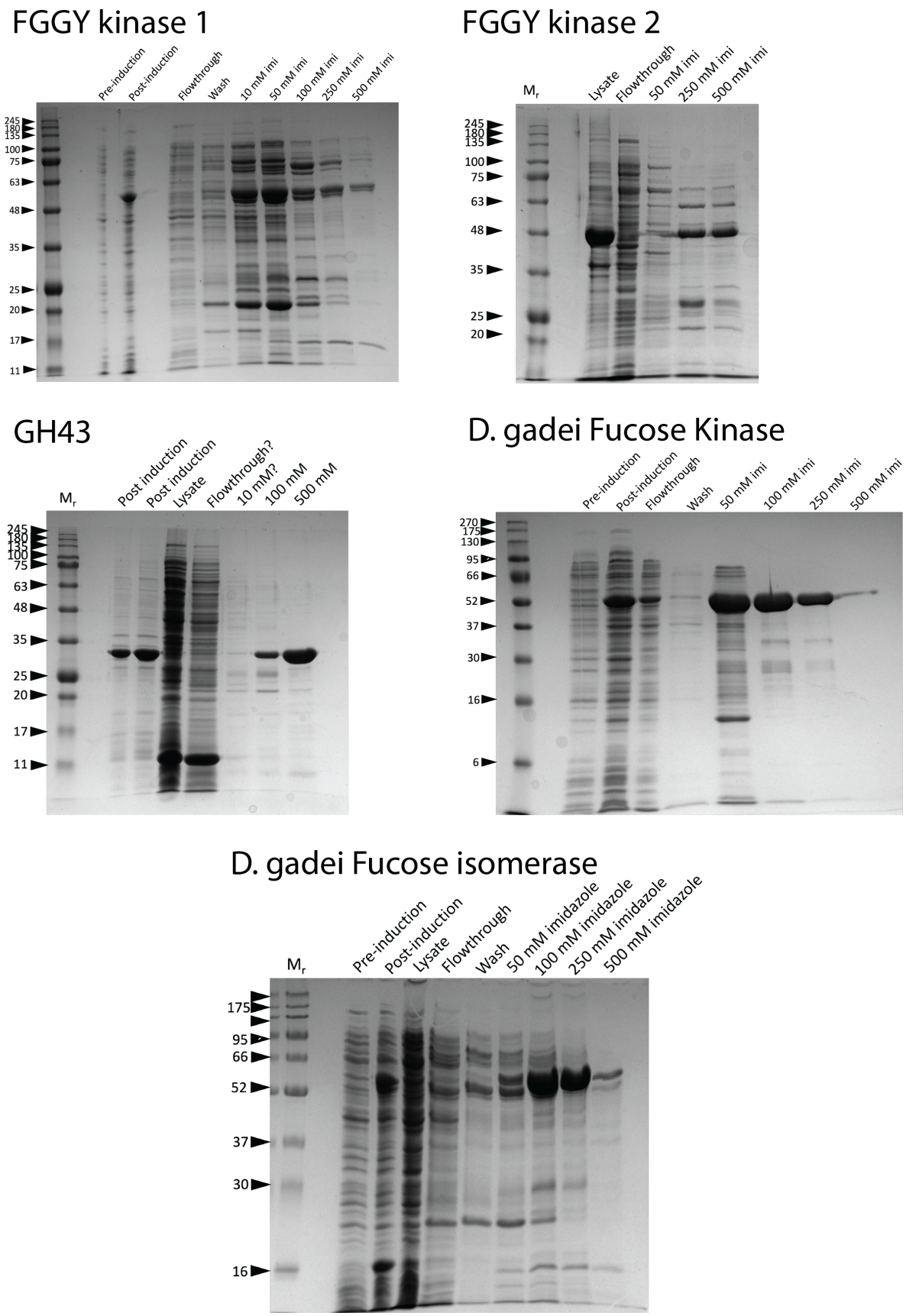


Figure 33. SDS-Page analysis of IMAC protein purification of metabolic enzymes used in this section.

7.3.12 FGGY Kinases 1 and 2 have differing substrate specificities

Following successful recombinant expression and purification of the *N. brasiliensis* FGGY kinase 1 and 2, shown in Figure 34, we carried out *in vitro* phosphorylation reactions using purified enzymes and pentoses. During initial attempts, we carried out the reactions in phosphate/citrate buffer. However, this inhibited the reactions due to the fact FGGY kinases are Mg^{2+} dependent, and the buffering salts were chelating metal ions (Kenyon *et al.*, 2012; Agarwal *et al.*, 2012).

Carrying out the reactions in 20 mM Tris-HCl pH 8 allowed analysis of the substrate preference for each of the kinases. As shown in Figure 35, products from each of the enzymes are highlighted with red arrows. FGGY kinase 1 is strongly active on D-ribulose, the ketopentose epimer of D-ribose, sharing stereochemistry at carbons 3 and 4 with D-ribose and D-arabinose. FGGY kinase 2 is weakly active on D-xylulose. Due to the limitations of TLC analysis, we cannot define the position or degree of phosphorylation. It is also important to note that both kinases are weakly active on D-arabinose and D-galactose. Based on our data, we concluded that the preferred substrates are ribulose for FGGY kinase 1 (which we will now refer to as RibuK) and xylulose for FGGY kinase 2 (which we will now refer to as XyluK). However, at present this model lacks a step catalysing the epimerisation of D-arabinose liberated from the cell wall into sugars that will serve as substrates for RibuK or XyluK for feedback into the pentose phosphate pathway.

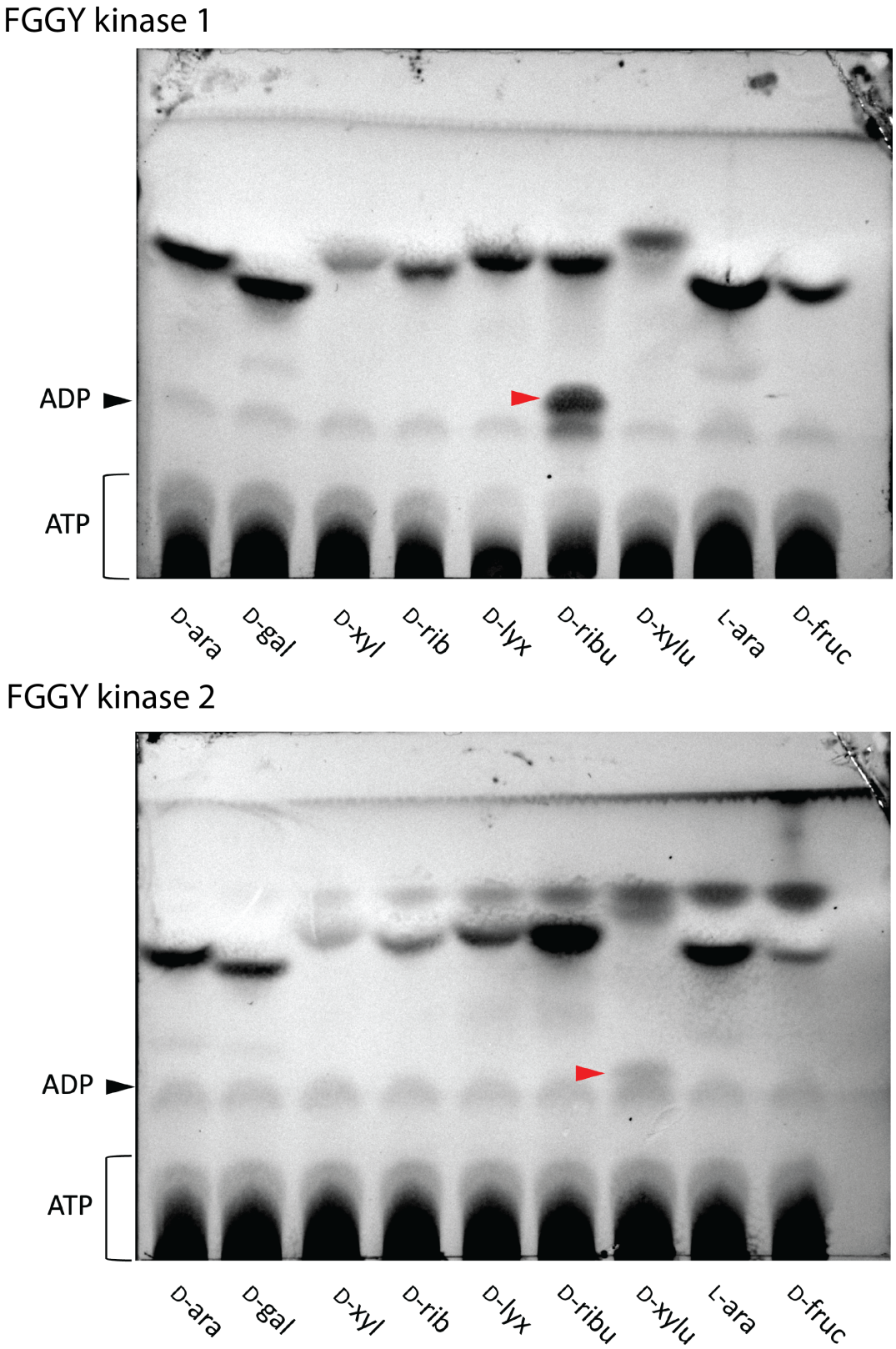


Figure 34. TLC analysis of FGGY kinase 1 and 2 substrate specificity. Enzyme was included at a concentration of 5 μm, and monosaccharide substrate at a concentration of 1 mM. Following 3 hr incubation of enzyme with appropriate substrate, reactions were boiled for 10 minutes to inactivate enzyme. 10 μL reaction mixture were then spotted at the base of a silica gel TLC foil and developed once in a system of 2:1:1 butan-1-ol:acetic acid:water. Carbohydrates were visualised using orcinol TLC stain.

7.3.13 Identification of a putative arabinose epimerase

Despite repeated attempts to purify the *N. brasiliensis* fucose isomerase, it remained completely insoluble. Fortunately, the ortholog from *D. gadei* was reasonably soluble and possessed reasonable sequence conservation to the isomerase encoded in the *N. brasiliensis* locus, and near-complete sequence identity in the predicted active site. Following recombinant expression and purification of the *D. gadei* homolog (HMPREF9455_02472), we once again carried out *in vitro* assays to identify its substrate and product.

As shown in Figure 36, the putative fucose isomerase catalyses the both the forward and reverse epimerisation of arabinose into ribulose. Due to the fact the reaction is first order to the amount of substrate, the reverse reaction can be catalysed with sufficient product. So, it is difficult to say which is the preferred direction of reaction. However, with this enzyme, the full pathway of recycling arabinogalactan to ribulose or xylulose phosphate for recycling through the pentose phosphate is complete. We therefore designate this protein as an arabinose isomerase (AraI).

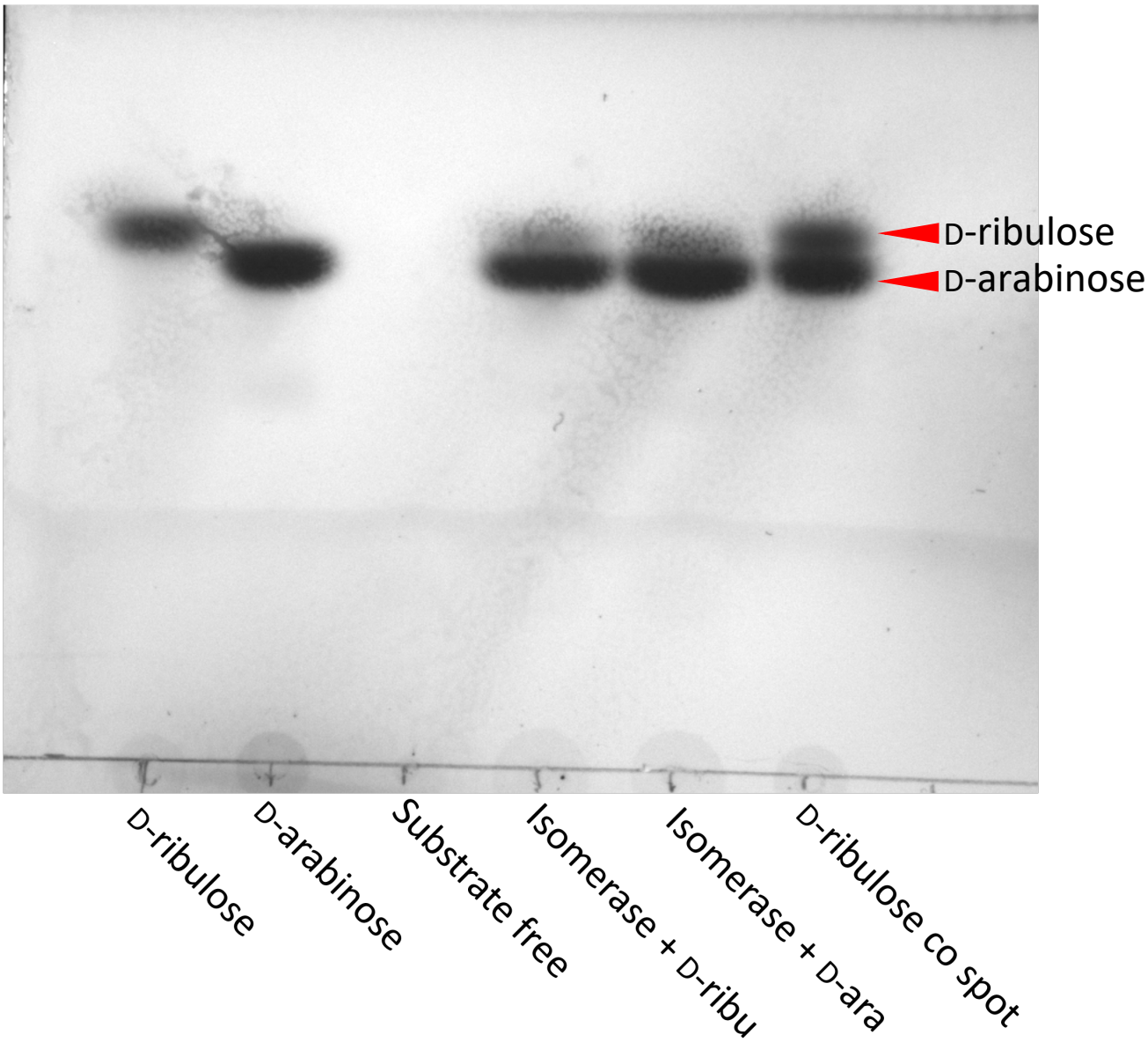


Figure 35. TLC analysis (system= 2:1:1 butan-1-ol:acetic acid:water) of 3 hr enzyme reaction using 5 µM Isomerase with ribulose and arabinose (1 mM) as substrate. Lanes (left to right) D-ribulose standard, D-arabinose standard; isomerase only; isomerase + D -ribulose, isomerase + D -arabinose; isomerase + D -arabinose co-spotted with D-ribulose

7.3.14 Characterisation of *N. brasiliensis* putative GH43

With the wide array of tools at our disposal for studying the specificity of D-arabinanases, we next set about characterising the annotated GH43 (this will now be referred to as $NBras_{GH43}$). $NBras_{GH43}$ is variably annotated as GH43 or GH32 in different nocardial genomes, suggesting that it does not sit fully within either family. Given its proximity to the GH172 with proven α-1,5-D-arabinofuranosides activity, we hypothesised that the enzyme might catalyse the hydrolysis of a different linkage within arabinogalactan (Goyal *et al.*, 2016; Farro *et al.*, 2018). Using recombinant purified enzyme, we carried out *in vitro* hydrolysis using a panel of D-arabinan-containing substrates and analysed reaction products by TLC (Figure 37).

On this TLC the D-arabinose standard appears to be contaminated with an unknown contaminant migrating with an Rf 0.45. We then used two arabinogalactans purified from *Corynebacterium glutamicum*, a close relative of *Mycobacterium smegmatis*. The WT *C. glutamicum and M. smegmatis* AG yield D-arabinose (Rf= 0.68) upon treatment with $NBras_{GH43}$. Incubation of AG purified from a Δ*aftB* strain of *C. glutamicum,* which lacks β-1,2 capping, does not appear to yield any product (Jankute, Alderwick, *et al.*, 2017). This therefore suggests that $NBras_{GH43}$ possesses exo-β-1,2-D-arabinofuranosidase activity. Subsequently, a disaccharide of β-1,2-D-arabinofuranose was provided to us by Prof. Todd Lowary. This was a glycoside of the disaccharide conjugated to an 8-carbon azide linker used in their synthesis. The rightmost lane is untreated substrate, whilst to the left of it can be seen the enzyme treated glycoside. Treatment of the glycoside with $NBras_{GH43}$ causes the production of D-arabinose. This confirms the $NBras_{GH43}$ as a β-1,2-D-arabinofuranosidase.

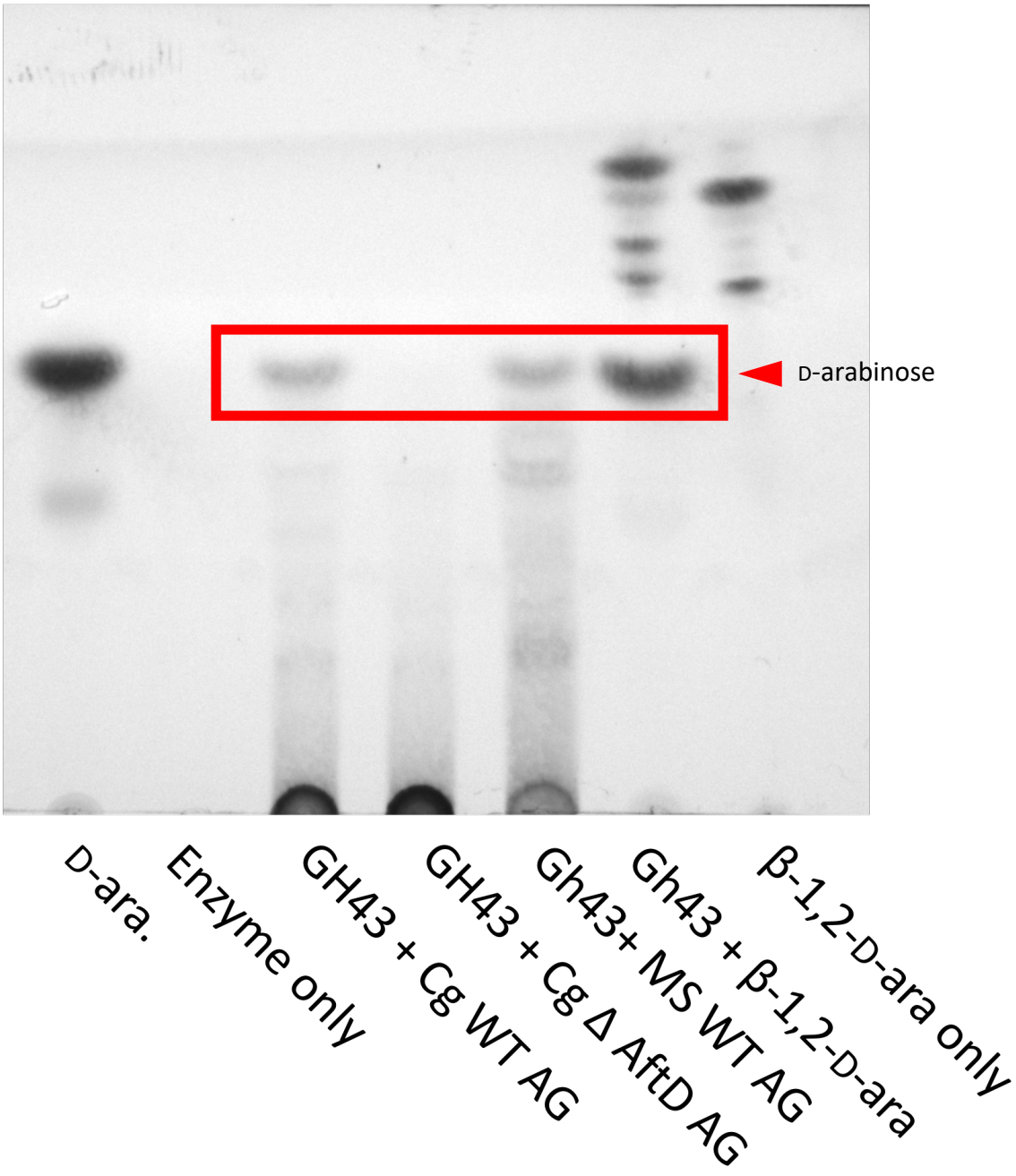


Figure 36. TLC analysis of in vitro hydrolysis assays of arabinan-containing substrates using 5 μM GH43 and 1 mg/ml of substrate. Following 3 hr incubation of enzyme with appropriate substrate, reactions were bioiled for 10 minutes to inactivate enzyme. 10 μL reaction mixture were then spotted at the base of a silica gel TLC foil and developed twice in a system of 2:1:1 butan-1-ol:acetic acid:water. Carbohydrates were visualised using orcinol TLC stain.

7.3.15 Mass spectrometry proteomics of *N. brasiliensis*

To test our hypothesis that the locus identified was implicated in recycling of arabinogalactan in *N. brasiliensis*, we carried out mass spectrometry proteomics of *N. brasiliensis* during growth on arabinogalactan, glucose, and glycerol in collaboration with Dr. Nichollas Scott (University of Melbourne). Bacteria were harvested following growth on each carbon source. Samples were then prepared for analysis using two methodologies. Information on each trial is summarised in Table 15.

Table 15. Sample information for mass spectrometry proteomics of *N. brasiliensis* grown on each carbon source.

Carbon source (attempt #)	Number of samples	Average OD_{600} at harvest (*n*=5)	Lysis method
Arabinogalactan 5 mg/ml (1)	5	0.4	None
Glucose 5 mg/ml (1)	5	0.4	None
Glycerol 0.5% (1)	5	0.4	None
Arabinogalactan 10 mg/ml (2)	5	0.6	Bead beating
Glucose 10 mg/ml (2)	5	1.2	Bead beating
Glycerol 0.5% (2)	5	1.2	Bead beating

As summarised in Table 15, the OD_{600} across the samples at the stop point of the experiment varied due to amount of available carbon in each sample. This is particularly notable comparing arabinogalactan sample set two to the other samples in the same run. The OD_{600} reached 0.4 the day before and overnight increased by 0.2.

Table 16. Twenty proteins with greatest change in abundance (log2 fold change) comparing growth on glucose to arabinogalactan. -log p value describes the statistical significance of this change in regulation, not degree. Signal peptide analysis was carried out using SignalP 5.0 using FASTA files for each protein obtained from uniprot using Gram positive signal peptides as the searches. Additional literature search was carried out in the cases of no secretion signal being observed through checking if homologs of genes found in bacteria are secreted through mycobacterial Type 7 secretion systems such as ESX-1.

Gene Name	Uniprot ID	Gene Function	Log2 fold change	-log p value	Signal peptide?
NCTC11294_03718	A0A379BKV2_NOCBR	Aminopeptidase S (M28 family)	4.3358	4.3358	Sec
NCTC11294_01895	A0A379A786_NOCBR	Uncharacterized protein conserved in bacteria	4.17985	5.5465	Sec
NCTC11294_02533	A0A379A8F0_NOCBR	Uncharacterized protein	4.1391	5.23835	Other
NCTC11294_03521	A0A379BJ01_NOCBR	MspA	4.11126	4.11126	Other
NCTC11294_06331	A0A379BY61_NOCBR	Tat pathway signal protein	3.99334	4.26539	Sec
NCTC11294_01902	A0A379A7J7_NOCBR	Uncharacterized protein	3.88798	3.25532	Sec
NCTC11294_07362	A0A379BZL3_NOCBR	Phenylalanine-4-hydroxylase	3.87046	4.42914	Other
NCTC11294_06441	A0A379BYG2_NOCBR	Tyrosine-protein phosphatase iphP_3	3.77937	3.80417	Sec
NCTC11294_01512	A0A379A7C5_NOCBR	Probable endopeptidase cgR_2070	3.746	3.8046	Sec
NCTC11294_00369	A0A379A2D6_NOCBR	LGFP repeat	3.70587	3.6973	Lipoprotein
NCTC11294_03282	A0A379BJ92_NOCBR	Metallophosphoesterase, PPA1498 family	3.58323	3.58323	TAT
NCTC11294_03801	A0A379BJC2_NOCBR	Aminopeptidase S	3.56726	5.17157	Sec
NCTC11294_01283	A0A379A4N5_NOCBR	DUF4185 domain-containing protein	3.53042	5.27287	Sec
NCTC11294_08121	A0A379C1E7_NOCBR	Uncharacterized protein	3.5215	4.07815	Other (membrane associated)
NCTC11294_02408	A0A379A7T1_NOCBR	Poly(3-hydroxybutyrate) depolymerase	3.43778	4.35882	Sec
NCTC11294_02528	A0A379A930_NOCBR	Uncharacterized protein	3.13921	5.40294	Other (ESX-1-associated)
NCTC11294_02323	A0A379A8P3_NOCBR	Aminopeptidase pepN_2,	3.07759	4.26431	Sec
NCTC11294_01988	A0A379A6M4_NOCBR	Glutathione import ATP-binding protein GsiA	3.0055	5.89469	Other
NCTC11294_02265	A0A379A7N9_NOCBR	Sulfurtransferase	-4.19305	5.80387	Other

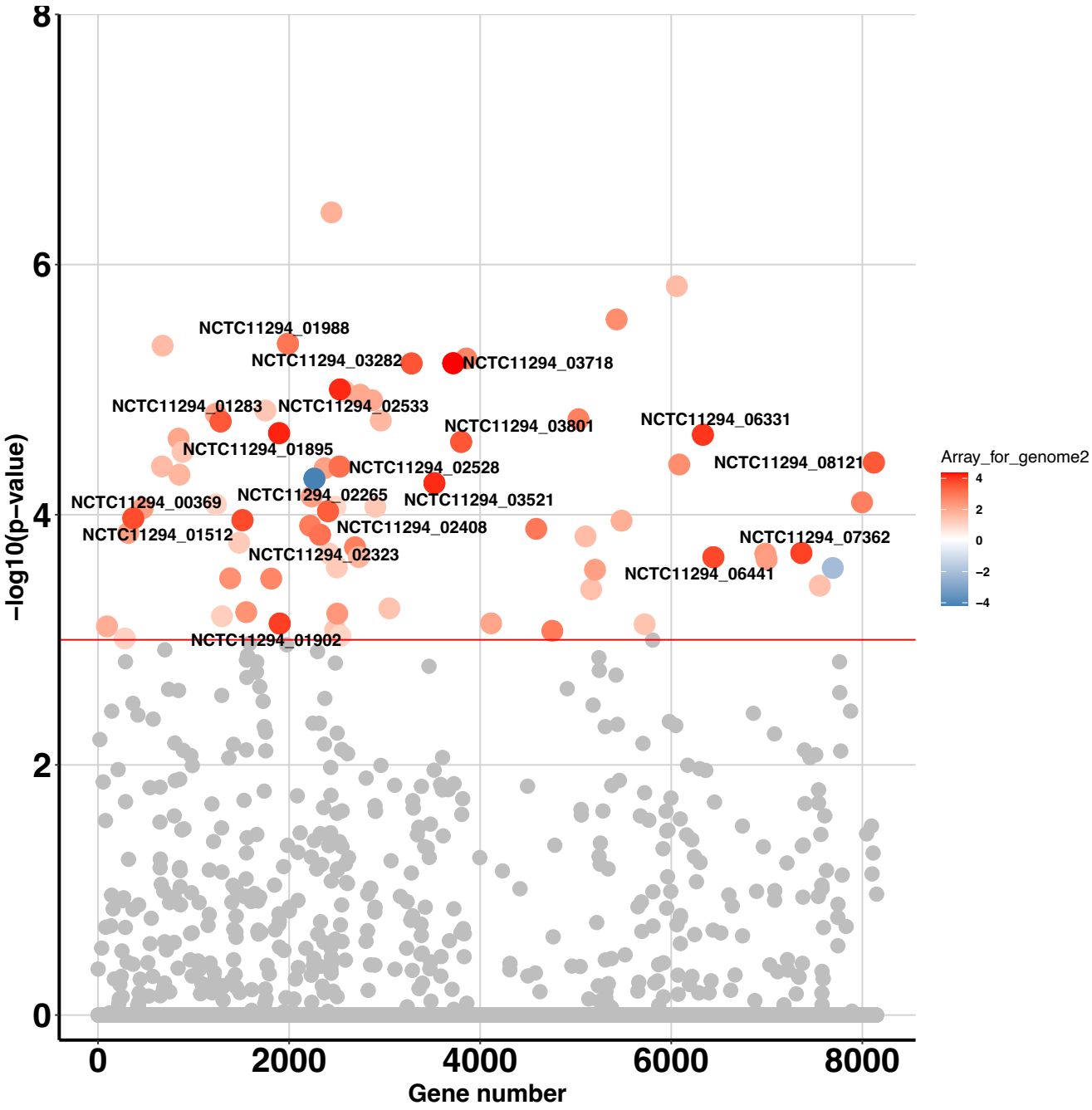


Figure 37. Manhattan plot summarising most significantly up or down hits found in analysis of the proteome of *N. brasiliensis* grown on arabinogalactan compared to those grown on glucose. Proteins are colured by log2 fold change, proteins with a -logP value of greater than 3 were considered significantly upregulated. Gene nuber referes to position in the genome based on gene number.

Cursory analysis of the dataset after processing revealed that the sample preparation methodology was suboptimal. *N. brasiliensis* has a proteome of over 8000 proteins, but only 538 were identified in the analysis. As shown in Table 16, of the 20 proteins with the greatest log-fold change, 14 (70%) possessed a secretion signal peptide based on SignalP 6.0 analysis, had been identified as membrane-associated, or were type VII secretion-associated (such as ESX-1). This strongly suggested that lysis was a limiting factor in our methodology. The lack of a mechanical lysis step meant only proteins that were external to the cell membrane were being extracted into the sample, biasing the sample for proteins that are secreted.

Despite this, many differences in the proteome in response to growth on arabinogalactan are apparent. Figure 38 is a Manhattan plot showing relative position of the ORF encoding the protein in the genome along the x axis, with individual dots representing proteins. -Log P is the significance of a proteins change in expression, and the colour of the dot represents degree of up- (red) or downregulation (blue). Values with a -log P value of greater than 3 were considered significantly upregulated. Table 15 displays the 20 proteins with greatest degree of change in expression in the dataset, which are also highlighted in Figure 38 with the locus tag.

The protein with the greatest degree of upregulation in the proteome is annotated as NCTC11294_03718 (Uniprot ID: A0A379A4N5_NOCBR). This is annotated as an M28 family aminopeptidase, a class of proteins which are hypothesised to be involved in the activation of alkaline phosphatases (Ishino *et al.*, 1987; Yeh and Trela, 1976). The protein with the greatest degree of downregulation in the dataset is a sulfurtransferase. PaperBlast identifies the protein as a homolog of Rv0815c, which encodes an essential protein mediating the

uptake of sulfur during infection, which mediates oxidative stress (Barth *et al.*, 2021; Meza *et al.*, 2019).

Multiple small molecule uptake systems are upregulated in response to arabinogalactan, including MspA and glutathione important ABC transporter GsiA. MspA is responsible for the uptake of many classes of molecules and is the primary hydrophilic route allowing the diffusion of sugars, ions, and other polar molecules (Stahl *et al.*, 2001; Wang *et al.*, 2017). GsiA once again implicates sulfur metabolism in facilitating growth on arabinogalactan. Most interestingly however is the significant upregulation of a DUF4185 containing protein. The previous chapter identifies this family of proteins as GH183, with multiple members catalysing the breakdown of arabinan containing polysaccharides. However, this DUF4185 enzyme is distant from the putative recycling locus.

7.3.16 Additional lysis step is required for representative proteomic extraction.

As demonstrated in Section 6.3.13, the existing methodology for proteome preparation was not suitable for *N. brasiliensis*, therefore for the second attempt, we included a bead-beating step before SDS boiling to break apart the mAGP complex and better release proteins from the cell mass. Figure 39 demonstrates the significant improvement in protein extraction the inclusion of a lysis step provides, increasing proteome representation from 537 proteins to 3378 of a total of 8810 genes, increasing percentage coverage from 6% to 38%, despite marginal loss of 99 proteins only found in Experiment 1. Additionally, the lysis step removed the need for imputation from most proteins identified, allowing more accurate quantification.

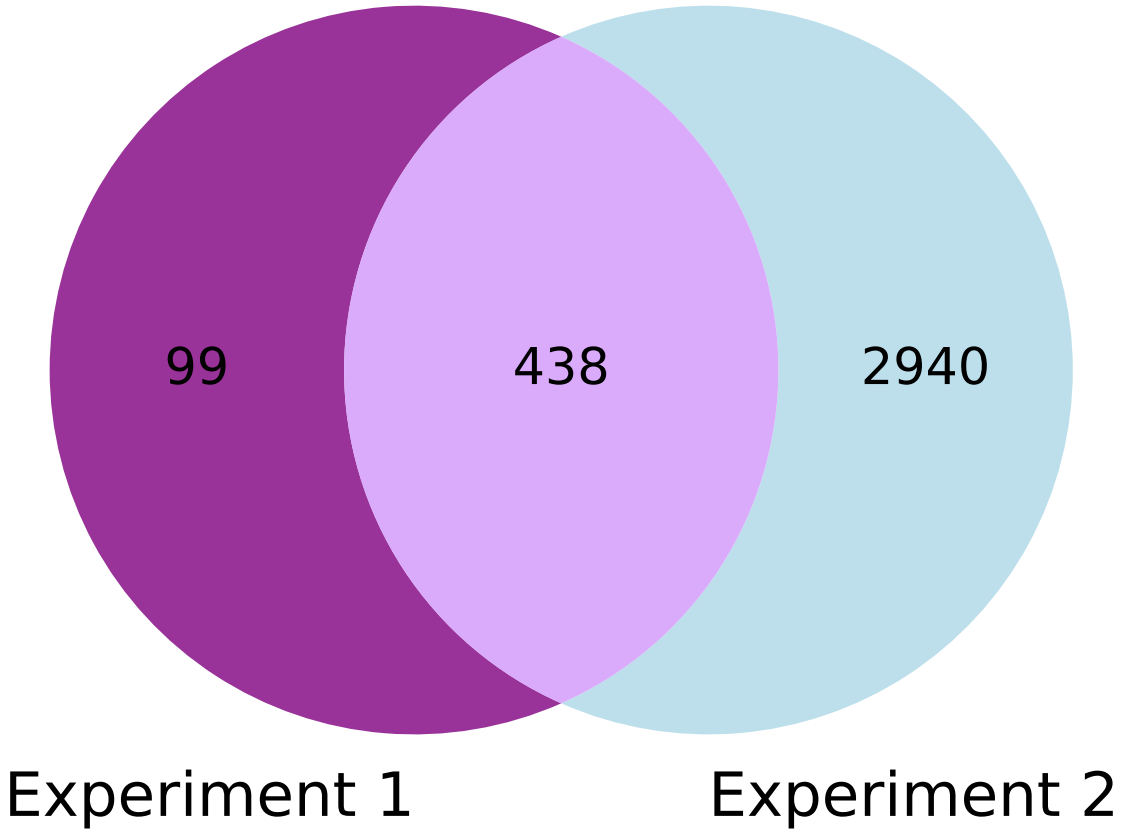


Figure 38. Venn diagram comparing number of proteins identified in Experiment 1 vs experiment 2. Experiment 1 made use of pretein samples prepared without a lysis step, whilst experiment 2 included a beadbeating step. Venn diagrams were generated using Python comparing unique and conserved proteins between the datasets with a -log(p-value) value of greater than 3. Inclusion of lysis step increased identified protein from 537 to 3378.

7.3.17 Improved proteomic analysis of *N. brasiliensis*

With a greatly improved extraction methodology, we repeated the experiment. Figure 40 is a Manhattan plot comparing relative protein abundance in the arabinogalactan condition compared to cells grown on glucose. Table 17 summarises proteins with greatest change in expression (up or down). With the improved resolution of the dataset, we can see operonic control of gene regulation, observing multiple components of the Ssu sulfonate transport complex, as well as multiple proteins from the *glp* operon, which is responsible for the glycerol-3-phosphate (G3P) shuttle. The G3P shuttle allows the reduction of a further NAD molecule using protons generated by glycolysis, which will be the primary mode of cellular energy production during growth on glucose (Yeh *et al.*, 2008), therefore downregulation of GlpD (log fold chang= -10) during growth on arabinogalactan compared to glucose is not unexpected.

Unfortunately, the translational response to arabinogalactan otherwise does not inform us of any arabinogalactan recycling pathway. However, a starvation response is apparent in the increased expression of trigger factor (A0A379BGM7_NOCBR) and RelA (A0A379C1Q8_NOCBR), mediators of central carbon stress and stringent response, respectively (Baram *et al.*, 2005; Kleine *et al.*, 2017; Danchik *et al.*, 2021). Due to the late harvesting of cell mass, we hypothesise that the bacteria were undergoing translational reprogramming to grow on another source of carbon in the arabinogalactan condition or undergoing starvation response. However, without further experimentation, it is difficult to identify what is supporting growth in the condition.

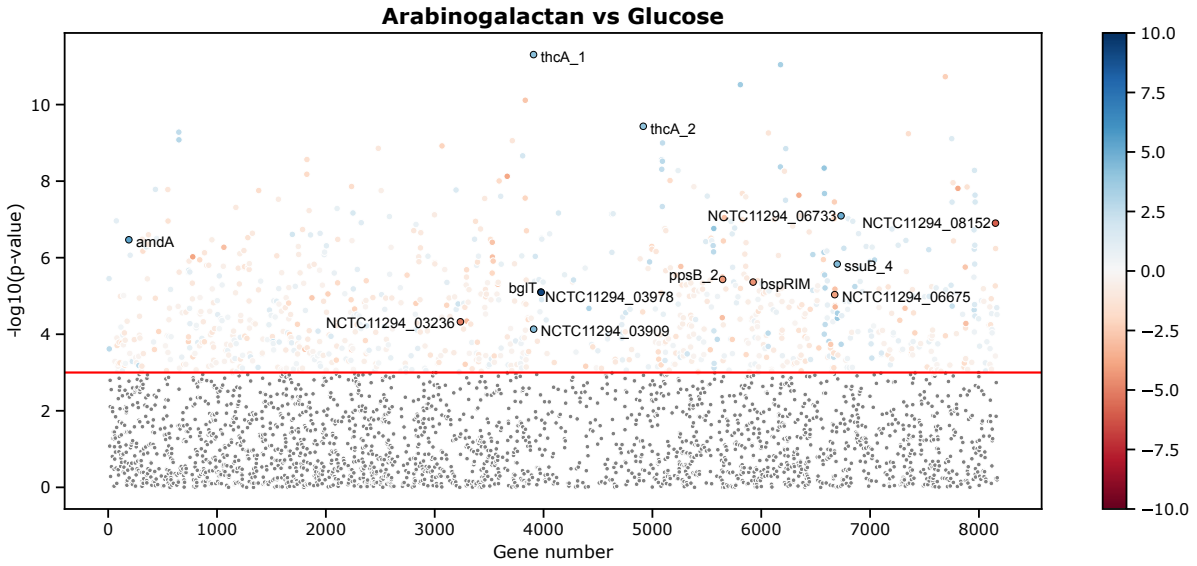


Figure 39. Manhattan plot comparing relative protein expression levels of *N. brasiliensis* grown on arabinogalactan compared to glucose. Gene number refers to position in the genome whilst the -log10(p-value) refers to the significance of gene expression change. Colour of the dot indicates log fold change, or degree up change in regulation, blue for upregulated and red for downregulated.

Table 17. Proteins from *N. brasiliensis* with the highest log-fold change in arabinogalactan vs glucose. Signal peptide prediction carried out using SignalP 6.0

Gene Name	Uniprot ID	Function	Log2-fold change	-log P value	Signal Peptide?
amdA	A0A379A2H2_NOCBR	Acetamidase	5.38558	9.34618	Other
bglT	A0A379BK68_NOCBR	Maltose-binding periplasmic proteins/domains	5.3042	4.5702	Other
thcA_2	A0A379BMW3_NOCBR	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (flavodoxin)	5.16278	8.90138	Other
NCTC11294_00650	A0A379A2Z0_NOCBR	MspA	5.04081	8.43128	Sec
NCTC11294_03909	A0A379BK21_NOCBR	6-phospho-beta-glucosidase BglT	5.02837	5.04632	Other
thcA_1	A0A379BLE5_NOCBR	Uncharacterized protein conserved in bacteria	5.00922	10.3062	Other
NCTC11294_05808	A0A379BV38_NOCBR	Isoxanthopterin deaminase	4.78524	5.88751	Lipoprotein
NCTC11294_03978	A0A379BLN4_NOCBR	EPTC-inducible aldehyde dehydrogenase	4.75765	6.49346	Lipoprotein
ssuB_4	A0A379BY85_NOCBR	Aliphatic sulfonates transport permease protein ssuC	4.4533	6.10759	Other
fabH_5	A0A379A2Z0_NOCBR	*EPTC-inducible aldehyde dehydrogenase*	4.39498	4.39498	Sec
NCTC11294_00649	A0A379A4X1_NOCBR	MspA	4.36066	8.71286	Sec
ispG_2	A0A379BVY4_NOCBR	Uncharacterized protein	4.34965	5.82614	TAT
NCTC11294_06733	A0A379BXT4_NOCBR	Roadblock/LC7 domain-containing protein	4.21117	5.17617	Other
ssuC_3	A0A379BY29_NOCBR	Aliphatic sulfonates transport permease protein ssuC	4.09585	5.42246	Other
NCTC11294_06581	A0A379BX33_NOCBR	Aliphatic sulfonates import ATP-binding protein SsuB	4.08019	7.90499	Other
gla	A0A379A2C3_NOCBR	Beta-glucosidase A	-0.667046	1.37126	Sec
glpK_3	A0A379C0L7_NOCBR	Glycerol kinase	-4.05123	8.26685	Other
accD5_3	A0A379BJM8_NOCBR	Probable propionyl-CoA carboxylase beta chain 5	-4.06329	5.44398	Other
NCTC11294_03230	*A0A379BI81_NOCBR*	Prokaryotic phospholipase A2	-4.06381	5.02891	Sec
F5X71_24245	A0A379BWY0_NOCBR	Alkyl hydroperoxide reductase subunit C	-4.27344	8.03706	Other
stp_24	A0A379C0S3_NOCBR	Spectinomycin tetracycline efflux pump	-4.34387	8.05356	Other
NCTC11294_03236	A0A379BIA4_NOCBR	Glycine zipper domain-containing protein	-4.37832	3.90621	Sec
bspRIM	A0A379BX37_NOCBR	*Cytosine-specific methyltransferase*	-4.6218	5.49983	Other
NCTC11294_06675	A0A379BZ75_NOCBR	Sterol carrier protein	-4.77981	5.71686	Other
ppsB_2	A0A379BWA7_NOCBR	Plipastatin synthase subunit B	-4.82695	5.25911	Other
NCTC11294_08152	A0A379C1T9_NOCBR	Preprotein translocase subunit YajC	-5.59952	6.90819	Sec
glpD_3	A0A379C2F4_NOCBR	Glycerol-3-phosphate dehydrogenase	-10.1711	7.07199	Other

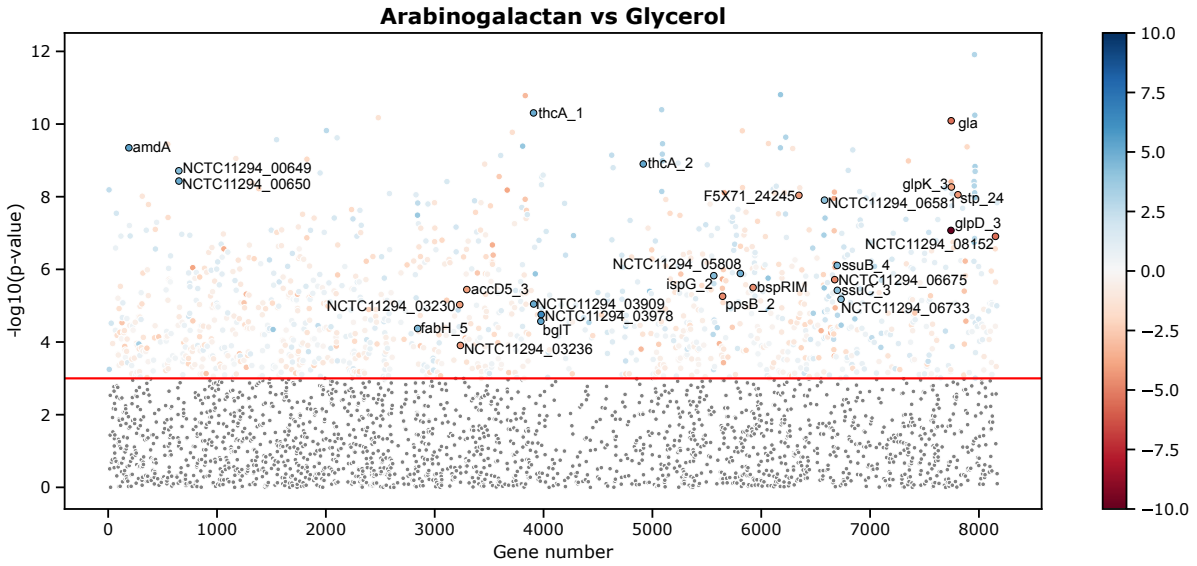


Figure 40. Manhattan plot comparing relative protein expression levels of N. brasiliensis grown on arabinogalactan compared to glycerol. Gene number refers to position in the genome whilst the -log10(p-value) refers to the significance of gene expression change. Colour of the dot indicates log fold change, or degree up change in regulation, blue for upregulated and red for downregulated.

Table 18. Proteins from *N. brasiliensis* with the highest log-fold change in arabinogalactan vs glycerol. Signal peptide prediction carried out using SignalP 6.0

Gene Name	Uniprot ID	Function	Log2-fold change	-log P value	Signal peptide?
NCTC11294_03978	A0A379BLN4_NOCBR	Maltose-binding periplasmic proteins/domains	9.16288	5.0964	Lipoprotein
bglT	A0A379BK68_NOCBR	6-phospho-beta-glucosidase BglT	5.7629	5.10236	Other
amdA	A0A379A2H2_NOCBR	Acetamidase	5.34026	6.46832	Other
NCTC11294_06733	A0A379BXT4_NOCBR	Roadblock/LC7 domain-containing protein	4.94222	7.09451	Other
ssuB_4	A0A379BY85_NOCBR	Aliphatic sulfonates import ATP-binding protein SsuB	4.39564	5.83391	Other
NCTC11294_03909	A0A379BK21_NOCBR	Uncharacterized protein conserved in bacteria	4.37082	4.13245	Other
thcA_1	A0A379BLE5_NOCBR	EPTC-inducible aldehyde dehydrogenase	4.17113	11.3059	Other
thcA_2	A0A379BMW3_NOCBR	EPTC-inducible aldehyde dehydrogenase	4.0118	9.43309	Other
NCTC11294_06675	A0A379BZ75_NOCBR	Sterol carrier protein	-4.07528	5.03461	Other
ppsB_2	A0A379BWA7_NOCBR	Plipastatin synthase subunit B	-4.34993	5.43193	Other
bspRIM	A0A379BX37_NOCBR	Cytosine-specific methyltransferase	-4.49524	5.36277	Other
NCTC11294_03236	A0A379BIA4_NOCBR	Glycine zipper domain-containing protein	-5.09264	4.3254	Sec
NCTC11294_08152	A0A379C1T9_NOCBR	Preprotein translocase subunit YajC	-6.08123	6.90251	Sec

7.4 Discussion

To date, only one other enzyme involved in turnover of arabinogalactan in Mycobacteriales has been identified. This is a galactofuranosidase enzyme, GlfH1 (Shen *et al.*, 2020). However, no further work on this enzyme has been reported since its publication in 2020. In this chapter, we present evidence for the existence of a putative arabinan recycling pathway in a pathogenic *Nocardia* species. In exploring this pathway, we have extensively characterised a GH172 enzyme from *N. brasiliensis*. During our biochemical analysis, we have identified a new oligomerisation state for the enzyme family, as well as the mechanism for this. Previously identified enzymes from this enzyme family hexamerise, whereas loss of a C-terminal helix in the $NBras_{GH172}$ enzyme prevents the face-to-face association of two trimers, preventing further oligomerisation and causing it to remain as a trimer. Additionally, we add to existing kinetic data of GH172 enzymes, indicating that our enzyme is somewhat lower in processivity than its homologue in *B. dentium* (Kashima *et al.*, 2021). A subject of further work will be whether these proteins can exist in different oligomerisation states, and why this is evolutionarily beneficial. We have also experimentally verified through our own methanolysis experiment the conclusion reached by Kashima and colleagues using molecular dynamics that GH172 enzymes are anomer-retaining (Kashima *et al.*, 2021). Another unique aspect of these enzymes is the localisation of the active site to the interaction face of two protomers, with residues from each contributing to stabilisation of the oligomer. Furthermore, through our Michaelis-Menten kinetic assays, we identify a putative catalytic residue, E245. Surprisingly, replacement of this residue with alanine does not cause a complete loss of catalytic activity. We hypothesise that other proximal carboxylate side chains may be allowing the enzyme to

suboptimally mediate catalysis, and further site directed mutagenesis is required to better ascertain these residues.

The first β-D-arabinofuranosidase was only recently reported and is a member of the GH116 family (Shimokawa *et al.*, 2023). Our data show that *O3I_017435* from *N. brasiliensis* encodes an enzyme with a similar activity. This protein does not, however, belong to the GH116 family and instead is likely the founding member of a new GH family (Shimokawa *et al.*, 2023a). This increases the number of GH families discovered during this project to three. We believe that this enzyme is responsible for specifically removing the capping motifs of D-arabinan in *Nocardia.* This will be a useful tool to the mycobacterial research community, permitting structural analyses of arabinogalactan in much greater detail. Furthermore, in the context of the recycling pathway, it allows complete depolymerisation of the arabinan to arabinose. This therefore reduces waste from cell wall remodelling through allowing access to α-linked arabinose.

We additionally present work showing growth of a microbial arabinogalactan-possessing organism able to grow on arabinogalactan. This raises interesting questions as to how the enzymes mediating this process are regulated and partitioned to not lyse cells through attack of structurally essential arabinogalactan. However, questions remain as to how this pathway functions fully *in vivo*. The presence of kinases specific for xylulose and ribulose confuses the data slightly. While both sugars are involved in the pentose phosphate pathway, production of both xylulose- and ribulose phosphates seems redundant. This pathway in *M. tuberculosis* has been reviewed in depth by Wolucka and colleagues (Wolucka, 2008). In their paper they show an alternative pathway by which exogenous D-arabinose is reduced into D-arabinitol,

which can then be dehydrogenated to xylulose, which can then be phosphorylated and reversibly epimerised into ribulose. Production of xylulose and ribulose can also allow for greater diversity of products produced in the pentose phosphate pathway, as each have distinct routes of enzymatic machinery, producing different products and thereby increasing the number of metabolites produced from one monosaccharide. Indeed, many nocardial species also include a dehydrogenase in this putative recycling locus, suggesting that flexibility in this pathway varies amongst species of *Nocardia*.

Finally, to the best of our knowledge, we present the first ever proteomics dataset of a *Nocardia* species, and one of only a handful of proteomics datasets available for Mycobacteriales, in doing so we have prepared an optimised protein extraction procedure for organisms with the mAGP complex, which will be of use to the wider mycobacterial research community. Unfortunately, due to experimental error, the datasets do not support our hypothesis that the locus identified is involved in arabinan recycling, however close analysis of the datasets indicates that many GH183 enzymes are upregulated, but not significantly so, this may indicate the end of the transition from active growth to stationary phase, giving hope for future proteomics experiments. Due to time restraints, repeats on the arabinogalactan samples were not possible but are a research priority in the short term. However, in the first dataset, the presence of a secreted DUF4185 enzyme indicates that arabinan recycling is biologically relevant. This is analogous to the selfish mechanism by which some gut Bacteroides grow on fungal mannan (Cuskin *et al.*, 2015). In that system, large polysaccharides are cleaved into oligosaccharides extracellularly, but are not turned into monosaccharides until they are within the cell and able to be phosphorylated, preventing

their return across the negatively charged plasma membrane (Briliūtė *et al.*, 2019; Crouch *et al.*, 2022).

8 Identification of a putative GH183 Type 7 effector-immunity pair in Mycobacteriales

8.1 Introduction

8.1.1 Interbacterial competition

We have learned an immense amount from the study of microbes in pure culture. However, the biological reality of microorganisms is that they grow in highly complex ecosystems such as the human gut microbiome or within soil (Velayati *et al.*, 2015; Arumugam *et al.*, 2011). With this complexity, competition for common goods such as carbon sources and essential nutrients inevitably arises. For example, in the human gut microbiome, *Bacteroides* species compete by diversifying the carbon sources they can use (Bjursell *et al.*, 2006; Briliūtė *et al.*, 2019; Chen *et al.*, 2017). This would be an example of exploitative competition. However, bacteria also compete directly in a process called interference competition, whereby members of a microbial community seek to actively harm non-related organisms to gain competitive advantage. One of humanity's first glimpses into this was the isolation of antibacterial secondary metabolites such as penicillin and vancomycin. These small molecules trigger the lysis of peptidoglycan-possessing organisms, thereby selecting strongly in favour of microorganisms without peptidoglycan (e.g. *Penicillin* molds) or those that possess resistance mechanisms (e.g. *Streptomyces* spp.) (Marshall and Wright, 1998; Yocum *et al.*, 1980). Further work on this has included proteinaceous mediators of interference competition.

At present, there are three main mechanisms by which these effectors are delivered to exert toxicity. The first is release by cell lysis, common in *E. coli* and *P. aeruginosa.* In this process,

subsets of the population trigger autolysis and release of large proteins which bind to and kill nearby competitor cells (Mavridou *et al.*, 2018; Oliveira *et al.*, 2022). This is the first known example of division of labour in bacteria. Additionally, *P. aeruginosa* presents a set of proteins, resembling phage tail fibre proteins, called tailocins on its cell wall. These physically bind non-like bacterial sugar moieties such as other species' O-antigens and teichoic acids and then punch through the membrane leading to lysis (Oluyombo *et al.*, 2019).

Arguably the best-studied of these proteinaceous interference mechanisms are the effectors of the Type VI secretion system (T6SS). These were first identified in *Vibrio cholerae* in 2006, and were initially proposed to be a protein secretion machinery for the delivery of host effectors into cells of the immune system (Mougous *et al.*, 2006; Pukatzki *et al.*, 2006). However, these systems were later shown to also mediate interbacterial competition (Hood *et al.*, 2010). The structure of the T6SS has been extensively elucidated and resembles phage tail fibre proteins, contractile in nature and delivering effectors in a manner similar to a syringe into the target organism (Bondage *et al.*, 2016; Bönemann *et al.*, 2009). Contact dependent lysis of competitors can be executed by T6SS in diderm species such as *V. cholerae* and *P. aeruginosa.* Monoderm bacteria, such as *S. aureus,* use a different protein secretion apparatus, Type VII, to carry out this contact-dependent toxin delivery. *S. aureus* was recently shown to make use of its T7SS system to deliver nuclease toxins to competitor bacteria (Whitney *et al.*, 2017; Cao *et al.*, 2016). However, to date, no such mechanisms of interspecies competition have been seen in Mycobacteriales.

8.1.2 Type VII secretion in Mycobacteriales species

Type VII secretion systems (T7SS) are a widespread family of protein translocation complexes found in Firmicutes and Mycobcteriales species such as *Staphylococcus aureus, Bacillus subtilis,* and *M. tuberculosis* (Abby *et al.*, 2016). T7SS were first discovered in *M. tuberculosis* 20 years ago and were quickly identified as critical virulence factors, with loss of a key T7SS as part of the region of difference 1 (RD1) in *M. bovis* BCG being responsible for the species' near-complete attenuation in humans (Stanley *et al.*, 2003; Mahairas *et al.*, 1996). Despite their initial discovery in *M. tuberculosis*, their homologues in *S. aureus* have become the model T7SS. Due to the differences in the cell wall between the two species, the secretion systems differ significantly in architecture leading to categorisation of those found in mycobacteria as T7SSa and those in Firmicutes as T7SSb (Klein *et al.*, 2021).

Mycobacteria encode up to five T7SS, annotated as ESX-1 to ESX-5. Evidence suggests that ESX-4 is the ancestral T7SS, with the others resulting from gene duplication or a form of horizontal gene transfer enabled by ESX-1 and 4 called distributive conjugal transfer which has been observed experimentally in *M. smegmatis* (Gray *et al.*, 2016). *M. marinum* possesses an additional plasmid-borne T7SS, ESX-P1. This secretion system acts as the conjugative mediator of the plasmid pRAW (Ummels *et al.*, 2014). ESX1,2, and 4 are nonessential for growth, whereas 3 and 5 are essential (Serafini *et al.*, 2009; Ates *et al.*, 2015).

In *S. aureus* they have been implicated in export of host targeting toxins, as well as toxins that mediate intra- and interspecies competition (Cao *et al.*, 2016). In mycobacteria, the five ESX systems carry out a diverse range of functions. In pathogenesis, ESX-1 is responsible for the

export of ESAT-6 and CFP-1, which cause rupture of phagolysosomes and escape of the bacilli into the cytosol of macrophage where they replicate freely (Gao *et al.*, 2004; Simeone *et al.*, 2012). Additionally, tuberculosis necrotising toxin is a NAD glycohydrolase, whose locus resembles a canonical Type VII toxin locus from *S. aureus*, where it is encoded proximally to a proteinaceous inhibitor within the bacillus prior to translocation by the type VII machinery into the host cell (Sun *et al.*, 2015). Furthermore, they are believed to act as channels allowing the passage of polar small molecules through the cell wall, which partially explains the essentiality of ESX-3 and 5 (Ates *et al.*, 2015; Stephan *et al.*, 2005). The ESX-family proteins are also responsible for the secretion of the PE/PPE proteins, which are proline and glutamic acid rich proteins that account for 10% of the coding capacity of the *M. tuberculosis* genome (Cole *et al.*, 1998). Therefore, these secretion systems represent a key facet of the cell wall of mycobacteria.

8.1.3 Mechanism of T7 secretion in mycobacteria

Despite the importance of T7SS in mycobacteria, studying them remains difficult due their status as large transmembrane complexes. This means many questions remain on the mechanism and functionality of these systems in mycobacteria. Recently, Beckham and colleagues solved the structure of the ESX-5 system from *Mycobacterium xenopi,* which constitutes a 1.8 MDa complex crossing the entire wall (Beckham *et al.*, 2017).

Our current knowledge of the mechanism of secretion comes from studies of *S. aureus* T7SS. The canonical secretory signal for T7 protein substrates is referred to as a WxG100 signal. The name refers to the conserved WxG motif and the length of the signal being approximately

100 amino acids in length (Poulsen *et al.*, 2014). This motif forms a heterodimer with another protein with the motif YxxxxD/E, sometimes called a carrier (Klein *et al.*, 2021). Structural analyses suggests that this pairing forms the active secretion signal which is a 4-helix bundle (Arbing *et al.*, 2010). Following this, the final 7 amino acids of the YxxxxD/E protein inserts into the N-terminal binding domain of the ATPase EccC. This forms multimerisation of the EccC ATPase domain. The remaining mechanism for translocation is unclear (Renshaw *et al.*, 2005). Despite this excellent work, questions remain on how the PE/PPE proteins are recognised, whether substrates are exported folded or unfolded, and whether the secretion is contact-dependent.

8.1.4 Identified T7SS toxins

The effectors of bacterial secretion systems represent an underexplored reservoir of biochemical potential, highlighting new mechanisms of triggering lysis within bacteria, their study can inform new treatment regimens and biotechnological applications. Type VI effectors are the best understood group of interbacterial interference toxins, with Type VII effectors now becoming similarly well characterised, as reviewed by the Palmer and Whitney laboratories (Bowman and Palmer, 2021; Tran *et al.*, 2021). Commonalities of the effectors of both systems is that they target conserved, essential substrates such as DNA, NAD^+ and PG. Toxins can broadly be grouped by locality of their substrate, either periplasmic or cytoplasmic.

Cytoplasmic toxins target DNA and cofactors found within the cytoplasm to cause lysis and death. *Serratia marscesans* and *S. aureus* secrete DNases to non-specifically cleave DNA to

prevent transcription through their Type 6 and Type 7 secretion systems, respectively (Alcoforado Diniz and Coulthurst, 2015; Cao *et al.*, 2016). A family of deaminase effectors can also catalyse the conversion of cytosine to uracil in double-stranded DNA to trigger rapid mutation and subsequent lysis, such as the Type 6 effector DddA from *Burkholderia cenocepacia* (Mok *et al.*, 2020). Tse6 from *P. aeruginosa* is an NAD-specific glycohydrolase which triggers the hydrolysis of NAD into nicotinamide and ADP-ribose, causing loss of the universal reducing agent of the cell and subsequent collapse of cellular processes (Whitney *et al.*, 2015).

Periplasmic toxins trigger lysis by direct hydrolysis of essential cell wall components, or by depolarisation of the cell membrane of competitor bacteria, which can cause uptake of further toxic small molecules through membrane channels. Tse4 from *P. aeruginosa* can permeabilise the membrane to ions, whilst VasX from *V. cholerae* even permits the uptake of larger molecules such as propidium iodide (LaCourse *et al.*, 2018; Miyata *et al.*, 2011). Both toxins are not toxic on their own, acting bacteriostatically, but can work in concert with other T6 effectors to neutralise competitors. Furthermore, PG hydrolases, both working on the glycan backbone similar to lysozyme, as well as peptidase and amidase toxins have been identified in diderm bacteria (Whitney *et al.*, 2013; Wang *et al.*, 2020).

Despite this extensive body of knowledge on Type 6 and 7 interbacterial competition effectors in other species, there is no evidence of any such mechanism in mycobacteria. This may be due to the significant structural differences in the mycobacterial cell wall to other monoderm or diderm species. However, based on the knowledge that AG is essential for the viability of mycobacteria, it represents an appealing periplasmic target of type 7 effectors

for mycobacteria. Despite the extensive involvement of Type 7 secretion systems in the homeostasis of these organisms, interference competition mediated by any ESX system has not been evidenced. This chapter aims to address this gap within the literature using the arabinan hydrolases discovered so far.

8.2 Results

8.2.1 Count of GH183 enzymes across the Mycobacteriales indicates gene duplication.

Table 19. Number of DUF4185-containing proteins encoded in the genomes of representative species from the order Mycobacteriales. Counts produced using taxonomic search of the DUF4185 protein family on PFAM.

Species	Strain	No. GH183 enzymes	Protein ID
M. smegmatis	MC^2155	5	MSMEG_2107; MSMEG_4352; MSMEG_4360; MSMEG_4365; MSMEG_6255
M. leprae	TN	2	ML2491; ML2321
M. tuberculosis	H37Rv	2	Rv1754c; Rv3707c
M. marinum	M	2	MMAR_2586; MMAR_5220
M. avium* subsp. *paratuberculosis	K-10	3	Q742I8_MYCPA; Q743L4_MYCPA; Q744N2_MYCPA
Mycobacterium phage Myrna	N/A	1	B5LJL4_9CAUD
Mycobacterium abscessus* subsp. *abscessus	L948	2	SAMEA2259684_01641; SAMEA2259684_08800
Mycobacterium abscessus* subsp. *massiliense	GO 06	3	A0A829I2D0_9MYCO; A0A829HSN6_9MYCO A0A829I1K4_9MYCO
Mycobacterium fortuitum	TMC 1529	5	XA26_21030; XA26_38440; XA26_56100; XA26_38350; XA26_38060

The absolute number of genes encoding GH183 enzymes varies widely across mycobacterial species, as summarised in Table 19. Genome reduction is a prominent driver of evolution in *M. tuberculosis* and *M. leprae*, due to their tropism now being obligate intracellular human pathogens, thereby reducing their need for as much genomic utility as soil-dwelling species

such as *M. smegmatis* (Vissa and Brennan, 2001; Chavarro-Portillo *et al.*, 2019; Nimmo *et al.*, 2022). This is apparent in the number of GH183 encoding genes in *M. smegmatis, M. tuberculosis,* and *M. leprae b*eing 5, 2, and 2, respectively. In many opportunistic pathogens, such as *M. avium* and *M. abscessus* species, there are 3. This therefore led us to hypothesise that 2 homologues are the minimum requirement for proper homeostasis of D-arabinan, and others operate in an auxiliary manner, and are possibly not required for pathogenesis.

At the start of this project, when Rv3707c remained insoluble despite extensive work, we used FDetect, a web server which predicts the solubility of proteins expressed recombinantly from *E. coli* to find more soluble GH183 homologues in Mycobacteriales (Meng *et al.*, 2018). This search suggested that Q744N2_MYCPA (now referred to as $MAP_{GH183)}$ and MYCMA_14050 (now referred to as MAB_{GH183}) from *M. avium* subsp. *paratuberculosis* and *M. abscessus* subsp. *massiliense*, respectively, would be most soluble *in vitro*.

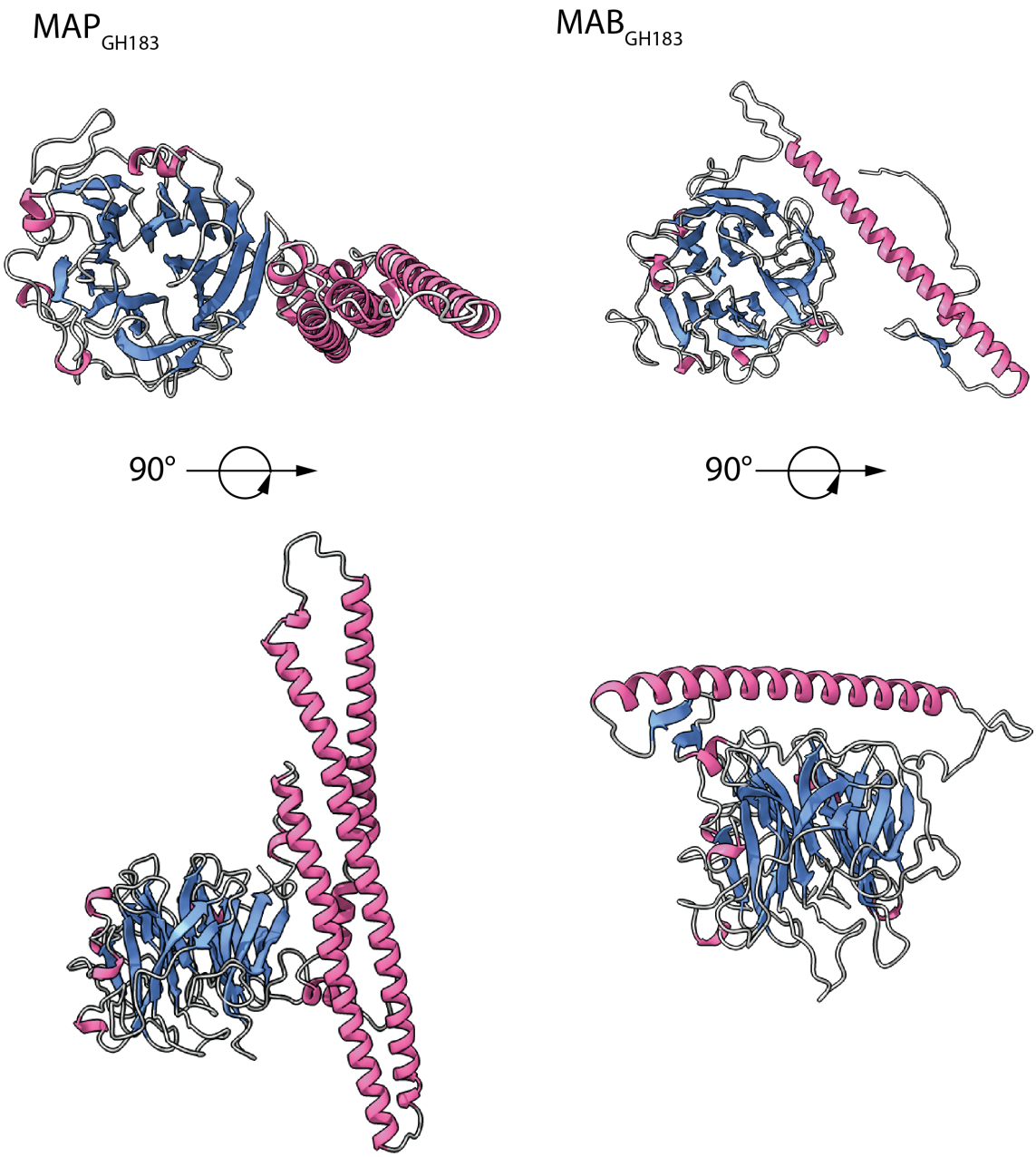
8.2.2 Alphafold analysis indicates possible type 7 secretion of MAP_{GH183} and MAB_{GH183}

Before carrying out purification of the new GH183 enzymes, we wanted to compare them to Rv3707c. Following multiple sequence alignment, as shown in Figure 42b, we could see that the two new proteins possessed low sequence similarity to Rv3707c, with MAP_{GH183} possessing 37% identity to Rv3707c at the amino acid level, and MAB_{Gh183} with 27%. MAP_{GH183} and MAB_{GH183} only possess 53% sequence identity to each other. This indicates the proteins are sequentially divergent. Notably, neither MAP_{GH183} nor MAB_{Gh183} contain any cysteine residues. Despite this, based on the Alphafold models, at the secondary structure level, the catalytic domain of the proteins superpose with an RMSD of 0.767 Å, indicating that they are

highly structurally similar. However, an unexpected observation in each structure was the presence of an N-terminal highly extended α-helical arrangement reminiscent of a Type 7 secretion signal (Figure 42). In MAP_{GH183}, this constitutes amino acids 1-195, whereas in MAB_{GH183}, this includes amino acids 1-102. We therefore hypothesised that these proteins were Type VII effectors and could be the first ever example of toxins used in interspecies competition in mycobacteria. To test this hypothesis, we carried out FoldSeek analysis of the N-terminal α-helical domains of each protein against the PDB. Encouragingly, MAP_{GH183} returned many high confidence hits against T7 substrates from their N-terminal domains such as PE-PPE proteins, as shown in Table 20.

Informed by the presence of a putative Type VII secretion signal on each of the proteins, we had each gene synthesised by Twist Biosciences in a pET28a backbone for use in recombinant protein expression from *E. coli*. The coding sequence used for each protein in the expression constructs excluded the putative secretion signals. In MAP_{GH183}, we excluded amino acids 1-195 and in MAP_{GH183}, we excluded amino acids 1-102 so that we could purify the GH183 domains of each protein.

A.



B.

	MABGH183	MAPGH183	Rv3707c
MABGH183		53.314%	27.273%
MAPGH183	53.314%		37.110%
Rv3707c	27.273%	37.110%	

Figure 41. A. Side by side comparison of the Alphafold predicted models for MAB_{GH183} and MAP_{GH183}. Ribbons of each model have been coloured by secondary structure (blue = sheet; pink = helix; grey = strand). B. Distance matrix comparing sequence similarity of MAB_{GH183}, MAP_{GH183}, and Rv3707c from *M. tuberculosis* calculated using a MUSCLE alignment.

Table 20. FoldSeek result for the N-terminal α-helical bundle from MAP_{GH183} (amino acids 1-195) against the PDB.

PDB ID	Description	Scientific Name	Prob.	Seq. Id.
6vj5_B	PE25-PPE41(A124L) in complex with EspG5	Mycobacterium tuberculosis H37Rv	1	13.5
4w4l_B	EspG5 in complex with PE25 and PPE41 from the ESX-5 type VII secretion system of M. tuberculosis	Mycobacterium tuberculosis H37Rv	1	13.6
2g38_D	A PE/PPE Protein Complex from Mycobacterium tuberculosis	Mycobacterium tuberculosis H37Rv	1	13.6
4w4k_D	Crystal structure of a PE25-PPE41 heterodimer from a type VII secretion system of M. tuberculosis	Mycobacterium tuberculosis H37Rv	1	13.3
6uuj_E	Structure of PE5-PPE4-EspG3 complex from the type VII (ESX-3) secretion system	Mycobacterium tuberculosis H37Rv	1	13.3
6uuj_K	Structure of PE5-PPE4-EspG3 complex from the type VII (ESX-3) secretion system	Mycobacterium tuberculosis H37Rv	1	12.8
6vhr_B	Structure of PE5-PPE4-EspG3 complex from the type VII (ESX-3) secretion system	Mycobacterium tuberculosis H37Rv	1	13.4
6uuj_H	Structure of PE5-PPE4-EspG3 complex from the type VII (ESX-3) secretion system	Mycobacterium tuberculosis H37Rv	0.99	12.9
5xfs_B	Crystal structure of PE8-PPE15 in complex with EspG5 from M. tuberculosis	Mycobacterium tuberculosis H37Rv	0.99	11.2

8.2.3 Purification of MAP_{GH183} and MAB_{GH183}

Following bioinformatic analysis of the putative GH183 Type VII effectors, we set out to biochemically characterise the enzymes. The proteins both expressed extremely well in *E. coli* and in high yield. The expected M_r of MAP_{GH183} and MAB_{GH183} were 35.505 and 37.151 kDA, respectively. As shown in Figure 43, MAP_{GH183} elutes with reasonable purity in the 10 and 100 mM imidazole fractions, with a ratiometric purity of 62.7% for MAP_{GH183} and 70% for MAB_{GH183}. Whilst this purity is slightly lower than that is usual for biochemical analysis, it was decided that these proteins were of sufficient quality for the analyses we were carrying out. Additionally, there are no D-arabinan hydrolases encoded in the genome of our recombinant expression strain, *E. coli* T7 SHuffle K-12. The presence of bands at a lower M_r than the protein, but not above, suggests that some of it was subject to proteolytic degradation. Unfortunately, inclusion of COMplete protease inhibitor tablets in lysis did not prevent this. MAB_{GH183}, as expected possesses a slightly higher M_r, and like MAP_{GH183}, there is significant proteolysis of the protein.

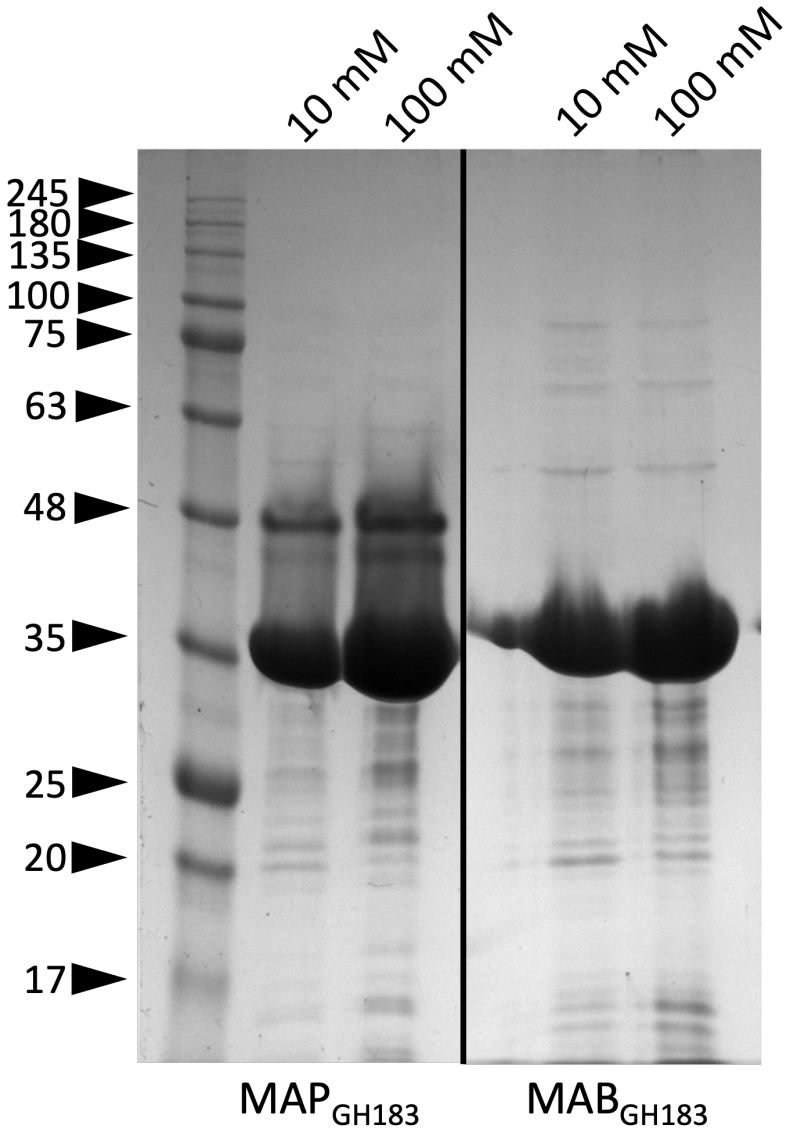


Figure 42. SDS-PAGE analysis of TALON IMAC purification of MAP_{GH183} and Mab_{GH183}. Cells were lysed by sonication in Buffer 1 before processing on a gravity column containing TALON cobalt resin. Elutions shown were in 10 and 100mM imidazole. Figure has been cropped from a larger gel for ease of understanding.

8.2.4 MAB_{GH183} is a retaining glycoside hydrolase

Due to the favourable traits of MAB_{GH183} compared to Rv3707c, we reasoned that methanolysis analysis would be informative. Catalytic mechanism in GH enzymes is typically conserved at the family level, therefore what is true for the MAB homologue is likely also true for Rv3707c. Given the challenges of working with Rv3707c, having a facile system to study will enable more sophisticated analysis of the enzyme family. This information is critical if we wish to go on to design inhibitors of Rv3707c as retaining glycoside hydrolases can be locked in their covalent enzyme-substrate complexes by inhibitors.

MAB_{GH183} causes extensive degradation of arabinogalactan, as shown by the release of four species of oligosaccharides in Figure 44. Based on the L-arabino standards, the products are primarily 3, 4, 6, and 7 monosaccharides in length, with larger oligosaccharides migrating below the standards. Compared to the product profiles of Rv3707c, the low sequence conservation between the two proteins clearly leads to significant alterations in product profile. Inclusion of methanol in the reaction also causes the appearance of methylated products above the standards as indicated by the red box in Figure 44. We therefore posit that GH183 enzymes are anomer-retaining.

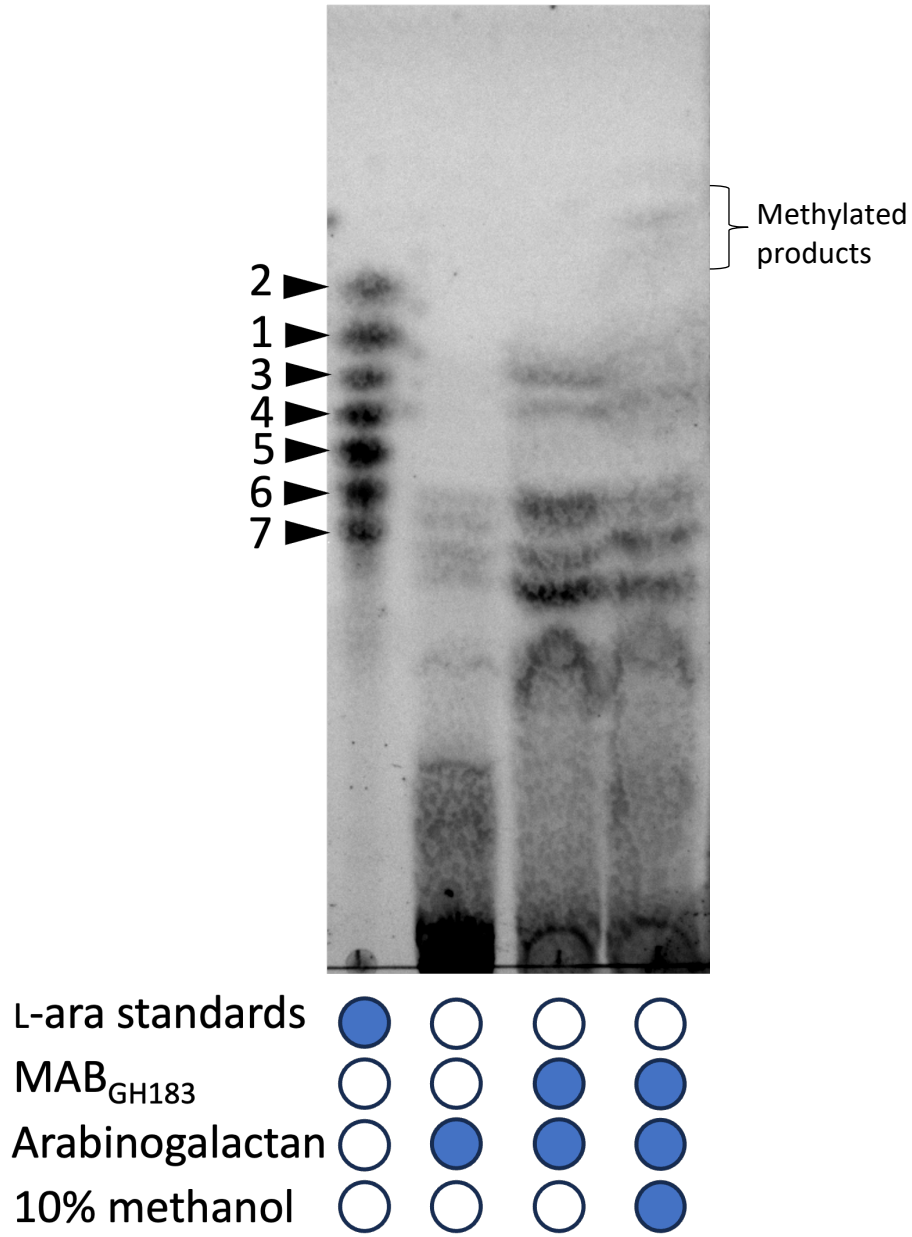


Figure 43. TLC analysis of reaction products produced using recombinant MAB_{4185} (5 µM) against *M. smegmatis* AG (1 mg/ml) for 3 hours. TLC system is 2:1:1 butan-1-ol:acetic acid: water, TLCs were developed twice in the system before visualisation using orcinol TLC stain. Inclusion of methanol in the reaction mixture results in species with a lower R_f, indicating a retaining reaction mechanism.

8.2.5 Structure of MAB_{GH183}

Following successful crystallisation of MAB_{GH183}, we collected a high-resolution dataset (resolution = 1.953 Å) on beamline I04 at Diamond Lightsource. As summarised in Table 19, following molecular replacement using the Alphafold structural prediction for $MAB_{GH183,}$ the structure solved with high confidence. The protein crystallised with two protomers in the asymmetric unit. Superposition of the two chains in the asymmetric unit results in an RMSD of 0.250 Å, indicating that there is very little structural flexibility captured in the crystal. The proteins differ in the coordination of a glycerol molecule in a surface pocket formed on Chain B. As shown in Figure 45b, the residues involved in coordination are R390 and D381, forming one and two hydrogen bonds, respectively. Additionally, it appears that W393 is posed to engage in π-stacking interactions with the glycerol, but it is at a distance of greater than 5 Å, which makes this unlikely. It is unclear if this coordination is biologically relevant or whether it is an artefact of the cryoprotection of the crystal with 30% glycerol.

As is the case with Rv3707c, the protein adopts a five-bladed β-propeller conformation in the crystalline lattice, with the five blades surrounding a central pore which penetrates the depth of the protein, this can be seen in Figure 45a. This appears to be a core structural feature of GH183 enzymes, with its conservation observed in both Rv3707c and MAB_{GH183}. Additionally, the five blades appear slightly degenerate with one blade, consisting of amino acids 293-328 forming a smaller blade than the other 5 β-sheets.

Table 21. Data collection and refinement statistics for MAB_{GH183}

	GH183 from *M. abscessus*
Wavelength	0.9795
Resolution range	64.51 - 1.953 (2.023 - 1.953)
Space group	C 2 2 21
Unit cell	88.007 94.837 175.803 90 90 90
Total reflections	72756 (290)
Unique reflections	36415 (145)
Multiplicity	2.0 (2.0)
Completeness (%)	67.85 (2.72)
Mean I/sigma(I)	6.63 (1.33)
Wilson B-factor	29.02
R-merge	0.07291 (0.5035)
R-meas	0.1031 (0.7121)
R-pim	0.07291 (0.5035)
CC1/2	0.993 (0.59)
CC*	0.998 (0.861)
Reflections used in refinement	36409 (145)
Reflections used for R-free	1820 (2)
R-work	0.2242 (0.3180)
R-free	0.2452 (0.4232)
CC(work)	0.938 (0.657)

CC(free)	0.920 (-1.000)
Number of non-hydrogen atoms	5361
macromolecules	5042
ligands	6
solvent	313
Protein residues	684
RMS(bonds)	0.009
RMS(angles)	1.20
Ramachandran favored (%)	96.76
Ramachandran allowed (%)	3.24
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.32
Clashscore	3.05
Average B-factor	23.62
macromolecules	22.91
ligands	43.51
solvent	34.66
Number of TLS groups	2

Statistics for the highest-resolution shell are shown in parentheses.

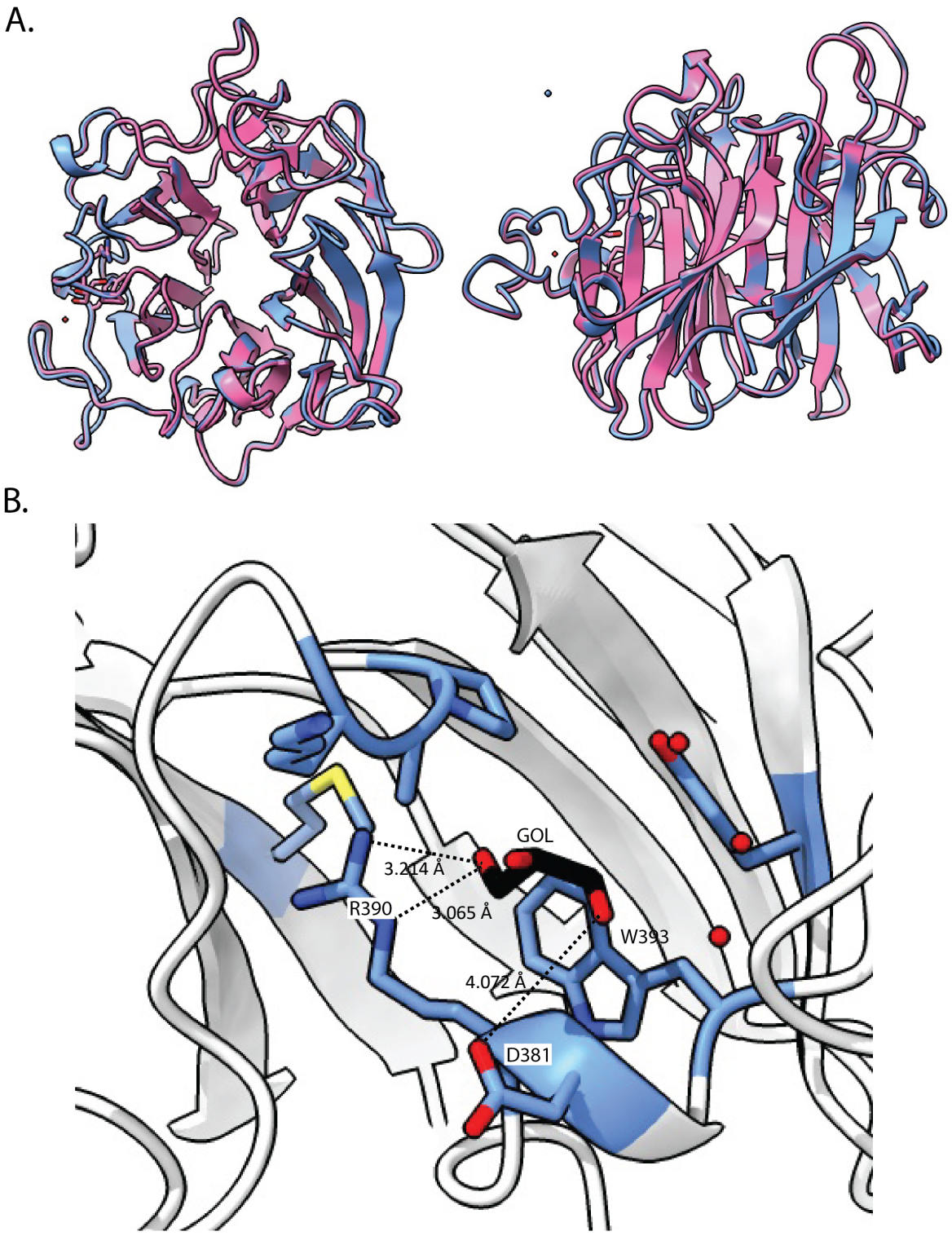


Figure 44. A. Superposition (RMSD = 0.254) of the two protomers of MAB_{GH183} found in the asymmetric unit of crystals obtained following molecular replacement phasing and manual solving. B. coordination site of glycerol (GOL) molecule in Chain B of the asymmetric by R390 and D381.

8.2.6 Identification of small ORFs upstream of MAB_{GH183} and MAP_{GH183} suggest formation of a complex

A hallmark of Type-7 toxins in Firmicutes is the presence of a small immunity protein usually encoded immediately upstream of the toxin. Initial analysis of the toxin locus failed to identify such a gene in the *M. absceussus* subsp. *massiliense* strain G0-06 genome. However, use of ORFinder identified a small ORF just downstream of the protein in the MAB genome, as shown in Figure 46 (Rombel *et al.*, 2002). The two loci differed in each species based on the arrangement of the Type 7 secretion machinery. In the case of MAP_{GH183}, the WxG100 motif was encoded within the same gene as the GH183. However, in MAB_{GH183}, we observed two small ORFs 5' of the GH183 which we hypothesised could be two carrier proteins which multimerise with the protein to cause the actively secreted ternary complex, which has been reported by the Palmer and Whitney laboratories (Whitney *et al.*, 2017; Cao *et al.*, 2016).

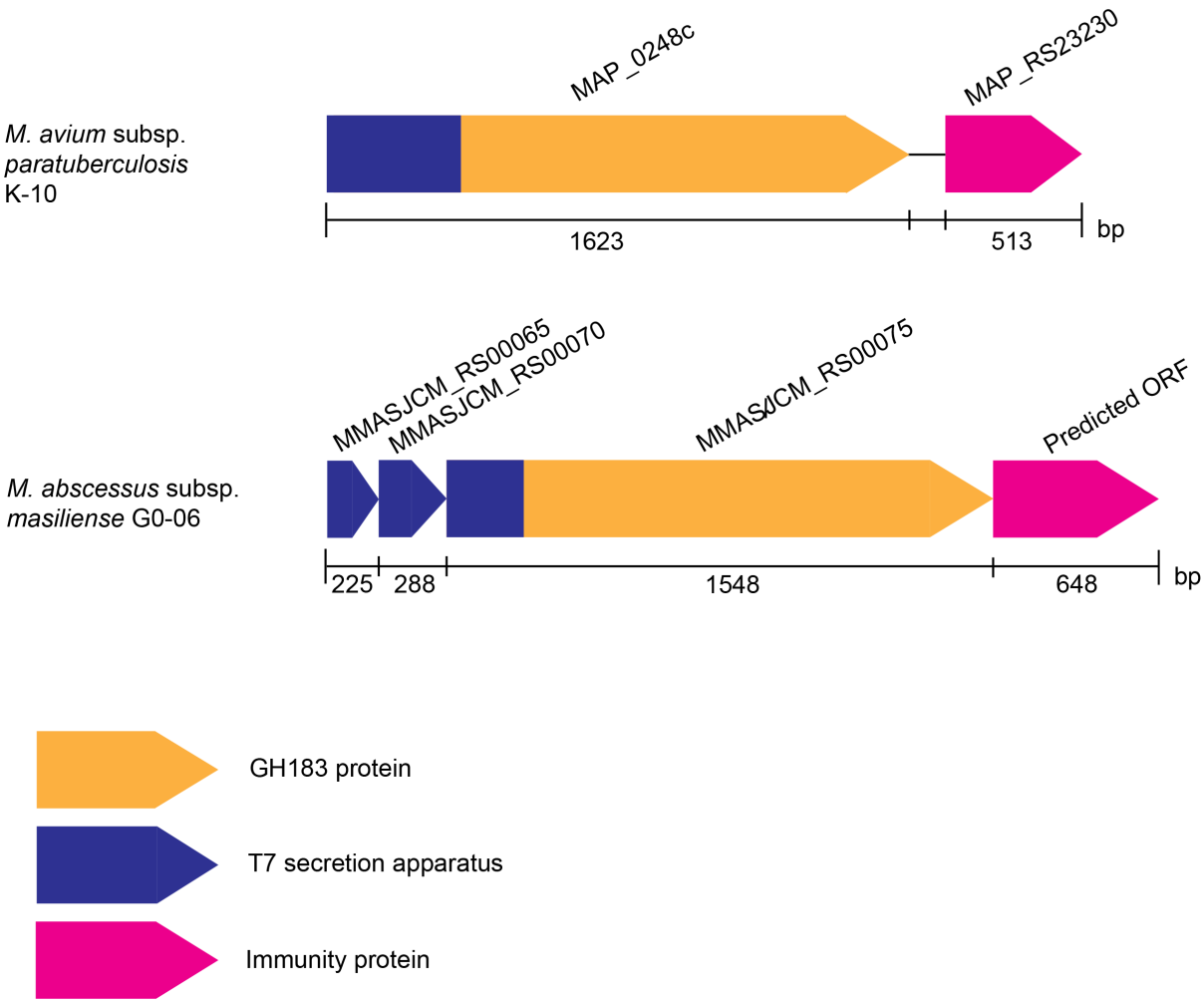


Figure 45. Schematic of the locus encoding MAB_{GH183} and MAP_{GH183} enzymes following analysis using ORFfinder.

8.2.7 Purification of $MAB_{Inhibitor}$ and $MAP_{inhibitor}$ proteins

Following identification of the putative immunity protein ORF downstream of the GH183 proteins, we expressed and purified the proteins following recombinant expression in *E. coli*. The proteins possessed expected molecular weights of 16.898 kDA ($MAP_{inhibitor}$) and 16.443 kDa ($MAB_{inhibitor}$). The proteins both expressed well and were approximately 90% ($MAP_{inhibitor}$), and 86% ($MAB_{inhibitor}$) pure as judged by SDS-PAGE (Fig 47). Migrating to a position between the 17 and 20 kDa markers there appeared to be higher molecular weight contaminating proteins. We hypothesised that this was a recalcitrant disulfide bond in the proteins not being fully reduced by the buffering conditions, thereby appearing to migrate at a position consistent with a higher molecular weight on the gel.

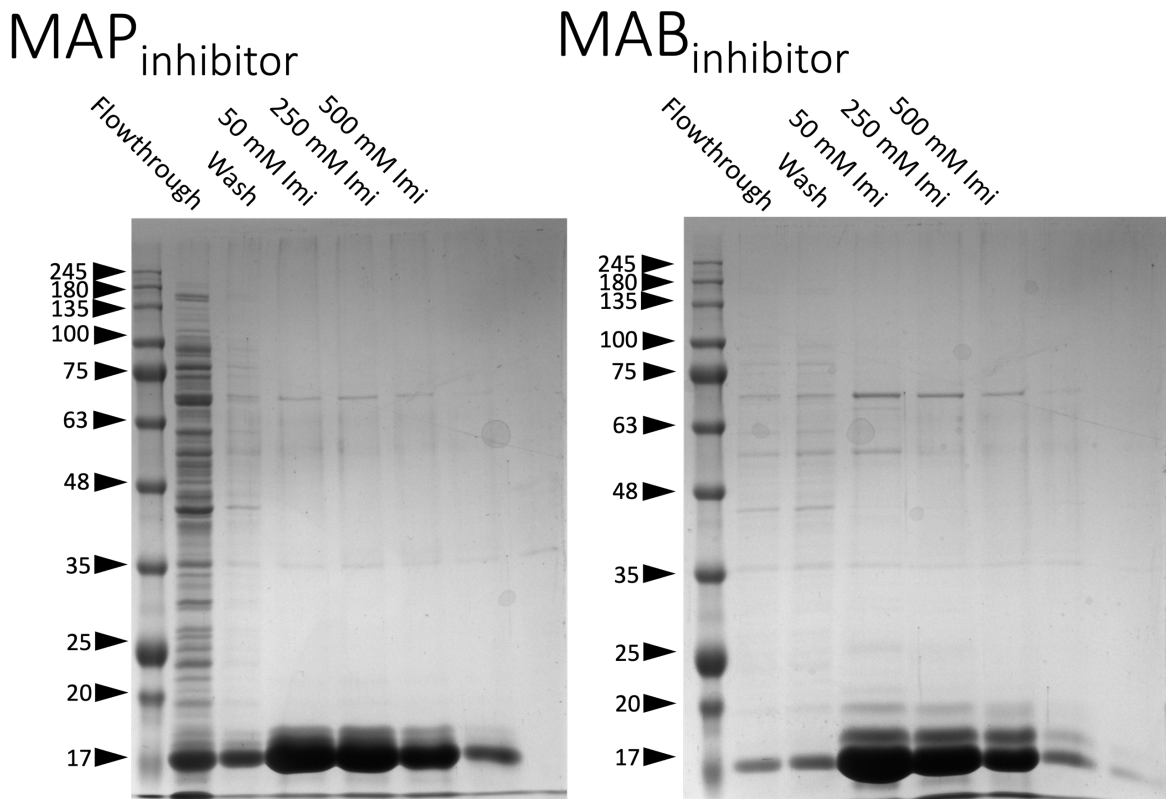


Figure 46. SDS-PAGE analysis of fractions resulting from purification of MAP_{inhib} and MAB_{inhib}. Proteins were purified from 1L pellets of induced T7 Shuffle cells containing appropriate plasmids. SDS PAGE analysis carried out on MINI-PROTEAN 4-20% precast gels.

8.2.8 MAB_{GH183} and MAP_{GH183} form complexes with their cognate inhibitors

With the putative inhibitor proteins in hand, we first wanted to test the hypothesis that the two proteins formed complexes with their cognate GH183 proteins. To assay this, we carried out analytical size exclusion chromatography on an AKTA Pure 25 system. Proteins were mixed in a 1:1 molar ratio and compared with single protein samples. Due to the similarity in size of the proteins, resolution is difficult but enough exists to see each peak in isolation. As shown in Figure 47, both GH183 proteins elute at approximately 12 mL, whilst the inhibitor proteins elute at approximately 14 mL. Upon mixing the two proteins with their cognate inhibitors (complex samples in green in Figure 48), a higher molecular weight species can be observed, eluting at approximately 11 mL. This supports the hypothesis that the proteins form a stable complex. The identity of the elution fractions was subsequently confirmed by SDS-PAGE analysis of the peak fractions (Figure 48).

Analysis of the fractions by SDS-PAGE supported the designation of the higher molecular weight peak as a complex between the GH183 domain and the inhibitor protein. In the complex sample we observed almost no inhibitor or GH183 proteins at elution volumes consistent with un-complexed proteins. This is consistent with a 1:1 stoichiometric ratio. Interestingly, when mixing non-cognate protein pairs (i.e. MAP_{GH183} with $MAB_{inhibitor}$), no complex is formed. This indicates that the proteins selectively form a complex with their syntenic neighbour. Protein complex formation can occur for multiple biological purposes, this can include both activation and inhibition. Given the hypothesis that the GH183 proteins are Type VII effectors, we presumed that the inhibitor proteins would act to prevent toxicity within the cell, and therefore we expected them to be inhibitory in arabinogalactan degradation assays.

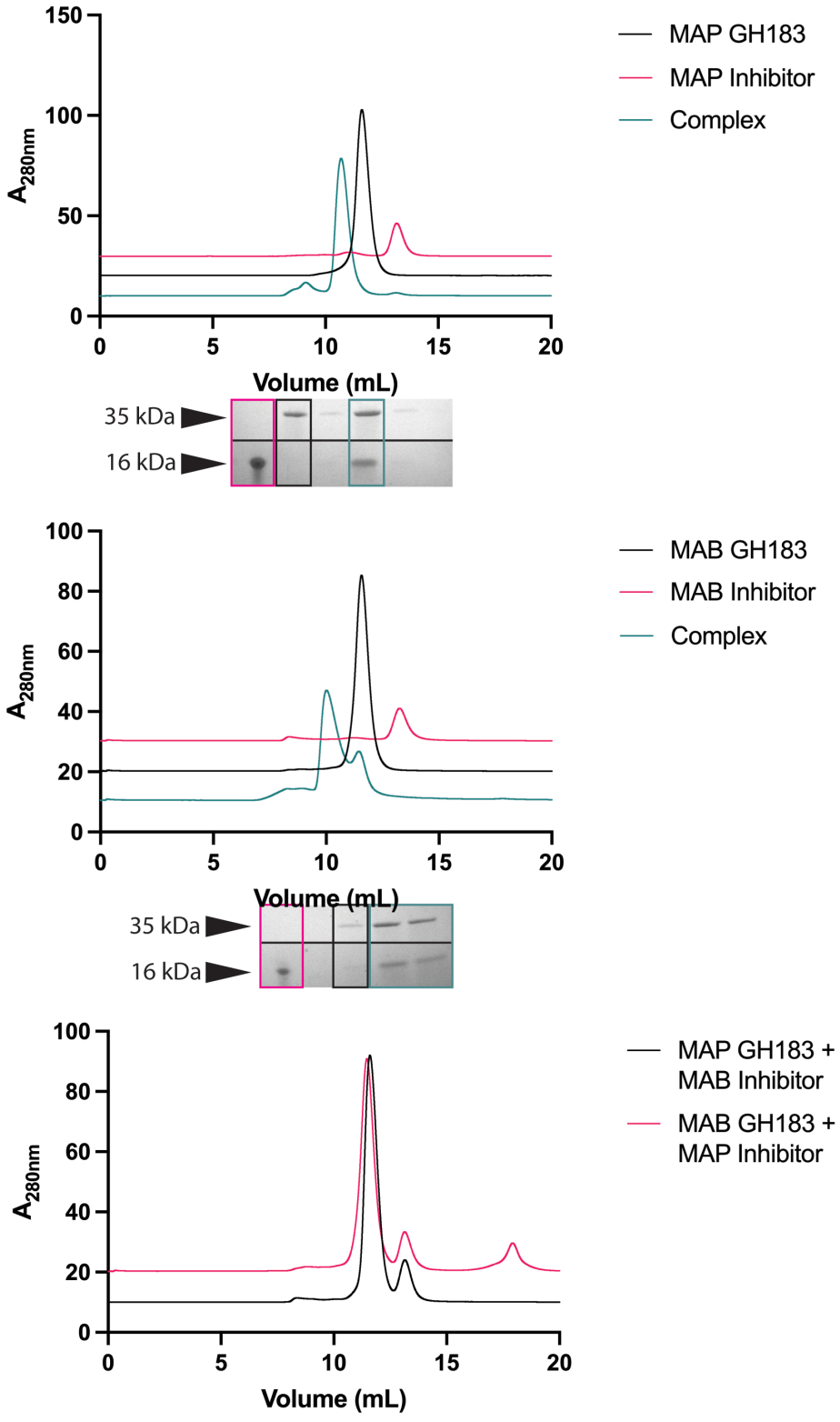


Figure 47. Analytical size exclusion chromatography analysis of MAB_{GH183}, MAP_{GH183}, and their inhibitor proteins. Earlier elution time of the mixed samples indicates an increase in molecular weight. Gels below chromatograms are fractions collected from the peaks of each run, coloured in the same way as the chromatograms. Pink = inhibitor protein only; Black = GH183 only; Green = complex of inhibitor and GH183 proteins. 100µL injections of protein samples were injected at a concentration of 75 µM were injected. SDS-PAGE lanes have been cropped from a larger gel for illustrative purposes.

8.2.9 Complexes formed between MAB_{GH183}, MAP_{GH183}, and upstream proteins are inhibitory

We next tested the hypothesis that these proteins are inhibitors of the GH183 using *in vitro* hydrolysis of AG and TLC analysis. The first panel of Figure 49a shows the TLC resulting from incubation of MAB_{GH183} with MAB_{inhib}. Unfortunately, MAB_{GH183} reached completion before the end of the assay, therefore meaning we cannot observe true dose dependence, but at a ratio of 1x and 2x (lanes 5 and 6), complete absence of D-arabinan hydrolysis is observed. Below this ratio, no visible change in hydrolytic activity can be observed. In panel 2 of Figure 49a, we can see that the MAP_{inhib} protein fails to prevent hydrolysis of the AG, even at a molar ratio of 2x. Figure 49b shows the same experiment but with MAP_{GH183}. In this case, dose dependence can be seen using MAP_{inhib}, with a gradual decrease in activity until the 1x molar ratio, where no activity can be observed. Once again, incubation with the non-cognate inhibitor, MAB_{inhib}, fails to produce any level of observable enzyme inhibition. Taken together, this indicates that the GH183 and inhibitor proteins can only form a complex with the cognate protein from the same species, and cross protection between the proteins tested does not exist. This is the first biochemical evidence of interspecies toxins observed in a Mycobacteriales member.

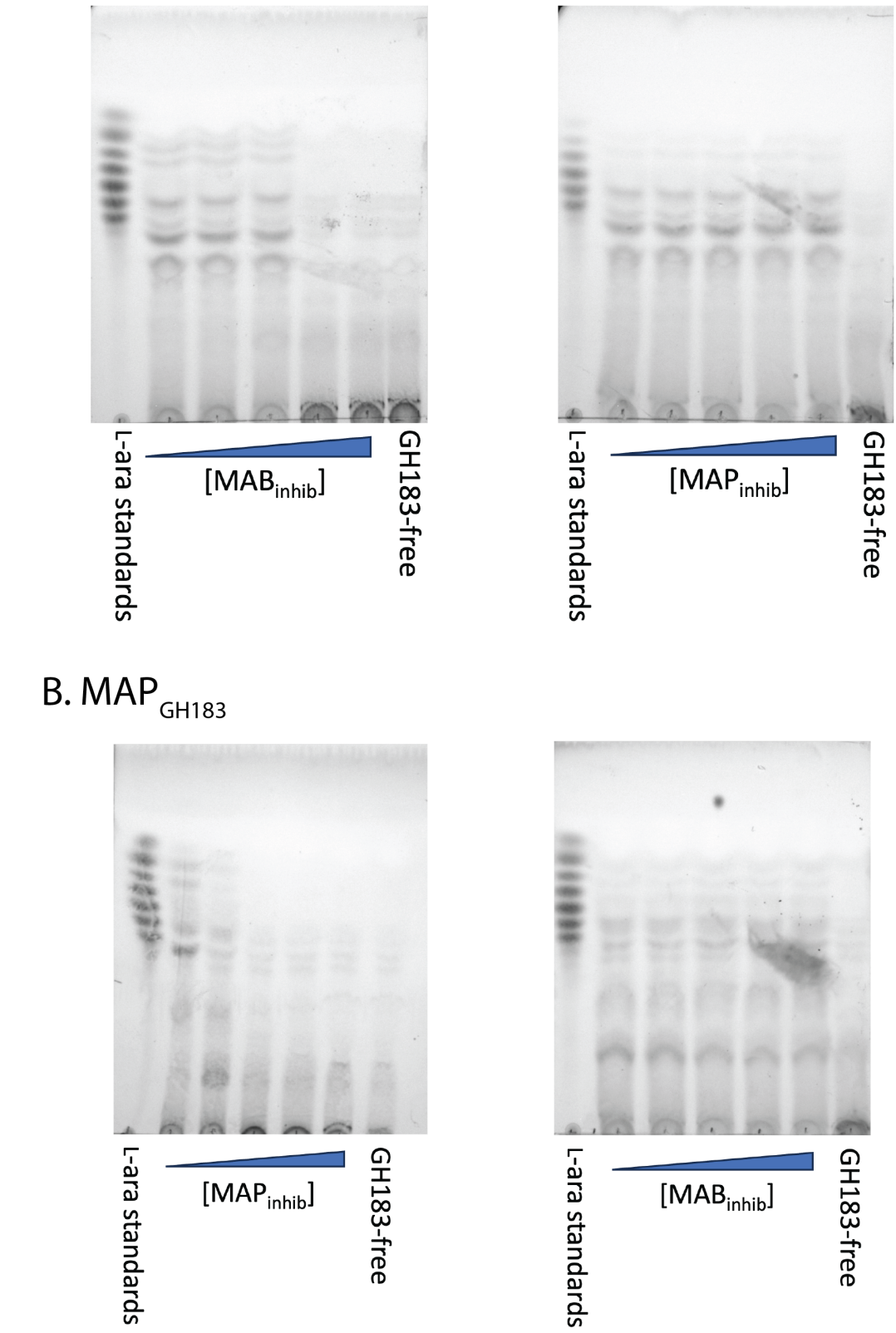
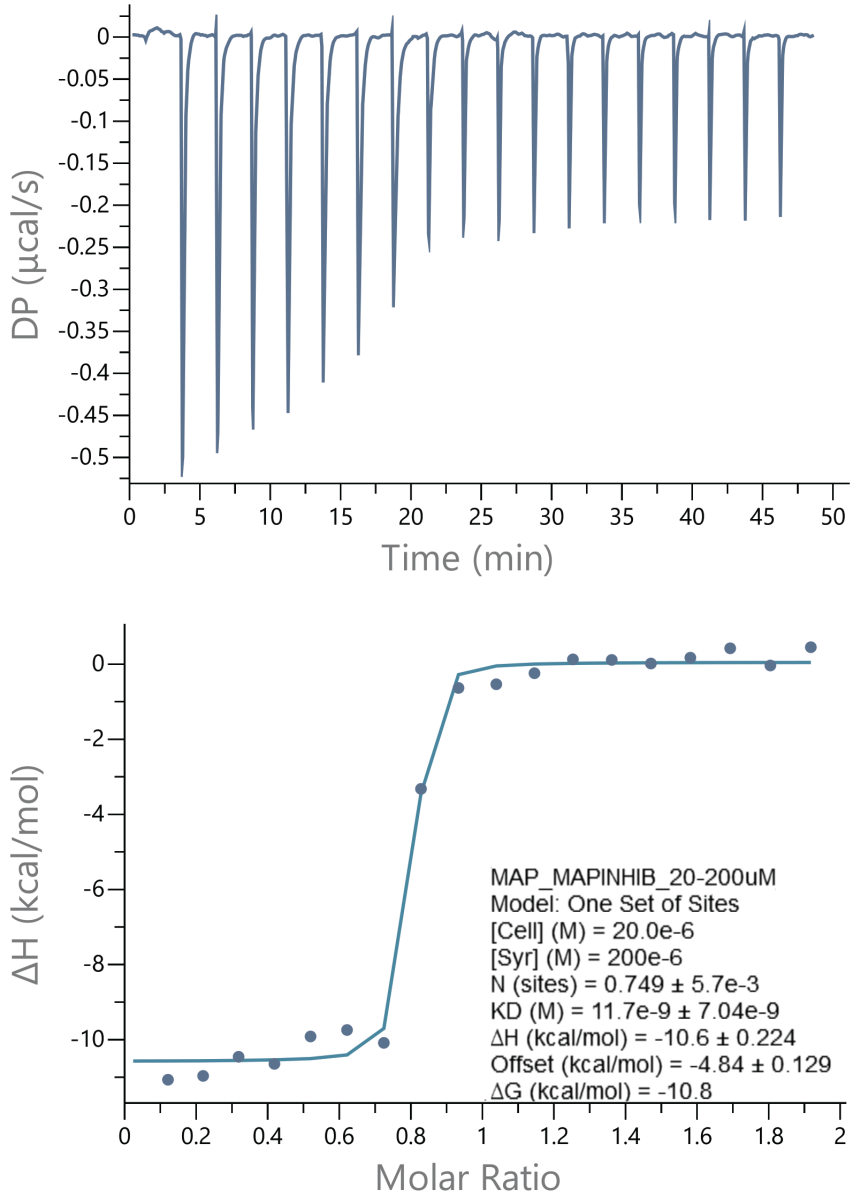


Figure 48. TLC analysis of *in vitro* hydrolysis reactions using GH183 enzymes (5 µM) with 1 mg/mL arabinogalactan from *M. smegmatis*. In lanes 3-6, inhibitor was included at a molar ratio of 0.25x, 0.5x, 1x, and 2x. the right most lane contains 10 µM of the inhibitor proteins without the GH183. TLCs were developed twice in 2:1:1 butan-1-ol: acetic acid: water and spots visualised using orcinol TLC stain

8.2.10 MAP_{GH183} forms an extremely high affinity complex with its inhibitor protein

Following the observation that complex formation leads to inhibition, we hypothesised that the interaction between inhibitor protein and its cognate GH183 enzyme would be high affinity, as poor association of the two would lead to toxicity in the cell. We therefore carried out isothermal titration calorimetry analysis of complex formation between MAP_{GH183} and $MAP_{inhibitor}$. The experiment was carried out in HEPES pH 8 with 200 mM NaCl to reduce buffering or ionic strength effects of the interaction. Following extensive technical and biological replicates, the affinity was shown to be in the low nanomolar range (K_D= 8-11 nM ± 7 nM), as shown in Figure 50. This makes the affinity comparable to that which is found between known bacteriocins and their proteinaceous inhibitors (Keeble *et al.*, 2006). Additionally, a change of Gibb's free energy (ΔG) of -10.8 kcal/mol indicates that the complex formation is highly thermodynamically favourable. Interestingly, despite multiple biological repeats, the stoichiometry of the interaction was consistently reported at 0.7-0.8 sites/protein. This may be due to experimental error with the true concentrations of the proteins varying due to pipetting error. Additionally, based on the size exclusion data in the previous section, the proteins seemed to complex in a 1:1 molar ratio. In comparison, no binding could be observed in experiments using MAP_{GH183} and $MAB_{inhibitor}$. This recapitulates the size exclusion data, and further reinforces our hypothesis that complex formation is the driver of protein inhibition. This highly specific, high affinity interaction comparable to known immunity mechanisms further supports our hypothesis that these enzymes likely play a role in interbacterial competition due to the fact the biochemical mechanisms of toxicity and detoxification remain well conserved across known systems (Granato *et al.*, 2019). Unfortunately, ITC analysis was unsuccessful with MAB_{GH183} as the complex was not stable in solution.



Experiment	Temperature (°C)	Bin	[Syr] (M)	[Cell] (M)	Control Type	N (sites)	KD (M)	ΔH (kcal/mol)	ΔG (kcal/mol)	-TΔS (kcal/mol)
MAP_MAPINHIB 1	25.1	Binding	2E-04	2E-05	Fitted Offset	0.833	8.25E-09	-10.8	-11.0	-0.241
MAP_MAPINHIB 2	25.1	Binding	2E-04	2E-05	Fitted Offset	0.718	3.55E-07	-16.4	-8.80	7.63
MAP_MAPINHIB 4	25.1	Binding	2E-04	2E-05	Fitted Offset	0.749	1.17E-08	-10.6	-10.8	-0.200
MAP_MAPINHIB 4	25.1	Binding	2E-04	2E-05	Fitted Offset	0.722	8.46E-09	-10.4	-11.0	-0.653

Figure 49. Isothermal titration calorimetry of the complex formation between MAP_{GH183} and $MAP_{inhibitor}$ at a temperature of 25 °C. Data analysed using a 1:1 binding model. Isotherm shown is representative of the repeats summarised in the table below.

8.2.11 Structure of MAP_{GH183} in complex with its inhibitor protein

As we were able to isolate and purify complexes of GH183 enzymes with their cognate inhibitor proteins, we attempted to crystallise both complexes using sitting drop vapour diffusion. Unfortunately, the $MAB_{complex}$ was unstable in solution if left for too long, therefore did not lend itself easily to crystallisation as it precipitated too rapidly. After many attempts, we obtained X-ray diffraction datasets for the $MAB_{complex,}$ but we have been unable to solve the phases for the data. However, the purified $MAP_{complex}$ was soluble and crystallised well. Unfortunately, the crystals diffracted sub-optimally, with ice rings present, significant anisotropy, and translational NCS meaning many reflections were not suitable for use in refinement, the statistics for refinement are summarised in Table 22. This is apparent in the very high R_{work} and R_{free} values, exceeding 10% difference from each other.

Post collection processing by DIALS starANISO brought completeness below 50%, making the datasets unusable for traditional analysis. We therefore used the unprocessed, anisotropic dataset for structural determination. The dataset solved with convincing metrics, and we reasoned that insight could still be drawn from the data we had, despite its poor quality given that the data were reasonably high resolution, at 2.49 Å and had CC1/2 values more than 90%. Additionally, the completeness was 80%. We therefore set out to tentatively draw insight from the sub-optimal solution but were cautious with interpretation.

Table 22. Merging and refinement statistics for MAP_{GH183} in complex with $MAP_{inhibitor}$

	***M. avium paratuberculosis* GH183 toxin in complex with inhibitor protein**
Wavelength	0.9578
Resolution range	60.7 - 2.49 (2.579 - 2.49)
Space group	P 21 21 21
Unit cell	67.751 104.923 121.403 90 90 90
Total reflections	55347 (4544)
Unique reflections	27677 (2272)
Multiplicity	2.0 (2.0)
Completeness (%)	88.98 (74.88)
Mean I/sigma(I)	4.51 (0.95)
Wilson B-factor	32.43
R-merge	0.1576 (0.633)
R-meas	0.2229 (0.8952)
R-pim	0.1576 (0.633)
CC1/2	0.983 (0.566)
CC*	0.996 (0.85)
Reflections used in refinement	27597 (2271)
Reflections used for R-free	1321 (109)
R-work	0.2888 (0.4109)
R-free	0.3656 (0.4512)

CC(work)	0.921 (0.601)
CC(free)	0.879 (0.333)
Number of non-hydrogen atoms	7081
macromolecules	6949
ligands	0
solvent	132
Protein residues	947
RMS(bonds)	0.022
RMS(angles)	1.53
Ramachandran favored (%)	91.04
Ramachandran allowed (%)	8.54
Ramachandran outliers (%)	0.43
Rotamer outliers (%)	5.24
Clashscore	15.96
Average B-factor	47.99
macromolecules	48.36
solvent	28.17

Statistics for the highest-resolution shell are shown in parentheses.

Figure 51a shows the asymmetric unit of $MAP_{complex}$ following solving and structural refinement. The protein complex crystallised with four proteins in the asymmetric unit, two copies of MAP_{GH183} and $MAP_{inhibitor}$. Figure 51b shows superposition of the two protein complexes, which have an RMSD of 0.231 Å. No clear differences between the two proteins

are obvious based on manual examination. The crystallisation of two complexes in the ASU further strengthens our hypothesis that the data are reasonably reliable as there are no significant differences between each protomer, indicating the solution is feasible.

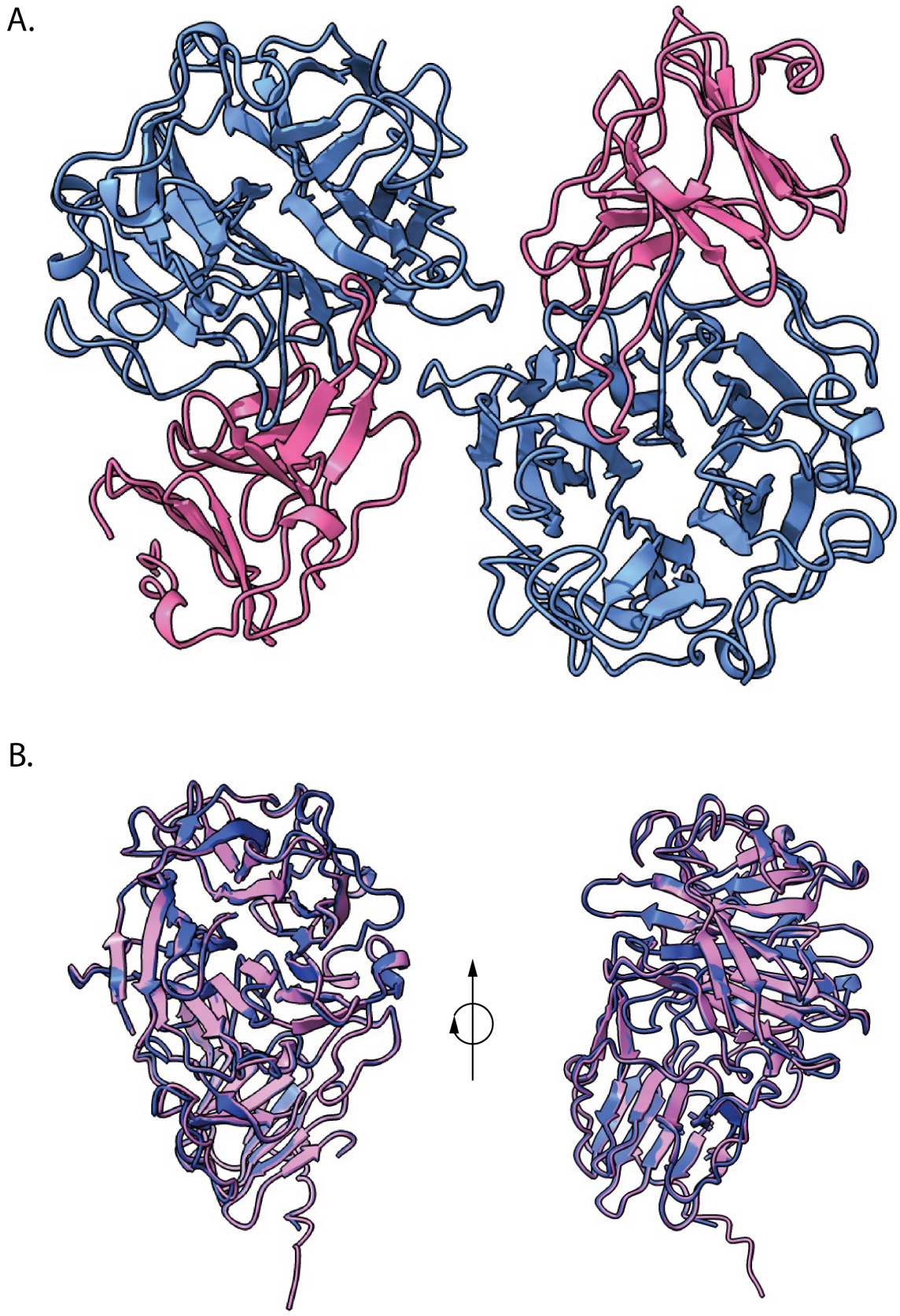


Figure 50. Asymmetric unit of crystal structure of MAP_{GH183} in complex with $MAP_{inhibitor}$ (termed $MAP_{complex}$). A. asymmetric unit consisting of two copies of the complex, coloured by protein. MAP_{GH183} is in blue, $MAP_{inhibitor}$ is in pink. B. superposition of the two copies in the asymmetric unit superposed using MatchMaker *(Meng* et al.*, 2023)*

8.2.12 Structure of $MAP_{inhibitor}$

As highlighted in Figure 52, the antitoxin protein exists as a 10-sheet β-sandwich protein made up of two β-meanders. Based on literature review, this is the first experimentally determined structure of this fold. The protein also possesses a β-hairpin motif upstream of the sandwich domain, extending 21 Å away from the protein mass.

A search of the PDB using Foldseek finds no proteins of less than 3e-3 with all hits obtained in the analysis corresponding to the helix-turn-helix motif, leading to structural similarities with DNA damage proteins such as UvrB (van Kempen *et al.*, 2023). This is because a β-hairpin domains are widespread amongst DNA-binding proteins (Aravind *et al.*, 2005). Higher confidence hits were obtained when searching against the Alphafold database, where significant similarity exists to the lipoprotein LppI from *N. brasiliensis* and *Mycobacterium tuberculosis*. The *M. tuberculosis* homolog, Rv2046, has been implicated in a link between mycolic acid homeostasis and export in a high throughput screen (Touchette *et al.*, 2017). The conservation of the fold indicates a possible evolutionary relationship from LppI to $MAP_{inhibitor}$ through gene duplication and evolution.

Table 23. Hit table resulting from a Foldseek search using the $MAP_{inhibitor}$ as a search model against the alphafold protein database, top 10 hits shown (van Kempen et al., 2023).

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query (aa)
AF-K0FD73-F1-model_v4	Lipoprotein LppI	Nocardia brasiliensis ATCC 700358	1	21.7	1.36E-05	12-147
AF-K0F2M3-F1-model_v4	Uncharacterized protein	Nocardia brasiliensis ATCC 700358	1	18.6	5.90E-05	1-147
AF-O53489-F1-model_v4	Probable lipoprotein LppI	Mycobacterium tuberculosis H37Rv	1	14.2	6.58E-05	1-147
AF-X8F7Y7-F1-model_v4	Uncharacterized protein	Mycobacterium ulcerans str. Harvey	1	20.8	7.74E-05	30-150
AF-K0F484-F1-model_v4	Uncharacterized protein	Nocardia brasiliensis ATCC 700358	1	20.8	2.29E-04	3-147
AF-K0ES25-F1-model_v4	Uncharacterized protein	Nocardia brasiliensis ATCC 700358	1	20.7	4.16E-04	26-146
AF-K0EV32-F1-model_v4	Uncharacterized protein	Nocardia brasiliensis ATCC 700358	1	19.8	8.88E-04	29-146
AF-K0F2A7-F1-model_v4	Uncharacterized protein	Nocardia brasiliensis ATCC 700358	1	15.8	2.00E-03	1-138
AF-K0EZB7-F1-model_v4	Uncharacterized protein	Nocardia brasiliensis ATCC 700358	0.18	16.6	8.23E-01	56-147
AF-E9PVG8-F1-model_v4	RIKEN cDNA 9530053A07 gene	Mus musculus	0.12	7.7	8.69E-01	34-141

Table 24. Hit table resulting from a Foldseek search using the $MAP_{inhibitor}$ as a search model against the PDB protein database, top 10 hits shown(van Kempen et al., 2023).

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query
4dfc_C	Core UvrA/TRCF complex	Escherichia coli K-12	0.06	12	9.57E+00	106-150
2gcu_C	X-Ray Structure of Gene Product from Arabidopsis Thaliana At1g53580	Arabidopsis thaliana	0.04	11.3	8.59E+00	98-134
3v4r_A	Crystal structure of a UvrB dimer-DNA complex	Bacillus subtilis subsp. subtilis str. 16	0.1	5.7	7.70E+00	108-138
4d4p_G	Crystal Structure of the Kti11 Kti13 heterodimer Spacegroup P65	Saccharomyces cerevisiae S288C	0.03	15.2	6.91E+00	34-150
4uzr_A	Crystal Structure of Pyrococcus horikoshii Ph1500	Pyrococcus horikoshii OT3	0.03	10.6	5.87E+00	92-147
6i1d_A	Structure of the Ysh1-Mpe1 nuclease complex from S.cerevisiae	Saccharomyces cerevisiae S288C	0.11	11.1	4.73E+00	103-132
3hih_B	Structure of human Plk1-PBD with glycerol and sulfate in the phophopeptide binding site	Homo sapiens	0.07	14.4	3.42E+00	94-148
7tdz_i	Cryo-EM model of protomer of the cytoplasmic ring of the nuclear pore complex from Xenopus laevis	Xenopus laevis	0.06	10.8	3.42E+00	33-146
6o8e_A	Crystal structure of UvrB bound to duplex DNA with ADP	[Bacillus] caldotenax	0.09	5	3.23E+00	106-149
7egt_A	The crystal structure of the C-terminal domain of T. thermophilus UvrD complexed with the N-terminal domain of UvrB	Thermus thermophilus HB8	0.08	8.3	3.23E+00	106-149

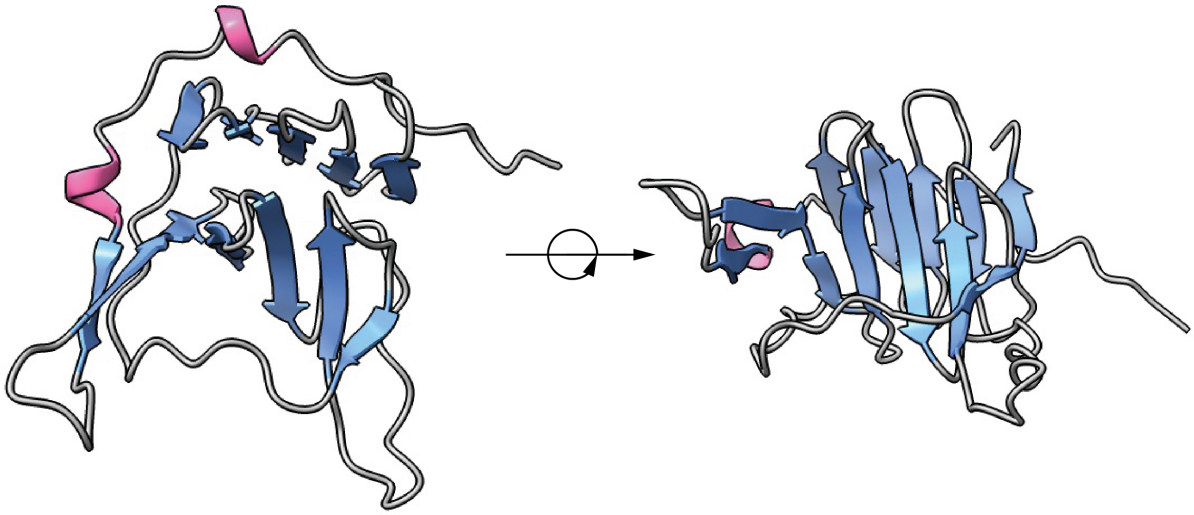


Figure 51. Top (left panel) and side view (right panel) of the $MAP_{inhibitor}$ (chains C and D from the ASU) crystal structure coloured by secondary structure. Blue = β-sheet, pink = α-helix, grey = coil. As can be seen, the protein forms two domains. The first is a 10-sheet beta sandwich motif consisting of a beta meander. Second domain is the helix-turn-helix domain.

8.2.13 $MAP_{complex}$ is a very closely association of the two proteins

Figure 53a shows the complex viewed form the top. In the left-hand panel, we can see the MAP_{GH183} in purple, and then the $MAP_{inhibitor}$ in beige. The inhibitor clasps over one blade of the propeller. In Figure 53b, the β-hairpin motif in red extends all the way into the active site, resembling a hand pointing a finger. In the right-hand panel of 53b, we can see the semi-transparent inhibitor protein overlaid with the hydrophobicity surface of the GH183. The surface is uniformly polar, and the interaction face does not appear to be notably non-polar. This indicates that the interaction is likely not due to hydrophobic interactions.

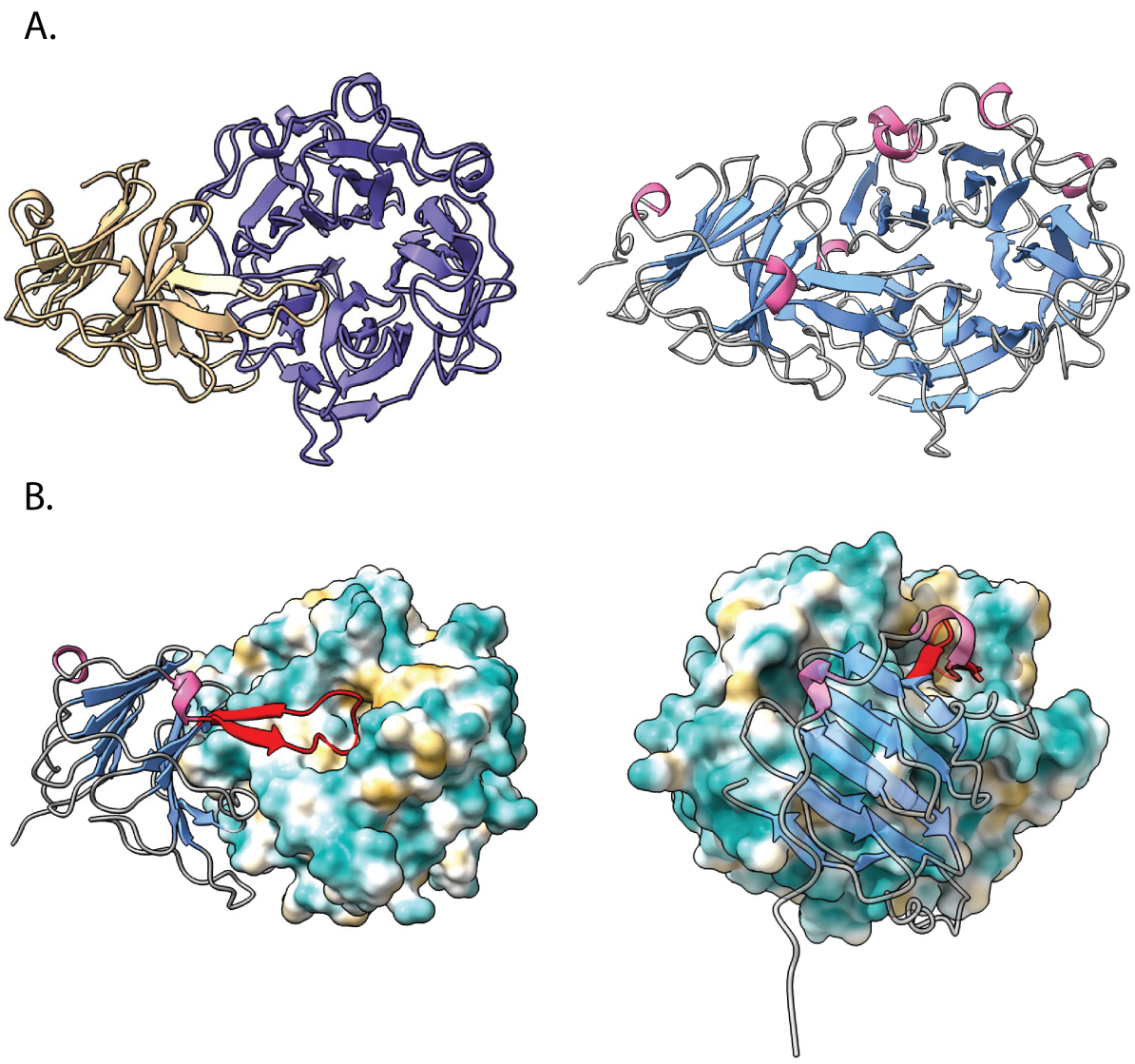


Figure 52. Structural overview of crystal structure of $MAP_{complex}$. A. complex coloured by chain with GH183 in purple and inhibitor in beige. B. both proteins coloured by secondary structure. Sheets coloured iin blue, alpha helix in pink and coil in grey. B. Hydrophobicity surface of MAP_{GH183} (green= hydrophilic; yellow = hydrophobic), with helix turn helix poking into the active site highlighted in red.

8.2.14 Mechanism of inhibition

Figure 54 shows the active site of MAP_{GH183} in complex with the inhibitor protein. MAP_{GH183} is shown in purple, whilst $MAP_{inhibitor}$ is shown in beige. As can be seen in Figure 53, the two active site residues from the GH183, D177 and E226, are posed pointed away from each other. This is likely due to the insertion of R32 from the turn of the inhibitor's helix-turn-helix domain. The crystal formed in a buffer at pH 7, at the pH the arginine bears a positive charge. This leads to H-bonding between E226, whilst R178 engages in a repulsive interaction with R32 of the inhibitor protein. This pair of interactions therefore will remodel the loop carrying D177 and cause this catalytically inactive pose to be adopted. This highly evolved interaction therefore potently inhibits GH183 activity upon complex formation. This mechanism of action is unlike any known inhibition mechanism of a glycoside hydrolase. Most known chemical inhibitors, including those of other glycoside hydrolases, such as acarbose, bind their target as a transition state mimic (Balaich *et al.*, 2021). This non-covalent, but potent and highly selective inhibition mechanism has evolved to protect cells from endogenous GH183 toxins. To fully confirm this hypothesis, it would have been helpful to obtain the apo structure of MAP_{GH183} and $MAP_{inhibitor}$ to observe conformational responses during binding. Unfortunately, despite multiple crystallisation attempts, this was not possible during the project.

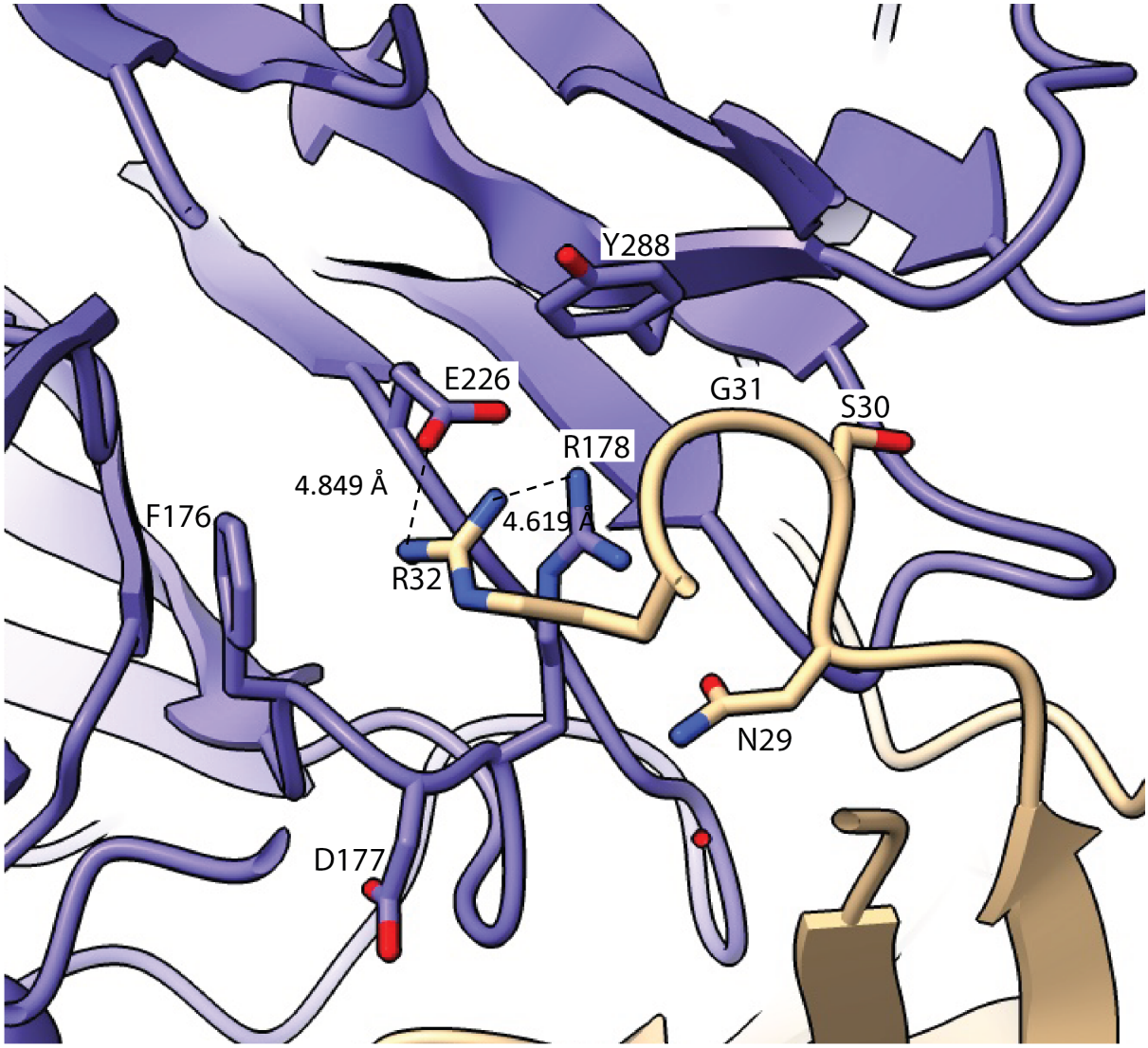


Figure 53. Active site of MAP_{GH183} (purple) in complex with $MAP_{inhibitor}$ (beige), with residues believed to occlude substrate and contribute to complex formation shown coloured by heteroatom. The left hand sheet of the hairpin in beige has been interrupted by a clipping plane to allow better view of the active site, there is no break in peptide in the loop.

8.2.15 Cross species inhibition is prevented by subtle conformational changes in the GH183 protein

Given we now understood with reasonable confidence how protein complex formation leads to enzyme inhibition, we set out to use our data to gain insight into the mechanism by which cross-species protection is prevented. Given the reasonable sequence conservation between MAP_{GH183} and MAB_{GH183}, we reasoned the structures would adopt a very similar conformation but would perhaps possess minor differences. We therefore superposed MAB_{GH183} with $MAP_{complex}$, hid the MAP_{GH183}, then displayed the hydrophobicity surface of MAB_{GH183} to produce Figure 55. What can be seen is that the MAB_{GH183} has evolved two lobes, highlighted in red, which induce steric clashing between the $MAP_{inhibitor}$ and MAB proteins and therefore prevent complex formation. These two lobes are amino acids 438-441, and the inclusion of a bulky tryptophan sidechain at position 462 in the MAB_{GH183}. Of course, this cannot be confirmed until apo and complexes for all the proteins have been solved, as this may be an artefact of crystallisation, or might not include conformational change upon binding.

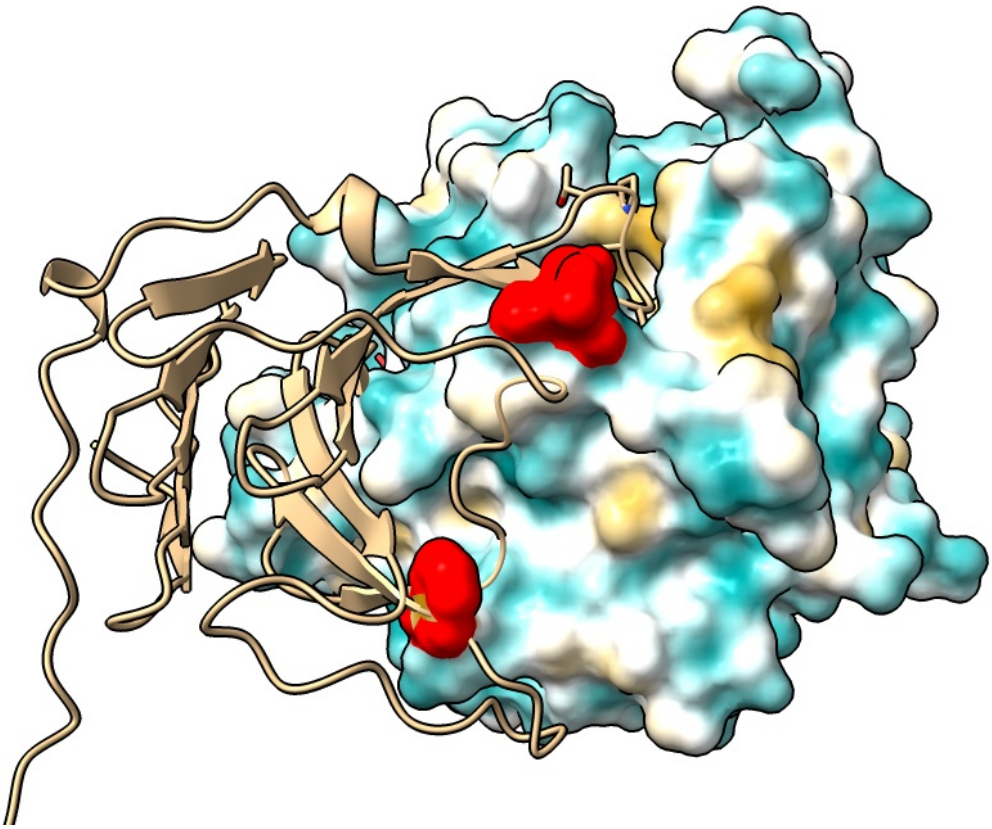


Figure 54. Hydrophobicity surface of MAB_{GH183} superposed with MAP_{GH183} in complex with its cognate inhibitor) shown in beige. Highlighted in red are the surface modifications preventing complex formation between non-cognate enzyme/inhibitor pairs through creation of steric clashing.

8.2.16 Exogenous expression of MAB_{GH183} in *M. smegmatis* leads to toxicity

Our biochemical analysis indicated that the two proteins formed a toxin-immunity complex. We therefore wanted to test that MAB_{GH183} is toxic when overexpressed in a mycobacterial species. For this purpose, we used the mycobacterial L-arabinose-inducible plasmid pMYBADc-KAN (Beckham *et al.*, 2020). We had two versions of the MAB_{GH183} construct produced by GeneWiz. One with the GH183 fused to a Tat signal peptide, and the other encoded the full genomic locus from *M. abscessus* subsp. *massiliense*, including the GH183 with its native N-terminal sequence and the two small ORFS upstream it, shown in Figure 56. Without L-arabinose, all transformants displayed similar fitness. In the presence of L-arabinose, the empty vector strains' survival was unchanged but there was a two-log reduction in viability of the two strains transformed with the plasmid containing a GH183 insert. In unpresented data, we found that growth of *M. smegmatis* is not inhibited by presence of up to 500 µM recombinant exogenous MAB_{GH183} included in the growth media. This therefore indicates that the protein must reach a specific compartment, likely the periplasm, which is inaccessible to proteins from the extracellular milieu to cause toxicity. We therefore show that the likely biological function of this GH183 homologue in *M. abscessus* subsp. *massiliense* is a mediator of interference competition between mycobacterial species.

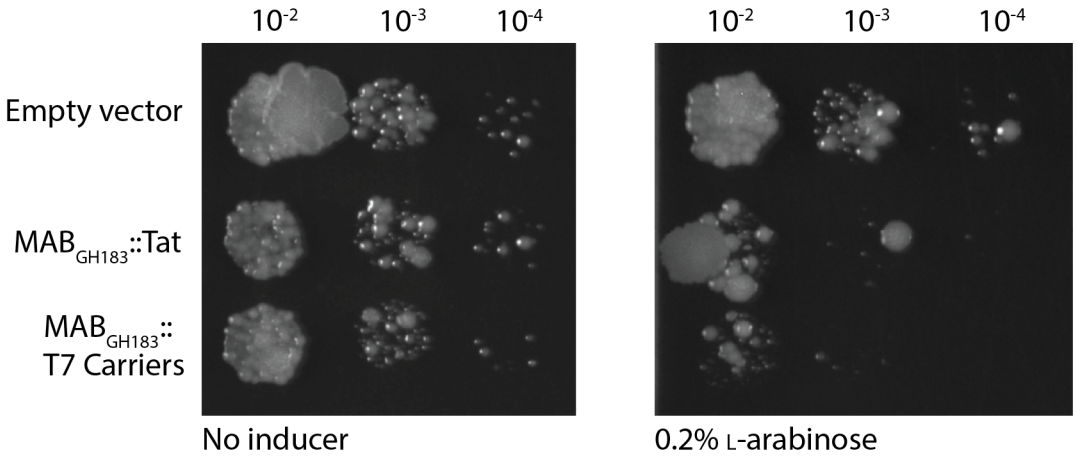


Figure 55. Overexpression of MAB_{GH183} results in a 2-fold-log decrease in viability. Spot plates of *M. smegmatis* transformed with kanamycin-selectable, L-arabinose-inducible plasmid pMyBADc-Kan with the following inserts: MAB_{GH183}::Tat contains the GH183 domain of MAB_{GH183} fused to a Tat signal peptide. MABGH183:: T7 carriers contains the full genomic locus of the putative GH183 toxin locus from *M. abscessus* subsp. *massiliense*, including the two carrier proteins encoded upstream of the MAB_{GH183}-encoding gene. Dilutions were spotted on LB agar containing 50 μg/mL kanamycin with or without 0.2% L-arabinose.

8.3 Discussion

The data presented in this chapter strongly implicates a sub-family of GH183 enzymes in inter-mycobacterial competition. Firstly, we serendipitously identified two distant homologs of Rv3707c in *Mycobacterium abscessus* subsp. *massiliense* and *Mycobacterium avium* subsp. *paratuberculosis* through use of a protein solubility prediction tool. Following biochemical analysis, we have experimentally verified that GH183 enzymes are retaining. Whilst not directly relevant to these homologues, Rv3707c and Rv1754c have been implicated in pathogenesis of *M. tuberculosis* and retaining glycoside hydrolases lend themselves to rational design of inhibitors (Perkowski *et al.*, 2017). Covalent glycoside hydrolase inhibitors such as acarbose for the treatment of diabetes mellitus and Tamiflu for influenza have been used safely and effectively used clinically for many years, indicating the efficacy of such treatments (Balaich *et al.*, 2021; Oo *et al.*, 2001; Babu *et al.*, 2000). Following further study into the efficacy of inhibition of GH183 enzymes in *M. tuberculosis*, such an avenue of enquiry would be potentially very helpful in treating tuberculosis and other mycobacterioses.

Upon the public release of Alphafold during the project, we noticed both proteins possessed N-terminal extensions which resembled Type 7 secretion signals from *Staphylococcus aureus*. This led us to hypothesise that the enzymes carried out an extracellular function, unrelated to Rv3707c and other housekeeping homologues. Given the availability of literature on the functions of Type 7 secretion systems in *S. aureus*, we hypothesised that this function may be interbacterial competition. This was supported by the presence of inhibitor proteins adjacent to the GH183 proteins, which through analytical size exclusion chromatography and ITC we have shown form a high affinity complex. Based on work from the Kleanthous laboratory, most colicins and their cognate inhibitors form complexes with an affinities in the range of

low nanomolar to picomolar, which is comparable to our K_d of 11 nM . Interestingly, colicins seem to still complex with non-(Wallis *et al.*, 1995; Housden and Kleanthous, 2011)cognate inhibitors from other species, but with an affinity roughly 10-log-fold lower, but our putative toxins cannot bind a non-cognate inhibitor protein (Blasco *et al.*, 2023; Keeble *et al.*, 2006).

Amongst glycoside hydrolases, proteinaceous inhibitors are rare, with two classes of glycoside hydrolase experimentally shown to possess proteinaceous non-covalent inhibitors. We have successfully crystallised both an apo form of MAB_{GH183} and a complex of MAP_{GH183} with $MAP_{inhibitor}$. The first example of proteinaceous glycoside hydrolase inhibitors we found in the literature were xylanase inhibitors produced by wheat and other monocots to prevent excessive breakdown of xylose-containing polysaccharides by commensal bacteria such *Agrobacterium nidulans* (Fierens *et al.*, 2007). Whilst the overall structure of these complexes are very different, a commonality exists in that both inhibitor proteins make use of a positively charged lysine or arginine residue inserted into the enzyme active site (Payan *et al.*, 2004). Ivy is another example of a proteinaceous GH inhibitor and encoded in the genome of *E. coli* among other bacteria (Abergel *et al.*, 2007). This protein acts on eukaryotic (Class C) lysozymes to cause their inhibition. Ivy possesses significant difference structurally to $MAP_{inhibitor}$, but once again makes use of positively charged residue, histidine, which interacts with the catalytic carboxylate side chains to inhibit the protein (Abergel *et al.*, 2007). Furthermore, it has been shown that *M. tuberculosis* encodes a lysozyme inhibitor, LprI, within its genome, which forms a complex with C type lysozymes with a 2 nM affinity (Sethi *et al.*, 2016). Whilst such a mechanism within itself is fascinating, it also may provide insight into novel mechanisms of glycoside hydrolase inhibition more broadly. Furthermore, the presence of toxin-immunity pairs in *M. tuberculosis* has been known for many years, following

the discovery of tuberculosis necrotizing toxin, which is an NAD glycohydrolase (Sun *et al.*, 2015). The protein was initially identified solely as a host-targeting factor (U *et al.*, 2019). However, given NAD is conserved in all living cells, we believe that the toxin may have initially evolved as an interbacterial competition effector which found utility in pathogenesis (Sasiak and Saunders, 1996; Spaans *et al.*, 2015; Sun *et al.*, 2015)

Finally, using a plasmid-based spotting assay, we have shown that overexpression of these enzymes in *Mycobacterium smegmatis* is toxic. Further experimentation is underway to demonstrate that this toxicity is relevant to interbacterial competition. Additionally, further work is underway to generate a catalytically inactive version of the GH183 enzyme for overexpression in *M. smegmatis* to show the toxicity is catalytic and not due to an artefactual effect such as overexpression causing amino acid starvation for other cellular processes leading to a drop in viability (Danchik *et al.*, 2021). We are also generating a system with a related plasmid from the pMY series which contains the cognate immunity protein under the control of an acetamide promoter to show that it can rescue viability (Beckham *et al.*, 2020).

Taken together, these data are persuasive in supporting our hypothesis that these enzymes cause toxicity in mycobacteria, likely through the hydrolysis of arabinogalactan. The linkages found in the substrate the toxins target are phylogenetically restricted to Mycobacteriales species and select other microbial species such as *P. aeruginosa* PA7 (Kus *et al.*, 2008). Furthermore, the presence of a putative type 7 secretion signal in both proteins indicates that these proteins can be translocated to competing organisms. This is the first reported potential mechanism of interspecies competition found in Mycobacteria, and the biological implications are unclear.

The question remains as to what purpose these proteins have *in vivo.* Unlike Rv3707c and Rv1754c, these proteins do not appear in high throughput pathogenicity screening of mycobacteria, indicating that whilst they might not contribute directly to pathogenesis, they may provide a selective advantage in soil and human microbiomes. One possible explanation is that most Tn-seq type experiments are not designed to address questions of competition with the microbiome. These proteins may allow the bacteria to specifically compete with related species for nutrients. Recent work from the Palmer laboratory indicates that loss of a type VII toxin in a monoculture infection model in zebrafish decreases bacterial burden (Ulhuq *et al.*, 2020). Drawing from existing knowledge of the function of these systems in *P. aeruginosa* Type 6 secretion system-mediated competition, these systems likely enable more efficient competition for available resources within sites of initial colonisation by mycobacterial pathogens, such as the lungs in the case of *M. abscessus* subsp. *massiliense*, and in the gut rumen for *M. avium* subsp. *paratuberculosis*. Both of these compartments contain commensal species of Corynebacteria, another genus which possesses microbial arabinogalactan, such as *Corynebacterium vitaeruminis*, which in many cases prevents colonisation by harmful bacteria (Colombo *et al.*, 2017; Kelly *et al.*, 2022). However, these Type 7 substrates would allow efficient killing of these Corynebacteria, enabling more efficient colonisation and pathogenesis. Evidence to support this hypothesis is the complete absence of Corynebacterial species in the gut rumen of cows recovering from Johne's disease (the disease resulting from *M. avium* subsp. *paratuberculosis*) up to 9 months following MAP infection (Matthews *et al.*, 2023). This loss of commensal corynebacteria following mycobacterial disease certainly bears further investigation, as it may indicate competition, or some other unexplained phenomenon.

Future work will include *in vivo* models of competition, to show that expression of these genes gives selective advantage against those lacking immunity mechanisms. Additionally, to verify our proposed mechanism for enzyme inhibition, we need a full set of apo and complex proteins, as well as site directed mutants within the interaction face, and in the warhead arginine to verify that molecular mechanisms presented are biologically relevant.

9 Discussion

Microbial arabinogalactans are rare in nature, only found in some members of the order Mycobacteriales (Daffe *et al.*, 1993; Jankute *et al.*, 2014). The polysaccharide plays a critical function covalently attaching the inner peptidoglycan sacculus to the mycolic acid outer membrane (Armitige *et al.*, 2000; Goins *et al.*, 2018). Up until this study, whilst significant evidence existed in the literature for release of arabinan oligosaccharides from growing mycobacteria, arabinogalactan was largely viewed as a static entity, whose covalent modifications were well characterised, but its turnover was under-studied (Dong *et al.*, 2006; Bhamidi *et al.*, 2008c; Angala *et al.*, 2017; Peng *et al.*, 2012).

In the first chapter of this thesis, we report the recombinant expression, purification, and characterisation of Rv3707c and MSMEG_2107. These two proteins, from *M. tuberculosis* and *M. smegmatis*, respectively, are conserved in all mycobacterial species. Following rigorous biochemical characterisation, including optimisation of porous graphitic carbon solid phase extraction (PGC-SPE) for the analysis of resulting *in vitro* hydrolysis assays, we demonstrate that mycobacteria encode D-arabinan hydrolases in their genomes. This activity had been known for close to 50 years, but the enzyme that catalysed this activity had not been determined (Dong *et al.*, 2006). The enzyme family enclosing these enzymes and homologues has since formed the basis for a new glycoside hydrolase family, GH183. Furthermore, we show X-ray crystallography that these enzymes adopt a 5-bladed β-propeller conformation, which is typical in carbohydrate-active enzymes such as glycoside hydrolases, lectins and glycosyltransferases (Meng and Fütterer, 2008; Bonnardel *et al.*, 2019; Afanasieva *et al.*, 2019). Through multiple sequence alignment and site directed, we successfully produced

catalytically null versions of the protein. Following exhaustive literature review of Rv3707c and MSMEG_2107's *M. tuberculosis* homolog, Rv1754c, we show that these proteins are critical in pathogenesis. Further work will be to identify the biological function of these enzymes, and how that utility benefits them in pathogenesis. We propose mechanisms for this. The first being that direct interaction between the products of these arabinan hydrolase directly interact with the human immune system to prevent pathogen clearance. The second is that the enzymes remodel the cell wall to better withstand the environment found within macrophage. Finally, the enzymes may remodel the cell wall to allow for the extrusion of type VII secretion systems which deliver virulence factors such as ESAT-6/CFP-10 to the host and trigger damage (Gao *et al.*, 2004; Renshaw *et al.*, 2005).

The second package of work carried out during this project followed from the identification of GH172 in the same PUL enabling growth of *D. gadei* in our paper published on arabinan hydrolases. We show that these GH172 enzymes were distributed in a different manner to GH183 enzymes, being particularly enriched in the genomes of *Nocardia* species. The family Nocardiaceae are a group of understudied but severe pathogens of the human central nervous system, responsible for Nocardiosis which can have a mortality rate of up to 80% (Colaneri *et al.*, 2021). These bacteria also can only be treated with Trimethoprim-sulfamethoxazole, a treatment regimen with an incidence of resistance at approximately 21% (Brown-Elliott *et al.*, 2012). Within this chapter we began with the identification of a conserved GH172 enzyme, which we extensively biochemically characterised using Michaelis-Menten kinetics, structural biology, and thin layer chromatographic analysis we identify that the $NBras_{GH172}$ enzyme adopts a significantly conformation to its homolog in *B. dentium*, with the loss of a short N-terminal α-helix abolishing dimerization of the two hexamers and causing

$NBras_{GH172}$ to remain as a trimer. Literature review indicated that the evolutionary origin of the protein family containing GH172 is an ancient pre-viral replicating organism, which then gradually became evolved to possess glycoside hydrolase activity (Krupovic *et al.*, 2022).

Syntenic analysis of the loci encoding the GH172 enzyme across *Nocardia* species showed that it was often encoded adjacent to genes associated with cycling of carbohydrates through the pentose phosphate pathway such as isomerases and FGGY kinases. Following successful recombinant expression of these enzymes, we believe that these enzymes function with the encoded glycoside hydrolases to passage endogenous polymeric arabinan back into the pentose phosphate pathway in times of carbon starvation to allow survival. During this enquiry, we also discover a new glycoside hydrolase family specific for β-1,2-linked D-arabinan, such as that which is found in the terminal caps of arabinogalactan.

Whilst the proteomic analysis we carried out was unsuccessful due to late harvest causing us to solely see mediators of starvation response significantly upregulated, we present the first known proteomic dataset of a *Nocardia* species, as well as an optimised proteomics preparation protocol which allows for recovery of many proteins encoded within the genome of *N. brasiliensis*. Given the recent interest in *Nocardia* species as secondary metabolite producers, we believe this dataset and protocol will be very helpful tools for the research community.

The final line of enquiry within this thesis is the characterisation of a less-well conserved group of GH183 enzymes found in environmental opportunistic mycobacterial pathogens such as *M. abscessus* subsp. *massiliense* and *M. avium* subsp. *paratuberculosis*. Their initial

identification was serendipitous due to their purported favourable characteristics predicted by FDetect (Meng *et al.*, 2018). Following the release of Alphafold, it's high-confidence structural prediction led us to hypothesise that the homologues identified, MAP_{GH183} and MAB_{GH183} were Type VII secreted, due to the presence of N-terminal α-helical bundles resembling WxG100 motifs found in Type VII substrates encoded in the genomes of *S. aureus* (Poulsen *et al.*, 2014). We then biochemically dissected the action of these homologs and then identified small proteins adjacent to the GH183 enzymes which we hypothesised to be inhibitor proteins, given Type VII substrates often require within-cell inhibition to prevent self-poisoning. We show that these proteins complex *in vitro*, and that the complex formed is inhibitory to the GH183 activity. This proteinaceous inhibition of a glycoside hydrolase is very rare in nature. However, amongst enzymatically distinct, but functionally related type VII effectors is very common to prevent toxicity within the cell and to related organisms (Ulhuq *et al.*, 2020). Finally, we show that cross-species inhibition of the GH183 enzymes is likely due to subtle structural variations which prevent enzyme-inhibitor complex formation. Due to the polytopic nature of the interaction formed, with multiple residues contributing to complex formation, there is a snapshot of a highly evolved relationship between GH183 protein and inhibitor. Finally, we show that overexpression of the identified toxin protein in trans in *M. smegmatis* causes significant toxicity *in vivo*. These data support the first proposal of interspecies competition between members of the Mycobacteriales.

To summarise, we show that D-arabinan hydrolysis, catalysed by GH172, GH183, and new enzyme families represents a new facet of cell wall homeostasis in Mycobacteriales organisms, disrupting the overarching belief that arabinogalactan in these organisms is a static entity, and that acid-fast organisms remodel their cell wall extensively. Within this

project we have gained an insight into mechanisms and significance of enzymes that cleave D-arabinan in maintaining the cell wall of these organisms, as well as mediating competition between species. Furthermore, we present tools enabling high-resolution analysis of arabinan polysaccharides, as well as tools that allow fine analysis of cell wall components of Mycobacteria and related organisms. As explained, we also believe that this project has wider implications for the understanding of the pathogenesis and homeostasis of the mycobacterial cell wall, and look forward to further work on this topic.

10 References

Abby, S.S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., Rocha, E.P.C. (2016) Identification of protein secretion systems in bacterial genomes. *Scientific Reports*. **6**(1), 23080.

Abergel, C., Monchois, V., Byrne, D., Chenivesse, S., Lembo, F., Lazzaroni, J.-C., Claverie, J.-M. (2007) Structure and evolution of the Ivy protein family, unexpected lysozyme inhibitors in Gram-negative bacteria. *Proceedings of the National Academy of Sciences of the United States of America*. **104**(15), 6394–6399.

Afanasieva, E., Chaudhuri, I., Martin, J., Hertle, E., Ursinus, A., Alva, V., Hartmann, M.D., Lupas, A.N. (2019) Structural diversity of oligomeric β-propellers with different numbers of identical blades. *eLife*. **8**, e49853.

Agarwal, R., Burley, S.K., Swaminathan, S. (2012) Structural Insight into Mechanism and Diverse Substrate Selection Strategy of L-Ribulokinase. *Proteins*. **80**(1), 261–268.

Ahamed, M., Jaisinghani, N., Li, M., Winkeler, I., Silva, S., Previti, M.L., Seeliger, J.C. (2022) Optimized APEX2 peroxidase-mediated proximity labeling in fast- and slow-growing mycobacteria. *Methods in Enzymology*. **664**, 267–289.

Alcoforado Diniz, J., Coulthurst, S.J. (2015) Intraspecies Competition in Serratia marcescens Is Mediated by Type VI-Secreted Rhs Effectors and a Conserved Effector-Associated Accessory Protein. *Journal of Bacteriology*. **197**(14), 2350–2360.

Alderwick, L.J., Harrison, J., Lloyd, G.S., Birch, H.L. (2015) The Mycobacterial Cell Wall--Peptidoglycan and Arabinogalactan. *Cold Spring Harbor Perspectives in Medicine*. **5**(8), a021113.

Alderwick, L.J., Lloyd, G.S., Ghadbane, H., May, J.W., Bhatt, A., Eggeling, L., Fütterer, K., Besra, G.S. (2011) The C-terminal domain of the Arabinosyltransferase Mycobacterium tuberculosis EmbC is a lectin-like carbohydrate binding module. *PLoS pathogens*. **7**(2), e1001299.

Alderwick, L.J., Lynn G. Dover, Seidel, M., Gande, R., Sahm, H., Eggeling, L., Besra, G.S. (2006) Arabinan-deficient mutants of Corynebacterium glutamicum and the consequent flux in decaprenylmonophosphoryl-D-arabinose metabolism. *Glycobiology*. **16**(11), 1073–1081.

Alderwick, L.J., Seidel, M., Sahm, H., Besra, G.S., Eggeling, L. (2006) Identification of a novel arabinofuranosyltransferase (AftA) involved in cell wall arabinan biosynthesis in Mycobacterium tuberculosis. *The Journal of Biological Chemistry*. **281**(23), 15653–15661.

Al-Jourani, O., Benedict, S.T., Ross, J., Layton, A.J., van der Peet, P., Marando, V.M., Bailey, N.P., Heunis, T., Manion, J., Mensitieri, F., Franklin, A., Abellon-Ruiz, J., Oram, S.L., Parsons, L., Cartmell, A., Wright, G.S.A., Baslé, A., Trost, M., Henrissat, B., Munoz-Munoz, J., Hirt, R.P., Kiessling, L.L., Lovering, A.L., Williams, S.J., Lowe, E.C., Moynihan, P.J. (2023) Identification of D-arabinan-degrading enzymes in mycobacteria. *Nature Communications*. **14**(1), 2233.

Angala, S. kumar, Palčeková, Z., Belardinelli, J.M., Jackson, M. (2018) Covalent modifications of polysaccharides in mycobacteria. *Nature chemical biology*. **14**(3), 193–198.

Angala, S.K., McNeil, M.R., Shi, L., Joe, M., Pham, H., Zuberogoitia, S., Nigou, J., Boot, C.M., Lowary, T.L., Gilleron, M., Jackson, M. (2017) Biosynthesis of the Methylthioxylose Capping Motif of Lipoarabinomannan in Mycobacterium tuberculosis. *ACS chemical biology*. **12**(3), 682–691.

Arahal, D.R., Busse, H.-J., Bull, C.T., Christensen, H., Chuvochina, M., Dedysh, S.N., Fournier, P.-E., Konstantinidis, K.T., Parker, C.T., Rossello-Mora, R., Ventosa, A., Göker, M. (2022) Judicial Opinions 112–122. *International Journal of Systematic and Evolutionary Microbiology*. **72**(8).

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS microbiology reviews*. **29**(2), 231–262.

Arbing, M.A., Kaufmann, M., Phan, T., Chan, S., Cascio, D., Eisenberg, D. (2010) The crystal structure of the mycobacterium tuberculosis Rv3019c-Rv3020c ESX complex reveals a domain-swapped heterotetramer. *Protein Science*. **19**(9), 1692–1703.

Ardèvol, A., Rovira, C. (2015) Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *Journal of the American Chemical Society*. **137**(24), 7528–7547.

Armitige, L.Y., Jagannath, C., Wanger, A.R., Norris, S.J. (2000) Disruption of the genes encoding antigen 85A and antigen 85B of Mycobacterium tuberculosis H37Rv: effect on growth in culture and in macrophages. *Infection and Immunity*. **68**(2), 767–778.

Aronson, J.D. (1926) Spontaneous Tuberculosis in Salt Water Fish. *The Journal of Infectious Diseases*. **39**(4), 315–320.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Doré, J., MetaHIT Consortium, Antolín, M., Artiguenave, F., Blottiere, H.M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K.U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Mérieux, A., Melo Minardi, R., M'rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S.D., Bork, P. (2011) Enterotypes of the human gut microbiome. *Nature*. **473**(7346), 174–180.

Ates, L.S., Ummels, R., Commandeur, S., Weerd, R. van der, Sparrius, M., Weerdenburg, E., Alber, M., Kalscheuer, R., Piersma, S.R., Abdallah, A.M., Ghany, M.A.E., Abdel-Haleem, A.M.,

Pain, A., Jiménez, C.R., Bitter, W., Houben, E.N.G. (2015) Essential Role of the ESX-5 Secretion System in Outer Membrane Permeability of Pathogenic Mycobacteria. *PLOS Genetics*. **11**(5), e1005190.

Attia, M.A., Nelson, C.E., Offen, W.A., Jain, N., Davies, G.J., Gardner, J.G., Brumer, H. (2018) In vitro and in vivo characterization of three Cellvibrio japonicus Glycoside Hydrolase Family 5 members reveals potent xyloglucan backbone-cleaving functions. *Biotechnology for biofuels*. **11**(1).

Aubry, A., Chosidow, O., Caumes, E., Robert, J., Cambau, E. (2002) Sixty-three Cases of Mycobacterium marinum Infection: Clinical Features, Treatment, and Antibiotic Susceptibility of Causative Isolates. *Archives of Internal Medicine*. **162**(15), 1746–1752.

Babu, Y.S., Chand, P., Bantia, S., Kotian, P., Dehghani, A., El-Kattan, Y., Lin, T.H., Hutchison, T.L., Elliott, A.J., Parker, C.D., Ananth, S.L., Horn, L.L., Laver, G.W., Montgomery, J.A. (2000) BCX-1812 (RWJ-270201): discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *Journal of Medicinal Chemistry*. **43**(19), 3482–3486.

Bacon, J., Alderwick, L.J., Allnutt, J.A., Gabasova, E., Watson, R., Hatch, K.A., Clark, S.O., Jeeves, R.E., Marriott, A., Rayner, E., Tolley, H., Pearson, G., Hall, G., Besra, G.S., Wernisch, L., Williams, A., Marsh, P.D. (2014) Non-Replicating Mycobacterium tuberculosis Elicits a Reduced Infectivity Profile with Corresponding Modifications to the Cell Wall and Extracellular Matrix. *PLOS ONE*. **9**(2), e87329.

Bakuła, Z., Napiórkowska, A., Bielecki, J., Augustynowicz-Kopeć, E., Zwolska, Z., Jagielski, T. (2013) Mutations in the embB Gene and Their Association with Ethambutol Resistance in Multidrug-Resistant Mycobacterium tuberculosis Clinical Isolates from Poland. *BioMed Research International*. **2013**.

Balaich, J., Estrella, M., Wu, G., Jeffrey, P.D., Biswas, A., Zhao, L., Korennykh, A., Donia, M.S. (2021) The human microbiome encodes resistance to the antidiabetic drug acarbose. *Nature*. **600**(7887), 110–115.

Bansal-Mutalik, R., Nikaido, H. (2014) Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(13), 4958–4963.

Bansal-Mutalik, R., Nikaido, H. (2011) Quantitative lipid composition of cell envelopes of Corynebacterium glutamicum elucidated through reverse micelle extraction. *Proceedings of the National Academy of Sciences of the United States of America*. **108**(37), 15360–15365.

Baram, D., Pyetan, E., Sittner, A., Auerbach-Nevo, T., Bashan, A., Yonath, A. (2005) Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proceedings of the National Academy of Sciences*. **102**(34), 12017–12022.

Baranowski, C., Welsh, M.A., Sham, L.-T., Eskandarian, H.A., Lim, H.C., Kieser, K.J., Wagner, J.C., McKinney, J.D., Fantner, G.E., Ioerger, T.R., Walker, S., Bernhardt, T.G., Rubin, E.J., Rego, E.H. (2018) Maturing Mycobacterium smegmatis peptidoglycan requires non-canonical crosslinks to maintain shape W. S. Garrett, B. D. Kana, & M. Arthur, eds. *eLife*. **7**, e37516.

Barth, V.C., Chauhan, U., Zeng, J., Su, X., Zheng, H., Husson, R.N., Woychik, N.A. (2021) Mycobacterium tuberculosis VapC4 toxin engages small ORFs to initiate an integrated oxidative and copper stress response. *Proceedings of the National Academy of Sciences of the United States of America*. **118**(32), e2022136118.

Basavannacharya, C., Robertson, G., Munshi, T., Keep, N.H., Bhakta, S. (2010) ATP-dependent MurE ligase in Mycobacterium tuberculosis: Biochemical and structural characterisation. *Tuberculosis*. **90**(1), 16–24.

Baughn, A.D., Garforth, S.J., Vilchèze, C., Jacobs, W.R. (2009) An anaerobic-type alpha-ketoglutarate ferredoxin oxidoreductase completes the oxidative tricarboxylic acid cycle of Mycobacterium tuberculosis. *PLoS pathogens*. **5**(11), e1000662.

Beaman, B.L., Beaman, L. (1994) Nocardia species: host-parasite relationships. *Clinical Microbiology Reviews*. **7**(2), 213–264.

Beckham, K.S.H., Ciccarelli, L., Bunduc, C.M., Mertens, H.D.T., Ummels, R., Lugmayr, W., Mayr, J., Rettel, M., Savitski, M.M., Svergun, D.I., Bitter, W., Wilmanns, M., Marlovits, T.C., Parret, A.H.A., Houben, E.N.G. (2017) Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. *Nature Microbiology*. **2**(6), 1–7.

Beckham, K.S.H., Staack, S., Wilmanns, M., Parret, A.H.A. (2020) The pMy vector series: A versatile cloning platform for the recombinant production of mycobacterial proteins in Mycobacterium smegmatis. *Protein Science: A Publication of the Protein Society*. **29**(12), 2528–2537.

Behrens, A.-J., Vasiljevic, S., Pritchard, L.K., Harvey, D.J., Andev, R.S., Krumm, S.A., Struwe, W.B., Cupo, A., Kumar, A., Zitzmann, N., Seabright, G.E., Kramer, H.B., Spencer, D.I.R., Royle, L., Lee, J.H., Klasse, P.J., Burton, D.R., Wilson, I.A., Ward, A.B., Sanders, R.W., Moore, J.P., Doores, K.J., Crispin, M. (2016) Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. *Cell Reports*. **14**(11), 2695–2706.

Beláňová, M., Dianišková, P., Brennan, P.J., Completo, G.C., Rose, N.L., Lowary, T.L., Mikušová, K. (2008) Galactosyl Transferases in Mycobacterial Cell Wall Synthesis. *Journal of Bacteriology*. **190**(3), 1141–1145.

Belisle, J.T., Vissa, V.D., Sievert, T., Takayama, K., Brennan, P.J., Besra, G.S. (1997) Role of the Major Antigen of Mycobacterium tuberculosis in Cell Wall Biogenesis. *Science*.

Ben Ami Pilo, H., Khan Khilji, S., Lühle, J., Biskup, K., Levy Gal, B., Rosenhek Goldian, I., Alfandari, D., Revach, O.-Y., Kiper, E., Morandi, M.I., Rotkopf, R., Porat, Z., Blanchard, V., Seeberger, P.H., Regev-Rudzki, N., Moscovitz, O. (2022) Sialylated N-glycans mediate

monocyte uptake of extracellular vesicles secreted from Plasmodium falciparum-infected red blood cells. *Journal of Extracellular Biology*. **1**(2), e33.

Bhamidi, S., Scherman, M.S., Rithner, C.D., Prenni, J.E., Chatterjee, D., Khoo, K.-H., McNeil, M.R. (2008a) The Identification and Location of Succinyl Residues and the Characterization of the Interior Arabinan Region Allow for a Model of the Complete Primary Structure of Mycobacterium tuberculosis Mycolyl Arabinogalactan. *Journal of Biological Chemistry*. **283**(19), 12992–13000.

Bhamidi, S., Scherman, M.S., Rithner, C.D., Prenni, J.E., Chatterjee, D., Khoo, K.-H., McNeil, M.R. (2008b) The Identification and Location of Succinyl Residues and the Characterization of the Interior Arabinan Region Allow for a Model of the Complete Primary Structure of Mycobacterium tuberculosis Mycolyl Arabinogalactan. *Journal of Biological Chemistry*. **283**(19), 12992–13000.

Bhamidi, S., Scherman, M.S., Rithner, C.D., Prenni, J.E., Chatterjee, D., Khoo, K.-H., McNeil, M.R. (2008c) The identification and location of succinyl residues and the characterization of the interior arabinan region allow for a model of the complete primary structure of Mycobacterium tuberculosis mycolyl arabinogalactan. *The Journal of Biological Chemistry*. **283**(19), 12992–13000.

Bhatt, A., Molle, V., Besra, G.S., Jacobs, W.R., Kremer, L. (2007) The Mycobacterium tuberculosis FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Molecular Microbiology*. **64**(6), 1442–1454.

Bi, E.F., Lutkenhaus, J. (1991) FtsZ ring structure associated with division in Escherichia coli. *Nature*. **354**(6349), 161–164.

Birch, H.L., Alderwick, L.J., Bhatt, A., Rittmann, D., Krumbach, K., Singh, A., Bai, Y., Lowary, T.L., Eggeling, L., Besra, G.S. (2008) Biosynthesis of mycobacterial arabinogalactan: identification of a novel alpha(1-->3) arabinofuranosyltransferase. *Molecular Microbiology*. **69**(5), 1191–1206.

Bisson-Filho, A.W., Hsu, Y.-P., Squyres, G.R., Kuru, E., Wu, F., Jukes, C., Sun, Y., Dekker, C., Holden, S., VanNieuwenhze, M.S., Brun, Y.V., Garner, E.C. (2017) Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science*. **355**(6326), 739–743.

Bjursell, M.K., Martens, E.C., Gordon, J.I. (2006) Functional Genomic and Metabolic Studies of the Adaptations of a Prominent Adult Human Gut Symbiont, Bacteroides thetaiotaomicron, to the Suckling Period *. *Journal of Biological Chemistry*. **281**(47), 36269–36279.

Blasco, L., de Aledo, M.G., Ortiz-Cartagena, C., Blériot, I., Pacios, O., López, M., Fernández-García, L., Barrio-Pujante, A., Hernández-Garcia, M., Cantón, R., Tomás, M. (2023) Study of 32 new phage tail-like bacteriocins (pyocins) from a clinical collection of Pseudomonas aeruginosa and of their potential use as typing markers and antimicrobial agents. *Scientific Reports*. **13**(1), 117.

Bondage, D.D., Lin, J.-S., Ma, L.-S., Kuo, C.-H., Lai, E.-M. (2016) VgrG C terminus confers the type VI effector transport specificity and is required for binding with PAAR and adaptor–effector complex. *Proceedings of the National Academy of Sciences*. **113**(27), E3931–E3940.

Boneca, I.G., Huang, Z.H., Gage, D.A., Tomasz, A. (2000) Characterization of Staphylococcus aureus cell wall glycan strands, evidence for a new beta-N-acetylglucosaminidase activity. *The Journal of Biological Chemistry*. **275**(14), 9910–9918.

Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H., Mogk, A. (2009) Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *The EMBO Journal*. **28**(4), 315–325.

Bonnardel, F., Kumar, A., Wimmerova, M., Lahmann, M., Perez, S., Varrot, A., Lisacek, F., Imberty, A. (2019) Architecture and Evolution of Blade Assembly in β-propeller Lectins. *Structure*. **27**(5), 764-775.e3.

Bowman, L., Palmer, T. (2021) The Type VII Secretion System of Staphylococcus. *Annual Review of Microbiology*. **75**, 471–494.

Brecik, M., Centárová, I., Mukherjee, R., Kolly, G.S., Huszár, S., Bobovská, A., Kilacsková, E., Mokošová, V., Svetlíková, Z., Šarkan, M., Neres, J., Korduláková, J., Cole, S.T., Mikušová, K. (2015) DprE1 Is a Vulnerable Tuberculosis Drug Target Due to Its Cell Wall Localization. *ACS Chemical Biology*. **10**(7), 1631–1636.

Breda, A., Martinelli, L.K.B., Bizarro, C.V., Rosado, L.A., Borges, C.B., Santos, D.S., Basso, L.A. (2012) Wild-type phosphoribosylpyrophosphate synthase (PRS) from Mycobacterium tuberculosis: a bacterial class II PRS? *PloS One*. **7**(6), e39245.

Brewer, T.E., Wagner, A. (2022) Translation stalling proline motifs are enriched in slow-growing, thermophilic, and multicellular bacteria. *The ISME Journal*. **16**(4), 1065–1073.

Brieger, E.M., Glauert, A.M. (1956) Sporelike structures in the tubercle bacillus. *Nature*. **178**(4532), 544.

Briliūtė, J., Urbanowicz, P.A., Luis, A.S., Baslé, A., Paterson, N., Rebello, O., Hendel, J., Ndeh, D.A., Lowe, E.C., Martens, E.C., Spencer, D.I.R., Bolam, D.N., Crouch, L.I. (2019) Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature Microbiology*. **4**(9), 1571–1581.

Broger, T., Sossen, B., du Toit, E., Kerkhoff, A.D., Schutz, C., Ivanova Reipold, E., Ward, A., Barr, D.A., Macé, A., Trollip, A., Burton, R., Ongarello, S., Pinter, A., Lowary, T.L., Boehme, C., Nicol, M.P., Meintjes, G., Denkinger, C.M. (2019) Novel lipoarabinomannan point-of-care tuberculosis test for people with HIV: a diagnostic accuracy study. *The Lancet. Infectious Diseases*. **19**(8), 852–861.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., Cole, S.T. (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proceedings of the National Academy of Sciences*. **99**(6), 3684–3689.

Brown, C.M., Corey, R.A., Grélard, A., Gao, Y., Choi, Y.K., Luna, E., Gilleron, M., Destainville, N., Nigou, J., Loquet, A., Fullam, E., Im, W., Stansfeld, P.J., Chavent, M. (2023) Supramolecular organization and dynamics of mannosylated phosphatidylinositol lipids in the mycobacterial plasma membrane. *Proceedings of the National Academy of Sciences*. **120**(5), e2212755120.

Brown-Elliott, B.A., Biehle, J., Conville, P.S., Cohen, S., Saubolle, M., Sussland, D., Wengenack, N., Kriel, K., Bridge, L., McNulty, S., Vasireddy, R., Wallace, R.J. (2012) Sulfonamide resistance in isolates of Nocardia spp. from a US multicenter survey. *Journal of Clinical Microbiology*. **50**(3), 670–672.

Buter, J., Cheng, T.-Y., Ghanem, M., Grootemaat, A.E., Raman, S., Feng, X., Plantijn, A.R., Ennis, T., Wang, J., Cotton, R.N., Layre, E., Ramnarine, A.K., Mayfield, J.A., Young, D.C., Jezek Martinot, A., Siddiqi, N., Wakabayashi, S., Botella, H., Calderon, R., Murray, M., Ehrt, S., Snider, B.B., Reed, M.B., Oldfield, E., Tan, S., Rubin, E.J., Behr, M.A., van der Wel, N.N., Minnaard, A.J., Moody, D.B. (2019) Mycobacterium tuberculosis releases an antacid that remodels phagosomes. *Nature Chemical Biology*. **15**(9), 889–899.

Camacho, L.R., Constant, P., Raynaud, C., Lanéelle, M.-A., Triccas, J.A., Gicquel, B., Daffé, M., Guilhot, C. (2001) Analysis of the Phthiocerol Dimycocerosate Locus ofMycobacterium tuberculosis: EVIDENCE THAT THIS LIPID IS INVOLVED IN THE CELL WALL PERMEABILITY BARRIER*. *Journal of Biological Chemistry*. **276**(23), 19845–19854.

Cambier, C., Banik, S.M., Buonomo, J.A., Bertozzi, C.R. (2020) Spreading of a mycobacterial cell-surface lipid into host epithelial membranes promotes infectivity C. L. Stallings, D. Soldati-Favre, & S. H. Oehlers, eds. *eLife*. **9**, e60648.

Cao, Z., Casabona, M.G., Kneuper, H., Chalmers, J.D., Palmer, T. (2016) The type VII secretion system of Staphylococcus aureus secretes a nuclease toxin that targets competitor bacteria. *Nature Microbiology*. **2**(1), 1–11.

Chaloupka, J., Kreckova, P., Rihova, L. (1962) The mucopeptide turnover in the cell walls of growing cultures of Bacillus megaterium KM. *Experientia*. **18**, 362–363.

Chaloupka, J., Křečková, P., Říhová, L., Халоуика, Ю., Кршечкоба, П., Ржпеоба, Л. (1962) Changes in the Character of the Cell Wall in Growth ofBacillus megaterium Cultures. *Folia Microbiologica*. **7**(5), 269–274.

Chang, S.F., Chang, L.L., Chow, T.Y., Wu, W.J., Chang, J.C. (1992) Prevalence of transposons encoding kanamycin, ampicillin and trimethoprim resistance in isolates from urinary tract infections detected using DNA probes. *Gaoxiong Yi Xue Ke Xue Za Zhi = The Kaohsiung Journal of Medical Sciences*. **8**(3), 141–147.

Chatterjee, D., Khoo, K.H., McNeil, M.R., Dell, A., Morris, H.R., Brennan, P.J. (1993) Structural definition of the non-reducing termini of mannose-capped LAM from Mycobacterium tuberculosis through selective enzymatic degradation and fast atom bombardment-mass spectrometry. *Glycobiology*. **3**(5), 497–506.

Chavarro-Portillo, B., Soto, C.Y., Guerrero, M.I. (2019) Mycobacterium leprae's evolution and environmental adaptation. *Acta Tropica*. **197**, 105041.

Chen, J.L., Steele, T.W.J., Stuckey, D.C. (2018) Metabolic reduction of resazurin; location within the cell for cytotoxicity assays. *Biotechnology and Bioengineering*. **115**(2), 351–358.

Chen, T., Long, W., Zhang, C., Liu, S., Zhao, L., Hamaker, B.R. (2017) Fiber-utilizing capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. *Scientific Reports*. **7**(1), 2594.

Chen, Y.-Y., Yang, F.-L., Wu, S.-H., Lin, T.-L., Wang, J.-T. (2016) Mycobacterium marinum mmar_2318 and mmar_2319 are Responsible for Lipooligosaccharide Biosynthesis and Virulence Toward Dictyostelium. *Frontiers in Microbiology*. **6**, 1458.

Choudhary, A., Patel, D., Honnen, W., Lai, Z., Prattipati, R.S., Zheng, R.B., Hsueh, Y.-C., Gennaro, M.L., Lardizabal, A., Restrepo, B.I., Garcia-Viveros, M., Joe, M., Bai, Y., Shen, K., Sahloul, K., Spencer, J.S., Chatterjee, D., Broger, T., Lowary, T.L., Pinter, A. (2018) Characterization of the Antigenic Heterogeneity of Lipoarabinomannan, the Major Surface Glycolipid of Mycobacterium tuberculosis, and Complexity of Antibody Specificities toward This Antigen. *The Journal of Immunology Author Choice*. **200**(9), 3053–3066.

Christodoulou, D., Kuehne, A., Estermann, A., Fuhrer, T., Lang, P., Sauer, U. (2019) Reserve Flux Capacity in the Pentose Phosphate Pathway by NADPH Binding Is Conserved across Kingdoms. *iScience*. **19**, 1133–1144.

Chung, B.C., Zhao, J., Gillespie, R.A., Kwon, D.-Y., Guan, Z., Hong, J., Zhou, P., Lee, S.-Y. (2013) Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. *Science (New York, N.Y.)*. **341**(6149), 1012–1016.

Colaneri, M., Lombardi, A., Morea, A., Monzillo, V., Mariani, B., Marone, M., Sciarra, M., Sambo, M., Brunetti, E., Bruno, R., Seminari, E. (2021) An eight-year experience of Nocardia infection in Italy: does immunosuppression matter? *The New Microbiologica*. **44**(2), 111–116.

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., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.-A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*. **393**(6685), 537–544.

Colombo, M., Castilho, N.P.A., Todorov, S.D., Nero, L.A. (2017) Beneficial and Safety Properties of a Corynebacterium vitaeruminis Strain Isolated from the Cow Rumen. *Probiotics and Antimicrobial Proteins*. **9**(2), 157–162.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris,

S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S. (2013) Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature Genetics*. **45**(10), 1176–1182.

Crouch, L.I., Urbanowicz, P.A., Baslé, A., Cai, Z.-P., Liu, L., Voglmeir, J., Diaz, J.M.M., Benedict, S.T., Spencer, D.I.R., Bolam, D.N. (2022) Plant N-glycan breakdown by human gut Bacteroides. , 2022.04.07.487459.

Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., Rogowski, A., Hamilton, B.S., Chen, R., Tolbert, T.J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munōz-Munōz, J.L., Day, A., Peña, M.J., McLean, R., Suits, M.D., Boraston, A.B., Atherly, T., Ziemer, C.J., Williams, S.J., Davies, G.J., Abbott, D.W., Martens, E.C., Gilbert, H.J. (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature*. **517**(7533), 165–169.

Daffé, M., Lanéelle, M.A., Lacave, C. (1991) Structure and stereochemistry of mycolic acids of Mycobacterium marinum and Mycobacterium ulcerans. *Research in Microbiology*. **142**(4), 397–403.

Daffe, M., McNeil, M., Brennan, P.J. (1993) Major structural features of the cell wall arabinogalactans of Mycobacterium, Rhodococcus, and Nocardia spp. *Carbohydrate Research*. **249**(2), 383–398.

Daffé, M., Quémard, A., Marrakchi, H. (2017) Mycolic Acids: From Chemistry to Biology. In O. Geiger, ed. *Biogenesis of Fatty Acids, Lipids and Membranes*. Handbook of Hydrocarbon and Lipid Microbiology. Cham: Springer International Publishing, pp. 1–36.

Dai, T., Xie, J., Zhu, Q., Kamariza, M., Jiang, K., Bertozzi, C.R., Rao, J. (2020) A Fluorogenic Trehalose Probe for Tracking Phagocytosed Mycobacterium tuberculosis. *Journal of the American Chemical Society*. **142**(36), 15259–15264.

Danchik, C., Wang, S., Karakousis, P.C. (2021) Targeting the Mycobacterium tuberculosis Stringent Response as a Strategy for Shortening Tuberculosis Treatment. *Frontiers in Microbiology*. **12**.

Davies, G., Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure (London, England: 1993)*. **3**(9), 853–859.

Davis, J.M., Ramakrishnan, L. (2009) The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell*. **136**(1), 37–49.

De Smet, K.A.L., Kempsell, K.E., Gallagher, A., Duncan, K., Young, D.B.Y. 1999 Alteration of a single amino acid residue reverses fosfomycin resistance of recombinant MurA from Mycobacterium tuberculosisThe EMBL accession number for the sequence in this paper is X96711. *Microbiology*. **145**(11), 3177–3184.

Delafont, V., Samba-Louaka, A., Cambau, E., Bouchon, D., Moulin, L., Héchard, Y. (2017) Mycobacterium llatzerense, a waterborne Mycobacterium, that resists phagocytosis by Acanthamoeba castellanii. *Scientific Reports*. **7**, 46270.

Deshmukh, H.S., Hamburger, J.B., Ahn, S.H., McCafferty, D.G., Yang, S.R., Fowler, V.G. (2009) Critical Role of NOD2 in Regulating the Immune Response to Staphylococcus aureus. *Infection and Immunity*. **77**(4), 1376–1382.

Dinadayala, P., Kaur, D., Berg, S., Amin, A.G., Vissa, V.D., Chatterjee, D., Brennan, P.J., Crick, D.C. (2006) Genetic Basis for the Synthesis of the Immunomodulatory Mannose Caps of Lipoarabinomannan in Mycobacterium tuberculosis. *Journal of Biological Chemistry*. **281**(29), 20027–20035.

Dinadayala, P., Sambou, T., Daffé, M., Lemassu, A. (2008) Comparative structural analyses of the alpha-glucan and glycogen from Mycobacterium bovis. *Glycobiology*. **18**(7), 502–508.

Dong, X., Bhamidi, S., Scherman, M., Xin, Y., McNeil, M.R. (2006) Development of a quantitative assay for mycobacterial endogenous arabinase and ensuing studies of arabinase levels and arabinan metabolism in Mycobacterium smegmatis. *Applied and Environmental Microbiology*. **72**(4), 2601–2605.

Dong, Y., Li, J., Qiu, X., Yan, C., Li, X. (2015) Expression, purification and crystallization of the (3R)-hydroxyacyl-ACP dehydratase HadAB complex from Mycobacterium tuberculosis. *Protein Expression and Purification*. **114**, 115–120.

Dover, L.G., Cerdeño-Tárraga, A.M., Pallen, M.J., Parkhill, J., Besra, G.S. (2004) Comparative cell wall core biosynthesis in the mycolated pathogens, Mycobacterium tuberculosis and Corynebacterium diphtheriae. *FEMS microbiology reviews*. **28**(2), 225–250.

Dubé, J.-Y., McIntosh, F., Zarruk, J.G., David, S., Nigou, J., Behr, M.A. (2019) Synthetic mycobacterial molecular patterns work synergistically to recapitulate complete Freund's adjuvant. , 699165.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. **5**(1), 113.

El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., Henrissat, B. (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews. Microbiology*. **11**(7), 497–504.

Elad, N., Baron, S., Peleg, Y., Albeck, S., Grunwald, J., Raviv, G., Shakked, Z., Zimhony, O., Diskin, R. (2018) Structure of Type-I Mycobacterium tuberculosis fatty acid synthase at 3.3 Å resolution. *Nature Communications*. **9**(1), 3886.

Emsley, P., Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*. **60**(12), 2126–2132.

Eniyan, K., Dharavath, S., Vijayan, R., Bajpai, U., Gourinath, S. (2018) Crystal structure of UDP-N-acetylglucosamine-enolpyruvate reductase (MurB) from Mycobacterium tuberculosis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. **1866**(3), 397–406.

van den Ent, F., Löwe, J. (2006) RF cloning: A restriction-free method for inserting target genes into plasmids. *Journal of Biochemical and Biophysical Methods*. **67**(1), 67–74.

Escombe, A.R., Moore, D.A.J., Gilman, R.H., Pan, W., Navincopa, M., Ticona, E., Martínez, C., Caviedes, L., Sheen, P., Gonzalez, A., Noakes, C.J., Friedland, J.S., Evans, C.A. (2008) The Infectiousness of Tuberculosis Patients Coinfected with HIV. *PLOS Medicine*. **5**(9), e188.

Escuyer, V.E., Lety, M.A., Torrelles, J.B., Khoo, K.H., Tang, J.B., Rithner, C.D., Frehel, C., McNeil, M.R., Brennan, P.J., Chatterjee, D. (2001) The role of the embA and embB gene products in the biosynthesis of the terminal hexaarabinofuranosyl motif of Mycobacterium smegmatis arabinogalactan. *The Journal of Biological Chemistry*. **276**(52), 48854–48862.

Farro, E.G.S., Leite, A.E.T., Silva, I.A., Filgueiras, J.G., de Azevedo, E.R., Polikarpov, I., Nascimento, A.S. (2018) GH43 endo-arabinanase from Bacillus licheniformis: Structure, activity and unexpected synergistic effect on cellulose enzymatic hydrolysis. *International Journal of Biological Macromolecules*. **117**, 7–16.

Fierens, E., Rombouts, S., Gebruers, K., Goesaert, H., Brijs, K., Beaugrand, J., Volckaert, G., Campenhout, S.V., Proost, P., Courtin, C.M., Delcour, J.A. (2007) TLXI, a novel type of xylanase inhibitor from wheat (Triticum aestivum) belonging to the thaumatin family. *Biochemical Journal*. **403**(Pt 3), 583.

Gachon, F., Leuenberger, N., Claudel, T., Gos, P., Jouffe, C., Fleury Olela, F., de Mollerat du Jeu, X., Wahli, W., Schibler, U. (2011) Proline- and acidic amino acid-rich basic leucine zipper proteins modulate peroxisome proliferator-activated receptor α (PPARα) activity. *Proceedings of the National Academy of Sciences*. **108**(12), 4794–4799.

Gao, L.-Y., Guo, S., McLaughlin, B., Morisaki, H., Engel, J.N., Brown, E.J. (2004) A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Molecular Microbiology*. **53**(6), 1677–1693.

García-Heredia, A., Kado, T., Sein, C.E., Puffal, J., Osman, S.H., Judd, J., Gray, T.A., Morita, Y.S., Siegrist, M.S. (2021) Membrane-partitioned cell wall synthesis in mycobacteria J. Xiao & G. Storz, eds. *eLife*. **10**, e60263.

Gavalda, S., Bardou, F., Laval, F., Bon, C., Malaga, W., Chalut, C., Guilhot, C., Mourey, L., Daffé, M., Quémard, A. (2014) The polyketide synthase Pks13 catalyzes a novel mechanism of lipid transfer in mycobacteria. *Chemistry & Biology*. **21**(12), 1660–1669.

Gilmore, S.A., Schelle, M.W., Holsclaw, C.M., Leigh, C.D., Jain, M., Cox, J.S., Leary, J.A., Bertozzi, C.R. (2012) Sulfolipid-1 Biosynthesis Restricts Mycobacterium tuberculosis Growth in Human Macrophages. *ACS Chemical Biology*. **7**(5), 863–870.

Girardin, S.E., Boneca, I.G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpott, D.J., Sansonetti, P.J. (2003) Nod2 Is a General Sensor of Peptidoglycan through Muramyl Dipeptide (MDP) Detection *. *Journal of Biological Chemistry*. **278**(11), 8869–8872.

Glauner, B., Höltje, J.V., Schwarz, U. (1988) The composition of the murein of Escherichia coli. *The Journal of Biological Chemistry*. **263**(21), 10088–10095.

Goins, C.M., Dajnowicz, S., Smith, M.D., Parks, J.M., Ronning, D.R. (2018) Mycolyltransferase from Mycobacterium tuberculosis in covalent complex with tetrahydrolipstatin provides insights into antigen 85 catalysis. *The Journal of Biological Chemistry*. **293**(10), 3651–3662.

Gopalaswamy, R., Shanmugam, S., Mondal, R., Subbian, S. (2020) Of tuberculosis and non-tuberculous mycobacterial infections – a comparative analysis of epidemiology, diagnosis and treatment. *Journal of Biomedical Science*. **27**(1), 74.

Gouzy, A., Healy, C., Black, K.A., Rhee, K.Y., Ehrt, S. (2021) Growth of Mycobacterium tuberculosis at acidic pH depends on lipid assimilation and is accompanied by reduced GAPDH activity. *Proceedings of the National Academy of Sciences of the United States of America*. **118**(32).

Goyal, A., Ahmed, S., Sharma, K., Gupta, V., Bule, P., Alves, V.D., Fontes, C.M.G.A., Najmudin, S. (2016) Molecular determinants of substrate specificity revealed by the structure of Clostridium thermocellum arabinofuranosidase 43A from glycosyl hydrolase family 43 subfamily 16. *Acta Crystallographica. Section D, Structural Biology*. **72**(Pt 12), 1281–1289.

Granato, E.T., Meiller-Legrand, T.A., Foster, K.R. (2019) The Evolution and Ecology of Bacterial Warfare. *Current Biology*. **29**(11), R521–R537.

Gray, T.A., Clark, R.R., Boucher, N., Lapierre, P., Smith, C., Derbyshire, K.M. (2016) Intercellular communication and conjugation are mediated by ESX secretion systems in mycobacteria. *Science*. **354**(6310), 347–350.

Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W., Brumer, H. (2017) Polysaccharide Utilization Loci: Fueling Microbial Communities. *Journal of Bacteriology*. **199**(15).

Guerardel, Y., Maes, E., Elass, E., Leroy, Y., Timmerman, P., Besra, G.S., Locht, C., Strecker, G., Kremer, L. (2002) Structural study of lipomannan and lipoarabinomannan from Mycobacterium chelonae. Presence of unusual components with alpha 1,3-mannopyranose side chains. *The Journal of Biological Chemistry*. **277**(34), 30635–30648.

Guerin, M.E., Kaur, D., Somashekar, B.S., Gibbs, S., Gest, P., Chatterjee, D., Brennan, P.J., Jackson, M. (2009) New Insights into the Early Steps of Phosphatidylinositol Mannoside Biosynthesis in Mycobacteria. *The Journal of Biological Chemistry*. **284**(38), 25687–25696.

Guerrant, R.L., Walker, D.H., Weller, P.F. (2011) *Tropical Infectious Diseases: Principles, Pathogens and Practice*. Saunders.

Gupta, R., Lavollay, M., Mainardi, J.-L., Arthur, M., Bishai, W.R., Lamichhane, G. (2010) The Mycobacterium tuberculosis protein LdtMt2 is a nonclassical transpeptidase required for virulence and resistance to amoxicillin. *Nature Medicine*. **16**(4), 466–469.

Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omaïs, B., Marmiesse, M., Supply, P., Vincent, V. (2005) Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis. *PLOS Pathogens*. **1**(1), e5.

Hansen, J.M., Golchin, S.A., Veyrier, F.J., Domenech, P., Boneca, I.G., Azad, A.K., Rajaram, M.V.S., Schlesinger, L.S., Divangahi, M., Reed, M.B., Behr, M.A. (2014) N-glycolylated peptidoglycan contributes to the immunogenicity but not pathogenicity of Mycobacterium tuberculosis. *The Journal of Infectious Diseases*. **209**(7), 1045–1054.

Harrison, J., Lloyd, G., Joe, M., Lowary, T.L., Reynolds, E., Walters-Morgan, H., Bhatt, A., Lovering, A., Besra, G.S., Alderwick, L.J. (2016) Lcp1 Is a Phosphotransferase Responsible for Ligating Arabinogalactan to Peptidoglycan in Mycobacterium tuberculosis. *mBio*. **7**(4).

Harwood, J.L., Nicholls, R.G. (1979) The plant sulpholipid-- a major component of the sulphur cycle. *Biochemical Society Transactions*. **7**(2), 440–447.

Healy, C., Gouzy, A., Ehrt, S. (2020) Peptidoglycan Hydrolases RipA and Ami1 Are Critical for Replication and Persistence of Mycobacterium tuberculosis in the Host. *mBio*. **11**(2).

Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical Journal*. **280 (Pt 2)**, 309–316.

Honda, J.R., Virdi, R., Chan, E.D. (2018) Global Environmental Nontuberculous Mycobacteria and Their Contemporaneous Man-Made and Natural Niches. *Frontiers in Microbiology*. **9**.

Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M.A., Trinidad, R.R.S., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Plemel, R.L., Li, M., Schwarz, S., Wang, W.Y., Merz, A.J., Goodlett, D.R., Mougous, J.D. (2010) A Type VI Secretion System of Pseudomonas aeruginosa Targets a Toxin to Bacteria. *Cell Host & Microbe*. **7**(1), 25–37.

Hoskisson, P.A. (2018) Microbe Profile: Corynebacterium diphtheriae – an old foe always ready to seize opportunity. *Microbiology*. **164**(6), 865–867.

Hostetler, K.Y., Landau, B.R. (1967) Estimation of the Pentose Cycle Contribution to Glucose Metabolism in Tissue in Vivo*. *Biochemistry*. **6**(10), 2961–2964.

Houben, E.N.G., Korotkov, K.V., Bitter, W. (2014) Take five — Type VII secretion systems of Mycobacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. **1843**(8), 1707–1716.

Housden, N.G., Kleanthous, C. (2011) Thermodynamic dissection of colicin interactions. *Methods in Enzymology*. **488**, 123–145.

Huang, H., Scherman, M.S., D'Haeze, W., Vereecke, D., Holsters, M., Crick, D.C., McNeil, M.R. (2005) Identification and Active Expression of the Mycobacterium tuberculosis Gene Encoding 5-Phospho-α-D-ribose-1-diphosphate: Decaprenyl-phosphate 5-Phosphoribosyltransferase, the First Enzyme Committed to Decaprenylphosphoryl-D-arabinose Synthesis*. *Journal of Biological Chemistry*. **280**(26), 24539–24543.

Huber, R.E., Kurz, G., Wallenfels, K. (1976) A quantitation of the factors which affect the hydrolase and transgalactosylase activities of beta-galactosidase (E. coli) on lactose. *Biochemistry*. **15**(9), 1994–2001.

Huc, E., Meniche, X., Benz, R., Bayan, N., Ghazi, A., Tropis, M., Daffé, M. (2010) O-Mycoloylated Proteins from Corynebacterium: AN UNPRECEDENTED POST-TRANSLATIONAL MODIFICATION IN BACTERIA*. *Journal of Biological Chemistry*. **285**(29), 21908–21912.

Indrigo, J., Hunter, R.L., Actor, J.K. (2003) Cord factor trehalose 6,6'-dimycolate (TDM) mediates trafficking events during mycobacterial infection of murine macrophages. *Microbiology*. **149**(8), 2049–2059.

Inohara, N., Ogura, Y., Fontalba, A., Gutierrez, O., Pons, F., Crespo, J., Fukase, K., Inamura, S., Kusumoto, S., Hashimoto, M., Foster, S.J., Moran, A.P., Fernandez-Luna, J.L., Nuñez, G. (2003) Host Recognition of Bacterial Muramyl Dipeptide Mediated through NOD2: IMPLICATIONS FOR CROHN'S DISEASE *. *Journal of Biological Chemistry*. **278**(8), 5509–5512.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*. **169**(12), 5429–5433.

Jankute, M., Alderwick, L.J., Moorey, A.R., Joe, M., Gurcha, S.S., Eggeling, L., Lowary, T.L., Dell, A., Pang, P.-C., Yang, T., Haslam, S., Besra, G.S. (2018) The singular Corynebacterium glutamicum Emb arabinofuranosyltransferase polymerises the α(1 → 5) arabinan backbone in the early stages of cell wall arabinan biosynthesis. *Cell Surface (Amsterdam)*. **2**, 38–53.

Jankute, M., Alderwick, L.J., Noack, S., Veerapen, N., Nigou, J., Besra, G.S. (2017) Disruption of Mycobacterial AftB Results in Complete Loss of Terminal β(1 → 2) Arabinofuranose Residues of Lipoarabinomannan. *ACS Chemical Biology*. **12**(1), 183–190.

Jankute, M., Grover, S., Birch, H.L., Besra, G.S. (2014) Genetics of Mycobacterial Arabinogalactan and Lipoarabinomannan Assembly. *Microbiology Spectrum*. **2**(4), MGM2-0013–2013.

Jankute, M., Nataraj, V., Lee, O.Y.-C., Wu, H.H.T., Ridell, M., Garton, N.J., Barer, M.R., Minnikin, D.E., Bhatt, A., Besra, G.S. (2017) The role of hydrophobicity in tuberculosis evolution and pathogenicity. *Scientific Reports*. **7**(1), 1315.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*. **30**(9), 1236–1240.

Joosten, R.P., Long, F., Murshudov, G.N., Perrakis, A. (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*. **1**(4), 213–220.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**(7873), 583–589.

Justen, A.M., Hodges, H.L., Kim, L.M., Sadecki, P.W., Porfirio, S., Ultee, E., Black, I., Chung, G.S., Briegel, A., Azadi, P., Kiessling, L.L. (2020) Polysaccharide length affects mycobacterial cell shape and antibiotic susceptibility. *Science Advances*. **6**(38), eaba4015.

Kalscheuer, R., Palacios, A., Anso, I., Cifuente, J., Anguita, J., Jacobs, W.R., Guerin, M.E., Prados-Rosales, R. (2019) The Mycobacterium tuberculosis capsule: a cell structure with key implications in pathogenesis. *The Biochemical journal*. **476**(14), 1995–2016.

Kalscheuer, R., Weinrick, B., Veeraraghavan, U., Besra, G.S., Jacobs, W.R. (2010a) Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*. **107**(50), 21761–21766.

Kalscheuer, R., Weinrick, B., Veeraraghavan, U., Besra, G.S., Jacobs, W.R. (2010b) Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences*. **107**(50), 21761–21766.

Kamariza, M., Shieh, P., Bertozzi, C.R. (2018) Imaging Mycobacterial Trehalose Glycolipids. *Methods in enzymology*. **598**, 355–369.

Kana, B.D., Mizrahi, V. (2010) Resuscitation-promoting factors as lytic enzymes for bacterial growth and signaling. *FEMS Immunology & Medical Microbiology*. **58**(1), 39–50.

Kang, P.B., Azad, A.K., Torrelles, J.B., Kaufman, T.M., Beharka, A., Tibesar, E., DesJardin, L.E., Schlesinger, L.S. (2005) The human macrophage mannose receptor directs Mycobacterium tuberculosis lipoarabinomannan-mediated phagosome biogenesis. *The Journal of Experimental Medicine*. **202**(7), 987–999.

Kashima, T., Okumura, K., Ishiwata, A., Kaieda, M., Terada, T., Arakawa, T., Yamada, C., Shimizu, K., Tanaka, K., Kitaoka, M., Ito, Y., Fujita, K., Fushinobu, S. (2021) Identification of difructose dianhydride I synthase/hydrolase from an oral bacterium establishes a novel glycoside hydrolase family. *Journal of Biological Chemistry*. **297**(5).

Kaur, D., Obregón-Henao, A., Pham, H., Chatterjee, D., Brennan, P.J., Jackson, M. (2008) Lipoarabinomannan of Mycobacterium: mannose capping by a multifunctional terminal mannosyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*. **105**(46), 17973–17977.

Keane, J., Balcewicz-Sablinska, M.K., Remold, H.G., Chupp, G.L., Meek, B.B., Fenton, M.J., Kornfeld, H. (1997) Infection by Mycobacterium tuberculosis promotes human alveolar macrophage apoptosis. *Infection and Immunity*. **65**(1), 298–304.

Keeble, A.H., Kirkpatrick, N., Shimizu, S., Kleanthous, C. (2006) Calorimetric Dissection of Colicin DNase−Immunity Protein Complex Specificity. *Biochemistry*. **45**(10), 3243–3254.

Kelly, M.S., Plunkett, C., Yu, Y., Aquino, J.N., Patel, S.M., Hurst, J.H., Young, R.R., Smieja, M., Steenhoff, A.P., Arscott-Mills, T., Feemster, K.A., Boiditswe, S., Leburu, T., Mazhani, T., Patel, M.Z., Rawls, J.F., Jawahar, J., Shah, S.S., Polage, C.R., Cunningham, C.K., Seed, P.C. (2022)

Non-diphtheriae Corynebacterium species are associated with decreased risk of pneumococcal colonization during infancy. *The ISME Journal*. **16**(3), 655–665.

van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., Steinegger, M. (2023) Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, 1–4.

Kenyon, C.P., Roth, R.L., van der Westhuyzen, C.W., Parkinson, C.J. (2012) Conserved phosphoryl transfer mechanisms within kinase families and the role of the C8 proton of ATP in the activation of phosphoryl transfer. *BMC Research Notes*. **5**, 131.

Khoo, K.H., Douglas, E., Azadi, P., Inamine, J.M., Besra, G.S., Mikusová, K., Brennan, P.J., Chatterjee, D. (1996) Truncated structural variants of lipoarabinomannan in ethambutol drug-resistant strains of Mycobacterium smegmatis. Inhibition of arabinan biosynthesis by ethambutol. *The Journal of Biological Chemistry*. **271**(45), 28682–28690.

King, C.H., Desai, H., Sylvetsky, A.C., LoTempio, J., Ayanyan, S., Carrie, J., Crandall, K.A., Fochtman, B.C., Gasparyan, L., Gulzar, N., Howell, P., Issa, N., Krampis, K., Mishra, L., Morizono, H., Pisegna, J.R., Rao, S., Ren, Y., Simonyan, V., Smith, K., VedBrat, S., Yao, M.D., Mazumder, R. (2019) Baseline human gut microbiota profile in healthy people and standard reporting template. *PLoS ONE*. **14**(9), e0206484.

Klein, T.A., Grebenc, D.W., Gandhi, S.Y., Shah, V.S., Kim, Y., Whitney, J.C. (2021) Structure of the Extracellular Region of the Bacterial Type VIIb Secretion System Subunit EsaA. *Structure (London, England: 1993)*. **29**(2), 177-185.e6.

Kleine, B., Chattopadhyay, A., Polen, T., Pinto, D., Mascher, T., Bott, M., Brocker, M., Freudl, R. (2017) The three-component system EsrISR regulates a cell envelope stress response in Corynebacterium glutamicum. *Molecular Microbiology*. **106**(5), 719–741.

Korduláková, J., Gilleron, M., Mikušová, K., Puzo, G., Brennan, P.J., Gicquel, B., Jackson, M. (2002) Definition of the First Mannosylation Step in Phosphatidylinositol Mannoside Synthesis: PimA IS ESSENTIAL FOR GROWTH OF MYCOBACTERIA *. *Journal of Biological Chemistry*. **277**(35), 31335–31344.

Korduláková, J., Gilleron, M., Puzo, G., Brennan, P.J., Gicquel, B., Mikusová, K., Jackson, M. (2003) Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of mycobacterium species. *The Journal of Biological Chemistry*. **278**(38), 36285–36295.

Krafczyk, R., Qi, F., Sieber, A., Mehler, J., Jung, K., Frishman, D., Lassak, J. (2021) Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria. *Communications Biology*. **4**(1), 1–11.

Kremer, L., Dover, L.G., Carrère, S., Nampoothiri, K.M., Lesjean, S., Brown, A.K., Brennan, P.J., Minnikin, D.E., Locht, C., Besra, G.S. (2002) Mycolic acid biosynthesis and enzymic characterization of the beta-ketoacyl-ACP synthase A-condensing enzyme from Mycobacterium tuberculosis. *The Biochemical Journal*. **364**(Pt 2), 423–430.

Kremer, L., Nampoothiri, K.M., Lesjean, S., Dover, L.G., Graham, S., Betts, J., Brennan, P.J., Minnikin, D.E., Locht, C., Besra, G.S. (2001) Biochemical characterization of acyl carrier protein (AcpM) and malonyl-CoA:AcpM transacylase (mtFabD), two major components of Mycobacterium tuberculosis fatty acid synthase II. *The Journal of Biological Chemistry*. **276**(30), 27967–27974.

Krupovic, M., Makarova, K.S., Koonin, E.V. (2022) Cellular homologs of the double jelly-roll major capsid proteins clarify the origins of an ancient virus kingdom. *Proceedings of the National Academy of Sciences*. **119**(5), e2120620119.

Kus, J.V., Kelly, J., Tessier, L., Harvey, H., Cvitkovitch, D.G., Burrows, L.L. (2008) Modification of Pseudomonas aeruginosa Pa5196 type IV Pilins at multiple sites with D-Araf by a novel GT-C family Arabinosyltransferase, TfpW. *Journal of Bacteriology*. **190**(22), 7464–7478.

LaCourse, K.D., Peterson, S.B., Kulasekara, H.D., Radey, M.C., Kim, J., Mougous, J.D. (2018) Conditional toxicity and synergy drive diversity among antibacterial effectors. *Nature Microbiology*. **3**(4), 440–446.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227**(5259), 680–685.

Larrouy-Maumus, G., Škovierová, H., Dhouib, R., Angala, S.K., Zuberogoitia, S., Pham, H., Villela, A.D., Mikušová, K., Noguera, A., Gilleron, M., Valentínová, L., Korduláková, J., Brennan, P.J., Puzo, G., Nigou, J., Jackson, M. (2012) A small multidrug resistance-like transporter involved in the arabinosylation of arabinogalactan and lipoarabinomannan in mycobacteria. *The Journal of Biological Chemistry*. **287**(47), 39933–39941.

Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Veziris, N., Blanot, D., Gutmann, L., Mainardi, J.-L. (2008) The Peptidoglycan of Stationary-Phase Mycobacterium tuberculosis Predominantly Contains Cross-Links Generated by l,d-Transpeptidation. *Journal of Bacteriology*. **190**(12), 4360–4366.

Lee, M.-R., Sheng, W.-H., Hung, C.-C., Yu, C.-J., Lee, L.-N., Hsueh, P.-R. (2015) Mycobacterium abscessus Complex Infections in Humans. *Emerging Infectious Diseases*. **21**(9), 1638–1646.

Li, M., Müller, C., Fröhlich, K., Gorka, O., Zhang, L., Groß, O., Schilling, O., Einsle, O., Jessen-Trefzer, C. (2019) Detection and Characterization of a Mycobacterial L-Arabinofuranose ABC Transporter Identified with a Rapid Lipoproteomics Protocol. *Cell Chemical Biology*. **26**(6), 852-862.e6.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Poon, B.K., Prisant, M.G., Read, R.J., Richardson, J.S., Richardson, D.C., Sammito, M.D., Sobolev, O.V., Stockwell, D.H., Terwilliger, T.C., Urzhumtsev, A.G., Videau, L.L., Williams, C.J., Adams, P.D. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*. **75**(10), 861–877.

Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P., Berkmen, M. (2012) SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*. **11**(1), 753.

Maeda, N., Nigou, J., Herrmann, J.-L., Jackson, M., Amara, A., Lagrange, P.H., Puzo, G., Gicquel, B., Neyrolles, O. (2003) The Cell Surface Receptor DC-SIGN Discriminates betweenMycobacterium Species through Selective Recognition of the Mannose Caps on Lipoarabinomannan. *Journal of Biological Chemistry*. **278**(8), 5513–5516.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., Stover, C.K. (1996) Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *Journal of Bacteriology*. **178**(5), 1274–1282.

Mahapatra, S., Scherman, H., Brennan, P.J., Crick, D.C. (2005) N Glycolylation of the Nucleotide Precursors of Peptidoglycan Biosynthesis of Mycobacterium spp. Is Altered by Drug Treatment. *Journal of Bacteriology*. **187**(7), 2341–2347.

Malone, K.M., Rue-Albrecht, K., Magee, D.A., Conlon, K., Schubert, O.T., Nalpas, N.C., Browne, J.A., Smyth, A., Gormley, E., Aebersold, R., MacHugh, D.E., Gordon, S.V. (2018) Comparative 'omics analyses differentiate Mycobacterium tuberculosis and Mycobacterium bovis and reveal distinct macrophage responses to infection with the human and bovine tubercle bacilli. *Microbial Genomics*. **4**(3).

Marshall, C.G., Wright, G.D. (1998) DdlN from Vancomycin-Producing Amycolatopsis orientalis C329.2 Is a VanA Homologue with d-Alanyl-d-Lactate Ligase Activity. *Journal of Bacteriology*. **180**(21), 5792–5795.

Marsollier, L., Brodin, P., Jackson, M., Korduláková, J., Tafelmeyer, P., Carbonnelle, E., Aubry, J., Milon, G., Legras, P., André, J.-P.S., Leroy, C., Cottin, J., Guillou, M.L.J., Reysset, G., Cole, S.T. (2007) Impact of Mycobacterium ulcerans Biofilm on Transmissibility to Ecological Niches and Buruli Ulcer Pathogenesis. *PLOS Pathogens*. **3**(5), e62.

Matthews, C., Walsh, A.M., Gordon, S.V., Markey, B., Cotter, P.D., O' Mahony, J. (2023) Differences in Faecal Microbiome Taxonomy, Diversity and Functional Potential in a Bovine Cohort Experimentally Challenged with Mycobacterium avium subsp. paratuberculosis (MAP). *Animals*. **13**(10), 1652.

Mavridou, D.A.I., Gonzalez, D., Kim, W., West, S.A., Foster, K.R. (2018) Bacteria Use Collective Behavior to Generate Diverse Combat Strategies. *Current Biology*. **28**(3), 345-355.e4.

McCarter, J.D., Withers, S.G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology*. **4**(6), 885–892.

McIntosh, L.P., Hand, G., Johnson, P.E., Joshi, M.D., Körner, M., Plesniak, L.A., Ziser, L., Wakarchuk, W.W., Withers, S.G. (1996) The pKa of the General Acid/Base Carboxyl Group of a Glycosidase Cycles during Catalysis: A 13C-NMR Study of Bacillus circulans Xylanase. *Biochemistry*. **35**(31), 9958–9966.

McNeil, M., Daffe, M., Brennan, P.J. (1990) Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls. *The Journal of Biological Chemistry*. **265**(30), 18200–18206.

Meeske, A.J., Riley, E.P., Robins, W.P., Uehara, T., Mekelanos, J.J., Kahne, D., Walker, S., Kruse, A.C., Bernhardt, T.G., Rudner, D.Z. (2016) SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature*. **537**(7622), 634–638.

Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., Ferrin, T.E. (2023) UCSF ChimeraX: Tools for structure building and analysis. *Protein Science*. **32**(11), e4792.

Meng, F., Wang, C., Kurgan, L. (2018) fDETECT webserver: fast predictor of propensity for protein production, purification, and crystallization. *BMC Bioinformatics*. **18**(1), 580.

Meng, G., Fütterer, K. (2008) Donor substrate recognition in the raffinose-bound E342A mutant of fructosyltransferase Bacillus subtilis levansucrase. *BMC Structural Biology*. **8**(1), 16.

Mengin-Lecreulx, D., Texier, L., Rousseau, M., van Heijenoort, J. (1991) The murG gene of Escherichia coli codes for the UDP-N-acetylglucosamine: N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. *Journal of Bacteriology*. **173**(15), 4625–4636.

Meza, A.N., Cambui, C.C.N., Moreno, A.C.R., Fessel, M.R., Balan, A. (2019) Mycobacterium tuberculosis CysA2 is a dual sulfurtransferase with activity against thiosulfate and 3-mercaptopyruvate and interacts with mammalian cells. *Scientific Reports*. **9**, 16791.

Mills, J.A., Motichka, K., Jucker, M., Wu, H.P., Uhlik, B.C., Stern, R.J., Scherman, M.S., Vissa, V.D., Pan, F., Kundu, M., Ma, Y.F., McNeil, M. (2004) Inactivation of the Mycobacterial Rhamnosyltransferase, Which Is Needed for the Formation of the Arabinogalactan-Peptidoglycan Linker, Leads to Irreversible Loss of Viability. *Journal of Biological Chemistry*. **279**(42), 43540–43546.

Minnikin, D.E., Goodfellow, M. (1980) Lipid composition in the classification and identification of acid-fast bacteria. *Society for Applied Bacteriology Symposium Series*. **8**, 189–256.

Minnikin, D.E., Minnikin, S.M., Parlett, J.H., Goodfellow, M., Magnusson, M. (1984) Mycolic acid patterns of some species of Mycobacterium. *Archives of microbiology*. **139**(2), 225–231.

Mishra, A.K., Alderwick, L.J., Rittmann, D., Tatituri, R.V.V., Nigou, J., Gilleron, M., Eggeling, L., Besra, G.S. (2007) Identification of an alpha(1-->6) mannopyranosyltransferase (MptA), involved in Corynebacterium glutamicum lipomanann biosynthesis, and identification of its orthologue in Mycobacterium tuberculosis. *Molecular Microbiology*. **65**(6), 1503–1517.

Mishra, A.K., Alderwick, L.J., Rittmann, D., Wang, C., Bhatt, A., Jacobs, W.R., Takayama, K., Eggeling, L., Besra, G.S. (2008) Identification of a novel alpha(1-->6) mannopyranosyltransferase MptB from Corynebacterium glutamicum by deletion of a

conserved gene, NCgl1505, affords a lipomannan- and lipoarabinomannan-deficient mutant. *Molecular Microbiology*. **68**(6), 1595–1613.

Mistretta, N., Brossaud, M., Telles, F., Sanchez, V., Talaga, P., Rokbi, B. (2019) Glycosylation of Staphylococcus aureus cell wall teichoic acid is influenced by environmental conditions. *Scientific Reports*. **9**(1), 3212.

Miyata, S.T., Kitaoka, M., Brooks, T.M., McAuley, S.B., Pukatzki, S. (2011) Vibrio cholerae requires the type VI secretion system virulence factor VasX to kill Dictyostelium discoideum. *Infection and Immunity*. **79**(7), 2941–2949.

Mok, B.Y., de Moraes, M.H., Zeng, J., Bosch, D.E., Kotrys, A.V., Raguram, A., Hsu, F., Radey, M.C., Peterson, S.B., Mootha, V.K., Mougous, J.D., Liu, D.R. (2020) A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature*. **583**(7817), 631–637.

Morita, Y.S., Sena, C.B.C., Waller, R.F., Kurokawa, K., Sernee, M.F., Nakatani, F., Haites, R.E., Billman-Jacobe, H., McConville, M.J., Maeda, Y., Kinoshita, T. (2006) PimE is a polyprenol-phosphate-mannose-dependent mannosyltransferase that transfers the fifth mannose of phosphatidylinositol mannoside in mycobacteria. *The Journal of Biological Chemistry*. **281**(35), 25143–25155.

Morrison, H.M., Craft, J., Rivera-Lugo, R., Johnson, J.R., Golovkine, G.R., Bell, S.L., Dodd, C.E., Van Dis, E., Beatty, W.L., Margolis, S.R., Repasy, T., Shaker, I., Lee, A.Y., Vance, R.E., Stanley, S.A., Watson, R.O., Krogan, N.J., Portnoy, D.A., Penn, B.H., Cox, J.S. (2023) Deficiency in Galectin-3, -8, and -9 impairs immunity to chronic Mycobacterium tuberculosis infection but not acute infection with multiple intracellular pathogens. *PLoS pathogens*. **19**(6), e1011088.

Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., Goodman, A.L., Joachimiak, G., Ordoñez, C.L., Lory, S., Walz, T., Joachimiak, A., Mekalanos, J.J. (2006) A Virulence Locus of Pseudomonas aeruginosa Encodes a Protein Secretion Apparatus. *Science (New York, N.Y.)*. **312**(5779), 1526–1530.

Moynihan, P.J., Cadby, I.T., Veerapen, N., Jankute, M., Crosatti, M., Mukamolova, G.V., Lovering, A.L., Besra, G.S. (2019) The hydrolase LpqI primes mycobacterial peptidoglycan recycling. *Nature Communications*. **10**(1), 1–11.

Nandakumar, M., Nathan, C., Rhee, K.Y. (2014) Isocitrate lyase mediates broad antibiotic tolerance in Mycobacterium tuberculosis. *Nature Communications*. **5**, 4306.

Nimmo, C., Millard, J., Faulkner, V., Monteserin, J., Pugh, H., Johnson, E.O. (2022) Evolution of Mycobacterium tuberculosis drug resistance in the genomic era. *Frontiers in Cellular and Infection Microbiology*. **12**.

Nygaard, R., Graham, C.L.B., Belcher Dufrisne, M., Colburn, J.D., Pepe, J., Hydorn, M.A., Corradi, S., Brown, C.M., Ashraf, K.U., Vickery, O.N., Briggs, N.S., Deering, J.J., Kloss, B., Botta, B., Clarke, O.B., Columbus, L., Dworkin, J., Stansfeld, P.J., Roper, D.I., Mancia, F. (2023) Structural basis of peptidoglycan synthesis by E. coli RodA-PBP2 complex. *Nature Communications*. **14**(1), 5151.

Ocaña-Ortega, A., Pérez-Flores, G., Torres-Tirado, D., Pérez-García, L.A. (2023) O-Linked Glycans of Candida albicans Interact with Specific GPCRs in the Coronary Endothelium and Inhibit the Cardiac Response to Agonists. *Journal of Fungi*. **9**(2), 141.

Ojha, A.K., Trivelli, X., Guerardel, Y., Kremer, L., Hatfull, G.F. (2010) Enzymatic Hydrolysis of Trehalose Dimycolate Releases Free Mycolic Acids during Mycobacterial Growth in Biofilms*. *Journal of Biological Chemistry*. **285**(23), 17380–17389.

Oliveira, N.M., Wheeler, J.H.R., Deroy, C., Booth, S.C., Walsh, E.J., Durham, W.M., Foster, K.R. (2022) Suicidal chemotaxis in bacteria. *Nature Communications*. **13**(1), 7608.

Oluyombo, O., Penfold, C.N., Diggle, S.P. (2019) Competition in Biofilms between Cystic Fibrosis Isolates of Pseudomonas aeruginosa Is Shaped by R-Pyocins. *mBio*. **10**(1), 10.1128/mbio.01828-18.

Oo, C., Barrett, J., Hill, G., Mann, J., Dorr, A., Dutkowski, R., Ward, P. (2001) Pharmacokinetics and dosage recommendations for an oseltamivir oral suspension for the treatment of influenza in children. *Paediatric Drugs*. **3**(3), 229–236.

O'Reilly, L.M., Daborn, C.J. (1995) The epidemiology of Mycobacterium bovis infections in animals and man: A review. *Tubercle and Lung Disease*. **76**, 1–46.

Pajic, P., Pavlidis, P., Dean, K., Neznanova, L., Romano, R.-A., Garneau, D., Daugherity, E., Globig, A., Ruhl, S., Gokcumen, O. (2019) Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *eLife*. **8**.

Palčeková, Z., Angala, S.K., Belardinelli, J.M., Eskandarian, H.A., Joe, M., Brunton, R., Rithner, C., Jones, V., Nigou, J., Lowary, T.L., Gilleron, M., McNeil, M., Jackson, M. (2019) Disruption of the SucT acyltransferase in Mycobacterium smegmatis abrogates succinylation of cell envelope polysaccharides. *The Journal of Biological Chemistry*. **294**(26), 10325–10335.

Palcekova, Z., Gilleron, M., Angala, S.K., Belardinelli, J.M., McNeil, M.R., Bermudez, L.E., Jackson, M.C. (2020) Polysaccharide succinylation enhances the intracellular survival of Mycobacterium abscessus. *ACS infectious diseases*.

Palčeková, Z., Obregón-Henao, A., De, K., Walz, A., Lam, H., Philp, J., Angala, S.K., Patterson, J., Pearce, C., Zuberogoitia, S., Avanzi, C., Nigou, J., McNeil, M., Muñoz Gutiérrez, J.F., Gilleron, M., Wheat, W.H., Gonzalez-Juarrero, M., Jackson, M. (2023) Role of succinyl substituents in the mannose-capping of lipoarabinomannan and control of inflammation in Mycobacterium tuberculosis infection. *PLoS pathogens*. **19**(9), e1011636.

Payan, F., Leone, P., Porciero, S., Furniss, C., Tahir, T., Williamson, G., Durand, A., Manzanares, P., Gilbert, H.J., Juge, N., Roussel, A. (2004) The dual nature of the wheat xylanase protein inhibitor XIP-I: structural basis for the inhibition of family 10 and family 11 xylanases. *The Journal of Biological Chemistry*. **279**(34), 36029–36037.

Pearce, O.M.T. (2018) Cancer glycan epitopes: biosynthesis, structure and function. *Glycobiology*. **28**(9), 670–696.

de Pedro, M.A., Quintela, J.C., Höltje, J.V., Schwarz, H. (1997) Murein segregation in Escherichia coli. *Journal of Bacteriology*. **179**(9), 2823–2834.

Peng, W., Zou, L., Bhamidi, S., McNeil, M.R., Lowary, T.L. (2012) The Galactosamine Residue in Mycobacterial Arabinogalactan Is α-Linked. *The Journal of Organic Chemistry*. **77**(21), 9826–9832.

Perkowski, E.F., Zulauf, K.E., Weerakoon, D., Hayden, J.D., Ioerger, T.R., Oreper, D., Gomez, S.M., Sacchettini, J.C., Braunstein, M. (2017) The EXIT Strategy: an Approach for Identifying Bacterial Proteins Exported during Host Infection. *mBio*. **8**(2), 10.1128/mbio.00333-17.

Pohane, A.A., Carr, C.R., Garhyan, J., Swarts, B.M., Siegrist, M.S. (2021) Trehalose Recycling Promotes Energy-Efficient Biosynthesis of the Mycobacterial Cell Envelope. *mBio*. **12**(1), e02801-20.

Poulin, M.B., Lowary, T.L. (2016) Chemical Insight into the Mechanism and Specificity of GlfT2, a Bifunctional Galactofuranosyltransferase from Mycobacteria. *The Journal of Organic Chemistry*. **81**(18), 8123–8130.

Poulsen, C., Panjikar, S., Holton, S.J., Wilmanns, M., Song, Y.-H. (2014) WXG100 Protein Superfamily Consists of Three Subfamilies and Exhibits an α-Helical C-Terminal Conserved Residue Pattern. *PLOS ONE*. **9**(2), e89313.

Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., Heidelberg, J.F., Mekalanos, J.J. (2006) Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proceedings of the National Academy of Sciences*. **103**(5), 1528–1533.

Raymond, J.B., Mahapatra, S., Crick, D.C., Pavelka, M.S. (2005) Identification of the namH Gene, Encoding the Hydroxylase Responsible for the N-Glycolylation of the Mycobacterial Peptidoglycan *. *Journal of Biological Chemistry*. **280**(1), 326–333.

Rego, E.H., Audette, R.E., Rubin, E.J. (2017) Deletion of a mycobacterial divisome factor collapses single-cell phenotypic heterogeneity. *Nature*. **546**(7656), 153–157.

Renshaw, P.S., Lightbody, K.L., Veverka, V., Muskett, F.W., Kelly, G., Frenkiel, T.A., Gordon, S.V., Hewinson, R.G., Burke, B., Norman, J., Williamson, R.A., Carr, M.D. (2005) Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *The EMBO Journal*. **24**(14), 2491–2498.

Rombel, I.T., Sykes, K.F., Rayner, S., Johnston, S.A. (2002) ORF-FINDER: a vector for high-throughput gene identification. *Gene*. **282**(1), 33–41.

Ruhl, C.R., Pasko, B.L., Khan, H.S., Kindt, L.M., Stamm, C.E., Franco, L.H., Hsia, C.C., Zhou, M., Davis, C.R., Qin, T., Gautron, L., Burton, M.D., Mejia, G.L., Naik, D.K., Dussor, G., Price, T.J., Shiloh, M.U. (2020) Mycobacterium tuberculosis Sulfolipid-1 Activates Nociceptive Neurons and Induces Cough. *Cell*. **181**(2), 293-305.e11.

Sáenz, B., Hernandez-Pando, R., Fragoso, G., Bottasso, O., Cárdenas, G. (2013) The dual face of central nervous system tuberculosis: a new Janus Bifrons? *Tuberculosis (Edinburgh, Scotland)*. **93**(2), 130–135.

Saha, C.K., Sanches Pires, R., Brolin, H., Delannoy, M., Atkinson, G.C. (2021) FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics*. **37**(9), 1312–1314.

Sasiak, K., Saunders (1996) Purification and properties of a human nicotinamide ribonucleoside kinase. *Archives of biochemistry and biophysics*. **333**(2).

Sassetti, C.M., Rubin, E.J. (2003) Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences of the United States of America*. **100**(22), 12989–12994.

Seidel, M., Alderwick, L.J., Birch, H.L., Sahm, H., Eggeling, L., Besra, G.S. (2007) Identification of a novel arabinofuranosyltransferase AftB involved in a terminal step of cell wall arabinan biosynthesis in Corynebacterianeae, such as Corynebacterium glutamicum and Mycobacterium tuberculosis. *The Journal of Biological Chemistry*. **282**(20), 14729–14740.

Serafini, A., Boldrin, F., Palù, G., Manganelli, R. (2009) Characterization of a Mycobacterium tuberculosis ESX-3 Conditional Mutant: Essentiality and Rescue by Iron and Zinc. *Journal of Bacteriology*. **191**(20), 6340–6344.

Sethi, D., Mahajan, S., Singh, C., Lama, A., Hade, M.D., Gupta, P., Dikshit, K.L. (2016) Lipoprotein LprI of Mycobacterium tuberculosis Acts as a Lysozyme Inhibitor. *The Journal of Biological Chemistry*. **291**(6), 2938.

Sharma, M., Lingford, J.P., Petricevic, M., Snow, A.J.D., Zhang, Y., Järvå, M.A., Mui, J.W.-Y., Scott, N.E., Saunders, E.C., Mao, R., Epa, R., da Silva, B.M., Pires, D.E.V., Ascher, D.B., McConville, M.J., Davies, G.J., Williams, S.J., Goddard-Borger, E.D. (2022) Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria. *Proceedings of the National Academy of Sciences*. **119**(4), e2116022119.

Shen, L., Viljoen, A., Villaume, S., Joe, M., Halloum, I., Chêne, L.P., Méry, A., Fabre, E., Takegawa, K., Lowary, T.L., Vincent, S.P., Kremer, L., Guérardel, Y., Mariller, C. (2020) The endogenous galactofuranosidase GlfH1 hydrolyzes mycobacterial arabinogalactan. *The Journal of Biological Chemistry*.

Sher, J.W., Lim, H.C., Bernhardt, T.G. (2020) Global phenotypic profiling identifies a conserved actinobacterial cofactor for a bifunctional PBP-type cell wall synthase. *eLife*. **9**.

Shimokawa, M., Ishiwata, A., Kashima, T., Nakashima, C., Li, J., Fukushima, R., Sawai, N., Nakamori, M., Tanaka, Y., Kudo, A., Morikami, S., Iwanaga, N., Akai, G., Shimizu, N., Arakawa, T., Yamada, C., Kitahara, K., Tanaka, K., Ito, Y., Fushinobu, S., Fujita, K. (2023a) *Identification and characterization of endo- and exo-hydrolases cleaving the α- and β-D-arabinofuranosidic bonds of lipoarabinomannan and arabinogalactan of Mycobacteria*. In Review.

Shimokawa, M., Ishiwata, A., Kashima, T., Nakashima, C., Li, J., Fukushima, R., Sawai, N., Nakamori, M., Tanaka, Y., Kudo, A., Morikami, S., Iwanaga, N., Akai, G., Shimizu, N., Arakawa, T., Yamada, C., Kitahara, K., Tanaka, K., Ito, Y., Fushinobu, S., Fujita, K. (2023b) Identification and characterization of endo-α-, exo-α-, and exo-β-D-arabinofuranosidases degrading lipoarabinomannan and arabinogalactan of mycobacteria. *Nature Communications*. **14**(1), 5803.

Shukla, D., Trout, B.L. (2011) Understanding the Synergistic Effect of Arginine and Glutamic Acid Mixtures on Protein Solubility. *The Journal of Physical Chemistry B*. **115**(41), 11831–11839.

Silva, C.L., Ekizlerian, S.M., Fazioli, R.A. (1985) Role of cord factor in the modulation of infection caused by mycobacteria. *The American Journal of Pathology*. **118**(2), 238–247.

Simeone, R., Bobard, A., Lippmann, J., Bitter, W., Majlessi, L., Brosch, R., Enninga, J. (2012) Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death. *PLoS pathogens*. **8**(2), e1002507.

Spaans, S.K., Weusthuis, R.A., Oost, J. van der, Kengen, S.W.M. (2015) NADPH-generating systems in bacteria and archaea. *Frontiers in Microbiology*. **6**.

Spratt, B.G. (1975) Distinct penicillin binding proteins involved in the division, elongation, and shape of Escherichia coli K12. *Proceedings of the National Academy of Sciences of the United States of America*. **72**(8), 2999–3003.

Stahl, C., Kubetzko, S., Kaps, I., Seeber, S., Engelhardt, H., Niederweis, M. (2001) MspA provides the main hydrophilic pathway through the cell wall of Mycobacterium smegmatis. *Molecular Microbiology*. **40**(2), 451–464.

Stanley, S.A., Raghavan, S., Hwang, W.W., Cox, J.S. (2003) Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. *Proceedings of the National Academy of Sciences*. **100**(22), 13001–13006.

Stephan, J., Bender, J., Wolschendorf, F., Hoffmann, C., Roth, E., Mailänder, C., Engelhardt, H., Niederweis, M. (2005) The growth rate of Mycobacterium smegmatis depends on sufficient porin-mediated influx of nutrients. *Molecular Microbiology*. **58**(3), 714–730.

Sterling, T.R., Lehmann, H.P., Frieden, T.R. (2003) Impact of DOTS compared with DOTS-plus on multidrug resistant tuberculosis and tuberculosis deaths: decision analysis. *BMJ : British Medical Journal*. **326**(7389), 574.

Stewart, G.R., Patel, J., Robertson, B.D., Rae, A., Young, D.B. (2005) Mycobacterial Mutants with Defective Control of Phagosomal Acidification. *PLOS Pathogens*. **1**(3), e33.

Sugiki, T., Fujiwara, T., Kojima, C. (2017) Cold-Shock Expression System in E. coli for Protein NMR Studies. *Methods in Molecular Biology (Clifton, N.J.)*. **1586**, 345–357.

Sun, J., Siroy, A., Lokareddy, R.K., Speer, A., Doornbos, K.S., Cingolani, G., Niederweis, M. (2015) The Tuberculosis Necrotizing Toxin kills macrophages by hydrolyzing NAD. *Nature structural & molecular biology*. **22**(9), 672–678.

Sutherland, I.W. (1995) Polysaccharide lyases. *FEMS Microbiology Reviews*. **16**(4), 323–347.

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. (2012) Probing the Mycobacterial Trehalome with Bioorthogonal Chemistry. *Journal of the American Chemical Society*. **134**(39), 16123–16126.

Tahlan, K., Wilson, R., Kastrinsky, D.B., Arora, K., Nair, V., Fischer, E., Barnes, S.W., Walker, J.R., Alland, D., Barry, C.E., Boshoff, H.I. (2012) SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*. **56**(4), 1797–1809.

Tan, Y.Z., Rodrigues, J., Keener, J.E., Zheng, R.B., Brunton, R., Kloss, B., Giacometti, S.I., Rosário, A.L., Zhang, L., Niederweis, M., Clarke, O.B., Lowary, T.L., Marty, M.T., Archer, M., Potter, C.S., Carragher, B., Mancia, F. (2020) Cryo-EM structure of arabinosyltransferase EmbB from Mycobacterium smegmatis. *Nature Communications*. **11**(1), 3396.

Tanaka, Y., Tao, W., Blanchard, J.S., Hehre, E.J. (1994) Transition state structures for the hydrolysis of alpha-D-glucopyranosyl fluoride by retaining and inverting reactions of glycosylases. *Journal of Biological Chemistry*. **269**(51), 32306–32312.

Taylor, N.R., von Itzstein, M. (1994) Molecular Modeling Studies on Ligand Binding to Sialidase from Influenza Virus and the Mechanism of Catalysis. *Journal of Medicinal Chemistry*. **37**(5), 616–624.

TeSlaa, T., Ralser, M., Fan, J., Rabinowitz, J.D. (2023) The pentose phosphate pathway in health and disease. *Nature Metabolism*. **5**(8), 1275–1289.

Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., Nielsen, H. (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*. **40**(7), 1023–1025.

Tian, J., Bryk, R., Shi, S., Erdjument-Bromage, H., Tempst, P., Nathan, C. (2005) Mycobacterium tuberculosis appears to lack alpha-ketoglutarate dehydrogenase and encodes pyruvate dehydrogenase in widely separated genes. *Molecular Microbiology*. **57**(3), 859–868.

Tindall, B.J. (2019) On the status of the names Corynebacteriales Goodfellow and Jones 2015, Mycobacteriales Janke 1924 (Approved Lists 1980) and Mycobacteriales Cavalier-Smith 2002. *International Journal of Systematic and Evolutionary Microbiology*. **69**(10), 3310–3312.

Torrelles, J.B., Sieling, P.A., Zhang, N., Keen, M.A., McNeil, M.R., Belisle, J.T., Modlin, R.L., Brennan, P.J., Chatterjee, D. (2012) Isolation of a distinct Mycobacterium tuberculosis

mannose-capped lipoarabinomannan isoform responsible for recognition by CD1b-restricted T cells. *Glycobiology*. **22**(8), 1118–1127.

Touchette, M.H., Van Vlack, E.R., Bai, L., Kim, J., Cognetta, A.B., Previti, M.L., Backus, K.M., Martin, D.W., Cravatt, B.F., Seeliger, J.C. (2017) A Screen for Protein-Protein Interactions in Live Mycobacteria Reveals a Functional Link Between the Virulence-Associated Lipid Transporter LprG and the Mycolyltransferase Antigen 85A. *ACS infectious diseases*. **3**(5), 336–348.

Tran, H.-K.R., Grebenc, D.W., Klein, T.A., Whitney, J.C. (2021) Bacterial type VII secretion: An important player in host-microbe and microbe-microbe interactions. *Molecular Microbiology*. **115**(3), 478–489.

Trémolières, J. (1952) L'alimentation Du Tuberculeux: Le Problème De L'alimentation Carnée Du Tuberculeux. *Annales de la nutrition et de l'alimentation*. **6**(4/5), C375–C385.

U, T., J, V., A, G.-G., D, W., O, D., Lps, de C., Js, S., M, N. (2019) The tuberculosis necrotizing toxin is an NAD+ and NADP+ glycohydrolase with distinct enzymatic properties. *The Journal of biological chemistry*. **294**(9).

Ulhuq, F.R., Gomes, M.C., Duggan, G.M., Guo, M., Mendonca, C., Buchanan, G., Chalmers, J.D., Cao, Z., Kneuper, H., Murdoch, S., Thomson, S., Strahl, H., Trost, M., Mostowy, S., Palmer, T. (2020) A membrane-depolarizing toxin substrate of the Staphylococcus aureus type VII secretion system mediates intraspecies competition. *Proceedings of the National Academy of Sciences of the United States of America*. **117**(34), 20836–20847.

Ummels, R., Abdallah, A.M., Kuiper, V., Aâjoud, A., Sparrius, M., Naeem, R., Spaink, H.P., van Soolingen, D., Pain, A., Bitter, W. (2014) Identification of a Novel Conjugative Plasmid in Mycobacteria That Requires Both Type IV and Type VII Secretion. *mBio*. **5**(5), 10.1128/mbio.01744-14.

Velayati, A.A., Farnia, P., Mozafari, M., Malekshahian, D., Farahbod, A.M., Seif, S., Rahideh, S., Mirsaeidi, M. (2015) Identification and Genotyping of Mycobacterium tuberculosis Isolated From Water and Soil Samples of a Metropolitan City. *Chest*. **147**(4), 1094–1102.

Villeneuve, M., Kawai, M., Kanashima, H., Watanabe, M., Minnikin, D.E., Nakahara, H. (2005) Temperature dependence of the Langmuir monolayer packing of mycolic acids from Mycobacterium tuberculosis. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1715**(2), 71–80.

Vissa, V.D., Brennan, P.J. (2001) The genome of Mycobacterium leprae: a minimal mycobacterial gene set. *Genome Biology*. **2**(8), reviews1023.1.

Wallis, R., Ky, L., Aj, P., H, V., Gr, M., R, J., C, K. (1995) Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry*. **34**(42).

Wang, R., Ehrt, S. (2021) Rv0954 Is a Member of the Mycobacterial Cell Division Complex. *Frontiers in Microbiology*. **12**.

Wang, T., Hu, Z., Du, X., Shi, Y., Dang, J., Lee, M., Hesek, D., Mobashery, S., Wu, M., Liang, H. (2020) A type VI secretion system delivers a cell wall amidase to target bacterial competitors. *Molecular Microbiology*. **114**(2), 308–321.

Wang, Z., Zhang, M., Shi, X., Xiang, Q. (2017) Purification and Characterization of an ATPase GsiA from Salmonella enterica. *BioMed Research International*. **2017**, 3076091.

Wardman, J.F., Bains, R.K., Rahfeld, P., Withers, S.G. (2022) Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nature Reviews Microbiology*, 1–15.

Whitney, J.C., Chou, S., Russell, A.B., Biboy, J., Gardiner, T.E., Ferrin, M.A., Brittnacher, M., Vollmer, W., Mougous, J.D. (2013) Identification, structure, and function of a novel type VI secretion peptidoglycan glycoside hydrolase effector-immunity pair. *The Journal of Biological Chemistry*. **288**(37), 26616–26624.

Whitney, J.C., Peterson, S.B., Kim, J., Pazos, M., Verster, A.J., Radey, M.C., Kulasekara, H.D., Ching, M.Q., Bullen, N.P., Bryant, D., Goo, Y.A., Surette, M.G., Borenstein, E., Vollmer, W., Mougous, J.D. (2017) A broadly distributed toxin family mediates contact-dependent antagonism between gram-positive bacteria M. T. Laub, ed. *eLife*. **6**, e26938.

Whitney, J.C., Quentin, D., Sawai, S., LeRoux, M., Harding, B.N., Ledvina, H.E., Tran, B.Q., Robinson, H., Goo, Y.A., Goodlett, D.R., Raunser, S., Mougous, J.D. (2015) An interbacterial NAD(P)(+) glycohydrolase toxin requires elongation factor Tu for delivery to target cells. *Cell*. **163**(3), 607–619.

WHO (2022) Global tuberculosis report 2022. *https://www.who.int/publications/i/item/9789240061729*. [online]. Available from: https://www.who.int/publications-detail-redirect/9789240061729 [Accessed January 27, 2023].

Wiersma, C.J., Belardinelli, J.M., Avanzi, C., Angala, S.K., Everall, I., Angala, B., Kendall, E., de Moura, V.C.N., Verma, D., Benoit, J., Brown, K.P., Jones, V., Malcolm, K.C., Strong, M., Nick, J.A., Floto, R.A., Parkhill, J., Ordway, D.J., Davidson, R.M., McNeil, M.R., Jackson, M. (2020) Cell Surface Remodeling of Mycobacterium abscessus under Cystic Fibrosis Airway Growth Conditions. *ACS Infectious Diseases*. **6**(8), 2143–2154.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*. **67**(4), 235–242.

Wolucka, B.A. (2008) Biosynthesis of D-arabinose in mycobacteria - a novel bacterial pathway with implications for antimycobacterial therapy. *The FEBS journal*. **275**(11), 2691–2711.

World Health Organisation (2018) *Global tuberculosis report 2018*.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F.D., Lewis, J.D. (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science (New York, N.Y.)*. **334**(6052), 105–108.

Wu, X., Wu, Y., Zheng, R., Tang, F., Qin, L., Lai, D., Zhang, L., Chen, L., Yan, B., Yang, H., Wang, Y., Li, F., Zhang, J., Wang, F., Wang, L., Cao, Y., Ma, M., Liu, Z., Chen, Jianxia, Huang, X., Wang, J., Jin, R., Wang, P., Sun, Q., Sha, W., Lyu, L., Moura-Alves, P., Dorhoi, A., Pei, G., Zhang, P., Chen, Jiayu, Gao, S., Randow, F., Zeng, G., Chen, C., Ye, X.-S., Kaufmann, S.H.E., Liu, H., Ge, B. (2021) Sensing of mycobacterial arabinogalactan by galectin-9 exacerbates mycobacterial infection. *EMBO reports*. **n/a**(n/a), e51678.

van Wyk, N., Drancourt, M., Henrissat, B., Kremer, L. (2017) Current perspectives on the families of glycoside hydrolases of Mycobacterium tuberculosis: their importance and prospects for assigning function to unknowns. *Glycobiology*. **27**(2), 112–122.

Xie, L., Liu, W., Li, Q., Chen, S., Xu, M., Huang, Q., Zeng, J., Zhou, M., Xie, J. (2015) First Succinyl-Proteome Profiling of Extensively Drug-Resistant Mycobacterium tuberculosis Revealed Involvement of Succinylation in Cellular Physiology. *Journal of Proteome Research*. **14**(1), 107–119.

Xin, Y., Huang, Y., McNeil, M.R. (1999) The presence of an endogenous endo-D-arabinase in Mycobacterium smegmatis and characterization of its oligoarabinoside product. *Biochimica Et Biophysica Acta*. **1473**(2–3), 267–271.

Xu, J., Zhang, H., Zheng, J., Dovoedo, P., Yin, Y. (2020) eCAMI: simultaneous classification and motif identification for enzyme annotation. *Bioinformatics*. **36**(7), 2068–2075.

Yamamoto, T., Nishiura, M., Harada, N., Imaeda, T. (1958) Electron microscopy of ultra-thin sections of lepra cells and Mycobacterium leprae. *International Journal of Leprosy*. **26**(1), 1–8.

Yeh, J.I., Chinte, U., Du, S. (2008) Structure of glycerol-3-phosphate dehydrogenase, an essential monotopic membrane enzyme involved in respiration and metabolism. *Proceedings of the National Academy of Sciences of the United States of America*. **105**(9), 3280–3285.

Yeh, M.F., Trela, J.M. (1976) Purification and characterization of a repressible alkaline phosphatase from Thermus aquaticus. *Journal of Biological Chemistry*. **251**(10), 3134–3139.

Yocum, R.R., Rasmussen, J.R., Strominger, J.L. (1980) The mechanism of action of penicillin. Penicillin acylates the active site of Bacillus stearothermophilus D-alanine carboxypeptidase. *The Journal of Biological Chemistry*. **255**(9), 3977–3986.

Youmans, G.P., Youmans, A.S. (1953) Studies on the metabolism of Mycobacterium tuberculosis. I. The effect of carbohydrates and alcohols on the growth of Mycobacterium tuberculosis var. hominis. *Journal of Bacteriology*. **65**(1), 92–95.

Zhang, B., Li, J., Yang, Xiaolin, Wu, L., Zhang, J., Yang, Y., Zhao, Y., Zhang, L., Yang, Xiuna, Yang, Xiaobao, Cheng, X., Liu, Z., Jiang, B., Jiang, H., Guddat, L.W., Yang, H., Rao, Z. (2019) Crystal Structures of Membrane Transporter MmpL3, an Anti-TB Drug Target. *Cell*. **176**(3), 636-648.e13.

Zhang, Jian, Angala, S.K., Pramanik, P.K., Li, K., Crick, D.C., Liav, A., Jozwiak, A., Swiezewska, E., Jackson, M., Chatterjee, D. (2011) Reconstitution of functional mycobacterial arabinosyltransferase AftC proteoliposome and assessment of decaprenylphosphorylarabinose analogues as arabinofuranosyl donors. *ACS chemical biology*. **6**(8), 819–828.

Zhang, L., Zhao, Y., Gao, Y., Wu, L., Gao, R., Zhang, Q., Wang, Y., Wu, C., Wu, F., Gurcha, S.S., Veerapen, N., Batt, S.M., Zhao, W., Qin, L., Yang, X., Wang, M., Zhu, Y., Zhang, B., Bi, L., Zhang, X., Yang, H., Guddat, L.W., Xu, W., Wang, Q., Li, J., Besra, G.S., Rao, Z. (2020) Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol. *Science*.

Zhang, Ying, Zagnitko, O., Rodionova, I., Osterman, A., Godzik, A. (2011) The FGGY Carbohydrate Kinase Family: Insights into the Evolution of Functional Specificities. *PLOS Computational Biology*. **7**(12), e1002318.

Zhang, Z., Bulloch, E.M.M., Bunker, R.D., Baker, E.N., Squire, C.J. (2009) Structure and function of GlmU from Mycobacterium tuberculosis. *Acta Crystallographica Section D: Biological Crystallography*. **65**(3), 275–283.

Zheng, S., Sham, L.-T., Rubino, F.A., Brock, K.P., Robins, W.P., Mekalanos, J.J., Marks, D.S., Bernhardt, T.G., Kruse, A.C. (2018) Structure and mutagenic analysis of the lipid II flippase MurJ from Escherichia coli. *Proceedings of the National Academy of Sciences*. **115**(26), 6709–6714.

Zhong, W., Guo, J., Cui, L., Chionh, Y.H., Li, K., El Sahili, A., Cai, Q., Yuan, M., Michels, P.A.M., Fothergill-Gilmore, L.A., Walkinshaw, M.D., Mu, Y., Lescar, J., Dedon, P.C. (2019) Pyruvate Kinase Regulates the Pentose-Phosphate Pathway in Response to Hypoxia in Mycobacterium tuberculosis. *Journal of Molecular Biology*. **431**(19), 3690–3705.

Zimhony, O., Cox, J.S., Welch, J.T., Vilchèze, C., Jacobs, W.R. (2000) Pyrazinamide inhibits the eukaryotic-like fatty acid synthetase I (FASI) of Mycobacterium tuberculosis. *Nature Medicine*. **6**(9), 1043–1047.