Protein O-glycosylation

in *Streptomyces*

Tessa Anne Keenan

Doctor of Philosophy

University of York

Biology

November 2016

Abstract

Previously, a protein O-mannosyl transferase (Pmt, SCO3154) and a polyprenol phosphate mannose synthase (Ppm1, SCO1423) were found to be required for the glycosylation of PstS, a phosphate binding protein, in the bacterium *Streptomyces coelicolor.* Bacteria in this genus are prolific producers of antibiotics and are often phenotypically resistant to multiple antibiotics. *S. coelicolor pmt* and *ppm1* deficient mutants were hypersusceptible to cell-wall active antibiotics, suggesting that the protein modification could be required for cell wall and membrane homeostasis. The aim of this project was to investigate the *S. coelicolor* glycoproteome in order to better understand the physiological role of protein O-glycosylation in this model actinobacterium. Glycoproteins were detected in, and enriched from the membrane and culture filtrate proteomes of the *S. coelicolor* parent strain, J1929 and these were absent from the glycosylation deficient *pmt* (DT1025) and *ppm1* (DT3017) mutants. Liquid chromatography coupled to mass spectrometry was used to characterise the membrane glycoproteome from the *S. coelicolor* parent strain, J1929 and 37 new glycoproteins were identified. Glycopeptides were modified on Ser/Thr residues with up to 3 hexoses; consistent with previous observations that the glycoprotein PstS is modified with a trihexose. The *S. coelicolor* glycoprotein glycans were shown to consist of Hex_2 and Hex_3 oligosaccharides. A carbohydrate linkage analysis led to the observation of 2-substituted, 4-substituted and terminal mannose residues, suggesting presence of (1->2) and (1->4) linkages in *S. coelicolor* glycoprotein glycans. The *S. coelicolor* glycoproteome comprises glycoproteins with various biological roles including solute binding, transport and cell wall biosynthesis. The genes encoding two *S. coelicolor* glycoproteins with putative roles in cell wall biosynthesis, an L, D transpeptidase (SCO4934) and a D-Ala-D-Ala, carboxypeptidase (SCO4847) were disrupted. Both mutants were hypersusceptible to β-lactam antibiotics, while the *sco4847* mutant was hypersusceptible to lysozyme. These findings suggest that both proteins could be required for cell wall biosynthesis. As the phenotypes of the knockout mutants are reminiscent of the glycosylation deficient strains, we propose that glycosylation might be required for enzyme function.

List of Contents

List of Figures

List of Tables

List of Accompanying Material

Appendix A.7 The best matching spectra for each glycopeptide identification – data included on a Flash drive.

Acknowledgements

Firstly, I would like to thank my supervisor Maggie Smith for her continued support and guidance throughout this project. I always left from our discussions inspired with new ideas or solutions to problems I was having in the laboratory, and for that I am really grateful. I am also very grateful for the extra funding that Maggie found for me at the end of my three years, to continue with some of the unanswered questions in my project.

A special thanks to Robert Howlett and Nicholas Read for your continued help and advice in the laboratory, especially with regard to the *Streptomyces* work. I would also like to thank my TAP members, Gavin Thomas and James Moir for their continued support throughout the project.

I would like to thank the Proteomics Laboratory of the York Bioscience Technology Facility for your help with the mass spectrometry analyses. Without the Bruker maXis HD LC-MS/MS and the Orbitrap Fusion instruments this project would not have been nearly as successful as it was. I would specifically like to thank Adam Dowle for his invaluable help and support with the glycoproteomics analyses, as well as answering all of my silly questions. I would also like to thank Jerry Thomas and Rachel Bates for collaborating with me in the Glycomics analyses and for your continuous support in the analysis of the data.

I would like to thank Anne Dell for advising on the proteomics analyses carried out in this project, as well as for allowing me to visit your laboratory at Imperial College to carry out some mass spectrometry analyses. Thank you also to Simon North for helping me with my sample analyses during my week at Imperial College.

Thank you to Jessica Lorraine and Steph Evans for the coffee time chat and for putting up with my complaints about difficulties in the lab. I would also likely to thank the other past and present members of the Smith and McGlynn laboratory groups for your help, advice and entertainment; both in and out of the lab.

Declaration

I hereby declare that this thesis is an original report of my research, has been written by me and has not been submitted for any previous degree at this or any other university. The experimental work described herein was performed by myself, except where otherwise indicated. The collaborative contributions have been indicated clearly and acknowledged. Due references have been provided on all supporting literatures and resources.

Chapter 1 - Introduction

Chapter 1 – Introduction

1.1 Protein glycosylation is present in all domains of life.

Protein glycosylation, which is defined as the co- or posttranslational modification of a protein by the addition of a glycan, is one of the most complex and diverse protein modifications in nature (Spiro 1973). A variety of carbohydrates, as well as a range of anomeric configurations have been identified in glycoproteins in all domains of life (Spiro 2002). Protein glycosylation is considered to be one of the most ubiquitous protein modifications in eukaryotes, with at least two thirds of all eukaryotic proteins estimated to be glycosylated (Dell et al. 2011). Most commonly, protein glycosylation occurs through the attachment of glycans to polypeptide chains through amide linkages to asparagine (Asn) residues (*N*-glycosylation), glycosidic linkages to serine(Ser)/threonine(Thr) residues (O-glycosylation) or C-C linkages to tryptophan (Trp) residues (C-mannosylation). *N*-glycosylation pathways have been characterised in all domains of life and despite some variations between species, involve the synthesis of the lipid linked oligosaccharide (LLO) and the *en bloc* transfer of the oligosaccharide onto acceptor polypeptides; a process mediated by an oligosaccharyltransferase (OST) (Dell et al. 2011). In contrast, protein O-glycosylation pathways tend to be processive, involving the sequential addition of monosaccharides to acceptor peptides. However, OST mediated O-glycosylation of bacterial pilins has been identified (discussed in 1.3.2). This introduction will focus on protein O-glycosylation pathways; mainly protein O-mannosylation and the role of the pathway in maintaining cell wall homeostasis.

1.2 Protein O-glycosylation in eukaryotes.

1.2.1 Mammalian protein O-glycosylation

The most well studied examples of mammalian glycoproteins are mucins. Mucins are high molecular mass proteins that are characteristically heavily glycosylated in tandem repeats of Ser/Thr/Pro residues, found as cell surface exposed glycoproteins and in mucous secretions (Varki et al. 2009). Encoded by the MUC genes, mucins have roles in the lubrication of epithelial cells, serving as cell receptors, in fertilisation and in the immune response. In

humans, many diseases have been associated with abnormal mucin glycosylation. For example, the mucin polypeptide encoded by the MUC1 gene in humans is abnormally glycosylated in breast and other carcinomas (Brockhausen et al. 1995). Aberrant O-glycosylation of cancer cells has been implicated in their attachment and invasion, as well as in their ability to survive in the blood stream (Brockhausen 1999).

In contrast to the *en bloc N*-glycosylation of eukaryotic proteins in the endoplasmic reticulum (ER), O-glycosylation occurs mostly in the Golgi apparatus and is thought to involve the sequential addition of glycans to amino acids with a hydroxyl group (Ser, Thr, Tyr, Hyp [hydroxyproline] and Hyl [hydroxylysine]) (Spiro 2002; Varki et al. 2009). Mucins can be O-glycosylated with a variety of monosaccharides including N-acetylgalactosamine (GalNAc), fucose (Fuc), mannose (Man), glucose (Glc), galactose (Gal), arabinose (Ara), xylose (Xyl), and N-acetylglucosamine (GlcNAc), in both α and β configurations. The first step in the glycosylation of mucin polypeptides is the covalent α-linkage of an O-GalNac moiety to the hydroxyl group of Ser/Thr, carried out by a polypeptide N-acetyl-galactosaminyltransferase (ppGalNAcT) (Varki et al. 2009). The sequential addition of other monosaccharides can result in the synthesis of a range of mucin structures (Figure 1.1). In contrast to the assembly of an oligosaccharide on a lipid linked carrier in *N*-glycosylation pathways, no lipid linked carrier is required in the synthesis of O-linked glycoproteins. Additionally, further processing of the O-GalNac glycans by glycosidases in the Golgi, as is observed in the processing of N-linked glycans, does not occur.

In addition to the O-glycosylation of secreted and membrane bound proteins in eukaryotes, the modification of nuclear and cytoplasmic proteins with simple O-GlcNac residues has been reported (Torres and Hart 1984; Wells et al. 2001; Hart 1997). Despite no consensus sequence having been identified, single O-GlcNac residues have been shown to modify Ser/Thr residues in cytoplasmic and nuclear proteins. O-GlcNAcylation is thought to have a role in the cellular signalling and regulation, T lymphocyte activation and the protection of proteins against cellular degradation (Kearse and Hart 1991; Han and Kudlow 1997; Swain et al. 2002).

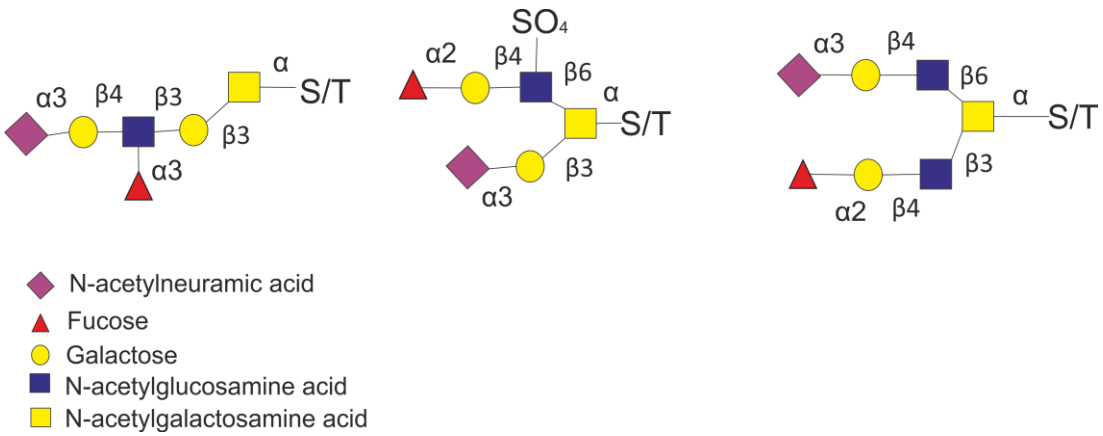


Figure 1.1 Representative examples of O-GalNAc glycans from human respiratory mucins composed of different cores, which can be further extended and branched. Image adapted from Varki et al. (2009).

1.2.2 Protein O-mannosylation in yeast.

Protein O-mannosylation was discovered in the cell-wall isolated proteins of *S. cerevisiae* and was long thought to be restricted to fungi (Sentandreu and Northcote 1968). While it is now clear that the modification is present in many eukaryotes and has also been identified in prokaryotes (discussed in section 1.3.3 and 1.5.2), many of the fundamental characteristics of the pathway have been elucidated in yeast (Loibl and Strahl 2013). Protein O-mannosylation involves the sequential attachment of mannose (Man) to the hydroxyl of Ser/Thr residues in polypeptide chains, via a glycosidic linkage (Spiro 2002). O-mannosyl glycan synthesis is initiated in ER and is catalysed by a conserved family of glycosytransferases, PMTs (Haselbeck and Tanner 1983). The protein O-mannosylation pathway involves two main steps: the synthesis of the mannosyl donor and the transfer of mannose onto target polypeptide chains (Figure 1.2).

The eukaryotic mannosyl donor, dolichol phosphate β-D-mannose (Dol-P-Man) is synthesised on the cytoplasmic face of the ER membrane where a glycosyltransferase GDP-α-D-Man:Dol-P ß-D-mannosyltransferase (Dpm1) catalyses the transfer of mannose from GDP-α-D-mannose (GDP-Man) to Dol-P (Lommel and Strahl 2009). The baker's yeast Dpm1 (Dpm1p) was first identified by Haselbeck (1989) and its catalytic activity demonstrated after the heterologous expression of the enzyme in *E. coli* (Orlean et al. 1988). Dol-P-Man serves as the primary mannosyl donor for the protein O-mannosylation pathway in eukaryotes, however it is also required for the extension of the LLO in the ER in the protein *N*-glycosylation pathway (Figure 1.2), as well as in GPI anchor biosynthesis (Kornfeld and Kornfeld 1985; Doering et al. 1990). Dol-P-Man was shown to be essential in yeast after *dpm1* knockout mutants were found to be lethal (Orlean 1990). This is not surprising considering the general requirement of the Dol-P-Man glycosyl donor for the synthesis of mannans in yeast. DPM1 homologues have been identified in other fungi, as well as in many other higher eukaryotes, including mice and humans (Zimmerman et al. 1996; Colussi et al. 1997; Tomita et al. 1998; Maeda and Kinoshita 2008). In humans, defects in *DPM1* leading to changes in protein O-glycosylation patterns, are known to cause metabolic disorders that often result in mental and psychomotor retardation, termed congenital disorders of glycosylation (CDGs).

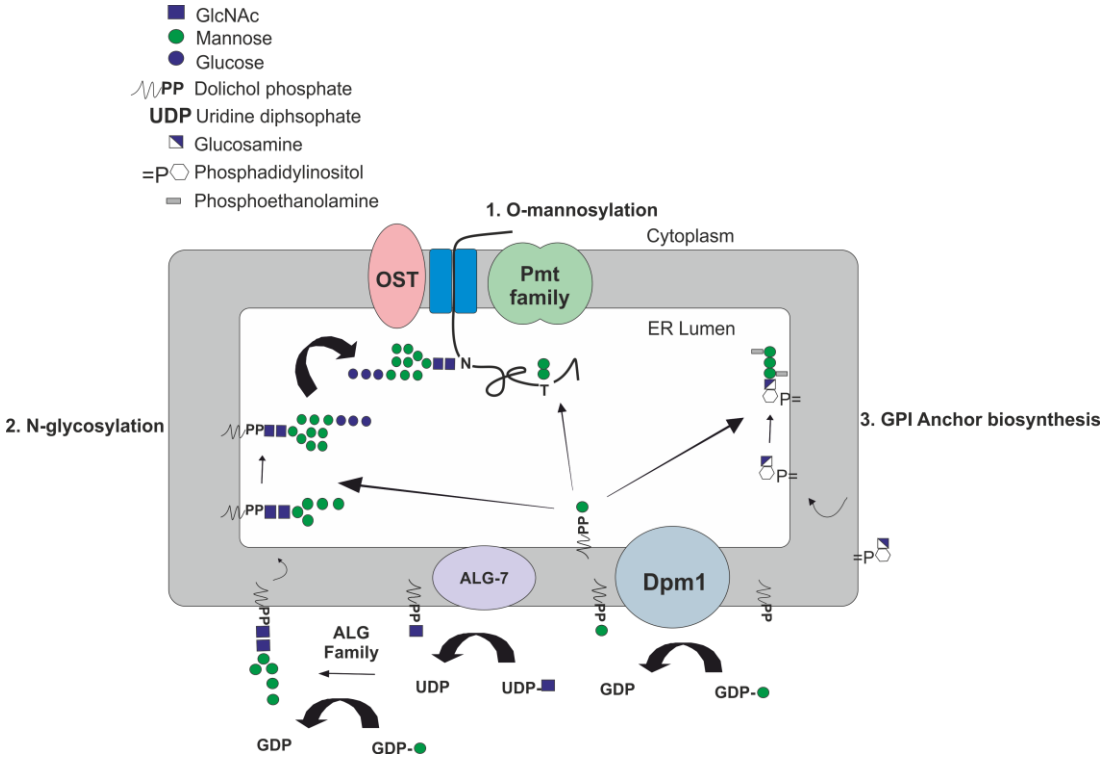


Figure 1.2 The protein O-mannosylation pathway in yeast shares dolichol phosphate-mannose with the *N*-glycosylation pathway and GPI anchor biosynthesis. Images adapted from Loibl and Strahl (2013).

Following the synthesis of Dol-P-Man by Dpm1, the Dol-P-Man is flipped into the ER lumen, most likely via the action of a flippase (Lommel and Strahl 2009). A second glycosyltransferase, Dol-P-Man:protein O-mannosyltransferases (Pmt) then catalyses the transfer of Man from Dol-P-Man onto Ser/Thr residues of target proteins (Figure 1.2). Early work on protein O-mannosylation in yeast led to the suggestion that protein O-mannosylation was coupled to translocation into the ER (Elorza et al. 1977). However, some studies have reported Pmt mediated O-mannosylation of proteins after translocation into the ER when they are misfolded, suggesting that the modification increases protein solubility and reduces the need for chaperones (Harty et al. 2001; Nakatsukasa et al. 2004).

After the transfer of mannose onto target proteins in the ER, the glycoproteins can be further modified in the Golgi apparatus. In *S. cerevisiae* for example, extension of the glycan is carried out by mannosyltransferases of the KTR- and MNN1-family that can generate α(1,2) and α(1,3) linkages (Lussier et al. 1999).

Pmts belong to the GT-C family of glycosyltransferases, that are integral membrane proteins with between 7 and 13 transmembrane helices (Lairson et al. 2008; Lommel and Strahl 2009). *S. cerevisiae* has at least six PMT family members (Pmt1p-Pmt6p) (Strahl-Bolsinger and Tanner 1991; Strahl-Bolsinger et al. 1993; Gentzsch and Tanner 1996). Strahl-Bolsinger and Scheinost (1999) proposed that *S. cerevisiae* Pmt1p consisted of seven transmembrane helices and that the N-terminus was localised in the cytoplasm, while the C-terminus was localised in the ER lumen (Figure 1.3). Further characterisation of Pmt1p demonstrated that the N-terminus interacts with Pmtp2 and that the central hydrophilic loop (loop 5) is required for mannosyltransferase activity of the enzyme (Girrbach et al. 2000). Additionally Girrbach et al. (2000) demonstrated that the amino acids Arg-64, Glu-78, Arg-138, and Leu-408 are required for Pmt1p activity (Figure 1.3).

The disruption of *PMT1* alone in *S. cerevisiae* demonstrated that there could be functional redundancy in Pmt activity and that PMT1 alone was not essential for growth (Strahl-Bolsinger et al. 1993). However, knockout mutants in certain combinations of three *PMTs* in *S. cerevisiae* were lethal, suggesting that the modification is essential in yeast (Gentzsch and Tanner 1996). Additionally, using several combinations of double *pmt* mutants it was shown that protein O-mannosylation was required to maintain cell wall rigidity and integrity. Similarly, in *S. pombe* double mutants in *PMT1* and *PMT4,* as well as a single knockout of *PMT2* were lethal (Willer et al. 2005). Additionally, *pmt1* and *pmt4* single mutants displayed abnormal cell wall and septum formation, further suggesting that protein O-mannosylation

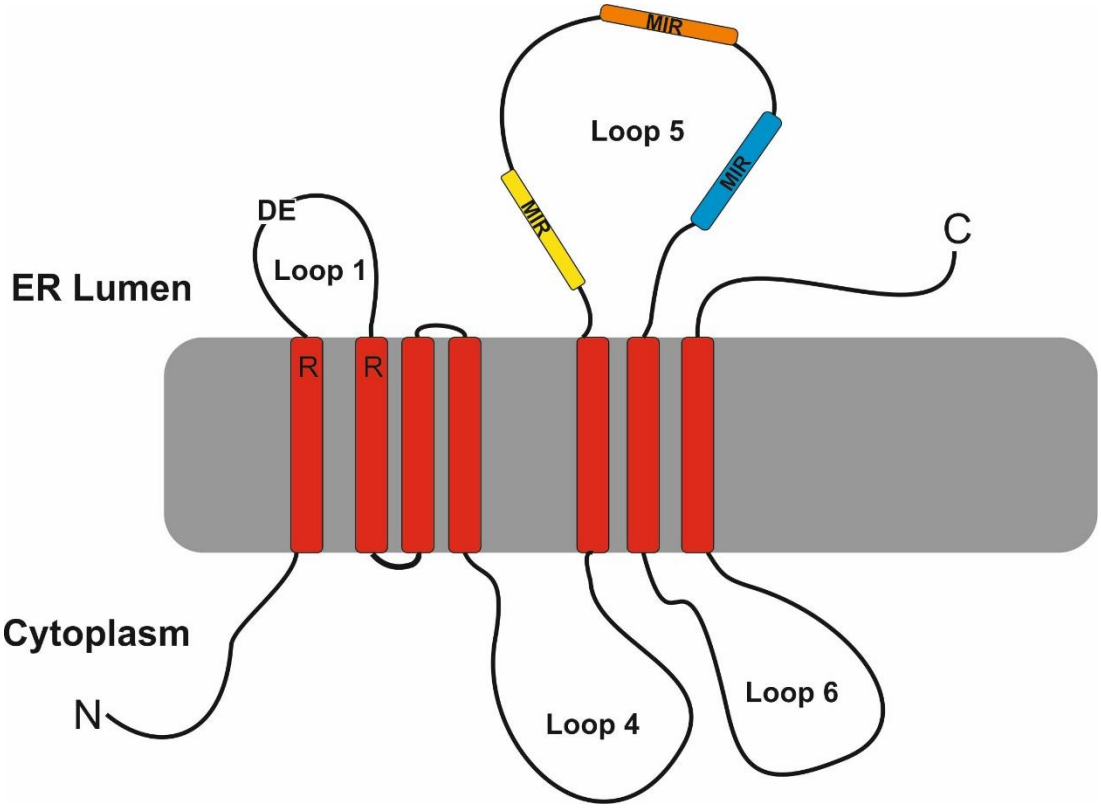


Figure 1.3 A schematic representation of *S. cerevisiae* Pmt1p, based on the work carried out by Girrbach et al. (2000). Pmt1p has seven transmembrane helices. Loop 5 was shown to be required for catalytic activity of the enzyme and three conserved motifs (MIR) were identified in as signatures of the PMT family proteins. The positions of the arginine (R), glutamic acid (D) and aspartic acid (E) residues required for activity enzyme activity are shown.

in yeast is required for stable cell wall formation. Since the identification of Pmts in yeast, PMT family proteins (termed POMTs) have been found in many higher eukaryotes (Lommel and Strahl 2009). The best studied O-mannosylated protein in higher eukaryotes is α-dystroglycan (α-DG), a fundamental component of the dystrophin–glycoprotein complex (DGC) in skeletal muscle (Barresi and Campbell 2006). Several studies in flies have demonstrated that POMTs were required for correct muscle development, while a POMT1 deficiency in mice was found to be lethal to embryonic development (Martín-Blanco and García-Bellido 1996; Ichimiya et al. 2004; Willer et al. 2004). In humans, the disruption of POMTs have been linked to changes in α-DG glycosylation, leading to congenital disorders of glycosylation (Muntoni and Voit 2004).

1.3 Protein glycosylation in prokaryotes

Prior to the mid-1970s, protein glycosylation was considered to occur exclusively in eukaryotes. The discovery of the first prokaryotic glycoprotein in the halophile *Halobacterium salinarium* (*H. salinarium*) by Mescher and Strominger (1976) challenged this idea. It is now clear that prokaryotes can modify proteins with both N- and O-linked glycans, and that these pathways have roles in pathogenesis, host invasion, maintaining cell wall integrity and survival in extreme environments (Nothaft and Szymanski 2010; Eichler 2013).

1.3.1 Protein *N*-glycosylation in bacteria

The first protein *N*-glycosylation pathway in bacteria was discovered in *Campylobacter jejuni* (*C. jejuni*) more than a decade ago and is arguably the most well characterised bacterial glycosylation system to date (Szymanski et al. 1999). Protein *N*-glycosylation takes place on the cytoplasmic face of the plasma membrane and is encoded by the *pgl* gene cluster (Szymanski et al. 1999; Linton et al. 2005). The *pgl* genes include five glycosyltransferases (*pglH, pglJ, pglI, pglA, pglC*), an oligosaccharyltransferase (*pglB*), a (UDP)-N-acetylglucosamine: glucosamine 4-epimerase (*gne*), a LLO flippase (*pglK*) and genes involved in the synthesis of the rare amino sugar UDP-2,4-diacetamido bacillosamine (UDP-dinAcBac) (*pglD, pglE, pglF*) (Nothaft and Szymanski 2010). The pathway starts with the sequential assembly of a heptasaccharide on a lipid linked carrier, undecaprenyl pyrophosphate (Und-PP) on the cytoplasmic face of the inner membrane (Linton et al. 2005). The lipid linked

heptasaccharide is translocated across the inner membrane into the periplasm, mediated by an ABC-type transporter with ATPase activity (PglK) (Alaimo et al. 2006). The heptasaccharide is transferred *en bloc* onto specific Asn residues on target proteins; a process mediated by PglB, a bacterial OST (Wacker et al. 2002). Bacteria have an *N*-glycosylation consensus sequence that is extended at the N-terminus, where a negatively charged amino acid (aspartic acid (D)/glutamic acid (E)) is present at the -2 position relative to the glycosylated Asn (N) residue (D/E-X-N-X-S/T, where X can be any amino acid except for proline) (Kowarik et al. 2006). Despite the full characterisation of the pathway, little is known about the role of the glycan. Glycosylation has been shown to be required for attachment to human epithelial cells, as well as for the colonisation in chickens (Jones et al. 2004; Karlyshev et al. 2004). One study found that N-linked glycans in *C. jejuni* bind to human macrophage galactose like-lectins and suggested that *N*-glycosylation may serve to modulate the host immune response (Van Sorge et al. 2009). Recently however, work carried out by Alemka et al. (2013) demonstrated that *N*-glycosylation of cell surface proteins in *C. jejuni* may be required for protection against proteolytic cleavage by proteinases in the chicken gut.

1.3.2 Pilin and flagellar O-glycosylation

Flagellar O-glycosylation has been widely reported in Gram negative bacteria. The modification in Gram positive genera is so far limited to *Clostridium* and *Listeria spp* (Schirm et al. 2004; Twine et al. 2008). While no consensus sequence has been identified, the modification occurs on Ser/Thr residues (Logan 2006). The O-glycosylation of flagella in *C. jejuni* is probably the best described example, where the flagellin A (FlaA) and flagellin B (FlaB) subunits are modified with pseudaminic acid (Pse) and legionaminic (Leg) acid derivatives, at up to 19 different sites (Thibault et al. 2001; Zebian et al. 2016). The genes required for flagellar glycosylation in *C. jejuni* are known, and the pathway for Leg biosynthesis has been reconstituted in *E. coli* (Schoenhofen et al. 2009). Glycosylation of flagella in *Campylobacter* spp. is required for flagellar assembly and motility, and mutants defective in Pse biosynthesis have been shown to lack flagella and were consequently immobile (Goon et al. 2003). Similar observations have been made in the human pathogen *Helicobacter pylori* (*H. pylori*), that glycosylates flagella in a similar way to *C. jejuni* (Schirm et al. 2003). However, in *H. pylori* the glycans display less heterogeneity than in *C. jejuni*.

Pilin glycosylation has been observed in some bacteria, and has been suggested to be involved in virulence. In *Pseudomonas aeruginosa* for example, the loss of pilin glycosylation

due to a deletion in the *pilO* gene led to reduced persistence in the lungs of mice (Smedley et al. 2005). Structural characterisation of the pilin glycans has demonstrated that they are composed of trisaccharides that are identical to the lipopolysaccharide O-antigen repeating unit in the strain (Castric et al. 2001). In contrast, the pilin glycans of *Neisseria meningitidis* (*N. meningitidis*) are modified with trisaccharides composed of Gal(β1-4)Gal(α1-3)2,4-diacetamido-2,4,6-trideoxyhexose (Stimson et al. 1995).

In contrast, the pilin glycans of *Neisseria gonorrhoeae* (*N. gonorrhoeae*) are thought to be disaccharides composed of an O-acetylated hexose residue linked to a 2,4-diacetamido-2,4,6-trideoxyhexose (HexDATDH) (Hegge et al. 2004; Aas et al. 2007). Several studies have identified genes that encode proteins required for pilin glycosylation (*pgl*) in *N. meningitides* (Jennings et al. 1998; Power et al. 2000). These include an O-oligosaccharyltransferase (O-OTase) PglL, which attaches the O-linked glycan to pili and was found to be part of a general O-glycosylation pathway in *Neisseria* spp. (Power et al. 2006; Ku et al. 2009). The general O-glycosylation pathway in *N. gonorrhoeae* has been shown glycosylate at least 11 membrane proteins and lipoproteins including solute binding proteins of ABC transporters and protein chaperones (Vik et al. 2009). The proteins were shown to be modified on Ser/Thr residues with O-acetylated HexDATDH glycans, in amino acid sequence regions abundant in proline, alanine and serine. Similarly in *N. meningitides,* the general O-glycosylation pathway responsible for pilin glycosylation has also been shown to glycosylate the surface exposed glycoprotein, AniA (Ku et al. 2009).

1.3.3 Protein O-mannosylation in mycobacteria

While protein O-mannosylation has been described in several actinobacteria, including *Corynebacterium* spp. and *Streptomyces* spp. (described in section 1.5.2), it is particularly well studied in mycobacteria (Lommel and Strahl 2009). The earliest reports of mannosylated proteins in mycobacteria were described after the binding of culture filtrate proteins to the mannose binding lectin concanavalin A (Con A) (Espitia and Mancilla 1989; Fifis et al. 1991). The first direct evidence for glycosylated residues on a mycobacterial protein was presented by Dobos et al. (1995), who demonstrated that a 45/47 kDa culture filtrate protein from *Mycobacterium tuberculosis* (*M. tuberculosis*) was modified with hexoses. Further characterisation of the 45/47 kDa glycoprotein revealed that the modifying hexose was mannose and that the glycopeptides were modified on Thr residues with α-D-Man, mannobiose, or mannotriose (Dobos et al. 1996). Additionally, the glycosylation sites were

found to be localised near the N- and C- termini of the glycoprotein, in proline rich sequences. A carbohydrate linkage analysis revealed that the mannobiose and mannotriose glycans were composed of α1,2 linkages. In contrast, the characterisation of MPB83, a secreted glycoprotein from *Mycobacterium bovis* revealed that the glycoprotein was modified on threonine residues with mannose and α1,3-linked mannobiose (Michell et al. 2003). In *Mycobacterium avium*, the 32 kDa glycoprotein SmT was shown to be glycosylated with a dihexose glycan (Taylor et al. 2006). A recent study of the *M. tuberculosis* culture filtrate glycoproteome revealed that glycosylation could occur on both serine and threonine residues (Smith et al. 2014). Smith et al. (2014) revealed that glycosylation sites were often localised near the C-terminus of the glycoprotein, and that sequences surrounding the glycosylation sites had a higher propensity for hydrophobic amino acids (Pro, Ala).

The protein O-glycosylation pathway in *M. tuberculosis* is distinctly similar to the protein O-mannosylation pathway in *S. cerevisiae* (Figure 1.4.A). Mannose is thought to be transferred from GDP-Man onto polyprenol phosphate, a functional analogue of dolichol phosphate by a polyprenol phosphate mannose (Ppm) synthase. In *M. tuberculosis*, the catalytic activity of the Ppm synthase *Mt*-Ppm1 was shown in a cell free assay, where [^{14}C]Man was transferred from GDP-Man onto various lipid monophosphate acceptors (Gurcha et al. 2002). Mt-Ppm1 has a two domain architecture (D1 and D2), where D2 is responsible for the catalytic activity of the enzyme and D1 contained several transmembrane domains (Figure 1.4.B). The Mt-Ppm1 homologues in other mycobacteria, such as *Mycobacterium smegmatis (M. smegmatis)* and *Mycobacterium. leprae* (*M. leprae*), exist as two separate proteins, Ppm1 and Ppm2 (Gurcha et al. 2002). In *M. smegmatis*, Ms-Ppm2 was shown to be an integral membrane protein that interacts with, and enhances the catalytic activity of Ms-Ppm1 (Baulard et al. 2003). These findings suggested that the two-domain architecture of Mt-Ppm1 had a role in anchoring the protein to the membrane.

Using a bioinformatics approach, a PMT homologue (*Rv1002c*) was identified in *M. tuberculosis* and shown to catalyse the first step of protein O-glycosylation of proteins that were translocated via the SEC pathway (VanderVen et al. 2005). An additional glycosyltransferase, a Ppm-dependent α1,2-mannosyltransferase (PimE) that was identified in *M. smegmatis,* is thought to be required for the elongation of the glycan chain (Morita et al. 2006; Liu et al. 2013a). Since α1,2-linked mannobiose and mannotriose glycans have been observed in *M. tuberculosis* glycoproteins, it is likely that a similar mechanism is present in other *Mycobacterium spp*. (Dobos et al. 1996).

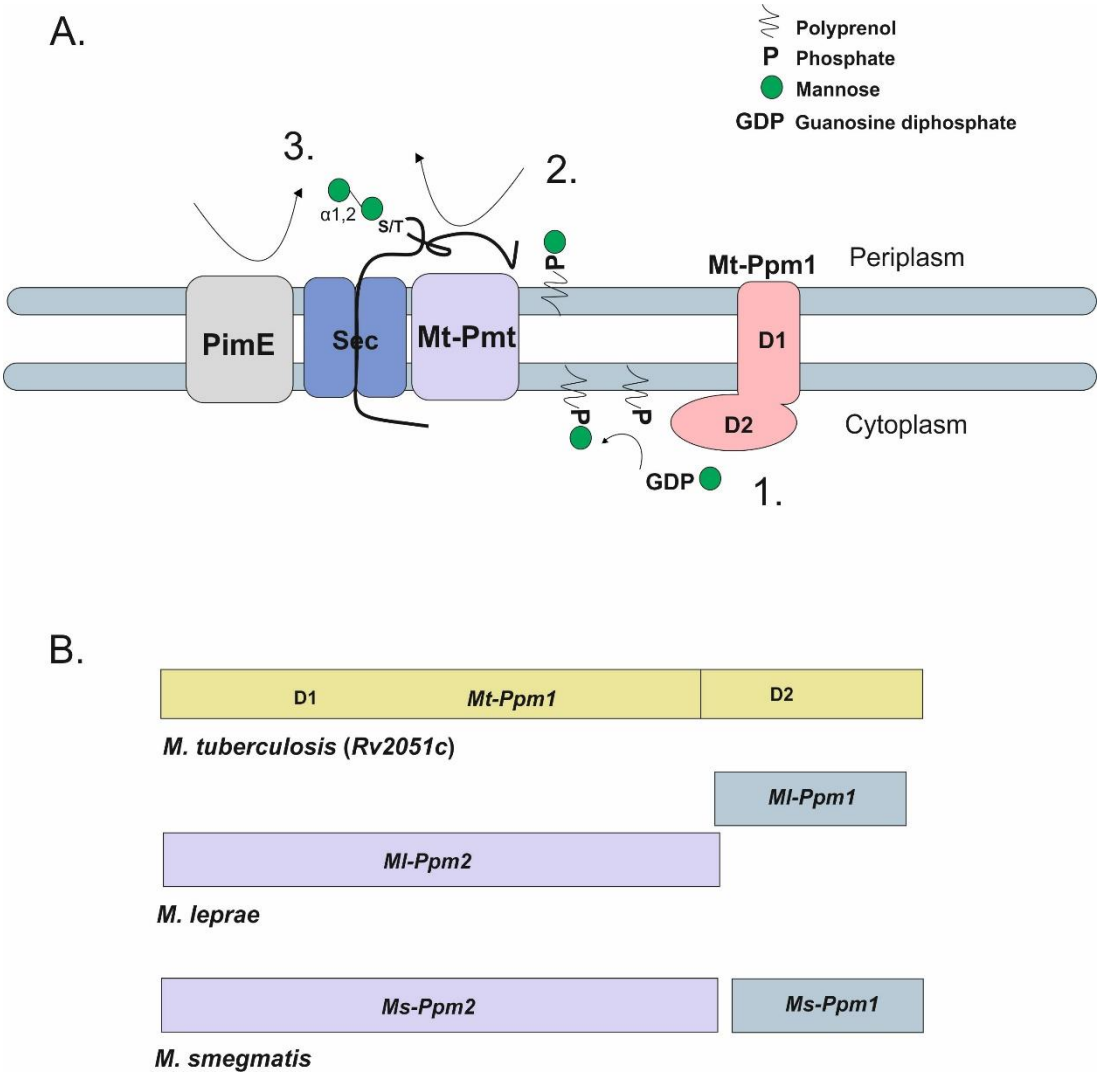


Figure 1.4 Protein O-mannosylation in *M. tuberculosis*. A. The D2 domain of Mt-Ppm1 catalyses the transfer of mannose from GDP-Man onto polyprenol phosphate (1). Mt-Pmt catalyses the first step of protein O-glycosylation by transferring mannose from polyprenol phosphate mannose onto target proteins, as they emerge from the SEC machinery (2). PimE further elongates the glycan chain by the addition of mannose with α1,2 linkages (3). B. A schematic representation of the *ppm1* locus in *M. tuberculosis*, compared to *M. leprae* and *M. smegmatis.* Figure adapted from (Gurcha et al. 2002).

In pathogenic mycobacteria, protein O-glycosylation is thought to have a role in pathogenicity. The glycosylation of secreted glycoproteins has been shown to be required for antigenicity and the stimulation of the T lymphocyte response (Horn et al. 1999; Romain et al. 1999). The altered glycosylation pattern of the 45/47 kDa glycoprotein (Apa) after its heterologous expression in *M. smegmatis* resulted in a reduced capacity to stimulate a T lymphocyte response *in vitro*, while unglycosylated Apa was unable to stimulate this response altogether (Horn et al. 1999). The glycosylation of Apa has been shown to be required to enable its interaction with C-type lectins of the host immune system, further suggesting that the protein is important for immune recognition (Ragas et al. 2007). Additionally, a secreted glycoprotein from *M. tuberculosis*, SodC was recognized by human antibodies and was shown to contribute towards the persistence of the bacterium in macrophages (Piddington et al. 2001; Sartain et al. 2006; Sartain and Belisle 2009). *M. leprae* was shown to bind in a carbohydrate specific manner to r-langerin, suggesting its role in langerin ligand binding and host immune response stimulation (Kim et al. 2015). These studies suggest that the immunogenicity of secreted antigens in pathogenic mycobacteria is linked to their correct glycosylation, alluding to the idea that a disruption of glycosylation could significantly affect pathogenesis. Indeed, the inactivation of *pmt* (*Rv1002c*) in *M. tuberculosis* and the subsequent loss of protein O-glycosylation was shown to significantly impair growth and reduce its pathogenicity in immunocompromised mice (Liu et al. 2013a). *M. smegmatis pmt* mutants display no changes in growth however, were hypersensitive to SDS-treatment, suggesting that a loss of Pmt may result in changes in the cell wall.

1.4 Bacterial cell walls and antibiotics

1.4.1 Peptidoglycan biosynthesis in bacteria

Peptidoglycan is the major polymer found in the cell walls of Gram positive and Gram negative bacteria. It has a crucial role in maintaining cell shape and counteracting turgor pressure (Typas et al. 2012). While variations exist, the basic peptidoglycan architecture in conserved in nearly all cell wall containing bacteria. In Gram positive bacteria, the cell wall is thick and multi-layered, where peptidoglycan is incorporated near the surface of the periplasmic membrane. Peptidoglycan biosynthesis starts in the cytoplasm with the biosynthesis of UDP-N-acetylmuramyl-pentapeptide (UDP-MurNAc-pentapeptide) and UDP-N-acetyl glucosamine (UDP-GlcNAc) (Figure 1.5) (Barreteau et al. 2008).

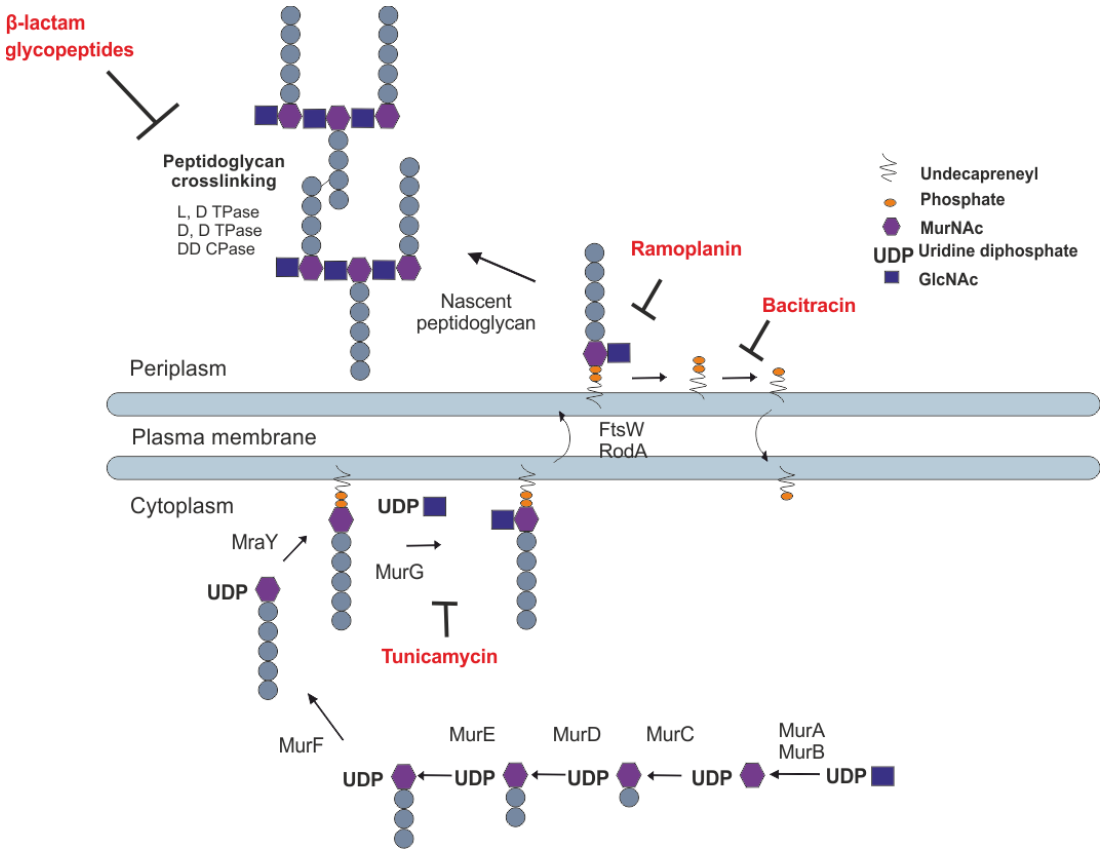


Figure 1.5 A general overview of peptidoglycan biosynthesis in Gram positive bacteria. Red text indicates antibiotics that inhibit different aspects of cell wall biosynthesis. Images adapted from Typas et al. (2012).

The UDP-MurNAc pentapeptide is synthesised through a series of enzymatic reactions that sequentially add amino acids to UDP-MurNAc. The most common muropeptide is seen as L-alanine-D-glutamic acid-meso diaminopimelic acid- (or L-lysine)-D-alanine-D-alanine. However variations in the peptide are widely reported in many bacterial species (Vollmer et al. 2008). At the cytoplasmic membrane, the MurNAc pentapeptide moiety is transferred from the nucleotide precursor onto a membrane lipid-linked acceptor, undecaprenyl phosphate, yielding undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide (Bouhss et al. 2008). In Gram positive bacteria, the GlcNAc moiety is added to the undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide and the lipid-linked stem pentapeptide is then flipped into the periplasm. Once facing the periplasm, the stem pentapeptide is incorporated into nascent peptidoglycan by penicillin binding proteins.

Penicillin binding proteins (PBPs) are a class of cell wall biosynthetic enzymes in bacteria that are required to catalyse various reactions, such as the polymerisation of the glycan strand (transglycosylation) and peptidoglycan crosslinking (transpeptidation) (Sauvage et al. 2008). Some PBPs recognise the terminal D-alanyl- D-alanine (D-Ala-D-Ala) of stem pentapeptides as part of their catalytic mechanism. The structural resemblance between the D-Ala-D-Ala of peptide stems and the β-lactam antibiotic penicillin, means that PBPs can bind penicillin irreversibly by forming an acyl-enzyme that inhibits further activity of the PBP in cell wall biosynthesis (Tipper and Strominger 1965). There are two main types of PBPs in bacteria: high molecular mass (HMM) PBPs and low molecular mass (LMM) PBPs (Sauvage et al. 2008). HMM PBPs are often multifunctional enzymes with transpeptidase activity and the ability to elongate glycan chains that are uncross-linked (transglycosylation). LMM PBPs are generally monofunctional enzymes. Streptomycetes produce large numbers of PBPs, most likely due to their complex life cycle and challenging environment. *S. coelicolor* for example, is predicted to make at least 20 PBPs. However, the functional roles of very few of these have been characterised (Sauvage et al. 2008).

1.4.2 Peptidoglycan crosslinking

There are two main types of peptidoglycan crosslinking in bacteria, 4->3 crosslinking and 3->3 crosslinking. Most commonly, peptidoglycan crosslinking occurs between the carboxyl group of D-Ala at position 4 and the amine group of the amino acid at position 3 (4->3 crosslinking), between two stem pentapeptides and is catalysed by D,D transpeptidases (Vollmer et al. 2008) (Figure 1.6). Alternatively, peptidoglycan crosslinking is catalysed

between the third position amino acids of two tetrapeptide stems (3->3 crosslinking). Peptidoglycan with 3->3 crosslinking was first identified in *M. smegmatis*. However the enzyme responsible for this catalytic activity, an L, D transpeptidase, was not identified until decades later in *Enterococcus faecium* (*E. faecium*) where 3->3 crosslinking is thought to have a role in antibiotic resistance (Wietzerbin et al. 1974; Mainardi et al. 2002). Peptidoglycan with 3->3 crosslinks has been identified in other Actinobacteria, including *S. coelicolor* (Hugonnet et al. 2014). In *M. tuberculosis*, the majority of peptidoglycan is thought to be 3->3 crosslinked and the chromosome encodes five L, D transpeptidases (Dubée et al. 2012b). Mutants defective in L, D transpeptidase activity were shown to have increased susceptibility to antibiotics, some of which target the cell wall, as well as displaying an increase in lysozyme sensitivity and altered colony morphology (Sanders et al. 2014; Schoonmaker et al. 2014). However, sensitivity to the antibiotics was mostly observed in the triple mutants, suggesting that there could be some overlap in function between L, D transpeptidases in mycobacteria. As in *M. tuberculosis*, there are five putative L, D transpeptidases annotated in the *S. coelicolor* genome (StrepDB).

1.4.3 Antibiotics that target cell wall biosynthesis.

Cell wall biosynthesis is an essential biological process in bacteria, and is the target of many antibiotics (Figure 1.5). The primary antibiotic targets in the cell wall are the enzymes that catalyse peptidoglycan biosynthesis. β-lactam antibiotics target multiple cell wall biosynthetic enzymes by mimicking the terminal D-Ala-D-Ala of stem pentapeptides (Sauvage et al. 2008). While peptidoglycan crosslinking (3->3) catalysed by L, D transpeptidases was initially thought to contribute to antibiotic resistance by bypassing classical PBP peptidoglycan crosslinking, L, D transpeptidases were recently shown to be inactivated by carbapenem antibiotics in *E. faecium* and *M. tuberculosis* (Dubée et al. 2012b; Dubée et al. 2012a). Glycopeptide antibiotics, such as vancomycin and teicoplanin act by a different mechanism to β-lactams, by targeting the peptidoglycan peptide stems. These antibiotics bind the terminal D-Ala-D-Ala of stem pentapeptides, preventing crosslinking of the peptidoglycan (Reynolds 1989). Some bacteria, such as the glycopeptide antibiotic producers *Streptomyces,* encode a vancomycin resistance pathway that results in the synthesis of pentapeptides ending in D-lactate, for which glycopeptide antibiotics have much lower affinity (Hong et al. 2004).

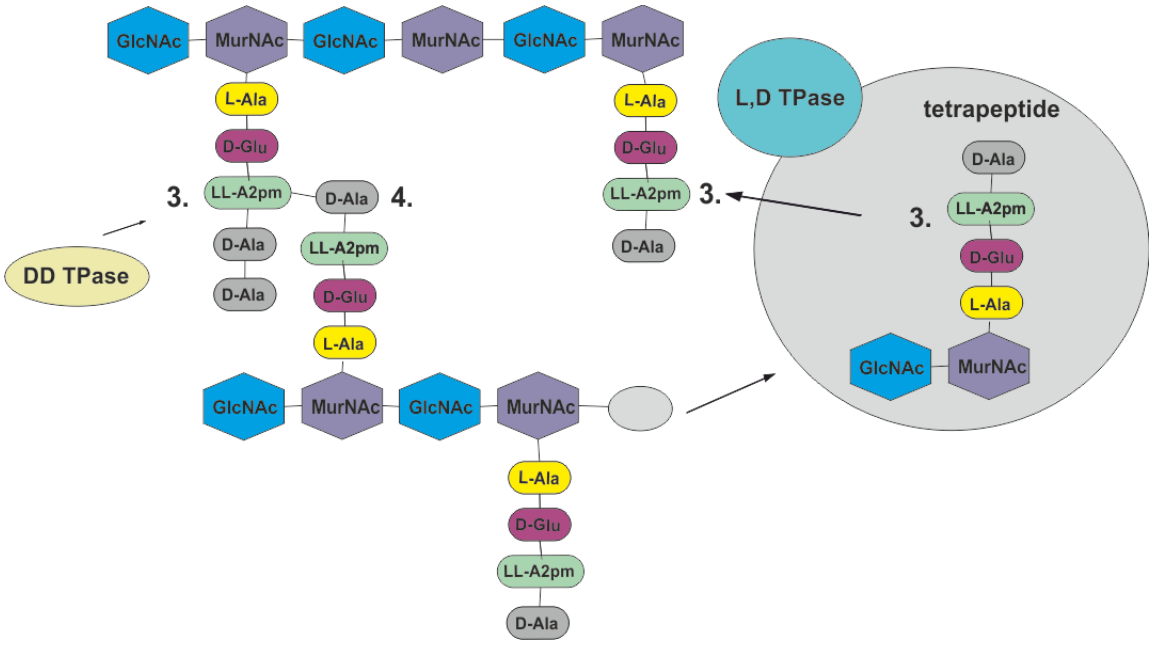


Figure 1.6 An overview of 4->3 and 3->3 peptidoglycan crosslinking in bacteria. 4->3 crosslinking of stem pentapeptides is catalysed by D, D transpeptidases (DD TPase), while 3->3 crosslinking of tetrapeptide stems is catalysed by L, D Transpeptidases (L, D TPase). LL-A2pm: 2,6 diaminopimelic acid.

Other cell-wall targets of antibiotics include the transglycosylation of peptide stems (ramoplanin), the dephosphorylation of the lipid carrier molecule involved in lipid carrier recycling (bacitracin) and the UDP-GlcNAc:dolichyl-phosphate GlcNAc-1-phosphate transferase that is involved to the addition of GlcNAc to the lipid-linked MurNAc-pentapeptide (tunicamycin) (Heifetz et al. 1979; Stone and Strominger 1971; Fang et al. 2006).

1.5 *Streptomyces*

1.5.1 A general introduction to Streptomycetes

Streptomycetes are a genus of high GC, Gram positive bacteria that belong to the phylum of Actinobacteria (Flärdh and Buttner 2009). Streptomycetes are mostly soil dwelling organisms, although they are known to occupy a wide range of niches (Hopwood 2007). They can metabolise a variety of carbon sources including sugars, organic acids, sugar alcohols and amino acids (Pridham and Gottlieb 1948) (Romano and Nickerson 1958). Streptomycetes have a complex lifecycle for a bacterium and are morphologically similar to filamentous fungi. The *Streptomyces* lifecycle begins under favourable conditions, when a spore germinates and germ tubes emerge to form filamentous hyphae (Figure 1.7) (Flärdh and Buttner 2009). Vegetative growth results in the formation of a network of hyphae called the substrate mycelium. The hyphae are often multinucleated since septation occurs only occasionally. When nutrients become depleted, aerial hyphae are generated and sporulation ensues. Due to their complex secondary metabolism, streptomycetes synthesise a multitude of natural products including immunosuppressants, insecticides, anti-tumour agents and around two thirds of clinically useful antibiotics (Hopwood 1999). Secondary metabolite production is initiated by nutrient depletion, and often coincides with morphological differentiation.

S. coelicolor is a model organism of the *Streptomyces* genus and was the first streptomycete to have its genome sequenced (Bentley et al. 2002). *S. coelicolor* has a single ~ 8,667 kb linear chromosome that contains 7,825 predicted genes. More than 20 gene clusters are thought to contain genes necessary for secondary metabolite biosynthesis. Among these are the well characterised clusters for the pigmented antibiotics, actinorhodin (ACT) and undecylprodigiosins (REDs) (Liu et al. 2013b).

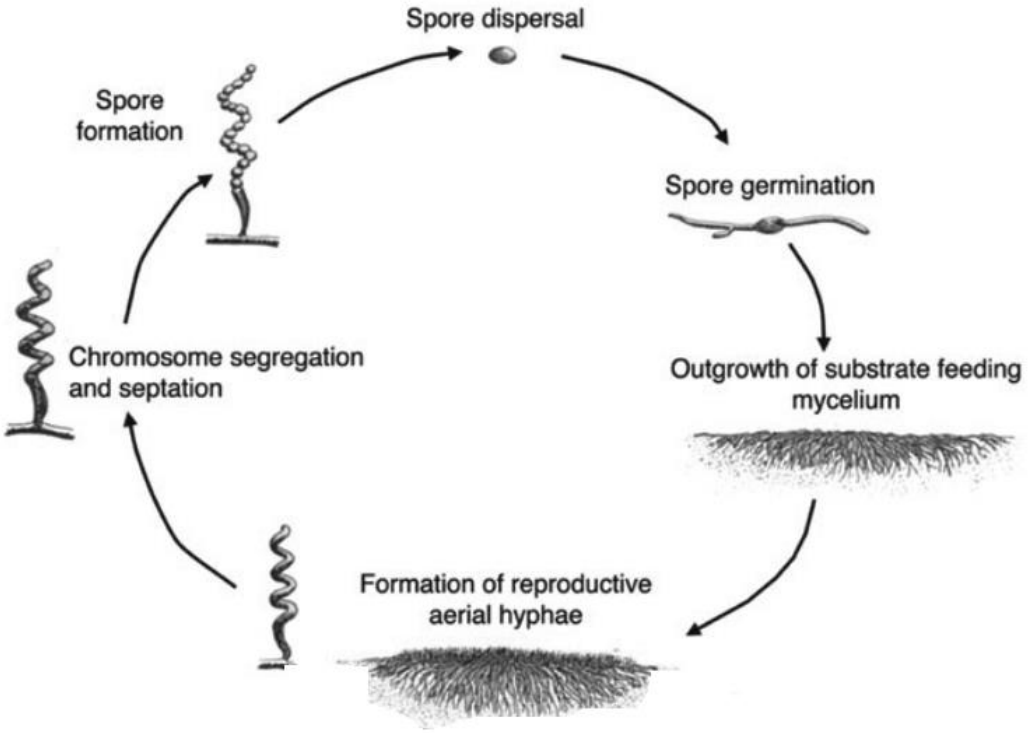


Figure 1.7 The life cycle of *Streptomyces*. Under favourable conditions, spore germination ensues and the growth of filamentous hyphae results in a network called the substrate mycelium. Nutrient limitation triggers differentiation, resulting in the formation of aerial hyphae and sporulation. Image adapted from Seipke et al. (2012).

ACT is a polyketide-derived benzoiso-chromanequinone that has a red/blue colour, depending on the environmental pH. REDs are pyrrole-based compounds that are synthesised by a pathway that overlaps with fatty acid biosynthesis.

1.5.2 Protein O-mannosylation in *Streptomyces*

The earliest reports of glycosylation in *Streptomyces spp*. were made after both native and heterologously expressed glycoproteins were characterised in *Streptomyces lividans* (Kluepfel et al. 1990; Ong et al. 1994). The first evidence of a protein O-glycosylation pathway was reported after an investigation into the nature of the phiC31 phage receptor in *S. coelicolor* (Cowlishaw and Smith 2001; Cowlishaw and Smith 2002). Cowlishaw and Smith (2001) isolated a collection of phiC31 phage resistant mutants that were deficient in the phage receptor. Several of the mutants were complemented by a homologue of fungal PMTs, a putative protein O-mannosyltransferase (*pmt*; *sco3154*) (Cowlishaw and Smith 2001). Additionally, glycoproteins were detected in the wild type strain J1929 by the presence of Con A reactivity and these were absent from the *pmt* strain. The phage receptor mutants that were not complemented by wildtype *pmt* were complemented by a gene with homology to DPM1 from *S. cerevisiae,* encoding a putative polyprenol phosphate mannose synthase (*ppm1*; *sco1423*) (Cowlishaw and Smith 2002). These studies suggested the presence of a protein O-glycosylation pathway in *S. coelicolor* that was similar to the protein O-mannosylation pathway in fungi, and that the phiC31 phage receptor was a glycoprotein. The fact that homologues of *pmt* (*sco3154*) and *ppm1* (*sco1423*) in several *Streptomyces spp,* have been annotated in the StrepDB, suggests that the pathway is highly conserved in streptomycetes.

The only glycoprotein to be characterised in *S. coelicolor* is PstS, a periplasmic phosphate binding protein that was shown to be glycosylated with a trihexose in the wild type strain but not in the *pmt* strain (Wehmeier et al. 2009). Additionally, two out of three serine-rich synthetic peptides designed using the PstS amino acid sequence were shown to be glycosylated by an extract containing wild type J1929 *S. coelicolor* membranes. These findings suggest that protein O-glycosylation in *S. coelicolor* occurs on specific serine or threonine residues, and is not a random process. Using a cell free assay, Wehmeier et al. (2009) demonstrated that Ppm1 was required for the transfer of mannose from a nucleotide activated sugar (GDP-mannose) onto polyprenol phosphate in *S. coelicolor* membrane

fractions. This study presented the first evidence that O-glycosylation in *S. coelicolor* required a lipid-linked carrier and that the modifying carbohydrate could be mannose.

Based on the pathway in mycobacteria, it was hypothesised that *S. coelicolor* Ppm1 was associated with a membrane protein that would serve as an anchor to the cytoplasmic face of the membrane (Varghese 2008). Anttonen (2010) presented evidence in support of this hypothesis by demonstrating that Ppm1 was present in both *S. coelicolor* cytosolic and membrane fractions. The *S. coelicolor* membrane protein SCO1014, that was identified as an orthologue of the transmembrane domain (D1) of Mt-Ppm1 in *M. tuberculosis* (Figure 1.4.B), was initially thought to be the membrane anchor. However, further studies found there to be no interaction between SCO1014 and Ppm1 in *S. coelicolor* (Anttonen 2010; Córdova-Dávalos et al. 2014). Additionally, SCO1014 (Lnt1) mutants were sensitive to phiC31 phage infection and could glycosylate the *M. tuberculosis* glycoprotein Apa, suggesting that SCO1014 is not required for glycosylation in *S. coelicolor* (Córdova-Dávalos et al. 2014).

The phage resistant *S. coelicolor pmt* and *ppm1* mutants previously described by Cowlishaw and Smith (2001) also had a slow growth phenotype and were hypersusceptible to a range of antibiotics that target cell wall biosynthesis (Howlett et al. 2016). Although the antibiotic susceptibility phenotype in the *pmt* strains was less extreme than in the *ppm1* strains, the *pmt* strains still displayed considerable sensitivity to β-lactam antibiotics and to vancomycin. This could suggest that glycosylation in *S. coelicolor* could be required for the correct functioning of periplasmic or membrane enzymes that are required for cell wall biosynthesis. Additionally, an RNAseq analysis of the *ppm1* strain compared to the wild type strain J1929 demonstrated that a loss of Ppm1 resulted in an upregulation of fatty acid biosynthetic genes presumably leading to overall changes in the membrane composition. Preliminary evidence showing changes in the membrane lipids of the *S. coelicolor ppm1* mutant has been gathered using Raman spectroscopy (Howlett et al. 2016). Additionally, Howlett et al. (2016) identified GDP-mannose pyrophosphorylases (ManC – encoded by *sco1388*, *sco3039* and *sco4238*) required for GDP-mannose biosynthesis in *S. coelicolor*. SCO3039 and SCO4238 together were found to have overlapping functions and strains with mutations in both genes were not obtained, suggesting lethality. However, mutants with depleted SCO3039 and SCO4238 levels were isolated and were phenotypically similar to *ppm1* mutants in antibiotic and phiC31 phage sensitivity. Additionally, it was shown that a mutation in *manB* (*sco3028*), encoding an enzyme with phosphomannomutase and phosphoglucomutase activity resulted in altered colony morphology, phiC31 phage resistance and an increase in antibiotic

sensitivity in *S. coelicolor*. These findings suggest that the activities of ManB and ManC are part of the protein O-glycosylation pathway in *S. coelicolor*.

The current model for the protein O-glycosylation pathway in *S. coelicolor* is summarised in Figure 1.8. GDP-mannose biosynthesis requires the activities of ManA (phosphomannose isomerase, SCO3025), ManB (phosphomannomutase, SCO3028) and ManC (GDP-Mannose pyrophosphorylases, SCO3039, SCO4238). Ppm1 catalyses the transfer of mannose from GDP-mannose onto polyprenol phosphate (PP) on the cytoplasmic face of the membrane. Ppm1 is thought to be localised to the membrane by a membrane anchor; however this has yet to be identified. PP-mannose is flipped from the cytoplasmic face of the membrane onto the periplasmic face of the membrane by an unknown mechanism. Pmt is thought to catalyse the first step of O-glycosylation by the transfer of mannose from PP-mannose onto specific Ser/Thr residues in target proteins. Based on the dogma in *M. tuberculosis* (section 1.3.3) this is thought to occur on proteins as they emerge from the SEC secretory pathway, and protein folding is thought to happen after glycosylation.

There are several unanswered questions regarding the mechanism and role of protein O-glycosylation in *S. coelicolor*. Mainly, what is the physiological role of protein O-glycosylation in *S. coelicolor*? The pleiotropic phenotypes observed previously in glycosylation deficient *pmt* and *ppm1* mutants suggests that glycosylation could be required for cell wall biosynthesis or membrane homeostasis (Howlett et al. 2016). A better understanding of the targets of O-glycosylation in *S. coelicolor* could help to underpin the role of this modification in maintaining cell wall homeostasis.

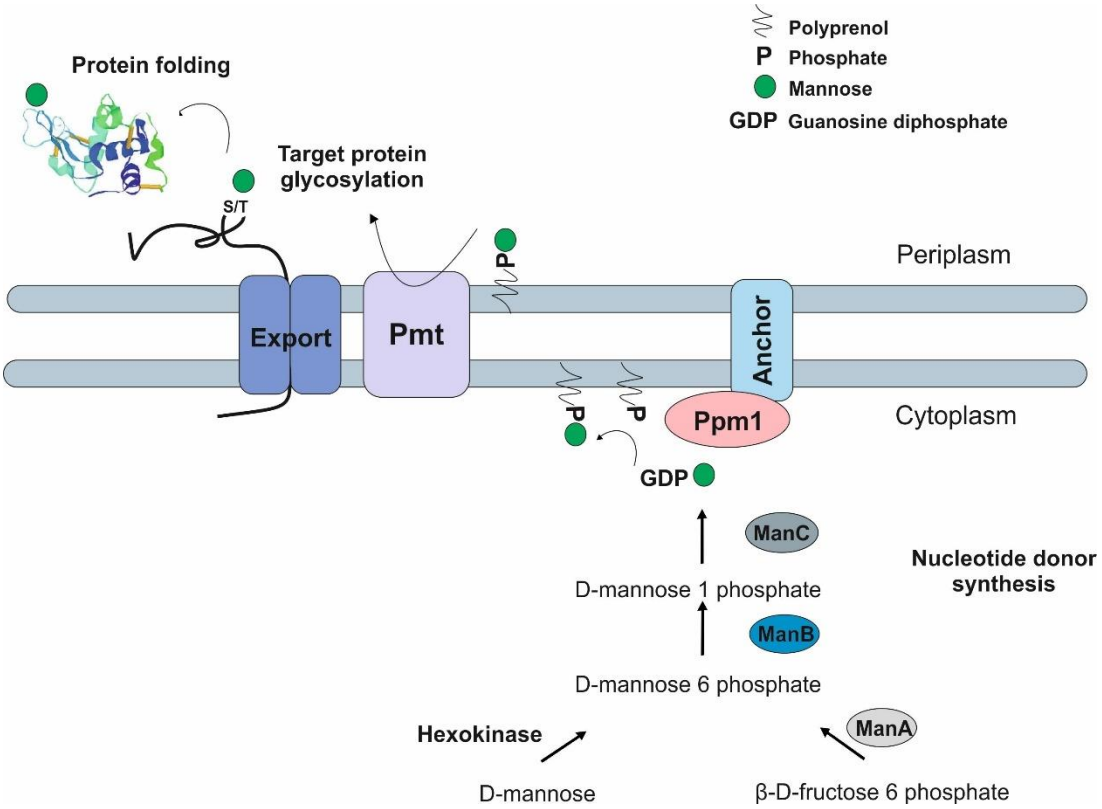


Figure 1.8 A model of protein O-glycosylation in *S. coelicolor*.

1.6 Aims

The aim of this project was to investigate the *S. coelicolor* glycoproteome in order to better understand the physiological role of protein O-glycosylation in a model actinomycete. A further aim was to identify glycoproteins that could be of relevance to the antibiotic hypersensitivity phenotypes observed previously in the *S. coelicolor pmt* and *ppm1* mutants (Howlett et al. 2016).

Specifically, the objectives were:

1. To characterise the *S. coelicolor* glycoproteome using biochemical and proteomics approaches.
2. To characterise the glycans that modify *S. coelicolor* glycoproteins and investigate the linkages present.
3. To investigate glycoproteins with a putative role in cell wall biosynthesis, by the characterisation of knockout mutants.

Chapter 2 - Materials and methods

Chapter 2 – Materials and methods

2.1 Materials

2.1.1 Chemicals

All chemicals were purchased from Sigma-Aldrich, Fisher Scientific, Thermo Fisher Scientific or VWR International unless otherwise stated. Difco nutrient broth (DNB) and Difco nutrient agar (Beckton Dickinson and Company) for the growth of *S. coelicolor* was purchased from Appleton Woods Ltd. Soya flour for the growth of *S. coelicolor* was purchased at Holland & Barratt. Casaminoacids, tryptone and yeast extract (Oxoid) were purchased from Thermo Fisher Scientific. All restriction enzymes, DNA polymerases and deoxynucleotides (dNTPs) were purchased from New England Biolabs (NEB). PCR primers were purchased from Integrated DNA Technologies. DNA sequencing was carried out by GATC BIOTECH.

2.1.2 Software

Snapgene viewer software (version 3.2) was used for the construction of plasmid maps and for the analysis of sequencing data. BLAST was used to carry out gene and protein sequence similarity and homology searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Clustal Omega (https://www.ebi.ac.uk/Tools/msa/ clustalo/) was used to perform multiple sequence alignments. Boxshade (version 3.21) was used to shade multiple sequence alignments (http://www.ch.embnet.org/software/ BOX_form.html). Weblogo software was used for the analysis of the *S. coelicolor* glycosylation motif (Crooks et al. 2004). Bruker Compass Data Analysis version 4.4 (Bruker), flexAnalysis version 3.0 (Bruker), MassHunter Software (Agilent) and Xcalibur software version 4 (Thermo Fisher Scientific) was used for the analysis of the mass spectrometry data. SigmaPlot version 13.0 (Systat software Inc) was used to produce the graphs reported in this study. Agarose gels were imaged using Quantity One (version 4.6.2; Basic) software. Transmembrane helices in proteins were predicted using the TMHMM Server v. 2.0 algorithm (http://www.cbs.dtu.dk/services/TMHMM/). Signal peptides were predicted using the SignalP 4.0 Server algorithm (Petersen et al. 2011). LipoP 1.0 and PRED_LIPO software were used for the prediction of lipoproteins (Juncker et al. 2003; Bagos et al. 2008). TatP 1.0 Server was used for the prediction of twin arginine (TAT) signal peptides (Bendtsen et al. 2005). Conserved domains in proteins were detected using the Conserved Domain Database CD-search tool (Marchler-Bauer and Bryant 2004).

2.1.3 Media and Buffers

Difco nutrient broth (DNB)

8 g of Difco nutrient broth powder was dissolved in 1000 mL of distilled H_2O (ddH_2O) and autoclaved.

Difco nutrient agar (DNA)

4.6 g of Difco nutrient agar solid was dissolved in 200 mL of distilled H_2O in a 250 mL DURAN bottle and autoclaved.

Soft nutrient agar (SNA) (Kieser et al. 2000).

8 g of Difco nutrient broth powder and 5 g of Difco Bacto agar solid were suspended in 1000 mL of ddH_2O and autoclaved (115 °, 15 min).

Mannitol soya flour agar (MSA) (Kieser et al. 2000)

Soya flour	20 g
Mannitol	20 g
Agar	20 g
Ultrapure H_2O	1000 mL

Mannitol was dissolved in the Ultrapure H_2O and 200 mL was added to DURAN bottles (250 mL) each containing 4 g soya flour and 4 g agar. The media was autoclaved twice (115 °, 15 min).

Luria Burtani (LB) broth and agar (Sandbrook et al. 1989)

Tryptone	10 g
Sodium chloride (NaCl)	10 g
Yeast extract	5 g
ddH_2O	1000 mL

The tryptone, NaCl and yeast extract were dissolved in 1000 mL of ddH_2O and autoclaved (115 °, 15 min). For LB agar, 1.5 g of agar per 100 mL of LB broth was added and autoclaved (115 °, 15 min).

2 X YT Medium (Sandbrook et al. 1989)

Bacto Tryptone	16 g
Bacto Yeast Extract	10 g
Sodium chloride	5 g

Made up to 1000 mL with ddH_2O and autoclaved (115 °, 15 min).

2 X Germination medium (Kieser et al. 2000)

Difco yeast extract 1 g

Difco casamino acids 1 g

Made up to 100 mL with ddH_2O and autoclaved to sterilise. 200 µL of sterile 5M $CaCl_2$ added.

TES buffer

TES 1.146 g

Made up to 100 mL in ddH_2O and the pH was adjusted to 8.0. The buffer was autoclaved to sterilise (115 °, 15 min).

Supplemented minimal medium, solid (SMMS) (Kieser et al. 2000)

Difco casamino acids 2 g

TES Buffer 5.73 g

Ultrapure H_2O to 1000 mL

The solution was made up and the pH was adjusted to 7.2. 200 mL was poured into a DURAN bottle (250 mL) containing 3 g of agar. The solution was autoclaved to sterilise and at the time of use, re-melted and the following was added:

$NaH_2PO_4 + K_2HPO_4$ (50 mM each) 2 mL

$MgSO_4$ (1 M stock) 1 mL

Glucose (50 % w/v) 3.6 mL

Trace elements* 0.2 mL

*Trace elements solution was prepared by adding 0.1 g. l^{-1} of the following: $ZnSO_4.7H_2O$, $FeSO_4.7H_2O$, $MncCl_2.4H_2O$, $CaCl_2.6H_2O$ and NaCl. The trace elements solution was stored at 4 °C for up to 2 weeks.

For the preparation of chemically competent cells:

TFB1 (Sandbrook et al. 1989)

Rubidium chloride	100 mM
Manganese chloride	50 mM
Potassium acetate	30 mM
Calcium chloride	10 mM
Glycerol	15 % (v/v)

The solution was made up to 100 mL with ddH_2O, the pH adjusted to 5.8 and filter sterilised.

TFB2 (Sandbrook et al. 1989)

Acid MOPS	10 mM
Rubidium chloride	10 mM
Calcium chloride	70 mM
Glycerol	15 % (v/v)

The solution was made up to 100 mL with ddH_2O, the pH adjusted to 8.0 and autoclave sterilised (115 °, 15 min).

F134 Medium (Nieselt et al. 2010)

F134 Main broth

L-glutamic acid monosodium salt	55.2 g
Ultrapure H_2O	865 mL

SMM 0.1 M PO_4 buffer

NaH_2PO_4	7.88 g
$K_2H\ PO_4$	8.70 g
Ultrapure H_2O	to 1000 mL

SMM TE

$FeSO_4$	100 mg	NaCl	100 mg
$ZnSO_4$	100 mg	Ultrapure H_2O to	950 mL
$MnCl_2$	100 mg	HCl	50 mL
$CaCl_2$	100 mg		

TMS1			
$FeSO_4 .7H_2O$	5000 mg	$MnSO_4$	150 mg
$ZnSO_4.7H_2O$	440 mg	$Na_2MoO_4.2H_2O$	10 mg
$CuSO_4.5H_2O$	390 mg	$CoCl_2.6H_2O$	20 mg
		Ultrapure H_2O	to 1000 mL

To make F134 liquid medium

F134 main broth	865 mL	Antifoam	100 µL.
$MgSO_4$ (1M stock)	2 mL	Autoclave sterilised. At the time of use 80 mL of glucose (50 % w/v stock, filter sterilised) and 5.6 mL of TMS1 (filter sterilised) was added.	
SMM TE	8 mL		
SMM 0.1 M PO_4 buffer	40 mL		

To make F134 agar

F134 main broth	865 mL	Agar	15 g
$MgSO_4$ (1M stock)	2 mL	Autoclave sterilised. At the time of use 80 mL of glucose (50 % w/v stock, filter sterilised) and 5.6 mL of TMS1 (filter sterilised) was added.	
SMM TE	8 mL		
SMM 0.1 M PO_4 buffer	40 mL		

To make F134 soft agar

F134 main broth	865 mL	Agar	5 g
$MgSO_4$ (1M stock)	2 mL	Autoclave sterilised. At the time of use 80 mL of glucose (50 % w/v stock, filter sterilised) and 5.6 mL of TMS1 (filter sterilised) was added.	
SMM TE	8 mL		
SMM 0.1 M PO_4 buffer	40 mL		

For *S. coelicolor* genomic DNA isolation

CTAB/NaCl solution (10 % w/v CTAB, 0.7 M NaCl) (Kieser et al. 2000)

4.1 g of NaCl was added to 80 mL of ddH_2O. 10 g of cetyltrimethylammonium bromide (CTAB) was slowly added whilst gently heating and mixing. The solution was made up to 100 mL with ddH_2O and autoclave sterilised.

TE25S buffer (Kieser et al. 2000)

Tris-HCl pH 8.0	25 mM
EDTA	25 mM
Sucrose	0.3 M

Dissolved in ddH_2O and autoclave sterilised.

TE buffer

Tris-HCl, pH 8	10 mM
EDTA, pH 8	1 mM

Dissolved in ddH_2O and filter sterilised.

For Southern blotting

20 X SSC

Sodium chloride	3M
Trisodium citrate	0.3 M

Dissolved in ddH_2O and the pH was adjusted to 7.0.

100 X Denhardt's solution

Bovine serum albumin (BSA)	1 g
Ficoll 400	1 g
Polyvinylpyrrolidone (PVP)	1 g

Dissolved in 50 mL of ddH_2O. The solution was filter sterilised and stored at – 20 °C.

Pre-hybridisation solution

20 X SSC	15 mL
100 X Denhardt's solution	2.5 mL
10 % (w/v) SDS	2.5 mL
Deionized formamide	25 mL

Made up to 50 mL with ddH_2O.

For agarose gel electrophoresis

10 X TBE buffer (Sandbrook et al. 1989)

Tris base	108.9 g
Boris acid	85.7 g
EDTA (0.5 M, pH 8)	40 mL

Made up to 1000 mL with ddH_2O.

For Western blotting

10 X Transfer buffer

Glycine 144 g

Tris base 30.2 g

Made up to 1000 mL with ddH_2O. 100 mL of the 10 X Transfer buffer stock was added to 100 mL of methanol and the solution was made up to 1000 mL for 1 X Transfer buffer solution.

10 X TBS buffer

Tris base 60.6 g

NaCl 87.6 g

Dissolved in 800 mL of ddH_2O. The pH was adjusted to 7.5 and made up to 1000 mL with ddH_2O

Chemiluminescent detection

Solution A (cover in foil to protect from the light)

Tris-HCl (pH 8.5) 5 mL

Distilled H_2O 45 mL

Coumaric acid 90 mM stock (0.15 g into 10 mL DMSO, stored at -20 °C) 110 µL

Luminol 250 mM stock (0.44 g in 10 mL DMSO, stored at -20 °C) 250 µL

Solution B

Hydrogen peroxide 30 % v/v stock 100 µL

Distilled H_2O 900 µL

For use, 15 µL of solution B was added to 5 mL of solution A in a Falcon tube covered with foil to protect from the light.

Table 2.1 List of bacterial strains

Strain	Description	Source
S. coelicolor J1929	*pglY*	Bedford et al. (1995)
S. coelicolor DT1025	*pmt*	Cowlishaw and Smith (2001)
S. coelicolor DT3017	*ppm1*	Cowlishaw and Smith (2002)
S. coelicolor TK005	*sco5204*	This work
S. coelicolor TK006	*sco4847*	This work
S. coelicolor TK008	*sco4934*	This work
S. coelicolor TK009	TK005: pTAK28	This work
S. coelicolor TK010	TK008: pTAK32	This work
S. coelicolor TK012	TK005: pTAK29	This work
S. coelicolor TK013	TK006: pTAK30	This work
S. coelicolor TK015	TK008: pIJ10257	This work
S. coelicolor TK016	TK006: pIJ10257	This work
S. coelicolor TK017	TK005: pIJ10257	This work
S. coelicolor TK018	TK005: pMS82	This work
S. coelicolor TK020	TK008: pTAK48.b	This work
S. coelicolor TK021	TK008: pTAK49.b	This work
S. coelicolor TK022	TK008: pTAK50.b	This work
S. coelicolor TK023	TK008: pTAk51.b	This work
S. coelicolor TK024	TK008: pTAK52.b	This work
S. coelicolor TK025	TK008: pTAK53.b	This work
Streptomyces phage	ΦC31cΔ25	Sinclair and Bibb (1988)
E. coli DH5α	*F– Φ80lacZΔM15 (Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA*	Invitrogen
E. coli ET12567 [pUZ8002] (ETZ)	*ET12567 - dam-13::Tn9, dcm-6, hsdM, hsdS; pUZ8002 – tra, neo, RP4*	Paget et al. (1999)

Table 2.2 List of cosmids. The Tn5062 transposon contains an Apramycin resistance marker.

Cosmid	Description	Tn*5062* position
2St3B6.1.G02	*sco5177-sco5206*, Tn*5062* in *sco5204*	5661778
2St3B6.1.G06	*sco5177-sco5206*, Tn*5062* in *sco5204*	5662683
2SCK31.2.F11	*sco4909-sco4945*, Tn*5062* in *sco4934*	5369107
5G8.1.A11	*sco4820-sco4860*, Tn*5062* in *sco4934*	5279744

Table 2.3 List of constructs

Name	Construct description	Source
pIJ10257	attP-int-derived integration vector for the conjugal transfer of DNA from *E. coli* to *Streptomyces spp.* Contains Hyg^R, oriT and *ermE*p* promoter.	Hong et al. (2005)
pGEM7	Cloning vector; fi oriC, SP6 and T7 RNA polymerase promoters , multiple cloning site, Amp^R, *lacZ* for blue/white screening	Promega
pMS82	attP-int-derived integration vector for the conjugal transfer of DNA from *E. coli* to *Streptomyces spp.* Contains Hyg^R and oriT	Gregory et al. (2003)
pTAK28	*sco5204* + 251 bp upstream in pMS82	This work
pTAK29	*sco5204* in pIJ10257	This work
pTAK30	*sco4847* in pIJ10257	This work
pTAK32	*sco4934* in pIJ10257	This work
pTAK47	*sco4934* in pGEM7	This work
pTAK48	*sco4934*_T(40)A in pGEM7	This work
pTAK49	*sco4934*_S(41)A in pGEM7	This work
pTAK50	*sco4934*_T(40)A, S(41)A in pGEM7	This work
pTAK51	*sco4934*_T(40)V in pGEM7	This work
pTAK52	*sco4934*_S(41)V in pGEM7	This work
pTAK53	*sco4934*_T(40)V, S(41)V in pGEM7	This work
pTAK48.b	*sco4934*_T(40)A in pIJ10257	This work
pTAK49.b	*sco4934*_S(41)A in pIJ10257	This work
pTAK50.b	*sco4934*_T(40)A, S(41)A in pIJ10257	This work
pTAK51.b	*sco4934*_T(40)V in pIJ10257	This work
pTAK52.b	*sco4934*_S(41)V in pIJ10257	This work
pTAK53.b	*sco4934*_T(40)V, S(41)V in pIJ10257	This work

Table 2.4 Selective antibiotics and concentrations used in this work

Antibiotic	Working Concentration (μg/mL)
Hygromycin B (HmB)	75-100
Ampicillin (Amp)	100
Kanamycin (Kan)	50
Chloramphenicol (Cml)	25
Nalidixic Acid (Nx)	25
Apramycin (Apr)	50

Table 2.5 Antibiotics used in the disc diffusion assays

Antibiotic	Solvent	Stock concentration (mg/mL)
Ampicllin	sterile ddH_2O	20
Penicillin	sterile ddH_2O	10
Imipenem	sterile ddH_2O	4
Meropenem	sterile ddH_2O	4
Vancomycin	sterile ddH_2O	4
Bacitracin	sterile ddH_2O	4
Rifampicin	DMSO	4
Gentomycin	sterile ddH_2O	4
Teicoplanin	sterile ddH_2O	4

2.2 Molecular procedures

2.2.1 Growth of *E. coli*

E. coli strains were cultivated on LB agar at 37 ^{o}C overnight. *E. coli* strains were cultured for 12 - 16 h at 37 ^{o}C (180 rpm) in LB broth. Where appropriate, the growth medium was supplemented with antibiotics (Table 2.4) to maintain selection. *E. coli* strains were maintained in 20 % (v/v) glycerol at -80 ^{o}C.

2.2.2 Plasmid DNA Isolation, gel extraction and DNA sequencing.

Plasmid and cosmid DNA was isolated from overnight *E. coli* cultures (2 – 5 mL) using the QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's instructions. For the extraction of DNA from agarose gels, the band of interest was excised using a sterile scalpel and the DNA was extracted using the QIAquick Gel Extraction Kit (QIAGEN) per manufacturer's instructions. Restriction enzymes were used per the manufacturer's instructions. For DNA sequencing, samples were sent to GATC Biotech (https://www.gatc-biotech.com/en/index.html). Sequencing primers are indicated in Table A.1.

2.2.3 PCR conditions

PCR to amplify DNA was carried out per the instructions provided by the manufacturer of the DNA polymerase. The DNA polymerases used in this work are Phusion, LongAmp® Taq, Taq and Q5 DNA polymerase; all purchased from NEB. PCR primers are listed in Table A.1.

2.2.4 Preparation of chemically competent cells

Competent *E. coli* cells were prepared by the rubidium chloride method (Sandbrook et al. 1989). Overnight cultures were diluted 1/100 into 50 mL of fresh LB liquid medium and grown to an OD_{600nm} of 0.4 – 0.5. Cells were placed on ice for 5 min. Cells were pelleted by centrifugation (10 min, 4000 rpm, 4 ^{o}C). Pellets were re-suspended in TFB1 (4 ^{o}C, 30 mL/100 mL culture) and kept on ice for 90 min. The cells were pelleted again by centrifugation (10

min, 4000 rpm, 4 ^{o}C). The pellet was re-suspended in TFB2 (4 ^{o}C, 4 mL/100 mL culture). Competent cells were snap-frozen in liquid nitrogen, and stored at -80 ^{o}C.

2.2.5 Transformation of *E. coli*

50-100 ng of plasmid DNA was incubated with 50 µL of competent cells on ice (20 min). The reactions were heat shocked (42 ^{o}C, 45 s) and then incubated on ice for 5 min. Reactions were allowed to recover at 37 ^{o}C for 1 h, after the addition of 500 µL of LB broth. The cells were plated onto LB agar plates, treated with appropriate selective antibiotics and incubated at 37 °C overnight.

2.2.6 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out as previously described by Sandbrook et al. (1989). 0.8 – 1% agarose was dissolved into 1 X TBE buffer. Ethidium bromide (15 µg/mL) was added directly to the molten agarose. Electrophoresis was carried out at 65 – 100V in 1 X TBE buffer. Gels were visualised under UV fluorescence and the images were recorded by a Biorad gel documentation system. Samples were run alongside suitable DNA markers (Figure 2.1).

2.2.7 In-Fusion Cloning

Infusion cloning was carried out per the manufacturer's instructions. The assembly of the reaction required 50 – 100 ng of both vector DNA and PCR product, 2 µL of HD Infusion enzyme and ddH_2O to make the reaction up to 10 µL. Reactions were incubated for 15 min (50 ^{o}C), after which 5 µL of the cloning product was transformed into *E. coli* competent cells.

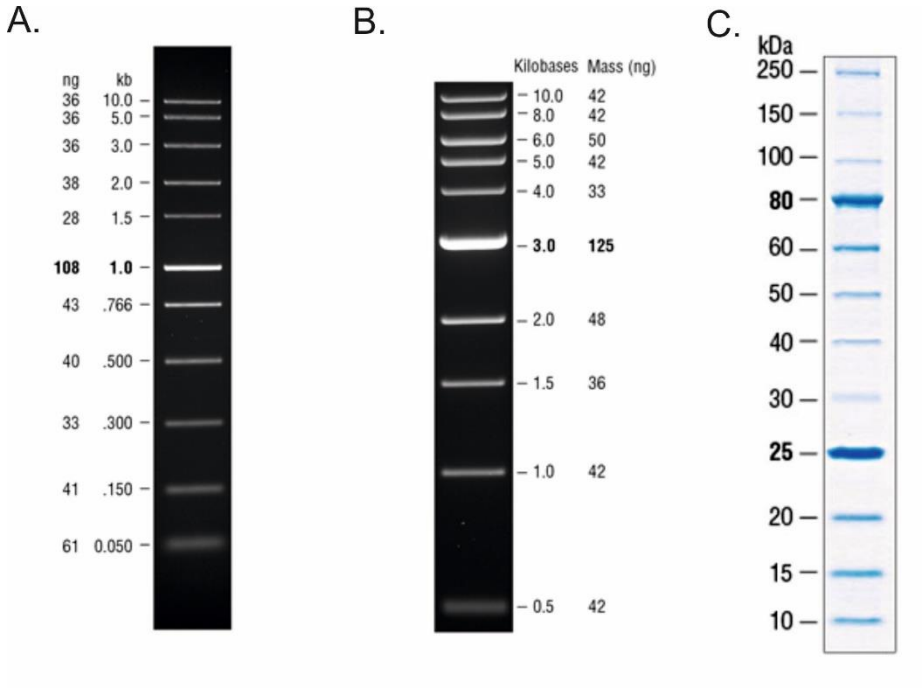


Figure 2.1 DNA and protein molecular weight markers used in this work. A. Fast DNA ladder (NEB). Mass values are for 0.5 µg/lane. B. 1 Kb DNA ladder (NEB). Mass values are for 0.5 µg/lane. C. 10 – 250 kDa unstained protein ladder (NEB).

2.3 *Streptomyces* work

2.3.1 *Streptomyces* culture conditions

Streptomyces culture conditions and genetic procedures used in this study have been previously described by Kieser et al. (2000) or (for F134 medium) by Nieselt et al. (2010). *Streptomyces* strains were grown on mannitol soya flour agar (MSA) for 7 days at 30 °C. If sporulation was not required, *Streptomyces* strains were grown on Difco nutrient agar at 30 °C for 48 h. For liquid cultures, *S. coelicolor* spores (~10^{10} mL^{-1}) were centrifuged (4000 rpm, 4 ^{o}C, 5 min) and resuspended in 5 mL TES Buffer. The spores were heat shocked at 50 ^{o}C for 10 min and then 5 mL of 2 x germination medium was added. The spore suspension was transferred to a sterile conical flask with glass beads and incubated for 6 - 12 h at 30 ^{o}C with shaking (180 rpm). The growth medium was inoculated to an OD_{450} of 0.03 – 0.05 and incubated at 30 ^{o}C with shaking (180 rpm) for up to 60 h.

2.3.2 Isolation of genomic DNA from *S. coelicolor*.

S. coelicolor genomic DNA was isolated according to the CTAB method as described by Kieser et al. (2000). The mycelium from a 30 mL culture grown in Difco nutrient broth was pelleted by centrifugation (4000 rpm, 4 ^{o}C, 5 min) and resuspended in 5 mL TE25S buffer (25 mM Tris-HCl pH 8, 25 mM EDTA pH 8, 0.3 M sucrose). 100 µl of lysozyme solution (100 mg mL^{-1} in ddH_2O) was added and the suspension was incubated at 37 ^{o}C for 60 min. 50 µl of proteinase K solution (20 mg mL^{-1} in ddH_2O) was added and the suspension was mixed by inversion. 300 µl of 10 % (w/v) SDS solution was added, the suspension was mixed by inversion and incubated at 55 ^{o}C for 60 min. 1 mL of 5 M NaCl solution was added and the suspension was mixed by inversion. 650 µl of CTAB/NaCl solution (10 % (w/v) CTAB in 0.7 M NaCl) was added, the suspension mixed by inversion and then incubated at 55 ^{o}C for 10 min. The suspension was cooled to 37 ^{o}C, 5 mL of chloroform/isoamyl alcohol (24:1) was added and the suspension was mixed by inversion for 30 min. Samples were centrifuged (12,000 rpm, 20 ^{o}C, 15 min), the supernatant transferred to a separate tube and 0.6 volume isopropanol was added. Samples were mixed by inversion. After 3 min the samples were centrifuged (14,000 rpm, 20^{o}C, 5 min). The DNA pellet was rinsed in 5 mL of 70 % ethanol, air dried and dissolved in TE Buffer at 55 ^{o}C for 5 min.

2.3.3 Preparation of *S. coelicolor* spores.

Approximately 50 – 100 µL of *S. coelicolor* spores (~10^{10}. mL $^{-1}$) were spread onto MSA plates and incubated for 5-6 days at 30 ^{o}C until sporulation ensued. 4mL sterile 20 % (v/v) glycerol in ddH_2O was added to each plate and the spores were gently removed using a sterile cotton bud. The spore stocks were stored at -30 ^{o}C. Spore stocks were titred by diluting spores in sterile ddH_2O and performing an appropriate dilution series. 5 µL aliquots of each dilution were plated onto Difco nutrient agar plates and the resulting colonies were counted after incubating the plates at 30 °C for 48 h.

2.3.4. Growth assays on agar in sterile Microwell plates

250 µl of sterile Difco nutrient broth or F134 medium + 1 % (w/v) agar was pipetted into the wells of a sterile Microwell plate. Approximately 10^7 *S. coelicolor* spores were added to the surface of the agar and the plates were dried under a lamina flow cabinet for 15 min. The plates were incubated at 30 ^{o}C for up to 47 h and the absorbance at 492 nm was measured at regular intervals.

2.3.5 Phage sensitivity tests

Nutrient agar (Difco nutrient agar or F134 agar) plates supplemented with 0.5 % (w/v) glucose, 10 mM $MgSO_4$ and 8 mM $Ca(NO_3)_2$ were prepared in 55 mm x 16 mm petri dishes. A tenfold serial dilution of the ϕC31*c*Δ25 phage (from 10^{-1} to 10^{-7}) was prepared in nutrient broth. 100 µl of diluted ϕC31*c*Δ25 phage was added to the surface of the nutrient agar plate. Approximately 5 x 10^6 *S. coelicolor* spores were suspended in 1 mL of soft agar (soft nutrient agar or F134 soft agar). The nutrient agar plate was overlaid with the soft top agar seeded with spores. The plates were incubated at 30 ^{o}C for 16 – 20 h.

2.3.6 Antibiotic production on SMMS

25 mL of SMMS agar was poured into a sterile petri dish. *S. coelicolor* spores were adjusted to 10^6 spores in a total volume 30 µL, diluted in MilliQ H_2O where necessary. The 30 µL spore

preparation was pipetted into the SMMS agar plate, and streaked out. The plates were incubated at 30 °C for 5 days.

2.3.7 Antibiotic sensitivity disc diffusion assays

20 mL Difco nutrient agar or F134 agar plates were prepared in 94 mm x 16 mm petri dishes. Approximately 10^7 *S. coelicolor* spores were suspended in 2.5 mL of soft nutrient agar or F134 soft agar, respectively. The nutrient agar plates were overlaid with the soft agar containing the spores. Antibiotics were prepared on the day of the assay to the concentrations indicated in Table 2.5. A tenfold serial dilution of the initial antibiotic stock was prepared to obtain 10^{-1}, 10^{-2} and 10^{-3} stocks in sterile water. 10 µl of each respective antibiotic concentration was added to 5 mm Whatman Grade 17 CHROMA filter paper discs (GE Healthcare Life Sciences). The filter paper discs were transferred onto the surface of the soft agar and the plates were incubated at 30 ^{o}C for 48 – 60 h. Zones of inhibition (measured in mm) were recorded.

2.3.8 Lysozyme sensitivity assays

A stock solution of lysozyme (Thermo Fisher Scientific) was prepared to a final concentration of 100 mg/mL in ddH_2O. 125 µL of lysozyme solution was added to 50 mL of Difco nutrient agar to give a final lysozyme concentration of 0.25 mg/mL and the agar was poured into a 10 x 10 cm petri dish. A 50 mL Difco nutrient agar plate without lysozyme was prepared for the negative control. *S. coelicolor* spores were adjusted to 10^8 spores/mL in ddH_2O. The adjusted spore stocks were diluted ten-fold in ddH_2O to prepare stocks from 10^7 spores/mL to 10^4 spores/mL. 5 µL of each spore stock was added to the agar plates and allowed to dry. The plates were incubated at 30 °C for 60 h.

2.3.9 Conjugation from *E. coli* into *Streptomyces*

The method of intergeneric conjugation used in this work was previously described by Gust et al. (2003). *E. coli* ET12567 [pUZ8002] competent cells were prepared after growth overnight in LB broth (37 ^{o}C, 180 rpm) containing kanamycin (50 µg/mL) and chloramphenicol (25 µg/mL) to select for pUZ8002 and the *dam* mutation respectively. The

E. coli ET12567 [pUZ8002] competent cells were transformed with ~ 100 ng of plasmid or cosmid DNA. The transformants were selected on LB agar plates containing the appropriate antibiotic for the incoming plasmid (Table 2.3), or apramycin (50 µg/mL) and carbenicillin (100 µg/mL) to select for the incoming Tn*5062* transposon insertion cosmid. Colonies were inoculated in 10 mL of LB broth containing kanamycin (50 µg/mL), chloramphenicol (25 µg/mL) and the appropriate antibiotic to select for the plasmid or cosmid, and grown overnight at 37 ^{o}C with shaking (180 rpm). The overnight culture was diluted 1/100 into 10 mL of fresh LB broth plus antibiotics and grown to an OD_{600nm} of ~ 0.4 at 37 ^{o}C for ~ 6 h. The cells were pelleted, washed twice with 10 mL of LB broth and re-suspended in 1 mL of fresh LB broth. Approximately 10^{8} cfu mL^{-1} *S. coelicolor* spores were heat shocked at 50 °C for 10 min. 500 µL of the heat shocked spores were added to 500 µL of the *E. coli* cells in LB broth. The mixture was centrifuged briefly, the supernatant was removed and the pellet was resuspended in 100 µL LB broth. A dilution series from 10^{-1} to 10^{-4} was prepared in sterile MilliQ H_2O, with each step in 100 µl. The 100 µL aliquots of each dilution were plated onto MSA + 10mM $MgCl_2$ and incubated at 30 °C for 16 – 20 h. The plates were overlayed with 1 mL of sterile MilliQ H_2O containing 0.5 mg of Nalidixic acid and the appropriate antibiotic for the selection of the plasmid or cosmid. The plates were further incubated for 5 days at 30 °C until antibiotic resistant colonies were observed.

For plasmid conjugation, the colonies were picked and crushed into 50 µL of sterile MilliQ H_2O. The crushed colonies were streaked for single colonies onto fresh MSA plates and incubated at 30 °C for 5 – 7 days until sporulation ensued. Single colonies were picked again, crushed into 50 µL of sterile MilliQ H_2O and the crushed colony was spread onto a fresh MSA plate. The plates were incubated at 30 °C for 5 – 7 days and the spores were harvested. Exconjugates were screened by PCR.

For cosmid conjugation, colonies were picked and crushed into 50 µL of sterile MilliQ H_2O. 5 µL of the crushed colony was plated onto Difco nutrient agar plates containing the transposon specific antibiotic apramycin, with and without kanamycin to identify colonies that had undergone double crossover events to lose the cosmid backbone. Kanamycin sensitive colonies were picked, crushed into 50 µL of sterile MilliQ and streaked for single colonies onto fresh MSA plates and incubated at 30 °C for 5 – 7 days until sporulation ensued. Single colonies were picked into 50 µL of sterile MilliQ, spread onto a fresh MSA plate and the plates were incubated at 30 °C for 5 – 7 days. Exconjugates were screened by PCR and southern blotting.

2.3.10 Generation of complementation plasmids

pTAK29 and pTAK32

The *sco5204* and *sco4934* coding sequences were amplified by PCR from *S. coelicolor* J1929 genomic DNA using primers TK93 and TK94, and TK101 and TK102 respectively (Table A.1). The PCR products were separated by agarose gel electrophoresis and the bands corresponding to the sizes of *sco5204* and *sco4934* respectively, were excised and gel extracted. pIJ10257 vector DNA was restricted with NdeI, separated by agarose gel electrophoresis, excised and gel extracted. The *sco5204* and *sco4934* coding sequences respectively, were cloned into NdeI digested pIJ10257 using the In-fusion Cloning. The resulting plasmids, pTAK29 and pTAK32 respectively were confirmed by sequencing.

pTAK28

The *sco5204* coding sequence + ~ 250 bp upstream was amplified by PCR from *S. coelicolor* J1929 genomic DNA using primers TK85 and TK86 (Table A.1). The PCR products were separated by agarose gel electrophoresis and the band corresponding to the size of *sco5204* + ~ 250 bp was excised and gel extracted. pMS82 vector DNA was restricted with HindIII, separated by agarose gel electrophoresis, excised and gel extracted. The *sco5204* coding sequence was cloned into HindIII digested pMS82 using the In-fusion Cloning. The resulting plasmid, pTAK28 was confirmed by sequencing.

pTAK30

Due to several repetitive sequences in *sco4847*, the gene could not be amplified by PCR directly from *S. coelicolor* J1929 genomic DNA. To simplify the template for PCR, the cosmid St5G8 (*sco4820-sco4860*) from the *S. coelicolor* genome library (kindly provided by Meirwyn Evans, University of Swansea) was used. The St5G8 cosmid DNA was restricted with BamHI. The digest was separated by agarose gel electrophoresis and a 2270 bp product containing the *sco4847* coding sequence was excised and gel extracted. The purified DNA was used as a template for PCR with primers TK97 and TK98 (Table A.1), to amplify the *sco4847* coding sequence. The PCR products were separated by agarose gel electrophoresis and the band corresponding to the size of *sco4847* was excised and gel extracted. pIJ10257 vector DNA was restricted with NdeI, separated by agarose gel electrophoresis, excised and gel extracted. The *sco4847* coding sequence was cloned into NdeI digested pIJ10257 using In-fusion Cloning. The resulting plasmid, pTAK30 was confirmed by sequencing.

2.3.11 Generation of mutated *sco4934* complementation plasmids

Using pTAK32 as a template, the *sco4934* coding sequence was amplified by PCR using primers TK160 and TK161 (Table A.1). The PCR products were separated by agarose gel electrophoresis and the band corresponding to the size of *sco4934* was excised and gel extracted. pGEM7 vector DNA was restricted with SmaI, separated by agarose gel electrophoresis, excised and gel extracted. The *sco4934* coding sequence was cloned into SmaI digested pGEM7 using In-fusion Cloning. The resulting plasmid, pTAK47 was confirmed by sequencing. Site directed mutagenesis, using the Q5 Site-directed Mutagenesis kit (NEB) per the manufacturer's instructions, was used to introduce point mutations into the *sco4934* coding sequence by PCR. pTAK47 was used as the template for the PCR reactions and the mutagenic primers indicated in Table A.1. The mutations are (T(40)A, S(41)A, T(40)A and S(41)A, T(40)V, S(41)V, and T(40) and S(41)V respectively) summarised in Chapter 6, Table 6.5. The resulting plasmids, pTAK48 – pTAK53 respectively were sequenced to confirm the correct mutations. Using each respective plasmid as a template, the mutant *sco4934* coding sequences were amplified by PCR using primers TK101 and TK102. The PCR products were separated by agarose gel electrophoresis and the band corresponding to the size of *sco4934* was excised and gel extracted. pIJ10257 vector DNA was restricted with NdeI, separated by agarose gel electrophoresis, excised and gel extracted. The mutant *sco4934* coding sequences were cloned into NdeI digested pIJ10257 using In-fusion Cloning. The resulting plasmids, pTAK48.b – pTAK53.b respectively were confirmed by sequencing.

2.4 Southern blotting

2.4.1 Alkaline transfer

Approximately 1-2 µg of genomic DNA was digested overnight with the appropriate restriction enzyme and separated in a 1 % agarose gel at 50 mV for ~ 4 h. The DNA was depurinated by soaking the gel in 0.25 M HCl for 10 min on an orbital shaker and then rinsed in ddH_2O. The membrane (Zeta -probe GT membranes, BIORAD) was cut to the size of the gel and equilibrated in ddH_2O for 5 min. The transfer was set up by placing a gel tray upside down in a deep dish. Four sheets of blotting paper were cut to the length of the gel tray + 5 cm at each end, and placed on the inverted gel tray. The blotting paper was saturated with

0.4 M NaOH and the bubbles were removed. The dish was filled with 0.4 M NaOH so that the ends of the blotting paper were immersed in NaOH. The blotting paper was further saturated with 0.4 M NaOH and the gel was placed carefully on top. A small amount of 0.4 M NaOH was poured on top of the gel. Plastic wrap was placed over the gel and a window was cut with a razor blade, allowing only the gel to be exposed. The equilibrated membrane was placed onto the gel surface whilst avoiding bubbles and then flooded with 0.4 M NaOH. Two pieces of blotting paper cut to the exact size of the gel, were soaked in ddH_2O and placed onto the gel/membrane stack; one at a time, whilst removing the bubbles. Precut paper towels (~ 8 cm high) were place on the gel/membrane stack. The stack was covered with a plastic plate, with a weight and left overnight for DNA to transfer to the membrane. After transfer, the paper towels were removed, the membrane was washed briefly in 2x SSC and the excess buffer was allowed to drip off.

2.4.2 Preparation of the probe

The probe was generated by PCR, using *S. coelicolor* J1929 genomic DNA as a template and the primers listed in Table A.1 The PCR product was separated in a 1 % agarose gel and the band corresponding to the size of the probe was excised and gel extracted. Probe DNA was labelled per the manufacturer's instructions in the DIG-High Prime DNA Labelling and Detection Starter Kit II (Roche Life Sciences) (*sco5204 mutants*) or the Biotin DecaLabel DNA Labelling Kit (Thermo Fisher Scientific) (*sco4934* mutants).

2.4.2 Southern blot hybridisation and detection

To confirm the generation of *sco5204* mutants, the Southern blot hybridisation and detection procedure was carried out using the DIG-High Prime DNA Labelling and Detection Starter Kit II (Roche Life Sciences) per the manufacturer's instructions. To confirm the generation of *sco4934* mutants, salmon sperm DNA (0.5 mg/mL) was denatured at 100 °C for 5 min and cooled on ice. The denatured DNA (50 µg/mL) was added to pre-hybridisation solution (0.2 mL/cm² membrane) and incubated with the membrane for 2 h at 42 °C. The biotin labelled probe was denatured at 100 °C for 5 min cooled on ice. The denatured probe (100 ng/mL pre-hybridisation solution) was added to the pre-hybridisation solution (60 µL/cm² membrane) and the solution was incubated overnight with the membrane at 42 °C

with shaking. The membrane was washed at room temperature for 2 x 10 min in 2x SSC, 0.1 % SDS (50 mL) and then at 65 °C for 2 x 20 min in 0.1x SSC, 0.1 % SDS (50 mL), both with shaking. The excess liquid was removed from the membrane by briefly placing it on blotting paper. The southern blot detection procedure was carried out using the Biotin Chromogenic Detection Kit (Thermo Fisher Scientific) per the manufacturer's instructions.

2.5 Protein methods

2.5.1 Ammonium sulfate precipitation of *S. coelicolor* culture filtrate proteins.

S. coelicolor spores were germinated (as described in 2.3.1) and inoculated into 2 x 500 mL of F134 liquid medium in 2 l baffled flasks. The cultures were cultivated for 43 h at 30 ^{o}C, with shaking at 180 rpm. The cultures were centrifuged at 4000 rpm for 10 min and the culture filtrate was separated from the pelleted cell mass. The culture filtrate was filtered through a 0.5 uM glass fibre Whatman filter (GE Healthcare Life Sciences) aided by a vacuum pump system at 4 ^{o}C. Per 450 mL of culture filtrate, 167.94 g of ammonium sulfate was added (to prepare a 60 % saturated solution) and the mixture was stirred continuously for 1 h at 4 ^{o}C. The mixture was centrifuged at 20 000 g at 4 ^{o}C for 20 min. The pellet was solubilised in 20 mM Tris-HCl buffer, pH 7.5. The culture filtrate protein was desalted using a HiPrep 26/10 Desalting column (GE Healthcare Life Sciences) on the AKTA Pure chromatography system. The column was washed with ddH_2O and then equilibrated 20 mM Tris-HCl, pH 7.5 buffer with 0.4 M NaCl, two column volumes (CV) each at a flow rate of 10 mL/min. The culture filtrate protein was loaded onto the column and eluted then in 1.5 x CV of 20 mM Tris-HCl, pH 7.5 buffer with 0.4 M NaCl at a flow rate of 10 mL/min.

2.5.2 Preparation of membrane proteins from *S. coelicolor*

The mycelia from *S. coelicolor* cultures that were cultivated in F134 liquid medium were harvested by centrifugation (5 min, 3500 *g,* 4°C) and washed in 20 mM Tris-HCl buffer (pH 8, 4°C). The mycelial pellets were re-suspended in twice the pellet volume of lysis buffer at 4°C [20 mM Tris-HCl pH 8, 4 mM $MgCl_2$, protease inhibitor tablet per volume (Roche) and 1 unit. mL^{-1} Benzonase (Sigma)]. The mycelia were lysed using a manual french press (Thermo Fisher Scientific) at 25 kPsi. The cell debris was removed by centrifugation (30 min, 5525 *g* followed

by 30 min at 12 000–15 000 *g*, 4 °C). Membranes in the supernatant were pelleted by ultracentrifugation (1 h, 100 000 *g*, 4 °C) in a Beckman L7-65 Ultracentrifuge. The membrane pellets were solubilised overnight on ice in 1% (w/v) dodecyl-β-D maltoside (Sigma) dissolved in 20 mM Tris-HCl buffer (pH 8). Any insoluble matter was separated by centrifugation (5 min, 5525 *g,* 4 °C). The solubilised membranes were stored in 50 % (v/v) glycerol at -80 °C.

2.5.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was carried out using the XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Fisher Scientific). Protein samples were prepared by adding 1 part of 4 x RunBlue LDS Sample Buffer (Expedeon) to 3 parts of protein sample. β-mercaptoethanol was added to a final concentration of 5 % (v/v) and the samples were boiled at 80 °C for 10 min. The samples were loaded onto precast RunBlue SDS Protein Gels 4 - 12% (Expedeon) and separated using RunBlue SDS Run Buffer (Expedeon) at 160 V for approximately 1 h. The 10 – 250 kDa protein ladder (NEB) (Figure 2.1) was run alongside the samples as a molecular mass standard. The gels were either prepared for western blotting (section 2.5.4) or stained in InstantBlue Protein Stain (Expedeon) for ~ 2 h. Gels were destained in ddH_2O for ~ 1 h.

2.5.4 Glycoprotein detection using Con A

Semi-dry western transfer, as described by Kurien and Scofield (2006), used in combination with chemiluminescent detection was performed following sample separation by SDS-PAGE. An Amersham Hybond P 0.2 PVDF membrane (GE Healthcare Life Sciences) cut to the size of the gel was prepared by soaking in methanol for 30 s and then rinsing in ddH_2O. The PVDF membrane and 3MM blotting papers were equilibrated in Transfer buffer for 15 min prior to blotting. Blotting was performed using the Trans-Blot® SD Semi-Dry Transfer Cell (15 V, 45 min). The membranes were blocked for 30 min in 10 mL of TBS + 2 % (v/v) Tween 20. The membranes were washed for 2 x 5 min in TBS. The membranes were incubated for 2 h in 12.5 mL of TBS + 0.05 % (v/v) Tween 20, 1 mM $MgCl_2$, 1 mM $MnCl_2$ and 1 mM $CaCl_2$ with Concanavalin A peroxidase conjugate (Con A-HRP) added to a final concentration of 5 µg/mL. For the inhibition of glycoprotein binding, membranes were incubated for 2 h in 12.5 mL of TBS + 0.05 % (v/v) Tween 20, 1 mM $MgCl_2$, 1 mM $MnCl_2$ and 1 mM $CaCl_2$ with Con A-HRP (5 µg/mL) and 200 mM methyl α-D-glucopyranoside. The membranes were washed for 2 x 10

min in TBS + 0.05 % (v/v) Tween 20 and 1 x 5 min in TBS. Under dark room conditions the membranes were incubated in chemiluminescent detection solution for 1 min. After exposure to the blot, X-ray film (GE Healthcare Life Sciences) was incubated for 3 – 5 min in Developer solution (Kodak) and 3 min in Fixer solution (Kodak). The film was rinsed in water and allowed to dry.

2.5.5 Glycoprotein enrichment using Con A affinity chromatography

Con A affinity chromatography was carried out on the on the AKTA Pure chromatography system. All steps were carried out at a flow rate of 5 mL/min. Approximately 10 mL of agarose bound Con A (Vector Laboratories) was packed into an empty XK 16/20 Column tube (GE Healthcare Life Sciences).). The column was washed with 20 mM Tris-HCl, pH 7.5 buffer with 0.4 M NaCl, 5 mM $MgCl_2$, 5 mM $MnCl_2$ and 5 mM $CaCl_2$ until the absorbance at 280 nm was zero. The column was then equilibrated in 5 x CV of binding buffer (20 mM Tris-HCl, pH 7.5, 0.4 M NaCl and 0.1 % (w/v) n-dodecyl β-D-maltoside). The sample was loaded onto the column and the column was washed with 16 CV of binding buffer. Glycoproteins were eluted from the column in 4 x CV of Glycoprotein Eluting Solution for mannose- or glucose-binding lectins (Vector Laboratories, ES-1100) or 200 mM methyl α-D-glucopyranoside. Elution fractions were stored in 50 % (v/v) glycerol at -80 °C.

2.6 Proteomics and Glycomics

2.6.1 Protein identification

The stained band was excised from an SDS-PAGE gel using a scalpel. The protein containing gel slice was cut into approximately 1 mm pieces and placed in a LoBind Eppendorf microfuge tube. The sample was sent to the Proteomics Laboratory of the York Bioscience Technology Facility to be identified by mass spectrometry.

2.6.2 In-gel digestion of *S. coelicolor* glycoproteins.

This method was provided by the Rachel Bates (Bioscience Technology Facility, University of York). *S. coelicolor* glycoproteins (quantities varying from 10 – 20 mg) were prepared in LDS sample buffer as described in section 2.5.3. The samples were separated in a NuPAGE™ 10

% Bis-Tris precast gel (1 mm x 10 well) (Thermo Fisher Scientific) for 7 min at 200 V constant. The gel was stained in InstantBlue Protein Stain for 2 h, and destained in ddH_2O for 1 h. Stained regions of the gel were cut into approximately 1 mm pieces for processing. Gel pieces were destained by washing for 2 x 20 min with 200 µL of 50 % (v/v) aqueous acetonitrile containing 25 mM ammonium bicarbonate ($(NH_4)HCO_3$), then once with 200 µL of acetonitrile for 5 min and dried in a vacuum concentrator for 20 min. The samples were reduced by adding 200 µL of 10 mM dithioerythritol (DTE) in 100 mM $(NH_4)HCO_3$ aq. and incubating 56 °C for 1 h. The supernatant was discarded and the gel pieces were allowed to return to room temperature. The samples were alkylated by adding 200 µL of 50 mM iodoacetamide in 100 mM $(NH_4)HCO_3$ aq. and incubating in the dark for 30 min at room temperature. The supernatant was discarded and the gel pieces were washed in 200 µL of 100 mM $(NH_4)HCO_3$ aq. for 15 min. After the supernatant was discarded, the gel pieces were washed in 50 % (v/v) aqueous acetonitrile containing 25 mM $(NH_4)HCO_3$ for 15 min. The supernatant was discarded and the gel pieces were dehydrated in 200 µL of acetonitrile for 5 min. The supernatant was removed and the gel pieces were dried in a vacuum concentrator for 20 min. For the removal of O-linked glycans, the gel pieces were subjected to a non-reductive β-elimination as described in Section 2.6.3. Following dehydration with acetonitrile, 0.2 µg of sequencing-grade, modified porcine trypsin (Promega) in 25 mM ammonium bicarbonate (enough to cover gel pieces) was added to the gel pieces, and the digest was incubated at 37°C overnight. The supernatant containing digested peptides was retained. The peptides from the residual gel were extracted by adding 200 µL of 50 % (v/v) aqueous acetonitrile for 15 min. The extracts were added to the retained supernatant and the extraction was repeated twice. The combined supernatant was dried in a vacuum concentrator (at a medium setting) and the peptides were reconstituted in 20 µL of 0.1 % trifluoroacetic acid (TFA) in ddH_2O.

For the glycosylation site analysis of SCO4471, 10 µL of the tryptic peptides (in 0.1 % TFA) were further dried in a vacuum concentrator and reconstituted in 10 µL 50 mM ammonium bicarbonate. Endoproteinase Asp-N from *Pseudomonas fragi* mutant strain (Sigma-Aldrich) was dissolved in water to give a concentration of 0.1 mg/mL. A 2 µL aliquot of Asp-N solution was added to the peptide mixture and allowed to incubate at 37 °C overnight. The digested peptides were then acidified to 0.1 % TFA.

2.6.3 In-gel non-reductive β-elimination release of O-linked glycans

This method was supplied by Rachel Bates (Bioscience Technology Facility, University of York). S*. coelicolor* glycoproteins (quantities varying from 10 – 20 mg) were prepared per the procedure for an in-gel trypic digest, to the point where the gel pieces were dried in a vaccum concentrator after reduction and alkylation (Section 2.6.2). For the removal of O-linked glycans as described by Taylor et al. (2006), 300 µL of 25 % (w/v) ammonium hydroxide was added to the dried gel pieces and they were incubated at 45 °C overnight. The supernatant (containing the glycans) was removed and retained, and the gel pieces were washed with 200 µL of 100 mM $(NH_4)HCO_3$ aq. for 15 min. The supernatant was retained and the gel pieces were washed with 200µL of 25 mM $(NH_4)HCO_3$ in 50 % (v/v) aqueous acetonitrile for 15 min. The supernatant was retained, and the gel pieces were washed in 200 µL acetonitrile for 5 min. The supernatant was retained and the gel pieces were dried in a vacuum concentrator for 20 min. The supernatants (containing the glycans) from each step were pooled and stored at – 80 °C. For the analysis of the deglycosylated protein, the in-gel digestion procedure (section 2.6.2) was continued.

2.6.4 Permethylation of *S. coelicolor* glycoprotein glycans

This method was supplied by Rachel Bates (Bioscience Technology Facility, University of York). Clean round bottom, screw-capped glass tubes were flamed using a Bunsen burner to remove any residual organic contamination. Maltose and cellobiose disaccharide standards were supplied in ddH_2O by the Proteomics Laboratory of the York Bioscience Technology Facility. *S. coelicolor* glycoprotein glycans isolated by non-reductive β-elimination (section 2.6.3), and the maltose and cellobiose standards were dried in separate round bottom, screw-capped glass tubes in a vacuum concentrator at a medium heat setting. The dried glycans were dissolved in 0.5 – 1 mL of dimethyl sulfoxide (DMSO) and 2 microspatulas of ground NaOH were added. Using a glass pipette, ~ 5 drops of CH_3I were added, the tubes were recapped and swirled to mix, and then allowed to stand for 10 min at rt. ~ 10 drops of CH_3I were added, the tubes were recapped and swirled to mix, and then allowed to stand for 10 min at rt. ~ 20 drops of CH_3I were added, the tubes were recapped and swirled to mix, and then allowed to stand for 1 h at rt. 1 mL of 100 mg/mL NaS_2O_3 in ddH_2O was added and then ~ 1 mL of CH_2Cl_2 was added. The tubes were mixed by shaking to form an emulsion and the caps were loosened to release gas build up. The upper (aqueous) phase was removed. The lower (CH_2Cl_2) phase was washed with ~ 0.5 mL of MilliQ H_2O three times. The CH_2Cl_2 was then dried under a stream of dry N_2. For analysis by MALDI-FT-ICR-MS or MALDI-TOF-

MS, the glycans were resuspended in 20 µL of 50 % (v/v) aqueous acetonitrile. For the carbohydrate linkage analysis, the *S. coelicolor* glycoprotein glycans were further processed per section 2.6.5.

2.6.5 Carbohydrate linkage analysis

This method was supplied by Rachel Bates (Bioscience Technology Facility, University of York). *S. coelicolor* glycans were isolated by non-reductive β-elimination and permethylated. For hydrolysis, 500 µL of 2M TFA was added to the N_2 dried sample and the tube was incubated at 120 °C for 1 h. After cooling, the sample was dried in a vacuum concentrator at a medium heat setting. 100 µL of propan-2-ol was added and the sample was re-dried in a vacuum concentrator at a medium heat setting. The hydrolysis procedure was repeated once. For reduction, 100 µL of 1M NH_4OH was added to the sample and mixed. 500 µL of 20 mg/mL $NaBD_4$ in DMSO was added, the tube was capped loosely and incubated at 40 °C for 90 min. The reaction was quenched after the addition of 100 µL of glacial acetic acid and contents was mixed until fizzing stopped. For acetylation, 100 µL of anhydrous 1-methylimidazole was added to the sample. 500 µL of Ac_2O was then added, the tube was capped and mixed, and allowed to stand for 10 min. To destroy unreacted Ac_2O, 1.5 mL of ddH_2O was added and the sample was cooled to rt. 1 mL of CH_2Cl_2 was added, the sample was vortexed and phases were allowed to separate for ~ 1 min. The upper phase was removed and the lower phase was washed with 500 µL of MilliQ H_2O; this was repeated a further 9 times. The CH_2Cl_2 was then dried under a stream of dry N_2. The sample was dissolved in 20 µL of CH_2Cl_2. 10 µL of dissolved sample was transferred to a screw-cap vial and analysed by GC-MS. Partially methylated alditol acetates of mannose and glucose (provided by the Proteomics Laboratory of the York Bioscience Technology Facility) were analysed as standards alongside the *S. coelicolor* glycans.

2.6.6 Analysis of di/tri saccharides as permethylated di/tri saccharide alditols

S. coelicolor glycans were isolated by non-reductive β-elimination. Maltose and 3α-mannobiose (Sigma-Aldrich) standards were provided in ddH_2O by the Proteomics Laboratory of the York Bioscience Technology Facility and were prepared alongside the *S. coelicolor* glycans. The samples were dried in round bottom, screw-capped glass tubes in a

vacuum concentrator at a medium heat setting. For reduction, 250 µL of 10mg/mL $NaBD_4$ in 0.5 M NH_4OH was added to the samples. The tubes were loosely capped and incubated at RT for 1 h. The reaction was quenched after the dropwise addition of glacial acetic acid until fizzing had stopped. The samples were dried in a vacuum concentrator at a medium heat setting. After the addition of 1 mL of 10 % (v/v) glacial acetic acid in methanol (MeOH), the samples were mixed and dried in a vacuum concentrator at a medium heat setting; this step was repeated three times. 1 mL of MeOH was added to the samples with mixing and they were dried in a vacuum concentrator at a medium heat setting; this step was also repeated three times. The samples were then methylated exactly as described in section 2.6.4 from the point where the glycans were dissolved in 0.5 – 1 mL DMSO, to the point of drying the CH_2Cl_2 under a stream of dry N_2. The samples were dissolved in 20 µL of CH_2Cl_2. 10 µL of dissolved sample was transferred to a screw-cap and analysed by GC-MS.

2.6.7 GC-MS analysis

This method was provided by Rachel Bates (Bioscience Technology Facility, University of York). GC-MS analysis was carried out on an Agilent Technologies 7890A GC system fitted with a DB-5 fused silica capillary column (30 m x 0.25 mm internal diameter). The carrier gas was helium. For the carbohydrate linkage analysis, the instrument settings were: temperature program: 50 °C (2 min), ramp to 130 °C at 40 °C/min, ramp to 230 °C at 4 °C/min; ionisation method: electron ionisation, +ve ion; electron energy: 70 eV; scanning: linear from *m/z* 50 – 800 over 2 s. For the analysis of partially methylated di/tri saccharide alditols the instrument settings were: temperature program: 100 °C (2 min), ramp to 325 °C at 20 °C/min (7 min); ionisation method: electron ionisation, +ve ion; electron energy: 70 eV; scanning: linear from *m/z* 50 – 800 over 2 s.

2.6.8 LC-ESI-CID-MS/MS analysis

This text was provided by Adam Dowle (Bioscience Technology Facility, University of York). Samples were loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C_{18}, 5 µm trap (180 µm x 20 mm Waters) and a nanoAcquity HSS T3 1.8 µm C_{18} capillary column (75 µm x 250 mm, Waters). The trap wash solvent was 0.1 % (v/v) aqueous formic acid and the trapping flow rate was 10 µL/min. The trap was washed

for 5 min before switching flow to the capillary column. The separation used a gradient elution of two solvents (solvent A: 0.1 % (v/v) formic acid; solvent B: acetonitrile containing 0.1% (v/v) formic acid). The flow rate for the capillary column was 300 nL/min. Column temperature was 60 °C and the gradient profile was liner 2 – 30 % B over 125 mins then linear 30-50 %B over 5 mins. All runs then proceeded to wash with 95 % solvent B for 2.5 min. The column was returned to initial conditions and re-equilibrated for 25 min before subsequent injections. The nanoLC system was interfaced with a maXis HD LC-MS/MS system (Bruker Daltonics) with a CaptiveSpray ionisation source (Bruker Daltonics). Positive ESI- MS & MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing were performed using Compass 1.7 software (microTOF control, Hystar and DataAnalysis, Bruker Daltonics). Instrument settings were: ion spray voltage: 1,450 V, dry gas: 3 L/min, dry gas temperature 150 °C, ion acquisition range: *m/z* 150-2,000, quadrupole low mass: 300 *m/*z, transfer time: 120 ms, collision RF: 1,400 Vpp, MS spectra rate: 5 Hz, cycle time: 3 s, and MS/MS spectra rate: 5 Hz at 2,500 cts to 20 Hz at 250,000 Hz. The collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table, absolute threshold 200 counts, preferred charge states: 2 – 4, singly charged ions excluded. A single MS/MS spectrum was acquired for each precursor and former target ions were excluded for 0.8 min unless the precursor intensity increased fourfold. Tandem mass spectral data were searched against a subset of the NCBInr database containing only *Streptomyces coelicolor* entries (8,578 sequences; 2,791,553 residues) using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.5), through the Bruker ProteinScape interface (version 2.1). Search criteria specified: Enzyme, trypsin or trypsin and Asp-N; Peptide tolerance, 10 ppm; MS/MS tolerance, 0.1 Da; Instrument, ESI-QUAD-TOF; Fixed modifications, carbamidomethyl (C); Variable modifications, oxidation (M) and deamidated (NQ). Samples without glycan removal included the variable modifications Hex_1 to Hex_5 (ST). Deglycosylated samples included the variable modifications: Ser->Dha (S), Ser->Diamino-propanoate (S), Thr->DAb (T), Thr->Diamino-butyrate (T). Results were filtered to accept only peptides with an expect score of 0.05 or lower.

2.6.9 LC-MS/MS analysis on the Orbitrap Fusion hybrid mass spectrometer

This text was provided by Adam Dowle (Bioscience Technology Facility, University of York). Samples were loaded onto an UltiMate 3000 RSLCnano HPLC system (Thermo) equipped with a PepMap 100 Å C_{18}, 5 µm trap column (300 µm x 5 mm Thermo) and an Acclaim PepMap

RSLC, 2 µm, 100 Å, C_{18} RSLC nanocapillary column (75 µm x 150 mm, Thermo). The trap wash solvent was 0.05% (v/v) aqueous trifluoroacetic acid and the trapping flow rate was 15 µL/min. The trap was washed for 3 min before switching flow to the capillary column. The separation used gradient elution of two solvents (solvent A: aqueous 1% (v/v) formic acid; solvent B: aqueous 80% (v/v) acetonitrile containing 1% (v/v) formic acid). The flow rate for the capillary column was 300 nL/min and the column temperature was 50°C. The linear multi-step gradient profile was: 3-10% B over 8 mins, 10-35% B over 125 mins, 35-65% B over 50 mins, 65-99% B over 7 mins and then proceeded to wash with 99% solvent B for 4 min. The column was returned to initial conditions and re-equilibrated for 15 min before subsequent injections. The nanoLC system was interfaced with an Orbitrap Fusion hybrid mass spectrometer (Thermo) with a Nanospray Flex ionisation source (Thermo). Positive ESI-MS and MS^2 spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were: ion spray voltage, 2,200 V; sweep gas, 0 Arb; ion transfer tube temperature; 275°C. MS^1 spectra were acquired in the Orbitrap with: 120,000 resolution, scan range: *m/z* 375-1,500; AGC target, $4e^5$; max fill time, 100 ms; data type, profile.

Four distinct MS^2 strategies were employed as detailed below:

ETD_IT

MS^2 spectra were acquired in the linear ion trap specifying: quadrupole isolation, isolation window, *m/z* 1.6; activation type, ETD; reaction time, 50 ms; reagent target, 1e6; maximum ETD reagent inject time, 200 ms; scan range, normal; scan rate, rapid; first mass, *m/z* 110; AGC target, $5e^3$; max injection time, 100 ms; data type, centroid. Data dependent acquisition was performed in top speed mode using a 1 s cycle, selecting the most intense precursors with charge states 3-8. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at $5e^4$.

EDT_OT

MS^2 spectra were acquired in the Orbitrap specifying: quadrupole isolation, isolation window, *m/z* 1.6; activation type, ETD; reaction time, 50 ms; reagent target, 1e6; maximum ETD reagent inject time, 200 ms; scan range, normal; Orbitrap resolution, 30,000; first mass, *m/z* 110; AGC target, $5e^3$; max injection time, 100 ms; data type, centroid. Data dependent acquisition was performed in top speed mode using a 3 s cycle, selecting most the intense precursors. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at $5e^4$.

HCD_IC

MS^2 spectra were acquired in the linear ion trap specifying: quadrupole isolation, isolation window, *m/z* 1.6; activation type, HCD; collision energy, 32%; scan range, normal; scan rate, rapid; first mass, *m/z* 110; AGC target, $5e^3$; max injection time, 100 ms; data type, centroid. Data dependent acquisition was performed in top speed mode using a 3 s cycle, with most intense precursors selected. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at $5e^3$.

HCD/ETD_IC

Precursors were sequentially selected and fragmented by both HCD and ETD. HCD spectra were acquired in the linear ion trap specifying quadrupole isolation, isolation window, *m/z* 1.6; activation type, HCD; collision energy, 30%; scan range, normal; scan rate, rapid; first mass, *m/z* 110; AGC target, $1e^4$; max injection time, 60 ms; data type, centroid. ETD spectra were acquired in the Orbitrap specifying: quadrupole isolation, isolation window, *m/z* 1.6; activation type, ETD; EThdD SA collision energy (15%), maximum ETD reagent inject time, 120 ms; scan range, normal; Orbitrap resolution, 60,000; first mass, *m/z* 120; AGC target, $5e^4$; max injection time, 200 ms; data type, centroid. Data dependent acquisition was performed in top N mode using a 20 precursor cycle for charge states 3-8. Highest charge state then most intense were set as selection priorities. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at $5e^3$.

Peak lists were generated in MGF format using Mascot Distiller (version 5, Matrix Science), stipulating a minimum signal to noise ratio of 2 and correlation (Rho) of 0.6. MGF files were searched against the *Streptomyces coelicolor* subset of the NCBInr database (8,578 sequences; 2,791,553 residues) using a locally-running copy of the Mascot search program (Matrix Science Ltd., version 2.5.1). Search criteria specified: Enzyme, trypsin; Fixed modifications, carbamidomethyl (C); Variable modifications, Hex (S,T), Hex_2 (S,T), Hex_3(S,T) and oxidation (M); Peptide tolerance, 10 ppm. MS/MS tolerance was set to 0.5 Da for linear ion trap data and 0.05 Da for Orbitrap data. Instrument type was set at ESI-TRAP, ETD-TRAP or CID + ETD as appropriate. Results were filtered to accept only peptides with expect scores of 0.05 or lower.

2.6.10 Assignment of O-glycosylation sites

All glycopeptide spectra with MASCOT expect scores of 0.05 or lower were manually validated. For glycopeptide spectra generated by CID and HCD fragmentation, glycosylation sites were only assigned in cases where only a single glycosylated residue was possible within the glycopeptide. For the site localisations of glycopeptides identified in the ETD_IT and ETD_OT acquisitions, an MD-score cut off of 10 was applied. In matches were the MD-score was greater than 10, the spectra were manually validated to confirm the site localisation.

2.6.11 MALDI analysis

This method was supplied by Rachel Bates (Bioscience Technology Facility). MALDI mass spectra were obtained on the Bruker solariX FT-ICR-MS or the Bruker ultraflex III MALDI-TOF/TOF instrument. 2,5-Dihydroxybenzoic acid (Sigma) made up in 50 % (v/v) acetonitrile, 0.1 % TFA was used as a matrix (20 mg/mL). The permethylated glycan solutions were diluted 1:1, 1:2, 1:5 and 1:10 in the matrix solution and 1 µL of each dilution was spotted onto a ground steel MALDI target plate and air dried. The Bruker solariX FT-ICR-MS instrument was operated in positive mode with a smartbeam™ laser operating at 355 nm. Each individual spectrum was acquired using 30 laser shots and an average spectra setting of 20. solariXcontrol software was used to obtain spectra. MS spectra were acquired over a mass range of *m/z* 154-3500. Precusor ions were selected manually adjusting the collision energy manually to generate fragmentation. Positive-ion MALDI mass spectra were obtained using a Bruker ultraflex III in reflectron mode, equipped with an Nd:YAG smartbeam™ laser. MS spectra were acquired over a mass range of *m/z* 100-3000. Spectra were externally calibrated against an adjacent spot containing 6 peptides (des-Arg^1-Bradykinin, *m/z* 904.681; Angiotensin I, *m/z* 1296.685; Glu^1-Fibrinopeptide B, *m/z* 1750.677; ACTH (1-17 clip), *m/z* 2093.086; ACTH (18-39 CLIP), *m/z* 2465.198; ACTH (7-38 CLIP), *m/z* 3657.929.).

Chapter 3 - Enrichment and detection of glycoproteins in *Streptomyces coelicolor*

Chapter 3 – Enrichment and detection of glycoproteins in *Streptomyces coelicolor*

The most commonly employed technique for glycoprotein characterisation is high performance liquid chromatography (HPLC) coupled to mass spectrometry (MS) (Ongay et al. 2012). Although recent technological advances in mass spectrometry enable the analysis of complex biological samples, the analysis of glycoproteins in complex mixtures remains challenging. This is often because glycopeptides are less abundant and the signal intensities are often lower, than those of non-glycosylated peptides (Geyer and Geyer 2006). Therefore, glycoprotein enrichment and detection strategies are often required prior to their analysis using high-throughput proteomics. Lectins are a diverse group of proteins that interact with carbohydrate moieties attached to glycoconjugates and are most commonly exploited in glycoprotein enrichment strategies (Cummings 1994). The carbohydrate binding specificities of many lectins are well characterised and they are commercially available as conjugates to a range of matrices. The mannose binding lectin concanavalin A (Con A) has been used successfully to detect glycoproteins in mycobacteria, as well as a glycoprotein enrichment strategy (Espitia and Mancilla 1989; Garbe et al. 1993; Gonzalez-Zamorano et al. 2009; Wehmeier et al. 2009).

S. coelicolor has a well characterised protein O-glycosylation pathway that is highly similar to the pathway in mycobacteria (discussed in Chapter 1) yet the glycoproteome is significantly less well-studied. Wehmeier et al. (2009) characterised a single glycoprotein in *S. coelicolor*, a periplasmic phosphate binding protein (SCO4142; PstS), after the enrichment of membranes using Con A sepharose. They suggested that other glycoproteins existed in *S. coelicolor* based on preliminary findings that alluded to the presence of ProQ Emerald staining of glycoproteins in the soluble and membrane associated protein fractions (Wehmeier unpublished).

The aim of this chapter was to develop and validate reproducible methods of enriching and detecting glycoproteins in the *S. coelicolor* culture filtrate and membrane. These methods would form the first steps of a glycoproteomics workflow to enable an in-depth characterisation of the *S. coelicolor* glycoproteome (discussed in Chapter 4).

3.1 Phenotypes of glycosylation deficient *S. coelicolor* strains in F134 medium

A glutamate based phosphate-limited medium (referred to as F134 medium) was selected for the cultivation of the *S. coelicolor* strains in this study. F134 medium was selected because phosphate depletion was previously shown to induce the expression of the *S. coelicolor* glycoprotein SCO4142 (PstS) (Nieselt et al. 2010; Thomas et al. 2012). SCO4142 could therefore serve as a positive control for the glycoprotein enrichment in this study.

The glycosylation deficient DT1025 (*pmt*) and DT3017 (*ppm1*) *S. coelicolor* strains were previously shown to have a small colony phenotype, and increased antibiotic susceptibility compared to the *S. coelicolor* M145 derivative J1929, when grown on Difco nutrient agar (DNA) (Howlett et al. 2016). Additionally, both glycosylation deficient strains were identified as being resistant to infection by the phage ϕC31cΔ25 suggesting that the receptor is a glycoprotein (Cowlishaw and Smith 2001; Cowlishaw and Smith 2002). Prior to attempting glycoprotein enrichment from these strains, I tested the phenotypes associated with the *pmt* and *ppm1* mutants on F134 medium.

S. coelicolor strains J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) were grown on both DNB + 1 % (w/v) agar and F134 + 1 % (w/v) agar in microwell plates for 47 h. On DNB + 1 % (w/v) agar a modest reduction in growth rate was observed for DT3017 (*ppm1*) when compared to J1929 and DT1025 (*pmt*) (Figure 3.1.a). All three strains appeared to have entered stationary phase after 35 h. In contrast, all three strains grew more slowly on F134 medium and appeared to be still growing after 47 h (Figure 3.1.b). J1929 and DT1025 (*pmt*) grew similarly while DT3017 (*ppm1*) displayed a reduced growth phenotype after 30 h. To investigate the colony morphology J1929, DT1025 and DT3017 spores were plated onto DNA (Figure 3.2.a) and F134 agar plates (Figure 3.2.b) and grown for 2-3 days. Colonies appeared on DNA after 2 days, while it took 3 days for colonies to reach the same size on F134 agar. DT3017 displayed a distinct small colony phenotype when compared to J1929 and DT1025 on both DNA and F134 agar. Sensitivity of the three strains to the ϕC31cΔ25 phage was tested on DNA (Figure 3.3.a) and F134 agar (Figure 3.3.b). Large clear plaques were observed on J1929 on both DNA and F134. Some plaques were obtained on the glycosylation deficient mutants on both DNA and F134 with a very low efficiency compared to J1929, and the plaque morphologies were small and turbid. These were seen previously on *pmt* and *ppm1* mutants and have been linked to compensatory host range mutants of the ϕC31cΔ25 phage (Cowlishaw and Smith 2001).

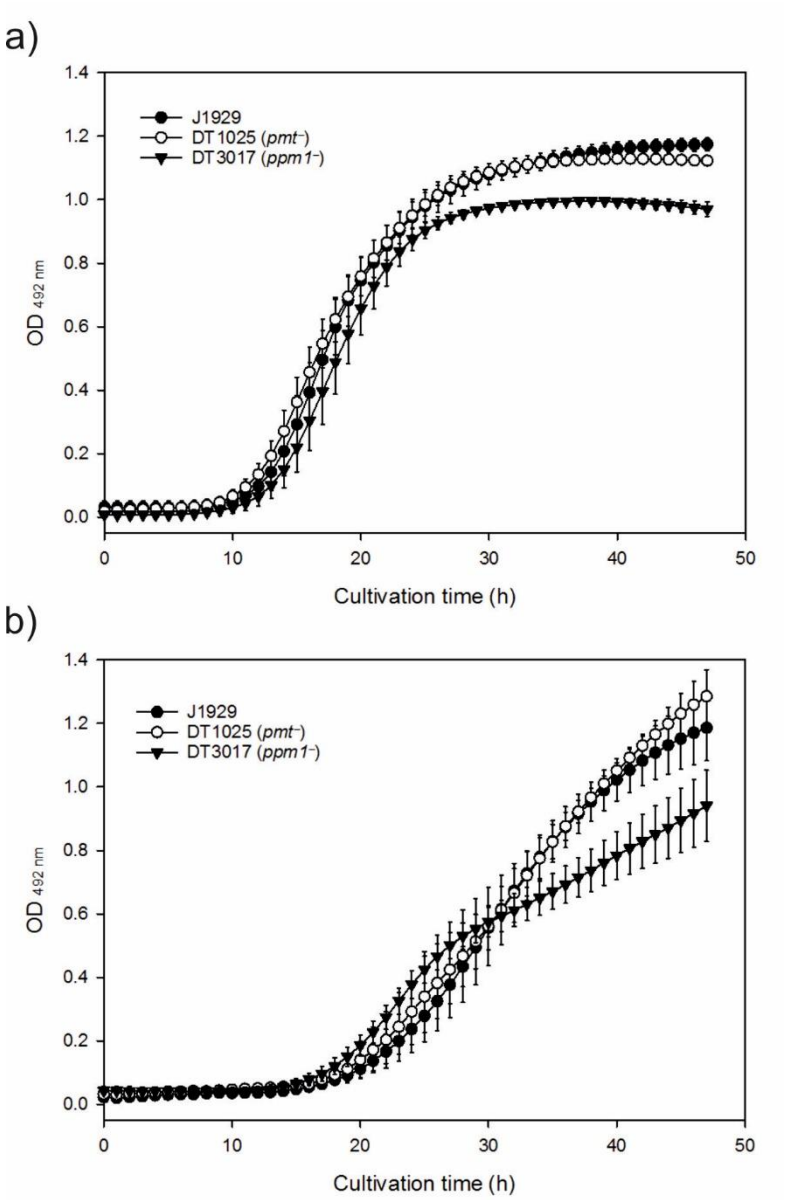


Figure 3.1 The effect of medium on the growth of *S. coelicolor* J1929 and the glycosylation deficient mutant strains DT1025 (*pmt*) and DT3017 (*ppm1*). J1929 (black circle), DT1025 (white circle) and DT3017 (black triangle) were cultivated on DNB + 1 % (w/v) agar (a) and F134 + 1 % (W/V) agar (b) respectively. The strains were cultivated in microwell plates over 47 h and the absorbance at 492 nm was measured on a microplate reader. Error bars represent the standard error of the mean of three biological replicates.

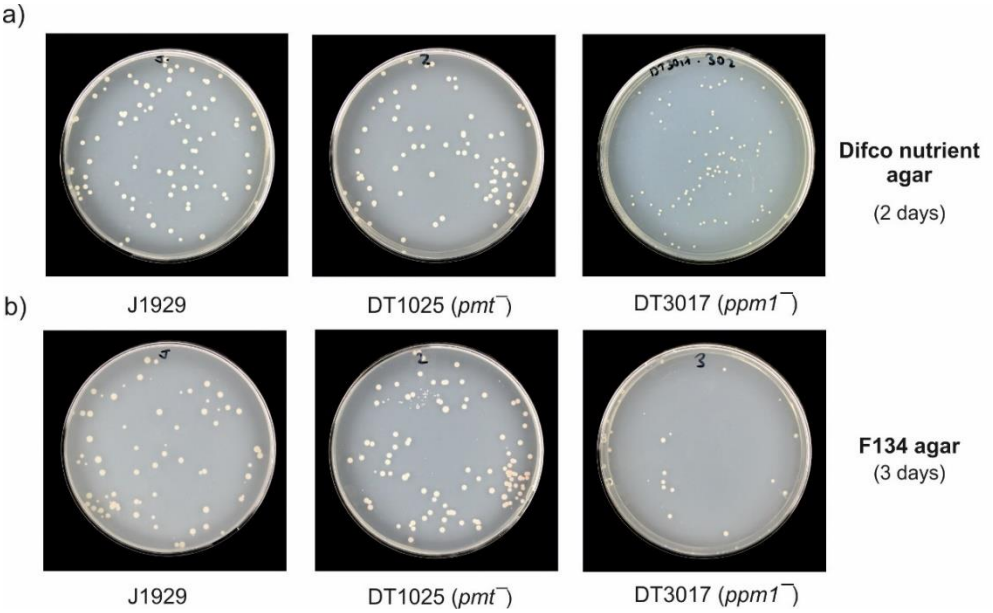


Figure 3.2 The effect of medium on the colony morphology of *S. coelicolor* J1929 and the glycosylation deficient mutants DT1025 (*pmt*) and DT3017 (*ppm1*). Spores were grown on DNA (a) for 2 days and F134 agar (b) for 3 days. Images representative of two biological and two technical replicates.

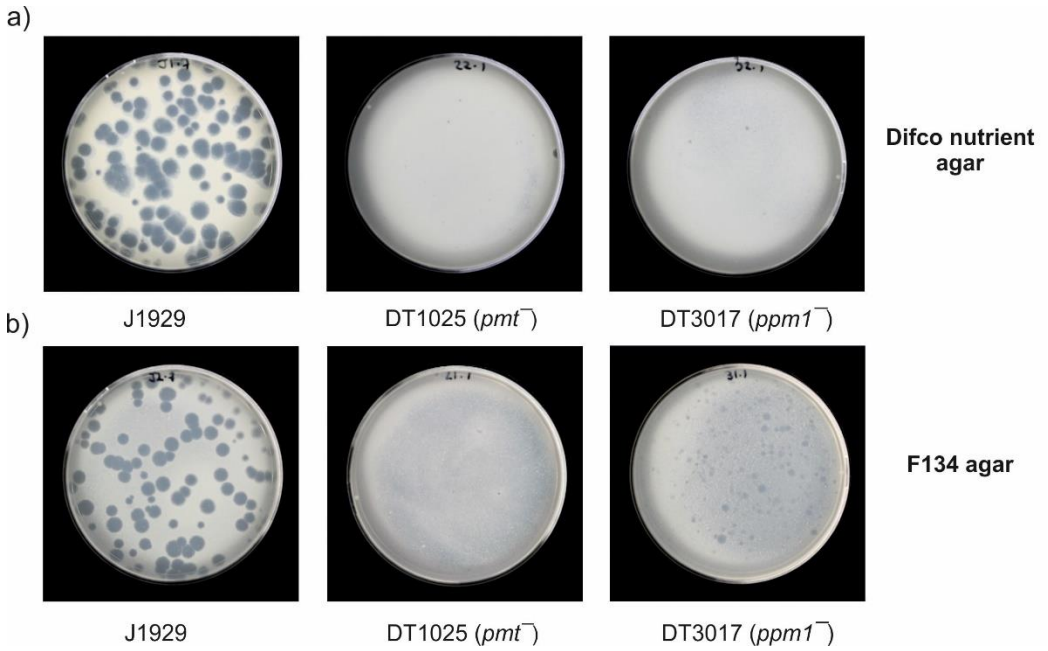


Figure 3.3 Plaque formation by *S. coelicolor* J1929 and the glycosylation deficient mutants DT1025 (*pmt*) and DT3017 (*ppm1*) with the ϕC31cΔ25 phage on DNA (a) and F134 agar (b). ϕC31cΔ25 (~ 1 x 10^3 pfu) was inoculated with J1929, whereas ~ 1 x 10^8 pfu of ϕC31cΔ25 was inoculated with DT1025 and DT3017. Images representative of at least two biological replicates and two technical replicates.

The antibiotic sensitivity of *S. coelicolor* J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) on F134 agar compared to DNA was measured using disc diffusion assays (Figure 3.4; Table A.2 and Table A.3). Three biological replicates were tested at least once against four different concentrations of each antibiotic. On DNA both DT1025 (*pmt*) and DT3017 (*ppm1*) mutants had increased susceptibilities to ampicillin, imipenem, meropenem and vancomycin compared to the parent strain J1929. DT3017 (*ppm1*) mutants also displayed increased sensitivity to teicoplanin, rifampicin, bacitracin and penicillin. F134 suppresses the antibiotic sensitivity phenotype seen with DT1025 (*pmt*) and DT3017 (*ppm1*) grown on DNA, with the exception of vancomycin and penicillin susceptibilities. In summary, the overall growth and phage resistance phenotypes are maintained on F134 medium. However, the antibiotic susceptibilities are remarkably different on F134 compared to DNA, possibly reflecting the defined nature of F134.

3.2 Detection and enrichment of glycoproteins in *S. coelicolor* J1929 that are absent from the glycosylation deficient strains

S. coelicolor J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) strains were cultivated in F134 medium for 25 h after a 12 h spore germination. The membrane proteins were isolated, analysed by SDS-PAGE, blotted onto polyvinylidene difluoride (PVDF) membrane and probed with Con A conjugated to horseradish peroxidase (Con A-HRP) (Figure 3.5). Strong bands of Con A reactivity were observed at ~ 100 kDa and ~ 55 kDa and fainter bands at ~ 90 kDa, ~ 50 kDa and ~ 45 kDa in J1929, which were absent from the glycosylation deficient DT1025 (*pmt*) and DT3017 (*ppm1*) (Figure 3.5.b). The Con A reactivity was lost in the present of methyl α-D glucopyranoside, a competitive inhibitor of mannose and glucose binding. The results suggest the presence of glycoproteins in the membrane of J1929, which are absent from the glycosylation deficient DT1025 (*pmt*) and DT3017 (*ppm1*). In order to identify glycoproteins in *S. coelicolor* lectin affinity chromatography was used enrich for glycoproteins from the membrane and culture filtrate of J1929. The glycosylation deficient DT1025 (*pmt*) served as a negative control for the glycoprotein enrichment. J1929 and DT1025 (*pmt*) were cultivated in F134 medium (2 x 500 mL cultures) for 43 h after a 6 h spore germination. The culture filtrate protein was isolated after ammonium sulfate precipitation and subjected to lectin affinity chromatography using agarose bound Con A (Figure 3.6).

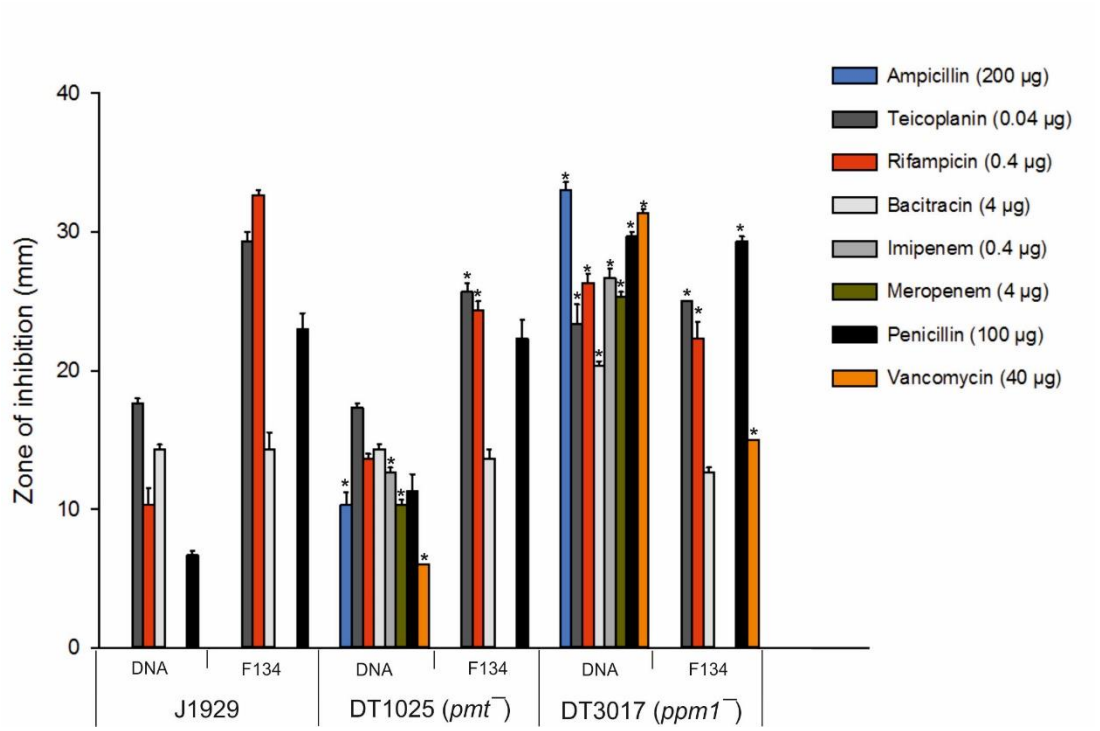


Figure 3.4 Antibiotic susceptibility of glycosylation deficient *S. coelicolor* strains on DNA and F134 agar. Shown are the diameters of the growth inhibition zones (mm) from disc diffusuion assays. Bars represent the mean of three biological replicates with error bars indicating SEM. * indicates $p < 0.05$ that the observed difference between the mutant strains and J1929 has occurred by chance. Only a single antibiotic concentration shown; the full data set is in Table A.2 and Table A.3.

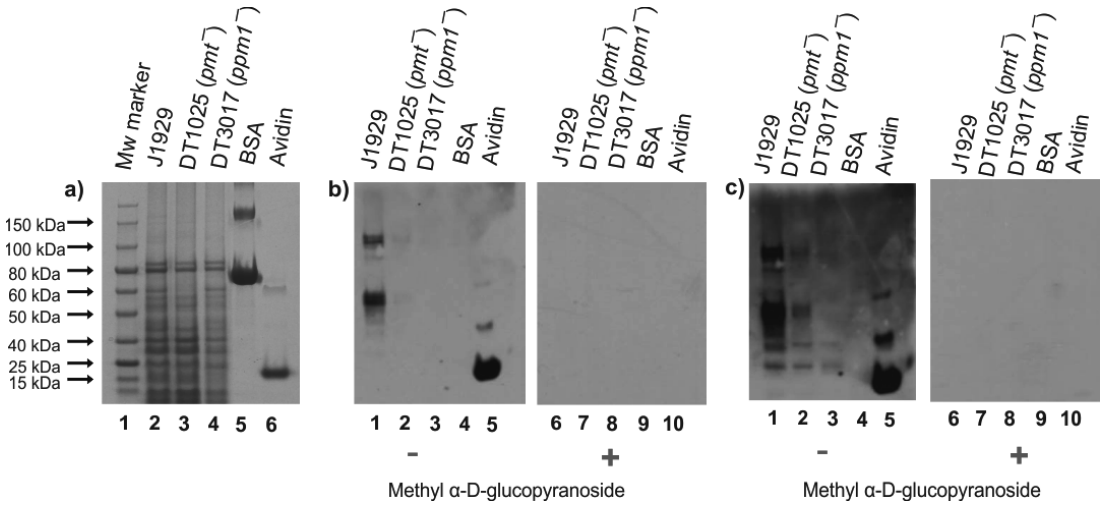


Figure 3.5 Detection of glycosylated proteins in the membrane of *S. coelicolor* J1929 using Con A-HRP. Protein loading was 17 µg for gels stained with protein stain (a) and 5 µg for western blots probed with Con A-HRP (b and c). a). Total membrane protein from J1929, DT1025 and DT3017 (lanes 2 – 4 respectively) was analysed by SDS-PAGE and stained with InstantBlue protein stain. The protein marker was the Broad range 10 – 250 kDa Mw marker (NEB) (lane 1). Bovine serum albumin was a negative control (lane 5) and Avidin was a positive control for the Con A-HRP reactivity. b) and c). Membrane protein was transferred to PVDF and probed with Con A -HRP, a 2 min (b) and 8 min (c) exposure to the membrane shown. Lanes indicate J1929 (lane 1 and 6), DT1025 (lane 2 and 7), DT3017 (lane 3 and 8), BSA (lanes 4 and 9) and avidin (lanes 5 and 10) in panel (b) and (c) respectively. This experiment was repeated at least 3 times with the same results.

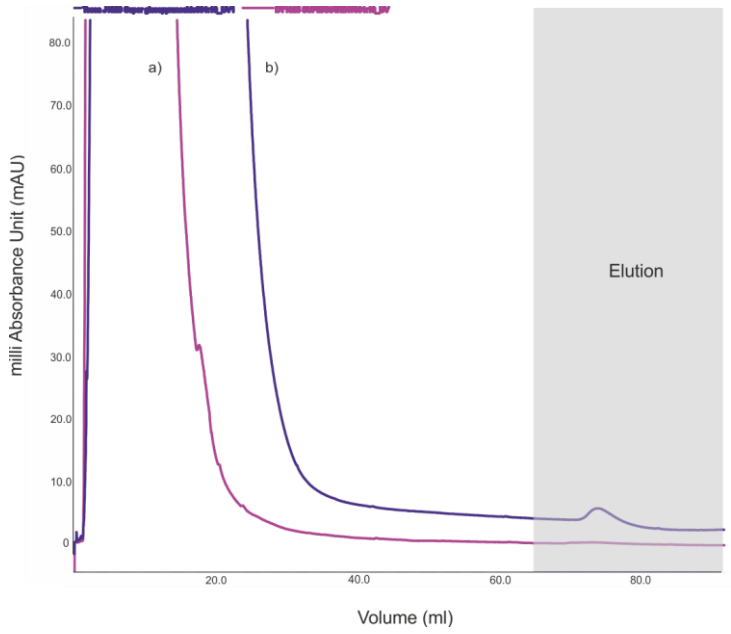


Figure 3.6 Overlayed chromatograms indicating the Con A agarose enrichment of the culture filtrates from J1929 (b) and DT1025 (*pmt*) (a). The grey shading indicates where elution with methyl α-D-glucopyranoside elution buffer was carried out. ~ 5 mg of total culture filtrate was loaded onto the column.

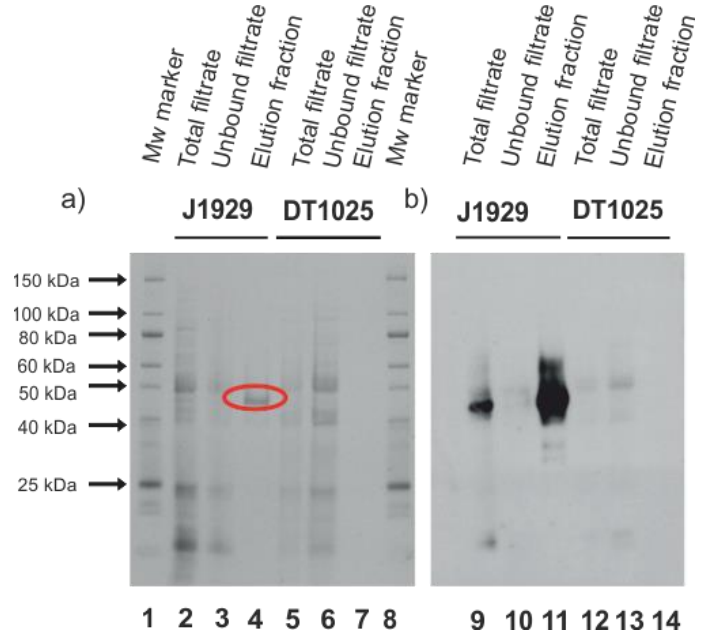


Figure 3.7 Analysis of J1929 and DT1025 (*pmt*) culture filtrate proteins after enrichment using Con A affinity chromatography and separated by SDS-PAGE, stained with Coomassie (a) and Con A-HRP (b). A single Coomassie stained protein was observed in the J1929 retained fraction (red circle). Lanes indicate: 10 – 250 kDa broad range protein ladder (NEB) (lane 1, 8), J1929 total filtrate (lanes 2, 9), J1929 unbound filtrate (lanes 3, 10), J1929 retained fraction (lanes 4, 11), DT1025 total filtrate (lanes 5, 12), DT1025 unbound filtrate (lanes 6, 13) and DT1025 retained fraction (lanes 7, 14).

After the elution of the glycoproteins from the column using methyl α-D-glucopyranoside, a small change in absorbance at 280 nm was observed in the J1929 sample. No change in absorbance at 280 nm was observed in the DT1025 sample. The retained fractions were subsequently pooled and concentrated. The total culture filtrate, unbound and retained fractions were fractionated by SDS-PAGE, blotted onto a PVDF membrane and probed with Con A-HRP (Figure 3.7). A single protein of ~ 45 kDa was observed by protein staining in the retained fraction of J1929 (Figure 3.7.a) which displayed significant Con A-HRP reactivity (Figure 3.7.b). Three additional bands at ~ 30 kDa, 40 kDa and 60 kDa were observed by Con A-HRP reactivity only in the J1929 retained fraction. There was no apparent enrichment of protein in the DT1025 sample. Some background Con A-HRP reactivity was observed in the DT1025 total and unbound fractions.

The glycoprotein enrichment was also carried out on the total membrane protein isolated from J1929 and DT1025 (*pmt*) cultures. Approximately 2 mg of total membrane protein was subjected to Con A affinity chromatography (Figure 3.8). After the elution of the glycoproteins from the column using methyl α-D-glucopyranoside, peaks corresponding to a change in absorbance at 280 nm were observed in both J1929 and DT1025 samples. The total soluble protein, total membrane protein, unbound membrane protein and retained fractions were fractionated by SDS-PAGE, blotted onto a PVDF membrane and probed with Con A-HRP (Figure 3.9). As observed in the culture filtrate enrichment of J1929, a single protein of ~ 45 kDa was observed by Coomassie staining in the J1929 membrane retained fraction. No single protein appeared to be enriched in the DT1025 retained fraction, despite the change in absorbance at 280 nm observed on the chromatogram of the DT1025 enriched membrane. This could suggest that other glycoconjugates are present in the DT1025 membrane. Multiple Con A-HRP reactive bands were observed in the J1929 soluble and total membrane fractions. Several Con A-HRP reactive bands were observed in the elution fraction of the J1929 membrane, including a highly reactive band observed at ~ 45 kDa. Taken together these data demonstrate the presence of a glycoproteome in the *S. coelicolor* membrane and culture filtrate, the synthesis of which requires Pmt.

3.3 Identification and glycosylation site analysis of SCO4471

The ~ 45 kDa proteins enriched from the J1929 culture filtrate and membrane were excised and identified by mass spectrometry (MALDI-TOF/TOF carried out by the Proteomics Department of the York Technology Facility). The protein identification summary of both

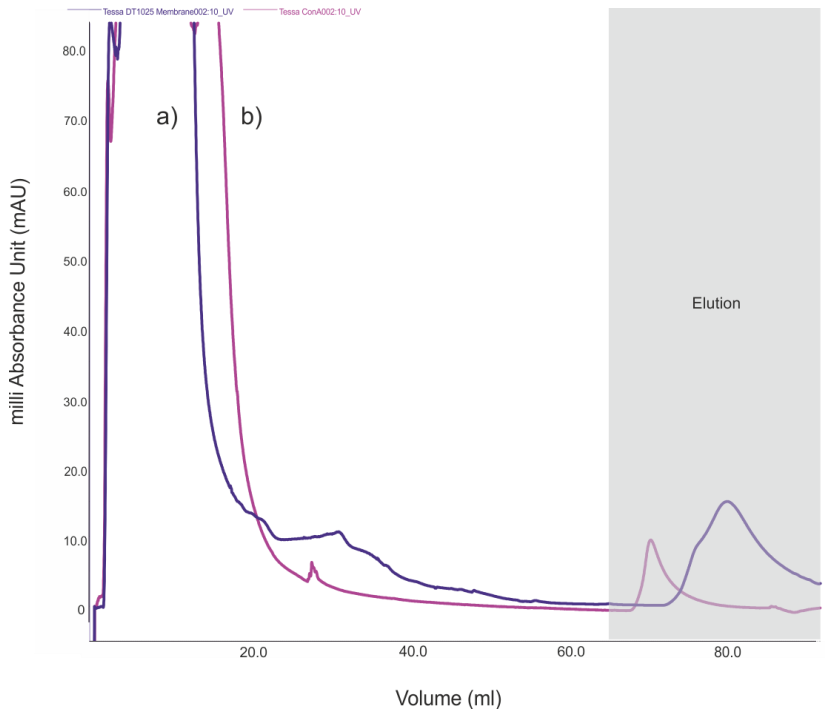


Figure 3.8 Overlayed chromatograms indicating the Con A agarose enrichment of the membrane proteins from J1929 (b) and DT1025 (*pmt*) (a). The elution volume is shaded in grey. ~ 2 mg of total membrane was loaded onto the column.

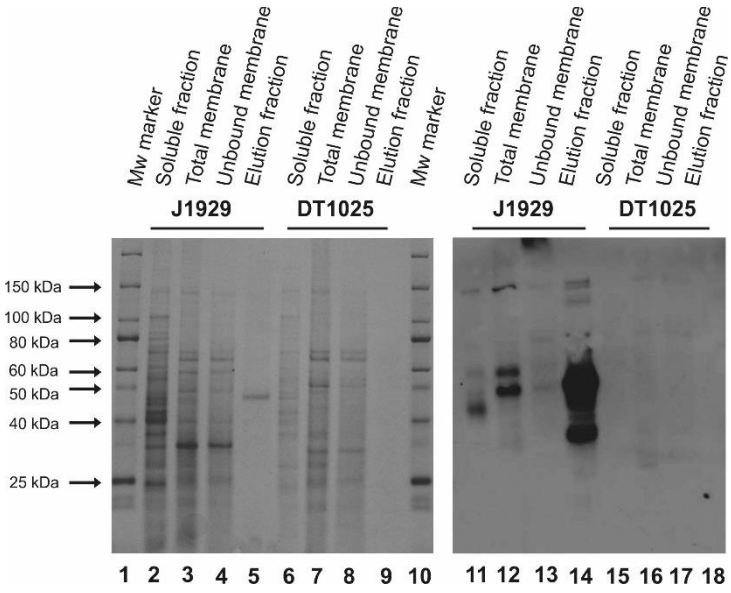


Figure 3.9 Analysis of J1929 and DT1025 (*pmt*) membrane proteins by SDS-PAGE (lanes 1-10) and Con A-HRP (lanes 11 - 18) after enrichment using Con A affinity chromatography. Lanes indicate: 10 – 250 kDa Broad range protein ladder (NEB) (lane 1, 10), J1929 total soluble (lanes 2, 11), J1929 total membrane (lanes 3, 12), J1929 unbound (lanes 4, 13), J1929 elution (lanes 5, 14) DT1025 total soluble (lanes 6, 15) DT1024 total membrane (lanes 7, 16), DT1025 unbound (lanes 8, 17) and DT1025 elution (lanes 9, 18).

bands is shown in Table A.10. Both bands were identified as SCO4471, a putative lipoprotein with a predicted Mw of 47 kDa. The MALDI-TOF/TOF data generated from the protein identifications was searched via MASCOT for glycopeptides modified with Hex, Hex_2 and Hex_3 ; however, none were identified. The protein sequence of SCO4471 was searched through the Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The only hit was the CE4_SF superfamily (e-value: 5.06e-120) which describes the Catalytic NodB homology domain of the carbohydrate esterase 4 superfamily. This family includes chitin deacetylases and bacterial peptidoglycan N-acetyglucosamine deacetylases. The CDD search results suggested that SCO4471 could have a role in the cell wall, making it an interesting target for characterisation. SCO4471 orthologues were found in *Streptomyces venezuelae (S. venezuelae)*, *Streptomyces scabies (S. scabies)*, *Streptomyces griseus (S. griseus)*, *Streptomyces avermitilis (S. avermitilis)* and *Streptomyces lividans (S. lividans)* (Figure 3.10).

Previous work has shown that the NetOGlyc algorithm could successfully predict possible glycosylated peptides in *M. tuberculosis* (Herrmann et al. 2000). NetOGlyc 4.0 was used to scan *S. coelicolor* SCO4471 for amino acids predicted to be O-glycosylated. 15 Putative O-glycosylation sites were predicted while six of the sites appeared to be on residues conserved in more than three *S. coelicolor* orthologues (Figure 3.10).

To identify the predicted O-glycosylation sites, SCO4471 was deglycosylated using non-reductive β-elimination. Non-reductive β-elimination of O-linked glycans from peptides using NH_4OH was previously shown to result in modified serine (S) or threonine (T) residues having a distinct mass (Rademaker et al. 1998). The modified masses can be used as indicators of a previously glycosylated residue. The release of O-linked glycan is a two-step process. First the glycan is released from the S/T residue to yield dehydroalanine (Dha)/dehyrobutyric acid (Dhb), observed as a mass decrease of 18 Da. Subsequently an amine group can be added to Dha/Dhb resulting in diamino-propanoate (S) or diamino-butyrate (T), observed as a mass decrease of 1 Da.

No single enzyme could generate peptides amenable to ionisation covering all of the predicted glycosylation sites in SCO4471. Therefore, the deglycosylated SCO4471 was digested with trypsin and then half of the tryptic digest was further digested with Asp-N. The two digests were pooled and analysed by LC-ESI-MS/MS. This analysis was carried out by the Proteomics Department of the Bioscience Technology Facility. The MASCOT search results are summarised in Table A.11.

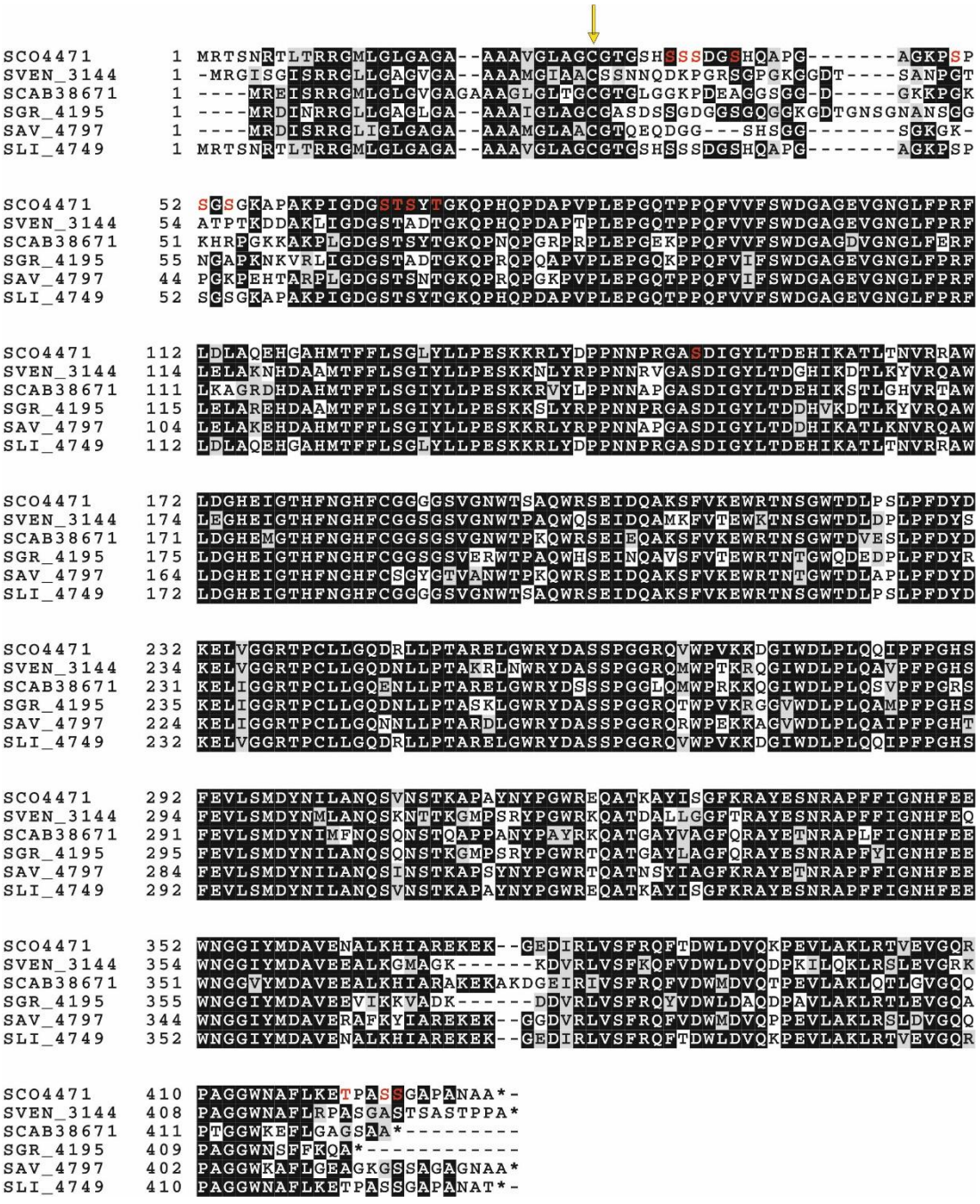


Figure 3.10 Clustal Omega alignment of SCO4471 and O-glycosylation sites predicted by NetOGlyc 4.0. *S. coelicolor* SCO4471 (sco4471) was aligned with orthologues from *S. venezuelae* (SVEN_3144), *S. scabies* (SCAB38671), *S .griseus* (SGR_4195), *S. avermitilis* (SAV_4797) and *S. lividans* (SLI_4749). Residues in red are predicted to be O-glycosylated by NetOGlyc. 4.0. Black and grey shading indicate 100 % amino acid identity and > 50 % amino acid similarity respectively. The lipid attachment site predicted by the LipoP 1.0 software is indicated by a yellow arrow.

Table 3.1 Peptides encompassing predicted glycosylation sites in SCO4471, observed by LC-MS after digestion with trypsin and Asp-N. Residues predicted to be O-glycosylated by NetOGlyc 4.0 are highlighted in red.

Amino acid range	Peptide monoisotopic mass	Peptide sequence	Modified S/T observed
57 - 69	1262.6156	K.APAKPIGDGSTSY.T	none
57 - 72	1548.778	K.APAKPIGDGSTSYTGK.Q	none
59 - 72	1380.6915	P.AKPIGDGSTSYTGK.Q	none
61 - 72	1181.5569	K.PIGDGSTSYTGK.Q	none
148 - 161	1517.7374	R.GASDIGYLTDEHIK.A	none
150 - 161	1389.6783	A.SDIGYLTDEHIK.A	none

Peptides encompassing five of the predicted O-glycosylation sites were observed by LC-MS, however these were unmodified (Table 3.1). These results could not demonstrate glycosylation of the predicted O-glycosylation sites in SCO4471.

3.4 Large scale glycoprotein enrichment

Con A affinity chromatography used in combination with western blotting using a Con A-HRP probe, appeared to be successful in the enrichment and detection of glycoproteins from the *S. coelicolor* J1929 membrane and culture filtrate proteomes in the small-scale trials (Section 3.2). The Con A enrichment was subsequently scaled up to increase the amount of material for downstream glycoproteomics analyses. The membrane proteome was selected for the large scale Con A enrichment because it was easier and more cost effective to isolate large amounts membrane protein, than to isolate the equivalent amount of protein from the culture filtrate. There also appeared to be more different Con A reactive proteins enriched from the J1929 membrane (Figure 3.9, lane 14) than the culture filtrate (Figure 3.7, lane 11) suggesting that an enrichment of the membrane proteome could result in greater glycoproteome coverage.

S. coelicolor J1929 was cultivated in F134 medium (4 x 500 mL cultures) for 43 h after a 6 h germination. The total membrane protein was isolated yielding 50 mg of protein. After the enrichment of the membranes using Con A affinity chromatography the total membrane protein, unbound membrane protein and retained fractions were separated by SDS-PAGE, blotted onto a PVDF membrane and probed with Con A-HRP (Figure 3.11). At least eight protein bands were clearly visible by protein staining in the elution fraction (Figure 3.11.a, lane 4) including a single band of ~ 45 kDa; mostly likely SCO4471. Of these, six bands appeared to be Con A-HRP reactive (Figure 3.11.b, lane 3) corresponding to the molecular masses of approximately 110 kDa, 65 kDa, 45 kDa, 40 kDa, 35 kDa and 30 kDa respectively. Con A reactivity was lost when the membranes were incubated in the presence of methyl α-D glucopyranoside indicating that the Con A-HRP was binding specifically to mannosyl or glucosyl residues.

The six proteins corresponding to the bands of Con A-HRP reactivity (Figure 3.11.a, labelled 1 – 6) were excised and identified by mass spectrometry (MALDI-TOF/TOF carried out by the Proteomics Department of the York Bioscience Technology Facility). The MASCOT search results are summarised in Table 3.2. As expected the ~ 45 kDa protein (band 3) was the

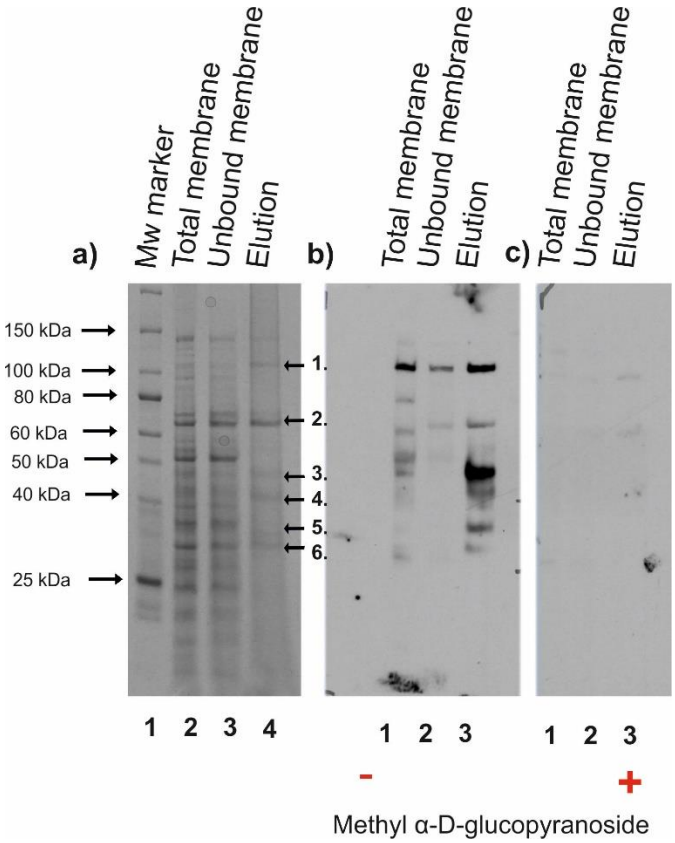


Figure 3.11 Large scale enrichment of *S. coelicolor* J1929 membrane proteins using Con A affinity chromatography. The samples were analysed by SDS-PAGE (a) and western blotting using Con A-HRP as a probe (b and c). Con A-HRP reactivity was lost in the presence of a competitive inhibitor of mannose binding, methyl α-D glucopyranoside (c). The 10 – 250 kDa broad range protein ladder (NEB) was used as a Mw standard (panel a, lane 1). Arrows 1 – 6 indicate protein bands excised and identified by mass spectrometry.

Table 3.2 MASCOT search results summary of the protein identifications of Con A enriched proteins excised in Figure 3.11 and analysed by mass spectrometry. Proteins were classified based on the bioinformatics database searches summarised in Table A.12.

Band no.	Gene Identifier	Genome annotation	Predicted Mw (kDa)	Protein classification	Number of peptides matched	Score (coverage)
1	SCO5204	Putative integral membrane protein	109.5	membrane protein	1	40 (1 %)
2	SCO4856	Putative succinate dehydrogenase flavoprotein subunit	64.7	cytoplasmic protein	4	202 (10 %)
3	SCO4471	Putative secreted protein	46	lipoprotein	3	153 (7 %)
4	SCO4142	Periplasmic phosphate-binding protein	38.3	lipoprotein	1	68 (4 %)
4	SCO6009	Solute-binding protein	39.3	lipoprotein	2	136 (8 %)
5	SCO1796	Putative secreted protein	34.6	membrane protein	1	53 (5 %)
6	SCO5776	Glutamate binding protein	29.5	lipoprotein	2	114 (11 %)

previously identified SCO4471. Band 4 was a mixture of two proteins, the previously characterised *S. coelicolor* glycoprotein SCO4142 (PstS) (Wehmeier et al. 2009) and SCO6009. The protein sequences were searched through various bioinformatics software to predict the presence of transmembrane domains, signal peptides and whether the proteins are lipoproteins (Table A.12). Four of the proteins identified are predicted to be lipoproteins (SCO6009, SCO5776, SCO4142 and SCO4471). All proteins identified are predicted to be localised in the membrane or the periplasm, with the exception of SCO4856. The Con A-HRP reactivity of band 2 (SCO4856) did not appear to increase in intensity in the Con A enriched fraction compared to the total membrane fraction (Figure 3.11.b). This might suggest that SCO4856 was not enriched by the Con A but is merely an abundant protein that bound non-specifically to the column.

3.5 Discussion

S. coelicolor has a well characterised protein O-glycosylation pathway, but little is known about the constitution of the glycoproteome. To better understand the physiological processes affected by protein O-glycosylation in *S. coelicolor* we decided that an in-depth characterisation of the glycoproteome was required. The first step towards glycoproteome characterisation was the optimisation of methods for glycoprotein detection and enrichment. Previous work has shown that a disruption of protein O-glycosylation in *S. coelicolor* results in changes in growth, colony morphology, ϕC31cΔ25 phage sensitivity and sensitivity to antibiotics, when grown on DNA (Cowlishaw and Smith 2001; Cowlishaw and Smith 2002; Howlett et al. 2016). F134 medium was chosen for the cultivation of strains in this study as the timing for the expression of the *S. coelicolor* glycoprotein SCO4142 (PstS) was well known (Nieselt et al. 2010; Thomas et al. 2012) This protein could therefore serve as a positive control for the glycoprotein enrichment in this study. We also hoped to further the characterisation of SCO4142 using glycoproteomics approaches, since glycopeptides mapping this glycoprotein were not previously observed by Wehmeier et al. (2009).

The phenotypes associated with defective glycosylation in *S. coelicolor* were studied on F134 agar to investigate whether these effects were growth medium dependent. We hypothesised that similar phenotypes on both growth media could suggest that glycoproteins associated with these phenotypes on DNA, are also expressed on F134 agar. The overall growth rates of all strains tested were reduced on F134 agar compared to DNA. However, the slow growth phenotype of DT3017 (*ppm1*) previously observed on DNA was

still visible on F134 agar, albeit with altered overall growth kinetics. The small colony phenotype of DT3017 (*ppm1*) compared to J1929 and DT1025 (*pmt*) was evident on F134 agar suggesting that this phenotype is not medium dependent. *S. coelicolor* J1929 was sensitive to infection by the phage ϕC31cΔ25 on both media, suggesting that the phage receptor glycoprotein is expressed when grown on DNA and F134 agar. Similarly, the glycosylation deficient strains were consistently resistant to infection by the phage ϕC31cΔ25 on both media, suggesting that the phage receptor is not glycosylated on DNA or F134 agar. The greatest phenotypic difference between the strains on F134 agar and DNA was observed in the differences in their antibiotic susceptibilities. The overall reduced sensitivity of the glycosylation deficient strains to ampicillin, imipenem, meropenem, bacitracin and vancomycin on F134 agar compared to DNA could be explained by the slow growth rate on F134 agar. β-lactam antibiotics (e.g. ampicillin, imipenem and meropenem) target proteins required for peptidoglycan crosslinking, while bacitracin inhibits steps in peptidoglycan biosynthesis (Stone and Strominger 1971; Waxman and Strominger 1983). Vancomycin acts by binding to the terminal D-alanyl-D-alanine dipeptide of peptidoglycan precursors and inhibiting their incorporation (Barna and Williams 1984). While a defective glycosylation system might alter the activity of cell-wall biosynthetic enzymes or limit cell-wall precursors leading to greater antibiotic sensitivity under rapid growth conditions, under slow growth conditions such as those observed on F134 medium these effects could be reduced. It is however unclear why J1929 and DT1025 (*pmt*) displayed increased sensitivity to penicillin, teicoplanin and rifampicin on F134 agar compared to DNA.

The analysis of the membrane using Con A-HRP after western blotting suggests that glycoproteins are present in *S. coelicolor* J1929, and are absent from the glycosylation deficient DT1025 (*pmt*) and DT3017 (*ppm1*). The concomitant loss of Con A-HRP reactivity in the presence of the competitive inhibitor of mannose binding, methyl α-D glucopyranoside suggests that Con A is binding to the glycan on glycoproteins in the membrane of J1929. This observation is further supported by the fact that glycoproteins were enriched from the *S. coelicolor* J1929 membrane and culture filtrate, but not from DT1025 (*pmt*) after Con A affinity chromatography. The peak eluted from the Con A column after the enrichment of DT1025 (*pmt*) membranes proteins (Figure 3.8), despite the apparent absence of Coomassie stained protein could be explained by the presence of other glycoconjugates in *S. coelicolor* membrane, such as teichoic acids. This is supported by the observation that it is also retained longer than the glycoprotein peak in the J1929 sample.

The Con A agarose (Vector Laboratories) used in this study was far more effective for glycoprotein enrichment than the Con A sepharose used previously by Wehmeier et al. (2009). The use of the latter resulted in a larger amount of non-specific binding, which was not observed with Con A agarose. In this work, the lack of non-specific binding to the Con A agarose was demonstrated in the Con A enrichments of the membrane and culture filtrate proteomes of DT1025 (*pmt*). The use of Con A agarose for glycoprotein enrichment was highly reproducible; all of the Con A-HRP reactive bands observed during the small-scale enrichments corresponded to the molecular masses of proteins observed by Coomassie staining and Con A-HRP reactivity after the large-scale enrichment.

Five candidate glycoproteins were successfully enriched during this study. Included in these was the previously characterised *S. coelicolor* glycoprotein SCO4142 (PstS) (Wehmeier et al. 2009). Four of the proteins are predicted to be lipoproteins, which are well known targets of O-glycosylation in *M. tuberculosis* (Dobos et al. 1995; Sutcliffe and Harrington 2004; Sartain and Belisle 2009). The homologue of SCO5776 in *M. tuberculosis* (Rv0411c) was previously identified by LC-ESI MS/MS after the enrichment of the culture filtrate by Con A affinity chromatography (Gonzalez-Zamorano et al. 2009). SCO4856, a putative succinate dehydrogenase flavoprotein subunit is part of a membrane-located complex facing the cytoplasm. This protein is homologous to the *E. coli* protein SdhA which has a role in the tricarboxylic acid (TCA) cycle (Yankovskaya et al. 2003). It is probable that the enrichment of this protein was non-specific due to its high abundance in the cell. However, the possibility that this protein is glycosylated cannot be excluded.

The identification of SCO4471 is exciting because of its possible role in cell wall biosynthesis. The characterisation of the glycosylated amino acids in SCO4471 using non-reductive β-elimination and mass spectrometry was unsuccessful, but glycosylation of SCO4471 cannot be ruled out. The O-glycosylation target site prediction algorithm NetOGlyc 4.0 identified seven putative glycosylated residues near the N-terminus, as well as three putative glycosylated residues near the C-terminus of SCO4471. It is possible that the C-terminal peptide (ETPASSGAPANAA) had difficulty ionising, because unlike most tryptic peptides the C-terminal amino acid is not basic. O-glycosylation sites in *M. tuberculosis* glycoproteins have been shown to occur more frequently at the C-terminal end of the protein (Smith et al. 2014).

These results provide further evidence of a glycoproteome in *S. coelicolor*. The combination of glycoprotein enrichment using Con A and Con A-HRP as a glycoprotein detection strategy

provide promising first steps towards characterising the *S. coelicolor* glycoproteome. The next step was to validate the glycoproteins using glycoproteomics approaches.

Chapter 4 - A glycoproteomics characterisation of the membrane glycoproteome in *Streptomyces coelicolor.*

Chapter 4 – A glycoproteomics characterisation of the membrane glycoproteome in *Streptomyces coelicolor*.

Glycoproteomics is the study of the modification of proteins with glycan moieties which can alter their activity and physico-chemical properties (Tissot et al. 2009). Mass spectrometry, because of its high sensitivity and selectivity, is arguably the most powerful and relevant analytical technology applied in glycoproteomics strategies. Glycoprotein characterisation by mass spectrometry most frequently involves the analysis of glycopeptides obtained after the proteolytic digestion of the glycoproteins (Wuhrer et al. 2007). Glycopeptide analysis can be complicated because of the combined information generated by the fragmentation of the peptide backbone, the glycan moieties and a combination of both. In order the gain information on the glycan, as well as being able to allocate the glycosylation site within a peptide it is often necessary to combine different mass spectrometry fragmentation techniques.

Collision induced dissociation (CID) fragmentation is the most widely used fragmentation technique for proteome identification. CID fragmentation of peptides occur when they undergo collisions with a neutral target gas and the vibrational energy produced results in ion dissociation, occurring most frequently at the amide bonds along the peptide backbone (Wells and McLuckey 2005). This fragmentation results predominantly in *y*- and *b*-ion series, as well as ions that have undergone a neutral loss ammonia or water (Figure 4.1). CID fragmentation can be used to identify glycopeptides; however the localisation of glycosylation sites using CID spectra is often difficult due to the preferential cleavage of the glycan moiety (Huddleston et al. 1993). This is because the glycan represents the lower energy fragmentation pathway and competes with the fragmentation of the peptide backbone. Higher energy collision dissociation (HCD) fragmentation is a higher energy form of the CID available on the Orbitrap mass spectrometer, that uses higher activation energy and a shorter activation time (Olsen et al. 2007). The fragmentation pattern is similar to that which is produced by CID fragmentation (i.e. *y*- and *b*-ions), but *b*-ions can be more frequently fragmented into *a*-ions and smaller species (Figure 4.1). The collision energies used by both HCD and CID on any instrument, are typically low energy (less than 100 eV) (Wells and McLuckey 2005; Olsen et al. 2007).

Complementary techniques to CID and HCD fragmentation, that are often more successful in the localisation of glycosylation sites, are electron based methods such as electron capture

dissociation (ECD) and electron transfer dissociation (ETD) fragmentation. ETD fragmentation of a peptide occurs when an electron is transferred from a radical ion to a multiply charged peptide cation, resulting in fragmentation at the N-Cα bonds in the peptide backbone. This fragmentation favours cleavage of the peptide backbone, leaving the glycan structure intact and results in the generation of c- and z-type product ions (Huang and McLuckey 2010) (Figure 4.1).

In Chapter 3 I presented further evidence of a glycoproteome in the *S. coelicolor* membrane and validated methods for glycoprotein enrichment and detection. The aim of this chapter was to use the methods developed in Chapter 3 as part of a glycoproteomics workflow to carry out an in-depth characterisation of the *S. coelicolor* membrane glycoproteome using mass spectrometry with CID, HCD and ETD fragmentation techniques. We hypothesised that some glycoproteins in *S. coelicolor* could be required for cell wall biosynthesis or maintaining membrane integrity, and could be of relevance to the antibiotic hypersensitivity phenotypes observed in the *pmt* and *ppm1* strains (discussed in Chapter 1).

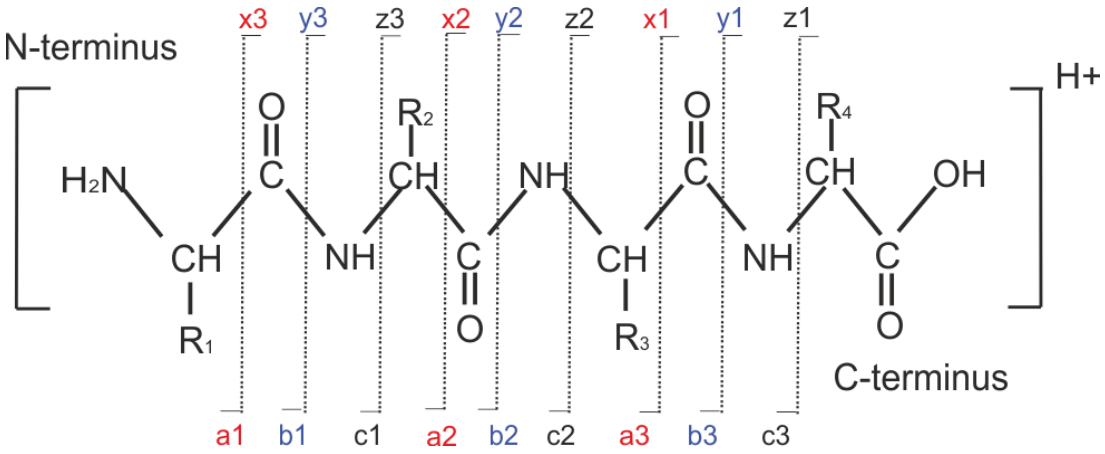


Figure 4.1 Peptide fragmentation as described by Roepstorff and Fohlman (1984). The fragment ions generated by peptide fragmentation observed in MS/MS spectra after using CID, HCD and ETD fragmentation techniques. y- and b-type ions, as well as a- and x-type ions, are generated after CID and HCD fragmentation. c- and z-type ions are observed after ETD fragmentation.

4.1 Time course enrichment of the *S. coelicolor* membrane glycoproteome

In order to maximise our characterisation of the membrane glycoproteome in *S. coelicolor*, I investigated whether the membrane glycoprotein profile changes during the different stages of *S. coelicolor* growth. This would ensure that we focussed on the growth stage where the most glycoproteins were produced. *S. coelicolor* J1929 spores were germinated for 6 h and then cultivated in defined phosphate limited (F134) liquid medium. The cultures were harvested after 20 h, 35 h, 43 h and 60 h of growth with shaking (4 x 500 mL cultures for each time-point). These time-points corresponded approximately to the transition from logarithmic growth into linear growth (20 h), mid-linear growth (35 h), the transition into stationary phase (43 h) and stationary phase (60 h) (Figure 4.2). After 60 h, the cultures were producing antibiotics, indicated by blue/red pigmentation.

The total membrane protein was isolated from the cultures harvested at each time point, yielding 52 mg (20 h), 140 mg (35 h), 106 mg (43 h) and 80 mg (60 h) of protein. After the enrichment of the solubilised membrane proteins using Con A affinity chromatography the total, unbound and elution fractions were separated by SDS-PAGE, blotted onto a PVDF membrane and probed with Con A-HRP (Figure 4.3).

Over the four time points, changes in the abundance and numbers of proteins enriched by Con A affinity chromatography were observed by Coomassie staining. This could suggest that the expression of some glycoproteins in *S. coelicolor* is growth stage dependent. The Con A-HRP reactivity profiles of the elution fractions, which appeared to change over the time course, are consistent with this observation. Only a single Con A-HRP reactive band of ~ 30 kDa was observed in the 20 h elution fraction, while numerous Con A-HRP reactive bands were observed in the 35 h and 43 h elution fractions. After 60 h, many of the Con A-HRP reactive bands observed after 35 h and 43 h were absent. For example, a Con A-HRP reactive protein with a Mw of ~ 110 kDa was enriched from the 35 h and 43 h membranes, but was absent from the enrichment of the 20 h and 60 h membranes. This might suggest that the protein has a role in the linear growth stage in *S. coelicolor*. Similarly, a Con A-HRP reactive protein of ~ 40 kDa appeared to be most highly enriched after 60 h. This protein could be PstS (SCO4142) which was previously shown to increase in expression as *S. coelicolor* M145 cultures approached stationary phase, when grown in F134 medium (Thomas et al. 2012). Some Con A-HRP reactive bands were observed in the unbound fractions and their relative abundances did not appear to change over the time course. It is possible that due to the proteins' tertiary structures, the glycans on these glycoproteins were not exposed to enable

efficient binding to the Con A agarose column. However, after denaturation and separation of the proteins by SDS-PAGE, the glycans could be exposed to enable binding by the Con A-HRP.

Taken together this data suggests that the membrane glycoproteome does change in *S. coelicolor* cultures as they grow. Therefore, to increase the coverage of the glycoproteome in the proteomics analyses, all four samples were selected for analysis.

4.2 A glycoproteomics characterisation of the *S. coelicolor* membrane glycoproteome.

4.2.1 Analysis of glycoproteins using mass spectrometry with CID fragmentation.

In order to demonstrate the presence of glycoproteins in the Con A enriched *S. coelicolor* membrane protein fractions that were isolated in the time course experiment, the proteins were subjected to in-gel tryptic digestion and analysed by mass spectrometry with CID fragmentation (LC-ESI-CID-MS/MS carried out on the Bruker maXis HD system by the Proteomics Laboratory of the York Bioscience Technology Facility). Since the previously characterised *S. coelicolor* glycoprotein PstS was shown to be modified with a trihexose (Wehmeier et al. 2009) and numerous glycoproteins with short mannose modifications have been previously described in the closely related *M. tuberculosis* (Dobos et al. 1996; Michell et al. 2003; Sartain and Belisle 2009), I focussed on short hexose modifications in my analyses.

Peak lists were submitted to two independent Mascot searches. One search allowed for the variable modification of Hex, Hex_2 and Hex_3 addition at S and T residues. The other search considered the variable Hex_4 and Hex_5 modification of the same residues. The searches were split because Mascot limits the number of variable modifications to six meaning that when oxidation (M) and deamidation (N/Q) are added a maximum of four Hex modifications can be included. Although it was anticipated that the modification of peptides would be limited to a maximum of three Hex per residue, the Hex_4 and Hex_5 modifications were incorporated to help aid identification of multiply glycosylated peptides where loss of the glycan is observed upon CID fragmentation.

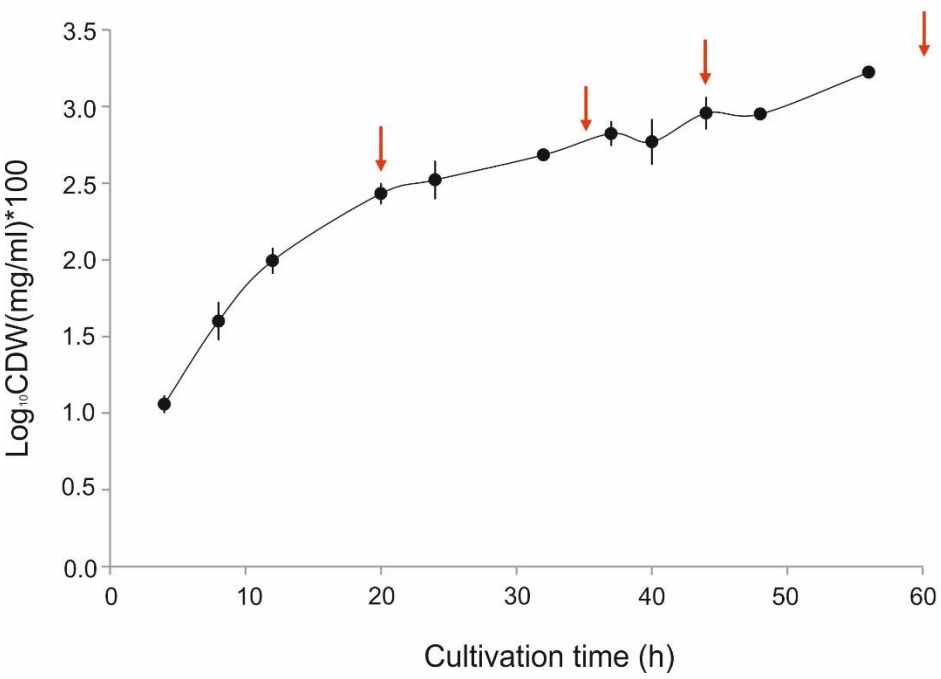


Figure 4.2 Growth curve of *S. coelicolor* J1929 in liquid F134 medium. *S. coelicolor* J1929 spores were germinated for 6 h and the cultures were grown for 56 h in F134 medium. Measurements of cell dry weight (CDW) were taken to monitor growth at regular intervals. The time points selected to harvest the cultures for glycoprotein isolation are indicated by red arrows. Error bars represent the standard error of the mean of three biological replicates.

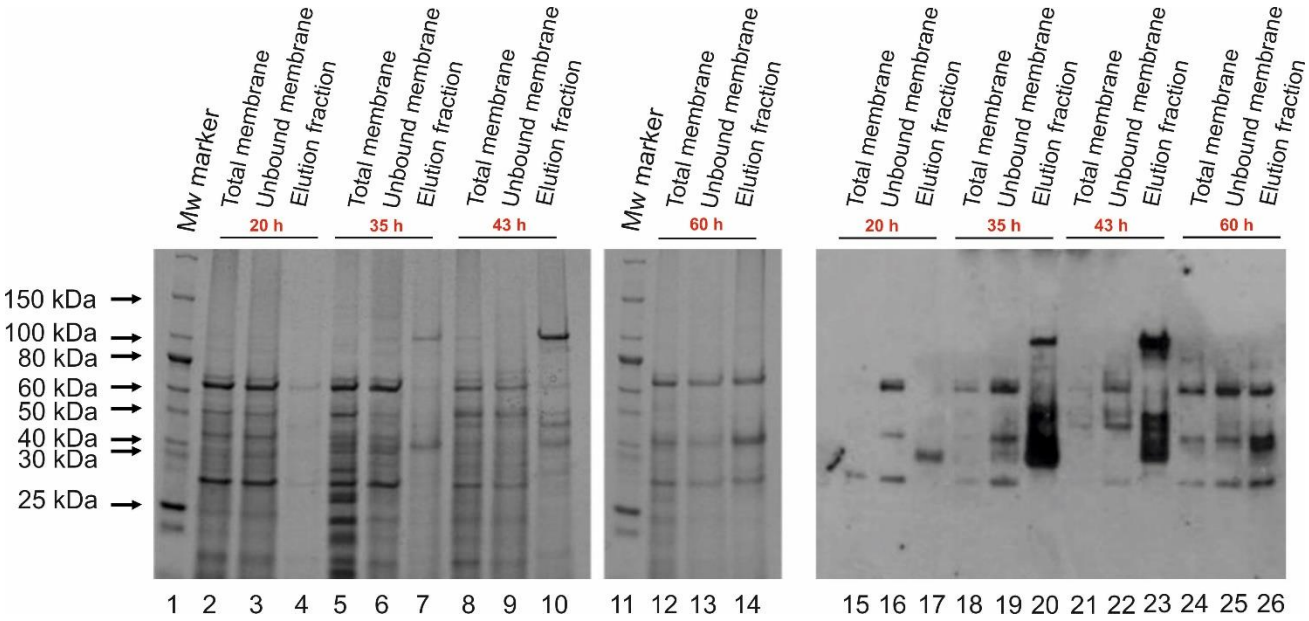


Figure 4.3 A time course glycoprotein enrichment of *S. coelicolor* J1929 membrane proteins using Con A affinity chromatography. The membrane fractions were analysed by SDS-PAGE (lanes 1 - 14) and western blotting using Con A-HRP as a probe (lanes 15 – 26). The protein mw marker (lane 1 and 11) was the Broad range protein ladder (NEB).

For example, if a peptide has two glycosylated amino acids each as Hex_2 then the precursor mass is isobaric with a single Hex_4 addition. The Mascot algorithm does not account for the loss of glycosylation in the product ion spectrum, meaning that there are situations where incorrect assignment of a single Hex_4 modification can generate a higher score than the correct assignment of Hex_2 at two positions. All glycopeptide identifications were manually validated to ensure correct site identification.

As summarised in Table 4.1, 24 *S. coelicolor* glycopeptides were identified. The full list of individual glycopeptides identified is shown in Table A.13. Annotated spectra of the glycopeptides where site allocations were made are shown in Figures A.2 - A.3. The best matching spectrum corresponding to each identification is included in Appendix A.7 (Digital appendix). Approximately 40 % of the glycopeptides identified were observed at multiple time points. Additionally, 60 % of the assignments were supported by more than one spectrum.

The most frequently observed glycopeptide was TEQSASAGGAEESAPAGK (aa 42 – 59) belonging to SCO4739, a putative lipoprotein. This glycopeptide was observed after 20 h, 35 h and 60 h of growth. There was heterogeneity in the number of hexoses modifying this glycopeptide, which was observed with up to nine hexoses. The spectra generated by CID fragmentation were mostly dominated by product ions formed because of the preferential cleavage of glycosidic bonds associated with a glycopeptide. In these cases, the glycopeptide was identified after the mass difference between the peptide backbone identified from the MS/MS spectra, and the precursor ion mass was equivalent to a hexose (162 Da) or multiples thereof. Since the unambiguous assignment of glycosylation sites relies on the observation of peptide product ions containing at least one hexose residue, in many cases it was not possible to map the glycosylation sites in the glycopeptides identified using CID fragmentation.

For example, Figure 4.4 shows the spectrum of the glycopeptide TEQSASAGGAEESAPAGK (aa 42 – 59) modified with 3 hexose residues. This glycopeptide is detected as the doubly charged precursor ion *m/z* 1067.459, corresponding to a mass of 2132.918 Da. The predicted mass of the unmodified peptide TEQSASAGGAEESAPAGK is 1646.738 Da, which is a difference of 3 hexose residues (486.180 Da) from the mass of the glycosylated precursor. A product ion corresponding to the mass of the precursor that has undergone a neutral loss of Hex_3 was observed ($[M+H - Hex_3]^+$ at *m/z* 1647.7495). The spectrum is dominated by the *y*-ion series without the glycan, which give the sequence of the peptide.

Table 4.1 High confidence glycopeptides identified over the four different time points, using mass spectrometry with CID fragmentation. The full list of glycopeptides identified is included in Table A.13. The best matching spectra for each glycopeptide is included in Appendix A.7.

SCO Number	Time point (h)	Peptide	Mascot expect value of the best matching spectrum	# Hex[1]	# MS/MS[2]	Site assignment in peptide
SCO0472	35	GGGSTPSATPAASVQDPLVATFDGGLYILDGK	2.10E-04	9	2	-
SCO0996	20	ATAPSAEGFPVTIDNCGVK	3.00E-05	3	1	-
SCO1714	35	TVTEPAADR	1.90E-02	3	1	-
SCO3357	43	DEGPAHADAVGGAGSASPAPAAK	1.30E-02	6	1	-
SCO3540	35, 43	ATPAELSPYYEQK	5.50E-05	2	2	-
SCO4141	60	TPQPPATEDTRPGR	4.00E-02	1	1	-
SCO4739	20	TEQSASAGGAEESAPAGK	1.10E-08	3	1	-
SCO4739	20	TEQSASAGGAEESAPAGK	4.10E-06	4	1	-
SCO4739	20	TEQSASAGGAEESAPAGK	2.50E-03	5	1	-
SCO4739	20	TEQSASAGGAEESAPAGK	2.50E-04	6	1	-
SCO4739	20	TEQSASAGGAEESAPAGK	1.20E-02	7	2	-
SCO4739	20, 35, 60	TEQSASAGGAEESAPAGK	1.10E-03	8	7	-
SCO4739	20, 35	TEQSASAGGAEESAPAGK	3.40E-04	9	2	-
SCO4847	20, 35	SATAASPSAEASGEAGGTGK	3.00E-04	9	2	-
SCO4905	60	ATEVPTDYGPAPSR	9.10E-03	3	1	-
SCO4905	20	ATPGLPAQVFLLCGSSLVAVDR	1.90E-04	2	1	-
SCO4905	20	ATPGLPAQVFLLCGSSLVAVDR	1.60E-08	3	1	-
SCO4934	35, 60	TSQAEVDEAAAK	1.90E-04	2	2	-
SCO4934	20, 35, 43, 60	TSQAEVDEAAAK	2.40E-04	3	4	-
SCO5115	60	AVDGLSFDLER	8.90E-03	1	1	S6

SCO Number	Time point (h)	Peptide	Mascot expect value of the best matching spectrum	# Hex[1]	# MS/MS[2]	Site assignment in peptide
SCO5204	43	QVQSQFNSEQDIAESIR	1.20E-02	1	1	-
SCO5736	20	EGDTGSPEVQVALLSR	5.00E-04	1	1	-
SCO5818	60	SPHAARLAALVTK	4.10E-02	5	1	S1, T12
SCO6558	43	IPDITLER	1.20E-02	1	2	T5

[1] Indicates the number of hexose residues on the peptide

[2] Indicates the number of high confidence MS/MS spectra confirming the identification.

The b_8 + Hex_3 and b_9(*) + Hex_3 ions suggest that the Hex_3 is in the TEQSASAGG part of the glycopeptide, in which three potential glycosylation sites reside (Thr42, Thr45, Ser47). However, it is unclear from this spectrum which residue/s are modified and how many hexoses reside on each residue. The b_{16} + Hex_2 ion is indicative of a product ion resulting from the neutral loss of a single Hex, since the Hex_3 on the peptide must reside in the TEQSASAGGAEESAPA part of the glycopeptide.

In cases where only a single glycosylated residue was possible within the glycopeptide (i.e. a single S/T) the glycosylation site was assigned by default. As demonstrated in Figure 4.5, the SCO5115 glycopeptide AVDGLSFDLER (aa 38 – 48) modified with a single Hex has only one possible modified residue (Ser43). The doubly charged precursor *m/z* 692.336 corresponds to a glycopeptide mass of 1382.672 Da, which is one hexose (162.052 Da) higher than the mass of the unmodified peptide (1220.603 Da). In this case the assignment is supported by the presence of y_6 + Hex, y_7 + Hex, y_8 + Hex, y_9 + Hex and b_9 + Hex product ions; b_9 + Hex was observed in low abundance. There were three higher intensity product ions in the spectrum that could not be assigned to the glycopeptide (*m/z* 781.381; *m/z* 710.348; *m/z* 595.325). This could indicate a chimeric spectrum, which arises when peptides with similar *m/z* ratios co-elute and are simultaneously fragmented (Houel et al. 2010). Despite this observation, most of the high intensity product ions were assigned to AVDGLSFDLER + Hex and therefore confidence in the assignment is maintained.

4.2.2 Analysis of glycoproteins using mass spectrometry with HCD and ETD fragmentation.

The mass spectrometry analysis using CID fragmentation was successful in the identification of numerous *S. coelicolor* glycopeptides. However, localisation of the glycosylation sites was not possible in many of the glycopeptides, as expected. To increase the number of glycosylation site localisations made, complementary experiments were carried out using both ETD and HCD fragmentation techniques. For this analysis, only membranes isolated after 43 h of growth were prepared and analysed. This time point was selected as a compromise between actively growing cultures and sufficient biomass generation. *S. coelicolor* J1929 was cultivated for 43 h in F134 liquid medium, after the spores were germinated for 6 h (3 x 300 mL cultures pooled). Glycoproteins were enriched from the total membrane protein using lectin affinity chromatography.

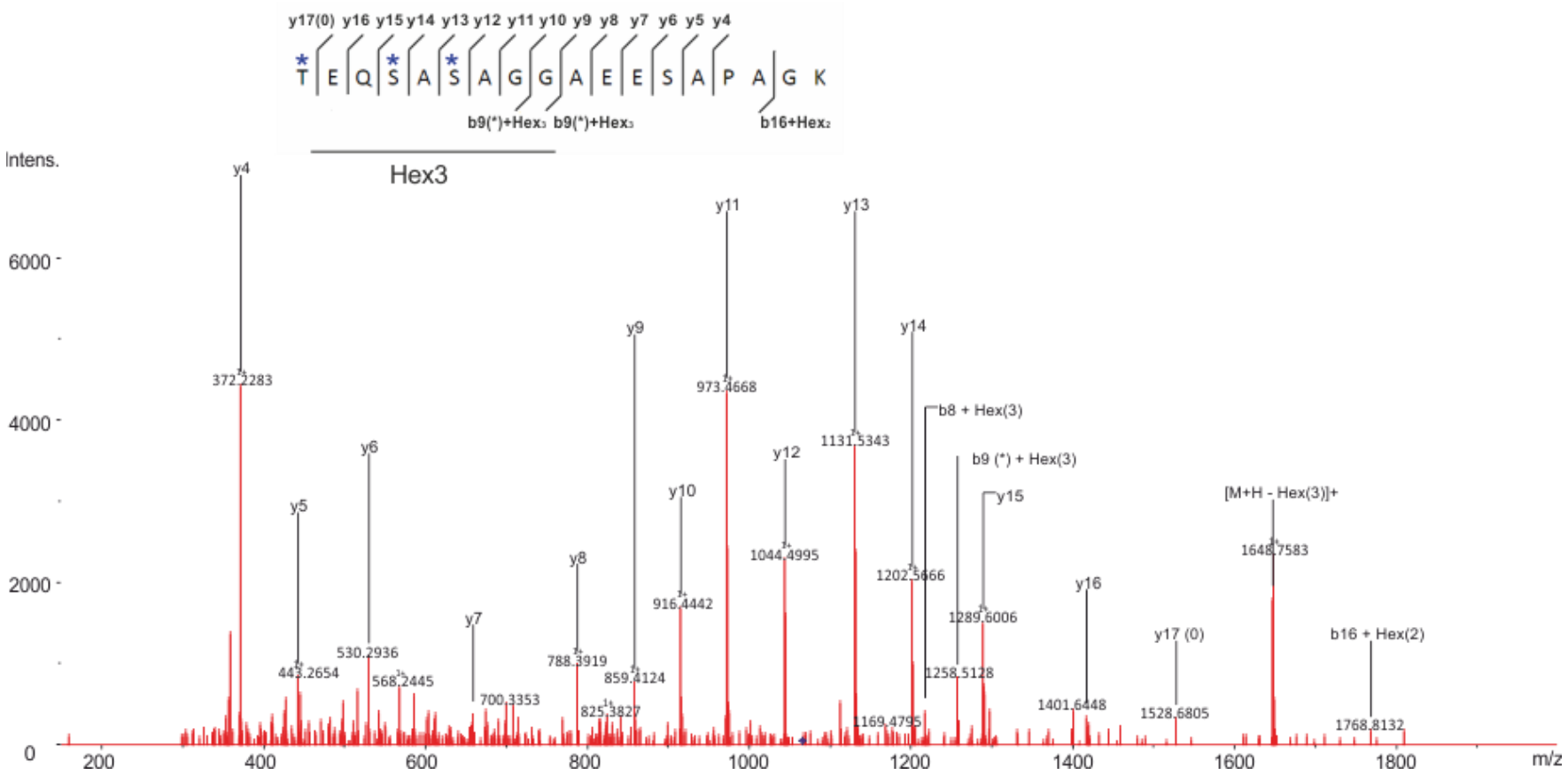


Figure 4.4 CID spectrum of the tryptic peptide TEQSASAGGAEESAPAGK (aa 42 – 59) + Hex$_3$ of the predicted lipoprotein SCO4739 observed at 20 h. Precursor *m/z* 1067.459; charge = 2+; retention time = 25.7 min; e-value = 1.10E-08. The underlined section indicates that the Hex$_3$ is located within the TEQSASAGG part of the glycopeptide. Possible glycosylated residues are indicated with a blue asterix. Product ions that contain hexose residues are indicted + Hex(n), where n = number of hexose residues. The precursor ion is indicated by a ◆ . Product ions exhibiting a neutral loss of H_2O and NH_3 are indicated by (0) and (*) respectively.

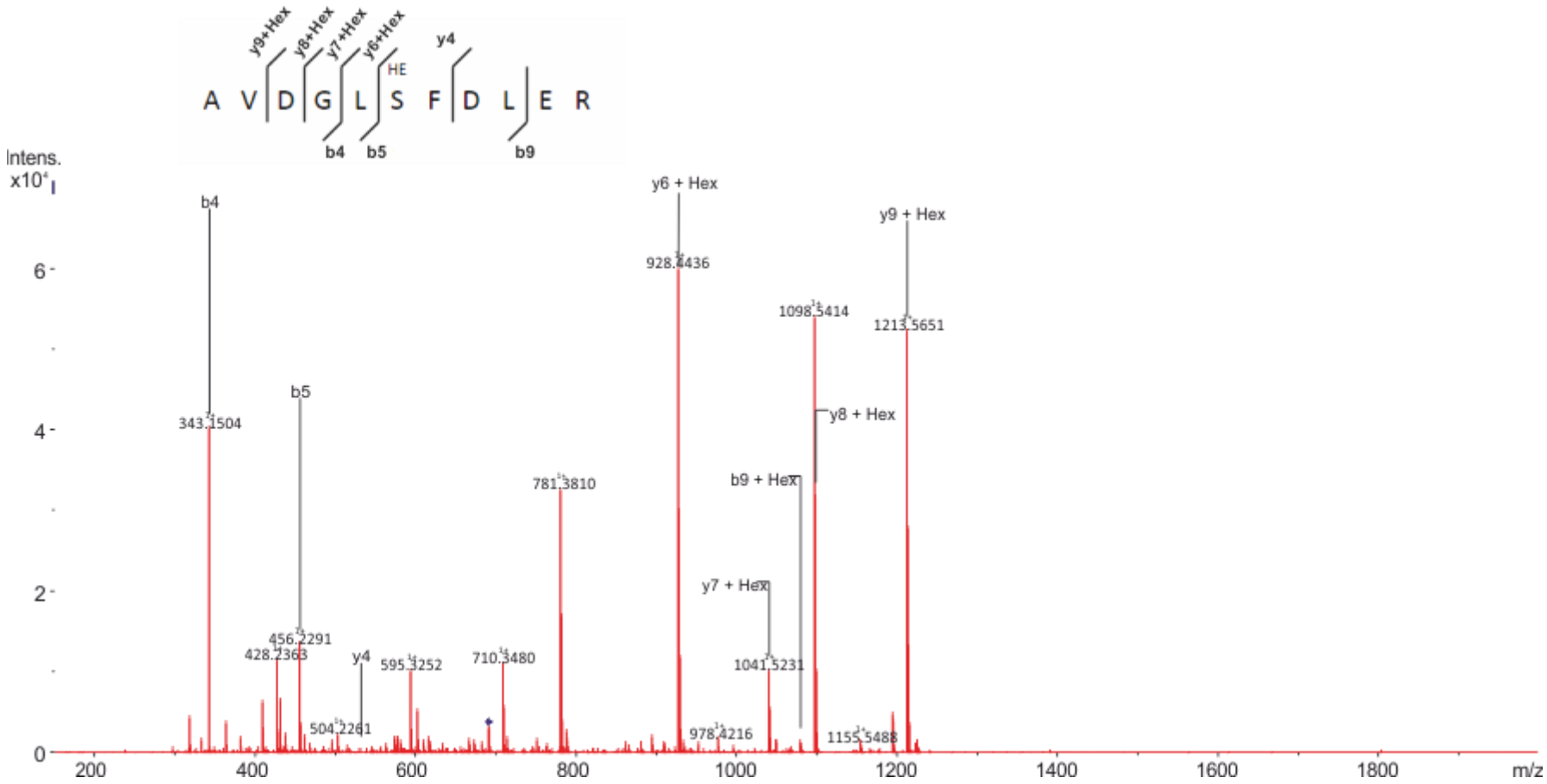


Figure 4.5 CID spectrum of the SCO5115 tryptic peptide AVDGLSFDLER (aa 38 – 48) indicates monohexosylation on Ser43 observed at 60 h. Precursor *m/z* = 692.336; charge = 2+; retention time = 103.6 min; e-value = 8.90E-03. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. The precursor ion is indicated by a ◆ .

Analysis of the elution fraction by SDS-PAGE confirmed the presence of enriched proteins in the elution fraction that approximated to the molecular masses of proteins enriched previously in the time course glycoprotein enrichment. The elution fraction was subjected to in-gel tryptic digestion after SDS-PAGE and analysed by mass spectrometry, using both HCD and ETD fragmentation techniques (carried out on the Thermo Orbitrap Fusion Tribrid mass spectrometer by Adam Dowle in the Proteomics Laboratory, in the York Bioscience Technology Facility).

As in the analysis of the data acquired using the maXis HD with CID fragmentation, we focussed on short hexose modifications (Hex, Hex_2 and Hex_3). Multiple data acquisition techniques are available on the Orbitrap Fusion mass spectrometer. The *S. coelicolor* glycopeptides in this experiment were analysed using both HCD and ETD fragmentation, and using both the ion trap (low resolution, high speed - mass accuracy < 0.5 Da) and Orbitrap (high resolution, lower speed - mass accuracy < 3 ppm) mass analysers. Four different data acquisitions were performed to include these options:

1) MS in Orbitrap, HCD measured in ion trap (HCD_IT)

2) MS in Orbitrap, ETD measured ion trap (ETD_IT)

3) MS in Orbitrap, ETD measured in Orbitrap (ETD_OT)

4) MS in Orbitrap, HCD measured in ion trap and ETD measured in Orbitrap (HCD_IT; ETD_OT)

The combined efforts of the four data acquisitions resulted in the identification of 44 *S. coelicolor* glycopeptides (Table 4.2). These included seven glycopeptides belonging to the previously characterised *S. coelicolor* glycoprotein, PstS (SCO4142). The full list of individual glycopeptides identified by all techniques is shown in Table A.14. Annotated spectra of the glycopeptides where site allocations were made are shown in Figures A.4 – A.15. The best matching spectra (i.e. the match with the lowest e-value) corresponding to each identification is included in Appendix A.7 (Digital appendix).

Approximately 30 % of the glycopeptide assignments were supported by multiple spectra. More than half of the glycopeptides identified (~60 %) were observed after HCD fragmentation only. Glycosylation sites were automatically assigned in peptides where only one possible site was present within the glycopeptide.

Table 4.2 High confidence glycopeptides identified using HCD_IT, ETD_OT and ETD_IT mass spectrometry. The full list of glycopeptides identified is included in Table A.14. The best matching spectra for each glycopeptide is included in Appendix A.7.

SCO Number	Method	Peptide	Mascot expect score of the best matching spectrum	# Hex[1]	# MS/MS[2]	Delta Mascot Score[3]	Site assignment in peptide
SCO0996	HCD_IT	ATAPSAEGFPVTIDNCGVK	8.20E-04	2	1	-	-
SCO0996	HCD_IT	ATAPSAEGFPVTIDNCGVK	2.10E-02	3	1	-	-
SCO2035	HCD_IT	DDGSESAGPVVAPSGAQGK	6.60E-03	2	1	-	-
SCO2096	HCD_IT	KLDACPNESAVAVPVTGDDGPK	5.90E-03	3	1	-	-
SCO2156	HCD_IT	EGTFLGKCAELCGVDHSR	1.50E-03	1	1	-	-
SCO2838	ETD_IT	AAGAGITQQPK	2.00E-03	2	1	Only possible site	T7
SCO2963	ETD_IT	GRGSSDADR	6.50E-03	1	1	0	-
SCO3044	HCD_IT	GDAGQPSDEPAADSEIGVLVQNATR	3.20E-05	3	2	-	-
SCO3046	HCD_IT, ETD_OT	VAKPTPNAAGQTPLNILVIGSDAR	3.10E-06	2	2	32	T5
SCO3184	ETD_IT, ETD_OT	ATVETAAPDRGDGYGVALR	1.10E-05	1	3	4.1	-
SCO3184	HCD_IT	KATVETAAPDRGDGYGVALR	1.40E-03	1	2	-	-
SCO3353	HCT_IT, ETD_IT, ETD_OT	KPSAPECGTPPAGSAK	1.60E-05	2	3	35.2	T9
SCO3353	HCT_IT, ETD_IT, ETD_OT	KPSAPECGTPPAGSAK	7.10E-04	3	3	15.5	T9
SCO3357	ETD_IT	ASPSKAPDRVDAVR	2.90E-02	6	1	0	-
SCO3357	ETD_OT	DEGPAHADAVGGAGSASPAPAAK	7.60E-04	6	1	Manual assignment	S15, S17
SCO3357	HCD_IT	QVYDKGDPVSSPSGENVITYR	2.30E-03	3	1	-	-
SCO3540	HCD_IT	AAGATEAATATLTPLPK	2.20E-05	3	2	-	-

SCO Number	Method	Peptide	Mascot expect score of the best matching spectrum	# Hex[1]	# MS/MS[2]	Delta Mascot Score[3]	Site assignment in peptide
SCO3540	HCD_IT	AAGATEAATATLTPLPK	2.00E-04	2	1	-	-
SCO3540	HCD_IT	ATPAELSPYYEQK	3.90E-04	2	1	-	-
SCO3848	HCD_IT	QGTDVDKESTVNLVVSTGAPK	4.70E-04	1	1	-	-
SCO3848	HCD_IT	QGTDVDKESTVNLVVSTGAPK	5.30E-04	2	2	-	-
SCO3891	HCD_IT	EQQTAIADTFSEGR	9.10E-04	2	1	-	-
SCO4013	HCD_IT	TDAVSPYPLPQSTNK	2.30E-02	1	1	-	-
SCO4130	HCD_IT	TSATAPSGTRPVQSGFAHDAQGA QSAAANYAVALGSDGMFDK	6.50E-05	2	3	-	-
SCO4141	HCD_IT, ETD_OT, ETD_IT	TPQPPATEDTRPGR	2.10E-04	1	4	13	T1
SCO4142	HCD_IT	ADTLPATKSFLNYMASEDGQG LLADAGYAPMPTEIITK	7.80E-04	1	1	-	
SCO4142	HCD_IT	CDDAKGQLQASGSSAQK	4.80E-02	1	1	-	
SCO4142	ETD_OT	DGIKTVDVK	8.10E-03	1	1	Only possible site	T5
SCO4142	HCD_IT	GGQSAQGSSGLAGQVKQTP GAISYFELSYAK	3.60E-05	1	1	-	
SCO4142	ETD_IT	QTPGAISYFELSYAKDGIK	2.80E-03	1	1	24.8	S12
SCO4142	ETD_IT, HCD_IT	TAAAEPVKATVENATAAIGAAK	1.90E-03	1	2	0.8	-
SCO4142	HCD_IT	VCKDGQAIDLPMVGGPIAVG FNVTGVDSLVLDAPTMAK	1.40E-02	1	1	-	-
SCO4256	ETD_OT	GGGGGGGGESKKPKPPVR	1.80E-02	3	1	Only possible site	S10

SCO Number	Method	Peptide	Mascot expect score of the best matching spectrum	# Hex[1]	# MS/MS[2]	Delta Mascot Score[3]	Site assignment in peptide
SCO4307	ETD_IT	LIYAGAGTAGR	6.00E-03	1	1	Only possible site	T8
SCO4548	HCD_IT	TTSSSSSTAPSAPSAPR	4.20E-04	1	1	-	-
SCO4885	HCD_IT	SDQAPEPGFADSPYITVTFR	8.10E-07	1	1	-	-
SCO4905	HCD_IT	ATPGLPAQVFLLCGSSLVAVDR	7.20E-05	2	1	-	-
SCO4905	HCD_IT	ATPGLPAQVFLLCGSSLVAVDR	2.40E-02	3	1	-	-
SCO4968	ETD_OT	VDFKEPAEQDASAGPEAKPQR	9.10E-06	1	2	Only possible site	S12
SCO5646	HCD_IT	AILTKDNPQGDVFFGVDNTLLSR	5.90E-03	1	1	-	-
SCO5751	ETD_OT	KPADPKPEPSDSAIAAAPADKVTVK	4.40E-04	6	2	31.8	S10, S12
SCO5776	HCD_IT	SEKVDFAGPYLLAHQDVLIR	1.30E-03	1	1	Only possible site	S1
SCO7218	ETD_OT	ASSGGHYPVTVENCGEK	6.60E-04	3	2	7.8	-
SCO7218	ETD_IT	ASSGGHYPVTVENCGEKLTFEK	4.70E-03	3	1	6.1	-

[1] Number of hexose residues on the peptide

[2] Number of high confidence MS/MS spectra confirming the match

[3] The difference between the Mascot ion scores for the two best alternative modification sites in the same peptide, assigned by the database search. The MD-score for HCD was not considered to be reliable and therefore is not reported.

For the localisation of sites in glycopeptide spectra generated by HCD fragmentation, a similar approach to the analysis of data acquired by CID fragmentation was taken. HCD yields a similar fragmentation pattern to that which is produced as a result of CID fragmentation on the maXis HD, namely the generation of *b*- and *y*-type product ions (Olsen et al. 2007) Spectra generated by HCD fragmentation were generally dominated by product ions formed as a result of the preferential cleavage of glycosidic bonds associated with a glycopeptide. Glycopeptide identifications were therefore manually validated to assure correct site identification. However, in most cases it was not possible to allocate the glycosylation sites in glycopeptides that were observed by HCD fragmentation only. Five glycopeptides previously identified after CID fragmentation on the maXis HD (Table 4.1) were observed after HCD fragmentation on the Orbitrap (Table 4.2).

For example, the glycopeptide ATPGLPAQVFLLCGSSLVAVDR belonging to SCO4905 (aa 53 – 74) modified with two hexose residues was observed by both HCD and CID fragmentation (Figure 4.6 A and B respectively). The triply charged precursor ions *m/z* 865.778 and *m/z* 865.783 (HCD (A) and CID (B) spectra respectively) are consistent with a glycopeptide mass of 2594.3 Da. The predicted mass of unmodified ATPGLPAQVFLLCGSSLVAVDR with cysteine carbamidomethylation is 2270.1879 Da, which is a difference of 2 hexose residues (324.105 Da) from the mass of the glycosylated peptide. Both spectra are dominated by the *y*-ion series, although none were observed with the glycan attached. In the HCD spectra, more *b*-series ions are observed (b_{10}, b_{11}, b_{12}) in comparison to the CID spectra (b_4, b_{10}), yet again none were observed with a hexose attached. Due to the absence of product ions with the hexose attached, the glycosylation site could not be assigned.

4.2.3 The identification of glycosylation sites

Savitski et al. (2011) previously validated the use of the Mascot Delta-Score (MD-score) as a method of localising phosphorylation sites within 180 synthetic phosphopeptides, in which the modification sites were known. The MD-score measures the difference between the Mascot ion scores for the two best alternative modification sites in the same peptide, assigned by the database search. They suggested using a false localisation rate (FLR) threshold of 1 %, which equated to an MD-score cut-off of 10. Due to the lack of similar validation studies carried out on synthetic glycopeptides with known glycosylation site localisations to date, the study by Savitski et al. (2011) was used as a guide for the analysis of site localisation in the spectra generated by ETD fragmentation in this study.

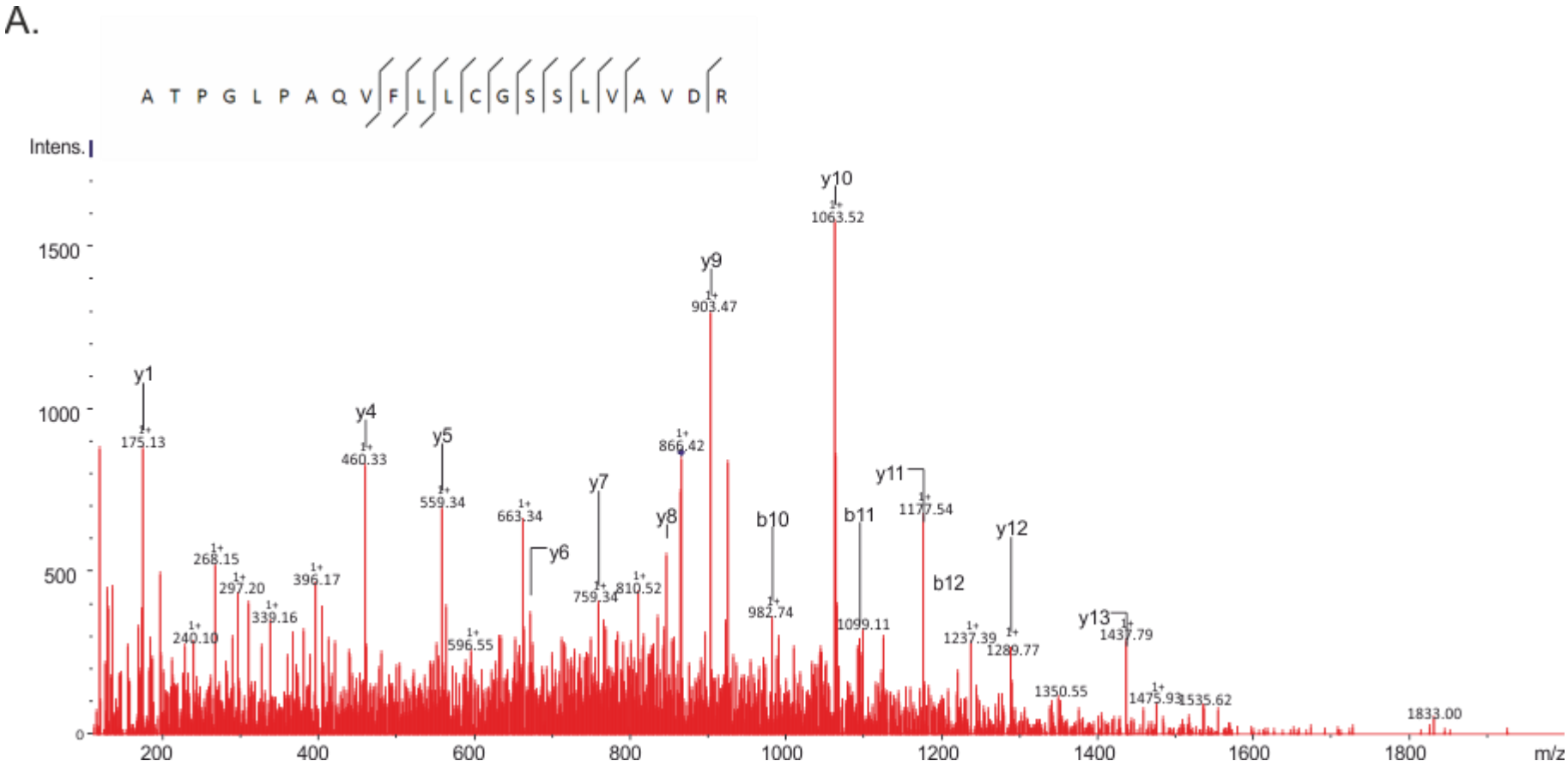


Figure 4.6 A. HCD spectrum of the SCO4905 glycopeptide ATPGLPAQVFLLCGSSLVAVDR (aa 53 – 74) modified with two hexoses. Precursor *m/z* 865.778; charge = 3+; retention time = 133.2 min; scan = 4335; e-value = 7.20E-05. Product ions bearing hexose were not observed. The precursor ion is indicated by a ◆ .

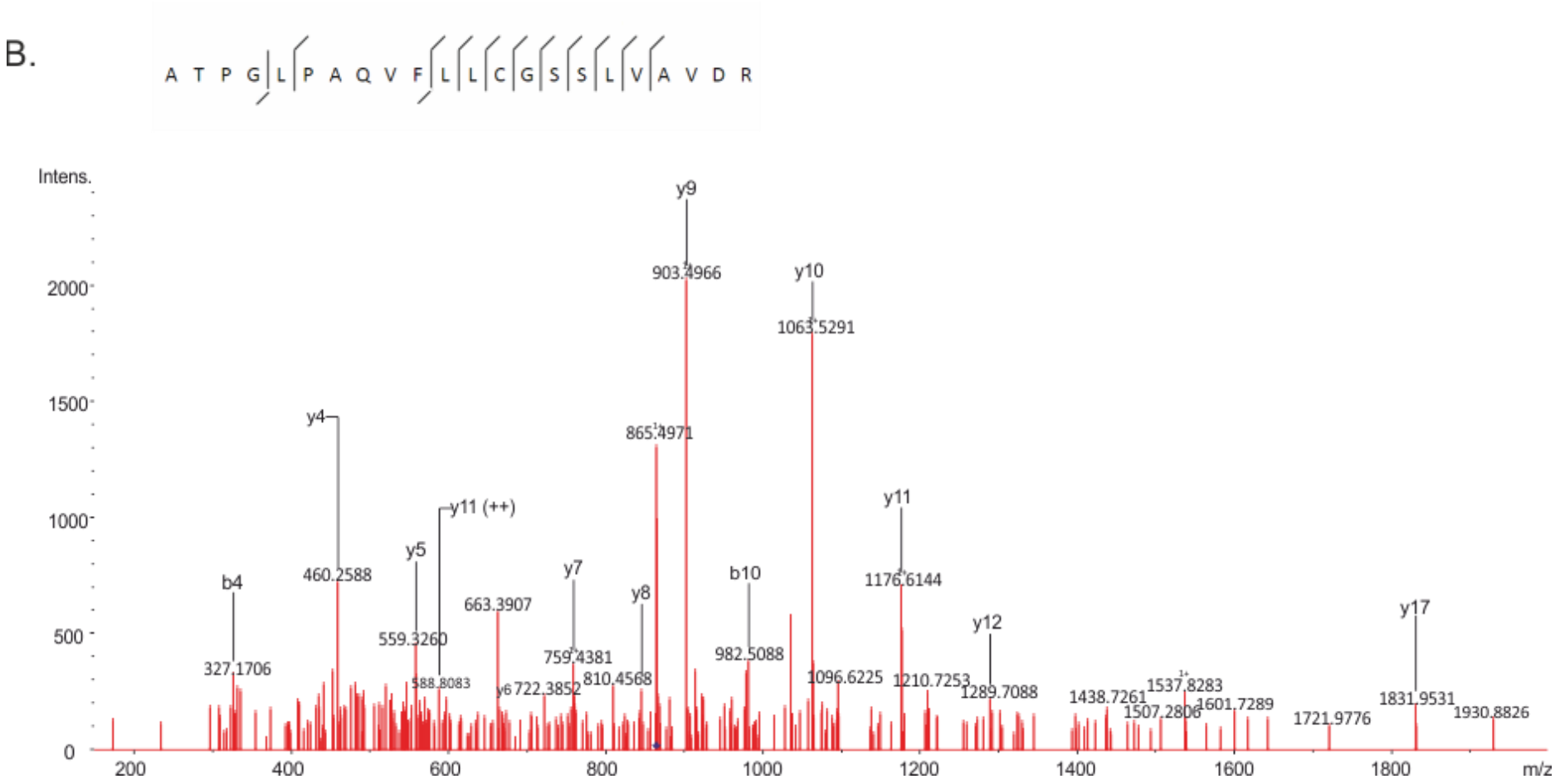


Figure 4.6 B. CID spectrum of the glycopeptide ATPGLPAQVFLLCGSSLVAVDR (aa 53 – 74) modified with two hexoses observed at 20 h. Precursor *m/z* 865.783; charge = 3+; retention time = 122.5 min; e-value = 1.90E-04. Product ions bearing hexose were not observed. The precursor ion is indicated by a ◆. (++) Indicates the doubly charged product ion series.

For the site localisations of glycopeptides identified in the ETD_IT and ETD_OT acquisitions, an MD-score cut-off of 10 was applied. In matches were the MD-score was greater than 10, the spectra were manually validated to confirm the site localisation assignment.

For example, Figure 4. 7 shows the ETD spectrum of the triply charged *m/z* 750.7115 of the SCO4142 (PstS) glycopeptide QTPGAISYFELSYAKDGIK (aa 240 – 258) which is modified with a single hexose. The MD-score for this glycopeptide is 24.8. The z_9 + Hex, z_{10} + Hex, z_{11} + Hex, z_{12} + Hex and z_{13} + Hex product ions, as well as c_{15} + Hex and c_{16} + Hex ion show that Ser251 is modified with a single hexose. Thus, ETD enabled the confident characterisation of the glycosylation site within this glycopeptide.

Despite the previous characterisation of PstS (SCO4142) as a glycoprotein by Wehmeier et al. (2009), this is the first direct evidence of *S. coelicolor* PstS glycopeptides. Wehmeier et al. (2009) previously demonstrated that two synthetic peptides belonging to PstS (SCO4142), PS2 – GQLQASGSSAQKNA and PS3 - PMPTEIITKVRETISGLS were glycosylated in a cell-free peptide glycosylation assay, but not PS1 – GSDDTGGNSGSDSSSAAANSNI. The data in this study is consistent with their findings as glycopeptides were identified that map within the regions covered by PS2 and PS3, but not PS1 (Figure 4.8). Glycopeptide coverage of PS3 was only partial. However, two potentially glycosylated threonine residues reside in the region covered by the glycopeptide. Neither of the glycosylation sites identified in this work resided within PS2 or PS3.

The mass spectrometry analysis using ETD fragmentation greatly increased the number of glycosylation site allocations in the *S. coelicolor* glycopeptides, when compared to the HCD and CID fragmentation techniques. However, HCD fragmentation performed better than ETD in the number of glycopeptide identifications made. ETD fragmentation is more efficient for higher multiply charged precursor ions (3+, 4+, 5+, etc.) and it is probable that some of the precursor ions with lower charged states that were observed after HCD fragmentation, were not selected for ETD fragmentation (Brodbelt 2015).

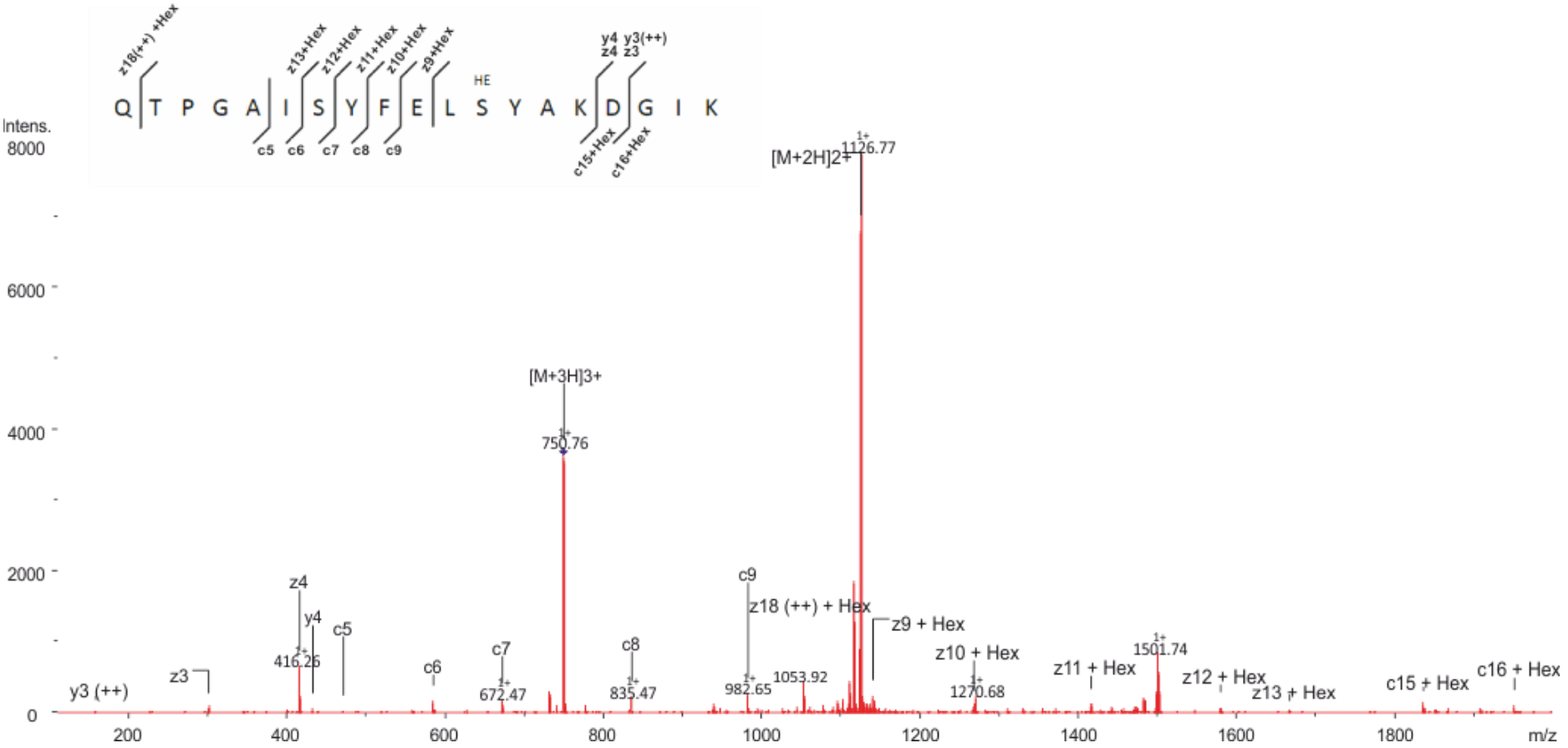


Figure 4.7 ETD spectrum of the SCO4142 glycopeptide QTPGAISYFELSYAKDGIK (aa 240 – 258) modified with a hexose on Ser251. Precursor *m/z* 750.7115; charge = 3+; retention time = 91.5 min; scan = 24203; e-value = 2.80E-03. The precursor ion is indicated by a ◆ . (++) indicates the doubly charged ion series.

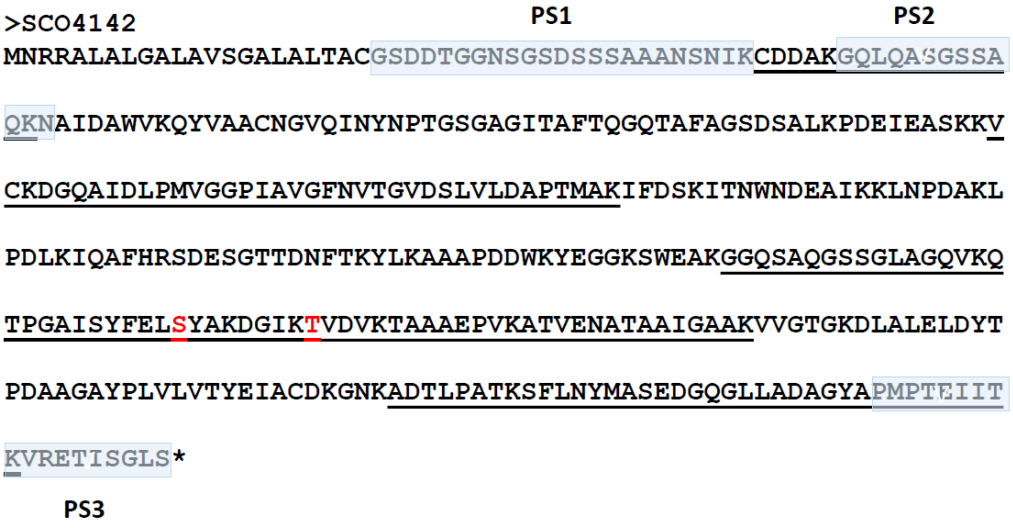


Figure 4.8 The amino acid sequence of PstS (SCO4142) indicating protein coverage by glycopeptides (underlined) in this study. Glycosylated residues identified in this study are highlighted in red. The blue boxes highlight the synthetic peptides (PS1, PS2 and PS3) that were used in the cell-free glycosylation assay by **Wehmeier et al. (2009)**.

4.2.4 The identification of 37 new glycoproteins in *S. coelicolor*

As a result of the combined approaches of using mass spectrometry with CID, HCD and ETD fragmentation techniques, 37 new *S. coelicolor* glycoproteins were identified, as well as the previously characterised *S. coelicolor* glycoprotein PstS (SCO4142). Database searches were carried out in order to classify the proteins as lipoproteins (Table 4.3), membrane proteins (Table 4.4) and secreted proteins (Table 4.5). Proteins that did not contain any predicted transmembrane domains or secretory signals were also grouped together (Table 4.6). Predicted transmembrane domains were identified using TMHMM server 2.0 (Krogh et al. 2001). Predicted lipoproteins were identified using the LipoP 1.0 server (Juncker et al. 2003). Signal peptides were predicted using SignalP 4.1 Server and the TatP 1.0 server (Bendtsen et al. 2005; Petersen et al. 2011). Proteins were functionally annotated using the combined efforts of the *Streptomyces* genome database (StrepDB; strepdb.streptomyces.org.uk/) and the Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al. 2014). In some cases, the literature was contradictory to the results observed after the database searches. For example, SCO7218 is annotated as a putative iron transport lipoprotein in the StrepDB. However, the LipoP 1.0 server did not predict a lipoprotein signal peptide (SpII) in this protein. SCO7218 is upstream of an ABC transporter (SCO7216/SCO7217) which is consistent with the known genome architecture of solute binding lipoproteins in *S. coelicolor* (Thompson et al. 2010). In these cases, the literature searches were considered to be more reliable in assigning a category to the proteins.

Three of the glycoproteins identified in this study (SCO5204, SCO5776, SCO4142) were identified previously as enriched proteins after the large scale Con A enrichment of the *S. coelicolor* membrane described in Chapter 3 (Table 3.2). More than a third of the newly identified glycoproteins are predicted lipoproteins and other secreted proteins. Among these are a number of substrate binding proteins predicted to interact with ABC transporters, for example SCO0472, SCO5776, SCO7218, SCO4885 and SCO4142 (PstS).

Nearly 50 % of the glycoproteins identified in this study are predicted membrane proteins, while a few had no predicted transmembrane domains or secretory signals at all. For example, the predicted ribosomal protein S15 is a homologue of the *E. coli* ribosomal protein RpsO (also known as SecC) that is a known component of the 30S ribosomal subunit (Zimmermann et al. 1972). Additionally, several the predicted lipoproteins and secreted proteins identified in this study are predicted to contain TAT pathway secretory signals.

Taken together, these results directly demonstrate the presence of a glycoproteome in the *S. coelicolor* membrane that includes membrane, secreted, lipoproteins and possibly some intracellular proteins.

4.3 Glycosylation site positional analysis

Several the glycoproteins identified in this work were predicted to contain numerous transmembrane helices (Table 4.4). To demonstrate that the glycosylation sites were not situated within transmembrane helices, the positions of the glycosylated peptides were mapped relative to the predicted transmembrane domains of each respective membrane glycoprotein (Figure 4.9). None of the glycosylated peptides mapped within a predicted transmembrane helix. A small number of glycopeptides mapped within regions predicted to be cytoplasmic facing (SCO2963, SCO4141, SCO4256, SCO4968 and SCO4548), however most were predicted on periplasmic facing parts of the glycoproteins. These results suggest that glycosylation may be possible on the cytoplasmic face of the membrane in *S. coelicolor.*

O-glycosylation sites in *M. tuberculosis* have been observed at a higher frequency nearer the N- or C-terminal parts of the glycoprotein (Dobos et al. 1996; Sartain and Belisle 2009; Smith et al. 2014). To investigate whether this was true for *S. coelicolor* glycoproteins, the positions of the O-glycosylation sites identified in this study were analysed relative to the total number of amino acids in the protein (Figure 4.10). Glycosylation sites were not found to be more frequently located nearer to the N- or C-terminal ends of the glycoproteins.

The amino acid sequence composition of the glycosylation motif was determined by submitting the sequences that contained the confidently allocated glycosylation sites to Weblogo (Crooks et al. 2004). The Weblogo software generates graphical representations of patterns observed within a multiple sequence alignment. The sequence logo resulting from the analysis of eighteen individual *S. coelicolor* glycosylation sites is shown in Figure 4.11, where the height of each amino acid indicates its relative frequency at that position. Glycosylation was observed more frequently on serine than threonine residues. There was a higher propensity for alanine, proline and glycine near the glycosylation site. This is consistent with what has been observed previously in *M. tuberculosis* glycopeptides (Smith et al. 2014). However, it is unclear from this data whether peptides rich in these amino acids are more likely to produce product ions resulting in confident site localisations.

Table 4.3 Predicted lipoproteins identified as *S. coelicolor* glycoproteins in this study

SCO Number	Genome annotation	Top hit in Conserved Domain Database	Accession	E-Value[1]	#TMHMM [2]	SignalP 4.1 [3]	TatP 1.0 [4]	LipoP 1.0 [5]
SCO0472	putative secreted protein	none	-	-	-	Y - 0.548	Y - 0.381	SpII - 22.2623
SCO0996	putative lipoprotein	TroA_a: predicted to function as initial receptors in ABC transport of metal ions in eubacteria	cd01148	4.09E-133	-	Y - 0.526	N	SpI - 11.5964
SCO1714	putative secreted protein	none	-	-	1	Y - 0.498	N	SpII - 12.878
SCO2838	putative secreted endoglucanase.	Glycosyl hydrolases family 6	pfam01341	7.32E-99	-	Y - 0.639	Y - 0.377	SpII - 32.6736
SCO3357	hypothetical protein	none	-	-	-	N	Y - 0.492	SpII - 17.3077
SCO4142	phosphate-binding protein precursor	PBP2_PstS: substrate binding domain of ABC-type phosphate transporter	cd13565	4.57E-80	-	Y - 0.595	N	SpII - 26.7983
SCO4739	putative lipoprotein	none	-	-	-	Y - 0.579	N	SpII - 20.7928
SCO4885	putative lipoprotein	PBP1_BmpA_PnrA_like: the PnrA lipoprotein (also known as Tp0319 or TmpC) represents a novel family of bacterial purine nucleoside receptor encoded within an ATP-binding cassette (ABC) transport system (pnrABCDE).	cd06354	2.40E-107	-	N	N	SpII - 23.8395
SCO4905	putative lipoprotein	none	-	-	-	Y - 0.574	N	SpII - 13.7291

SCO Number	Genome annotation	Top hit in Conserved Domain Database	Accession	E-Value[1]	#TMHMM [2]	SignalP 4.1 [3]	TatP 1.0 [4]	LipoP 1.0 [5]
SCO4934	putative lipoprotein	LDT_IgD_like_2 : IgD-like repeat domain of mycobacterial L,D-transpeptidases	cd13432	2.81E-39	-	Y - 0.571	Y - 0.483	SpII - 24.1553
SCO5646	putative solute binding lipoprotein	PBP2_TbpA: substrate binding domain of thiamin transporter, a member of the type 2 periplasmic binding fold superfamily.	cd13545	2.50E-115	-	N	Y - 0.468	SpII - 13.5061
SCO7218	putative iron transport lipoprotein	TroA_a: predicted to function as initial receptors in ABC transport of metal ions in eubacteria	cd01148	3.57E-137	-	Y - 0.632	N	SpI - 14.1761

[1] E-value of the top hit in the CDD database

[2] The number of transmembrane helices predicted by the TMHMM 2.0 server (http://www.cbs.dtu.dk/services/TMHMM/).

[3] SignalP 4.1 software predicts the presence of a signal peptide (http://www.cbs.dtu.dk/services/SignalP/). D-score is a score used to discriminate signal peptides from non-signal peptides. Scores > 0.450 indicate a signal peptide.

[4] TatP 1.0 predicts the presence of twin arginine (TAT) signal peptides. D-sore > 0.36 predicts the presence of a TAT pathway signal.

[5] LipoP 1.0 software produces predictions of lipoproteins (http://www.cbs.dtu.dk/services/LipoP/). SpI denotes SEC signal peptide; SpII denotes lipoprotein

Table 4.4 Predicted membrane proteins identified as *S. coelicolor* glycoproteins in this study

SCO Number	Genome annotation	Top hit in Conserved Domain Database	Accession	E-Value[1]	#TMHMM [2]	SignalP 4.1 [3]	TatP 1.0 [4]	LipoP 1.0 [5]
SCO2096	putative membrane protein	Transglutaminase-like superfamily: family includes animal transglutaminases and other bacterial proteins of unknown function	pfam01841	7.25E-26	6	Y - 0.529	N	SpII - 8.2333
SCO2035	putative membrane protein	Thioredoxin_like Superfamily: protein disulfide oxidoreductases and other proteins with a Thioredoxin fold	cl00388	2.10E-17	1	N	N	N
SCO2156	putative cytochrome c oxidase subunit II	CyoA: Heme/copper-type cytochrome/quinol oxidase, subunit 2	COG1622	4.29e-53	3	N	N	N
SCO2963	putative membrane protein	PHA03249 Superfamily: DNA packaging tegument protein UL25	cl19799	9.21E-05	1	N	N	N
SCO3044	conserved hypothetical protein	LytR_cpsA_psr : cell envelope-related transcriptional attenuator domain	pfam03816	2.04E-53	1	N	N	N
SCO3046	conserved hypothetical protein	LytR_cpsA_psr Superfamily: cell envelope-related transcriptional attenuator domain	cl00581	3.79E-47	1	N	N	N
SCO3184	putative penicillin acylase (EC 3.5.1.11).	Ntn_PGA_like: Penicillin G acylase belongs to a family of beta-lactam acylases that includes cephalosporin acylase and aculeacin A acylase.	cd03747	2.81E-118	1	N	Y - 0.366	N

SCO Number	Genome annotation	Top hit in Conserved Domain Database	Accession	E-Value[1]	#TMHMM [2]	SignalP 4.1 [3]	TatP 1.0 [4]	LipoP 1.0 [5]
SCO3848	putative serine/threonine protein kinase	STKc_PknB_like: catalytic domain of bacterial Serine/Threonine kinases, PknB and similar proteins (PASTA Domain containing)	cd14014	3.26E-128	1	N	N	N
SCO3891	putative membrane protein	none	-	-	1	N	N	N
SCO4013	putative secreted penicillin-binding protein	FtsI: cell division protein FtsI; penicillin-binding protein 2 [Cell cycle control; cell division; chromosome partitioning; Cell wall, membrane and envelope biogenesis]	COG0768	1.74E-89	1	N	N	N
SCO4130	putative integral membrane protein	none	-	-	1	N	N	N
SCO4141	phosphate ABC transport system permease protein	Phosphate_pstC: phosphate ABC transporter permease protein PstC	TIGR02138	1.48E-91	5	N	N	N
SCO4256	putative hydrolytic protein	ChW: Clostridial hydrophobic, with a conserved W residue, domain.	smart00728	3.04E-12	1	N	N	N
SCO4548	putative integral membrane protein	BTAD: Bacterial transcriptional activator domain	smart01043	8.46e-14	3	N	Y - 0.479	N
SCO4968	putative membrane protein	none	-	-	1	N	N	N
SCO5204	integral membrane protein	UPF0182: uncharacterized protein family	pfam03699	0.00E+00	7	N	N	N

SCO Number	Genome annotation	Top hit in Conserved Domain Database	Accession	E-Value	#TMHMM [1]	SignalP 4.1 [2]	TatP 1.0 [3]	LipoP 1.0 [4]
SCO5751	putative membrane protein	HTH_25: Helix-turn-helix domain	pfam13413	1.21E-20	1	N	N	N
SCO5818	putative ABC transporter	NatA: ABC-type Na+ transport system, ATPase component NatA [Energy production and conversion, Inorganic ion transport and metabolism]	COG4555	3.92e-35	5	N	N	N

[1] E-value of the top hit in the CDD database

[2] The number of transmembrane helices predicted by the TMHMM 2.0 server (http://www.cbs.dtu.dk/services/TMHMM/).

[3] SignalP 4.1 software predicts the presence of a signal peptide (http://www.cbs.dtu.dk/services/SignalP/). D-score is a score used to discriminate signal peptides from non-signal peptides. Scores > 0.450 indicate a signal peptide.

[4] TatP 1.0 predicts the presence of twin arginine (TAT) signal peptides. D-sore > 0.36 predicts the presence of a TAT pathway signal.

[5] LipoP 1.0 software produces predictions of lipoproteins (http://www.cbs.dtu.dk/services/LipoP/). SpI denotes SEC signal peptide; SpII denotes lipoprotein

Table 4.5 Predicted secreted proteins identified as *S. coelicolor* glycoproteins in this study

SCO Number	Genome annotation	Top hit in Conserved Domain Database	Accession	E-Value	#TMHMM [1]	SignalP 4.1 [2]	TatP 1.0 [3]	LipoP 1.0 [4]
SCO3540	proteinase (putative secreted protein)	Abhydrolase_4: family of putative bacterial peptidases and hydrolases that bear similarity to a tripeptidyl aminopeptidase isolated from *Streptomyces lividans.*	pfam08386	5.21E-40	1	Y - 0.627	Y - 0.700	SpI - 18.2099
SCO4847	putative D-alanyl-D-alanine carboxypeptidase	DacC: D-alanyl-D-alanine carboxypeptidase	COG1686	3.80E-49	1	Y - 0.711	Y - 0.427	SpI - 27.3476
SCO5776	glutamate binding protein	PBP2_GluB: substrate binding domain of ABC glutamate transporter; the type 2 periplasmic binding protein fold	cd13690	1.79E-93	-	Y - 0.618	N	SpI - 21.8509

[1] E-value of the top hit in the CDD database

[2] The number of transmembrane helices predicted by the TMHMM 2.0 server (http://www.cbs.dtu.dk/services/TMHMM/).

[3] SignalP 4.1 software predicts the presence of a signal peptide (http://www.cbs.dtu.dk/services/SignalP/). D-score is a score used to discriminate signal peptides from non-signal peptides. Scores > 0.450 indicate a signal peptide.

[4] TatP 1.0 predicts the presence of twin arginine (TAT) signal peptides. D-sore > 0.36 predicts the presence of a TAT pathway signal.

[5] LipoP 1.0 software produces predictions of lipoproteins (http://www.cbs.dtu.dk/services/LipoP/). SpI denotes SEC signal peptide; SpII denotes lipoprotein

Table 4.6 *S. coelicolor* glycoproteins identified in this study with no predicted transmembrane domains or secretory signals.

SCO Number	Genome annotation	Top hit in Conserved Domain Database	Accession	E-Value	#TMHMM [1]	SignalP 4.1 [2]	TatP 1.0 [3]	LipoP 1.0 [4]
SCO3353	hypothetical protein	none	-	-	-	N	N	N
SCO4307	conserved hypothetical protein SCD95A.40c	murQ: N-acetylmuramic acid-6-phosphate etherase (cell wall recycling)	PRK05441	4.41E-159	-	N	N	N
SCO5115	BldKD, putative ABC transporter intracellular ATPase subunit	ABC_NikE_OppD_transporters: ATP-binding cassette domain of nickel/oligopeptides specific transporters	cd03257	4.48E-122	-	N	N	N
SCO5736	30S ribosomal protein S15	rpsO: 30S ribosomal protein S15	PRK05626	1.36E-53	-	N	N	N
SCO6558	putative protein associated with oxidoreductase activity	PRK00724: formate dehydrogenase accessory protein	PRK00724	9.29E-128	-	N	N	N

[1] E-value of the top hit in the CDD database

[2] The number of transmembrane helices predicted by the TMHMM 2.0 server (http://www.cbs.dtu.dk/services/TMHMM/).

[3] SignalP 4.1 software predicts the presence of a signal peptide (http://www.cbs.dtu.dk/services/SignalP/). D-score is a score used to discriminate signal peptides from non-signal peptides. Scores > 0.450 indicate a signal peptide.

[4] TatP 1.0 predicts the presence of twin arginine (TAT) signal peptides. D-sore > 0.36 predicts the presence of a TAT pathway signal.

[5] LipoP 1.0 software produces predictions of lipoproteins (http://www.cbs.dtu.dk/services/LipoP/). SpI denotes SEC signal peptide; SpII denotes lipoprotein

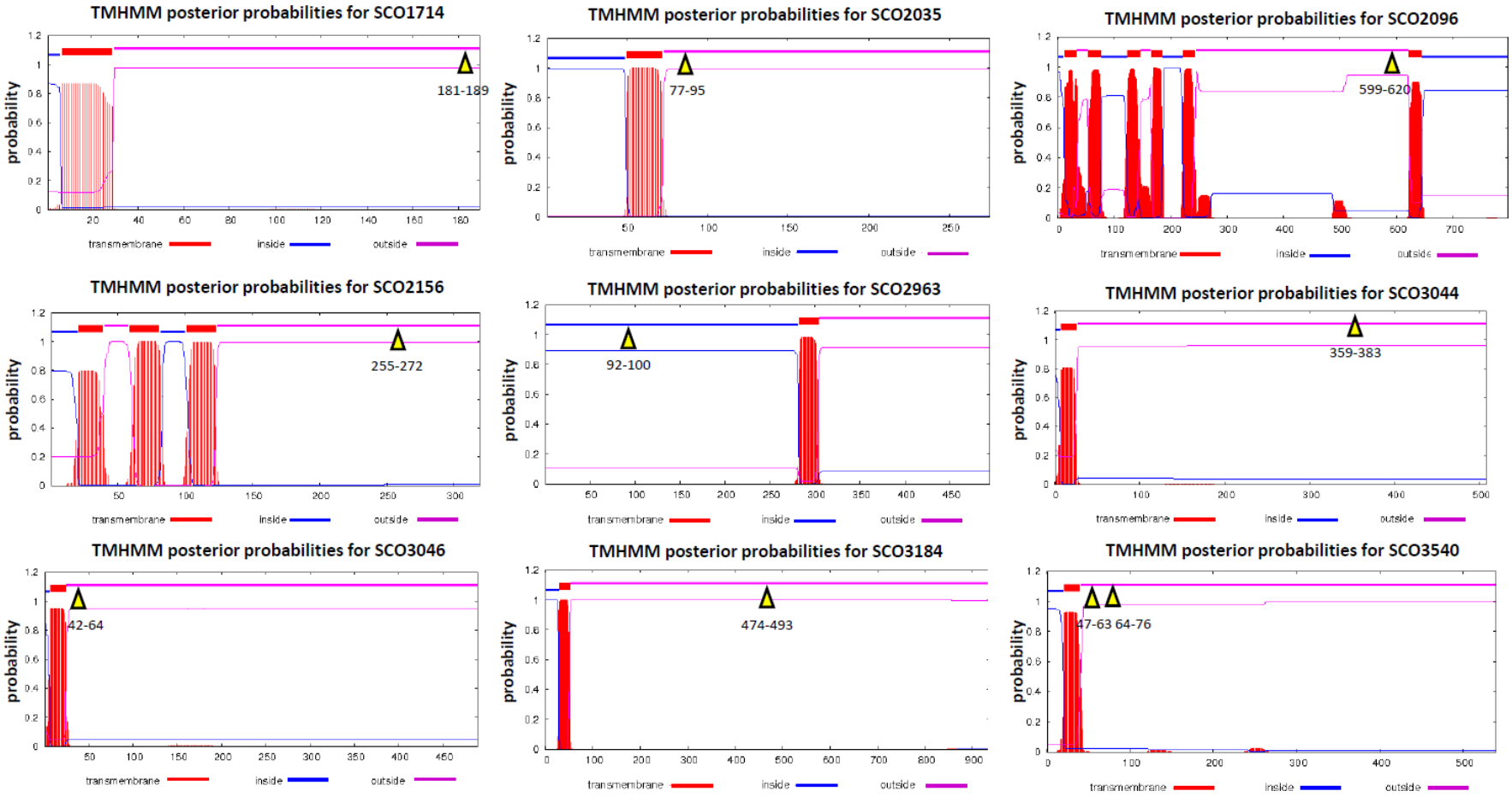


Figure 4.9 The positions of glycosylated peptides relative to transmembrane domains predicted using the TMHMM server v 2.0 in *S. coelicolor* membrane glycoproteins. Transmembrane domains indicated in red. The positions of the glycosylated peptides are indicated with yellow arrowheads. The y-axis indicates the probability localisation of the transmembrane domain. Pink indicates outside and blue indicates inside.

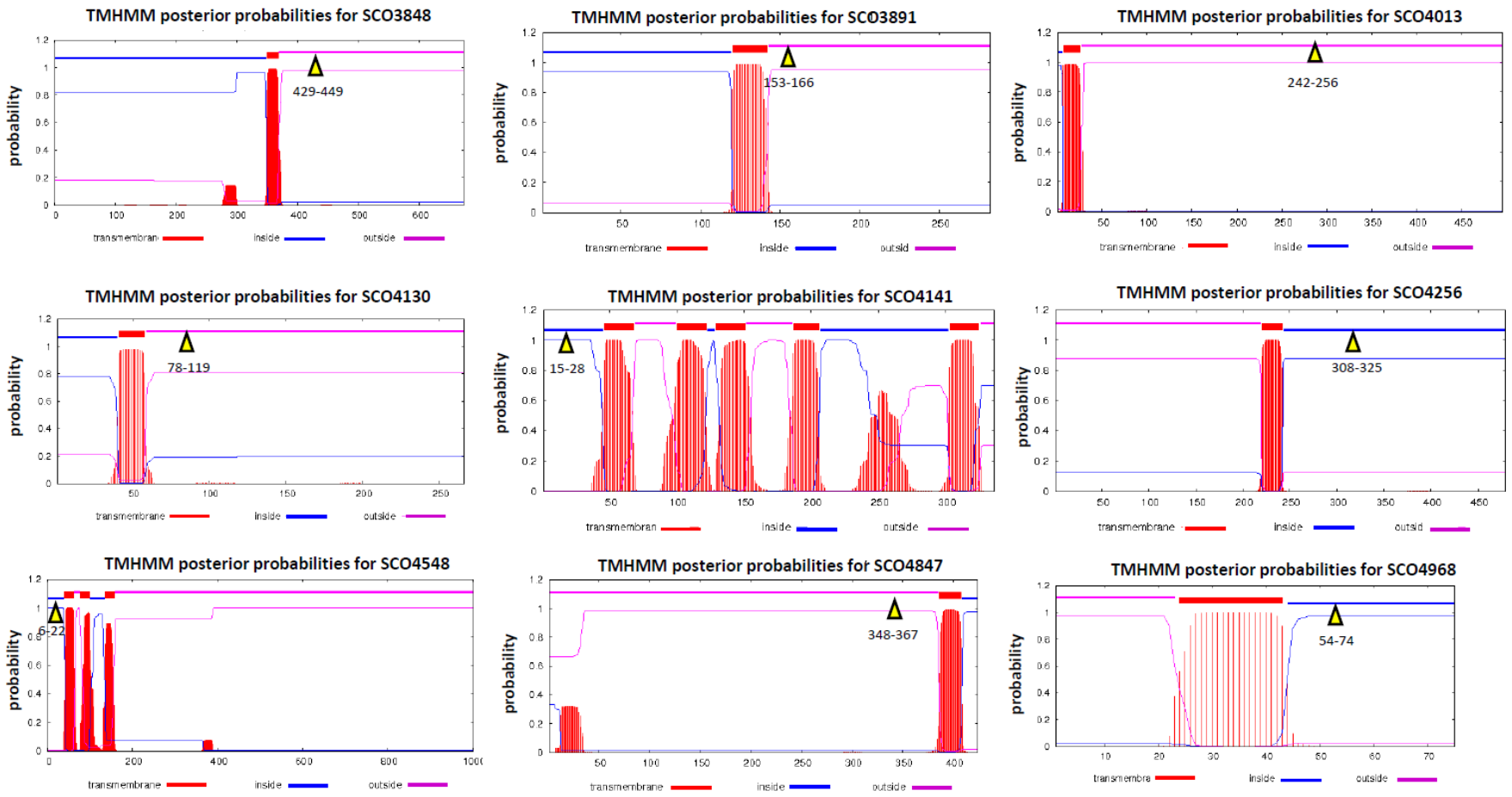


Figure 4.9 continued. The positions of glycosylated peptides relative to transmembrane domains predicted using the TMHMM server v 2.0 in *S. coelicolor* membrane glycoproteins. Transmembrane domains indicated in red. The positions of the glycosylated peptides are indicated with yellow arrowheads. The y-axis indicates the probability localisation of the transmembrane domain. Pink indicates outside and blue indicates inside.

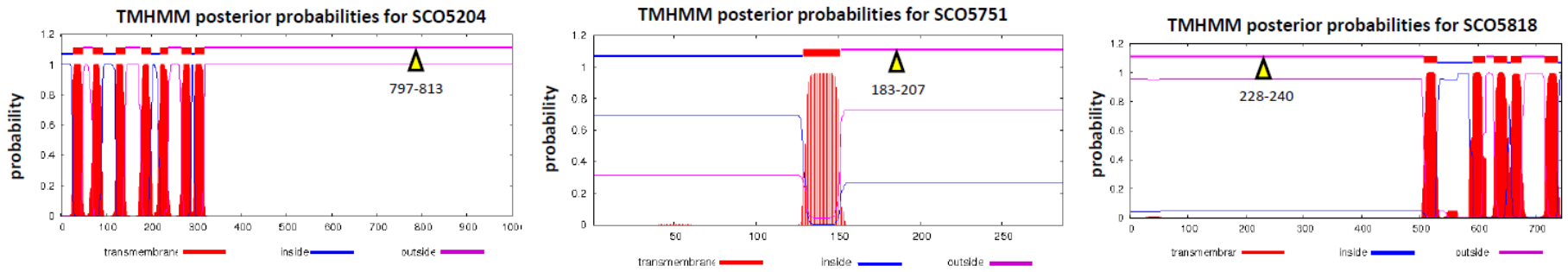


Figure 4.9 continued. The positions of glycosylated peptides relative to transmembrane domains predicted using the TMHMM server v 2.0 in *S. coelicolor* membrane glycoproteins. Transmembrane domains indicated in red. The positions of the glycosylated peptides are indicated with yellow arrowheads. The y-axis indicates the probability localisation of the transmembrane domain. Pink indicates outside and blue indicates inside.

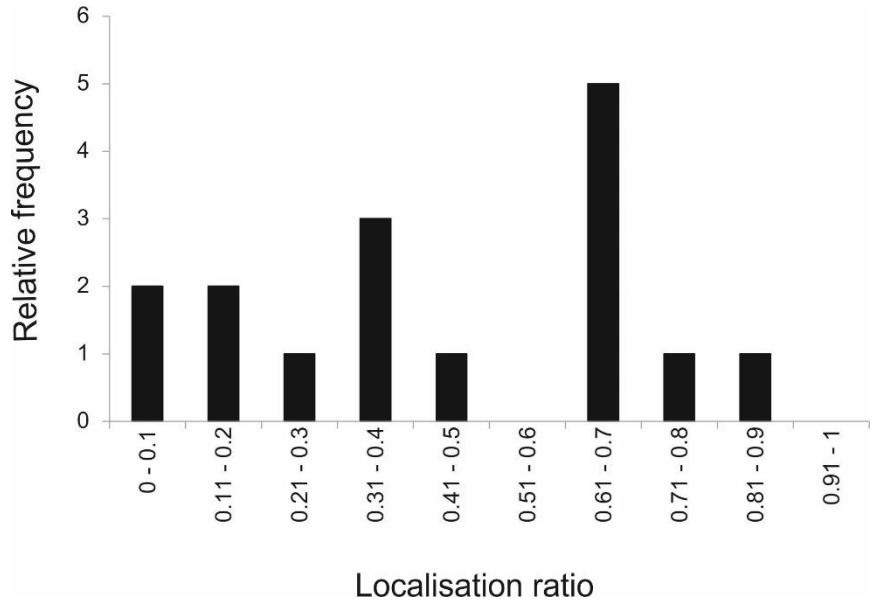


Figure 4.10 O-glycosylation site positional analysis in *S. coelicolor* glycoproteins. The O-glycosylation site localisation ratio = glycosylated amino acid position/total number of amino acids in the protein. A localisation ratio closer to 0 indicates a position closer the N-terminus, while a ratio closer to 1 indicates a position closer the C-terminus.

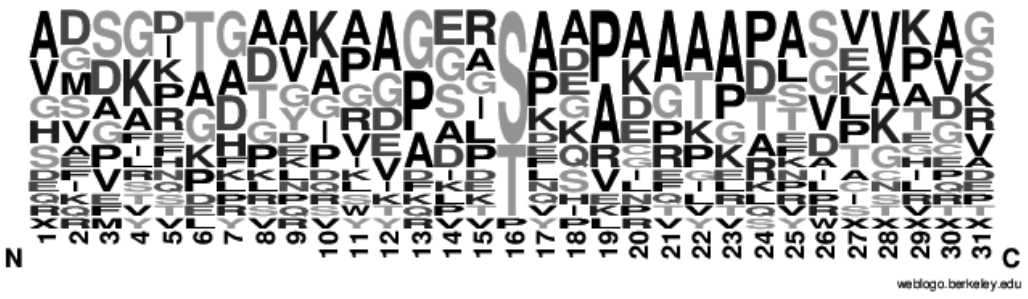


Figure 4.11 O-glycosylation site motif generated from the confident *S. coelicolor* O-glycosylation sites confirmed in this study. Shown are fifteen amino acids both before and after the glycosylated serine or threonine residue.

4.4 Discussion

Using a glycoproteomics strategy, the in-depth characterisation of the *S. coelicolor* membrane glycoproteome in this study resulted in the identification of 37 new *S. coelicolor* glycoproteins. In addition, this study enabled the further characterisation of the previously identified *S. coelicolor* glycoprotein PstS (SCO4142) by the observation of glycopeptides, which were not seen previously in the work carried out by Wehmeier et al. (2009).

The time-course glycoprotein enrichment and analysis using mass spectrometry with CID fragmentation on the maXis HD resulted in the identification of glycopeptides belonging to 15 *S. coelicolor* glycoproteins. However, the localisation of glycosylation sites in these glycopeptides was not possible in many cases. Glycosylation sites could be assigned in glycopeptides where only a single glycosylated residue was possible. However, in some cases the assignments were supported by product ions with the glycan attached. In many cases, there was ambiguity over the localisation of the glycan in the glycopeptide, especially in multiply glycosylated glycopeptides. As expected, spectra generated by CID fragmentation were often dominated by product ions formed as a result of the preferential cleavage of glycosidic bonds associated with a glycopeptide (Huddleston et al. 1993). When the hexose modifications fragment more easily than the peptide backbone, glycosylation site positional information is often lost.

The glycopeptide analysis by mass spectrometry on the Orbitrap Fusion with ETD and HCD fragmentation, using the Ion trap and Orbitrap mass analysers, demonstrated the merit of performing complementary peptide fragmentation techniques in glycoproteomics strategies. The glycopeptide analyses carried out using ETD fragmentation greatly increased the number of glycosylation sites localised, over the acquisitions carried out using HCD fragmentation. The confidence in the glycosylation site assignments made in this study are supported by the MD-scores and assignments of the respective product ion series. The increased number of glycopeptide identifications made after HCD fragmentation compared to ETD fragmentation could be explained by the fact that ETD fragmentation is more efficient for precursor ions with higher charged states (Brodbelt 2015).

The localisation of glycosylation sites in the glycoproteins identified in this study enabled the analysis of a glycosylation sequence motif in *S. coelicolor*. While no distinct consensus sequence was identified, a high propensity for hydrophobic amino acid (e.g. Ala, Pro, Gly) signatures was observed. This feature is reminiscent of sequences surrounding O-

glycosylation sites in other Actinobacteria (Dobos et al. 1995; Herrmann et al. 1996; Michell et al. 2003; Smith et al. 2014).

In some cases, there was heterogeneity in the number of glycans attached to the glycopeptide. For example, the SCO3353 glycopeptide N- KPSAPECGTPPAGSAK-C was observed with both Hex_2 and Hex_3 modifications on the same threonine residue (Thr9). Glycosylation heterogeneity has been described previously in the *M. tuberculosis* 45 kDa glycoprotein characterised by Dobos et al. (1996). Sequential protein O-glycosylation in *M. tuberculosis* has been suggested to occur in a similar fashion to the eukaryotic PMT-mediated process, with Pmt catalysing the first step of protein O-mannosylation (VanderVen et al. 2005). Liu et al. (2013a) demonstrated that a second glycosyltransferase, PimE was required for the elongation of oligomannosyl residues on the glycoprotein FasC. It is possible that glycoproteins are modified by a similar process in *S. coelicolor* and that the glycan heterogeneity observed in this study is a result thereof. However, this is purely speculative and remains to be experimentally validated.

Protein O-glycosylation by Pmt was shown to be coupled to protein secretion via the SEC pathway in *M. tuberculosis,* suggesting that protein O-mannosylation should only affect extracellular proteins (VanderVen et al. 2005). Glycoproteomics studies to date on the *M. tuberculosis* glycoproteome have therefore focussed on characterisation of the culture filtrate (Gonzalez-Zamorano et al. 2009; Smith et al. 2014). Many of the glycoproteins identified here (Tables 4.3 – 4.6) were predicted to be lipoproteins, membrane proteins and secreted proteins, which would agree with the dogma that protein O-glycosylation in Actinobacteria is coupled to protein secretion via the SEC pathway. However, a small number of the glycoproteins identified had no predicted transmembrane domains or secretory signals suggesting that they could be intracellular proteins. Similarly, some glycoproteins identified here were predicted to be exported by the TAT system. These “atypical” glycoproteins that to do not conform with the existing dogma will be discussed further in the general discussion (Chapter 7).

Upon characterising the membrane glycoproteome in *S. coelicolor*, we were particularly interested in proteins that could help to explain the antibiotic hypersensitivity phenotypes observed previously in the *pmt* and *ppm1 S. coelicolor* strains (Howlett et al. 2016). It was hypothesised that the *S. coelicolor* glycoproteome could contain proteins that are important in cell wall biosynthesis or for maintaining membrane integrity. In this study, at least seven glycoproteins have been identified that have predicted functions in the cell wall (SCO4934,

SCO4847, SCO3044, SCO3046, SCO3184, SCO4013, SCO4307). SCO4847, for example is a putative D-Ala-D-Ala carboxypeptidase and low molecular weight penicillin binding protein. These proteins are thought to catalyse the hydrolysis of the terminal D-alanine from the peptidoglycan stem peptide (Pratt 2008). SCO4013 is another predicted penicillin binding protein, while SCO4934 is a predicted L, D transpeptidase. L, D transpeptidases catalyse an alternative type of peptidoglycan crosslinking between the third position amino acids of tetrapeptide stems, termed 3->3 crosslinking. L, D transpeptidases have been identified in *M. tuberculosis* and were shown to be important for maintaining cell shape, virulence and resistance to β-lactam antibiotics (Schoonmaker et al. 2014). SCO3044 and SCO3046 both belong to the LytR-CpsA-Psr (LCP) family of proteins, that were first shown to catalyse the ligation of wall teichoic acids (WTA) to the N-acetylmuramic acid (MurNAc) units of peptidoglycan in *Bacillus subtilis* (Kawai et al. 2011). Other studies have demonstrated that LCP proteins are required to attach the capsular polysaccharide to peptidoglycan in both *Staphylococcus aureus* and *Streptococcus pneumoniae* (Eberhardt et al. 2012; Chan et al. 2014). Recently however, an LCP protein in *M. tuberculosis* (Lcp1) was shown to be required for cell viability and to attach arabinogalactan to peptidoglycan in a cell free assay (Harrison et al. 2016).

These results demonstrate the presence of a glycoproteome in the *S. coelicolor* membrane that is composed of many proteins that have predicted functions in cell wall biosynthesis.

Chapter 5 - Structural characterisation of the glycans modifying *S. coelicolor* glycoproteins.

Chapter 5 - Structural characterisation of the glycans modifying *S. coelicolor* glycoproteins.

Wehmeier et al. (2009) previously demonstrated that the *S. coelicolor* glycoprotein PstS was glycosylated with a trihexose by analysing the derivatised glycan using electrospray ionisation (ESI) mass spectrometry (MS). Pmt (SCO3154) was shown to be required for the glycosylation of PstS, while Ppm1 (SCO1423) was shown to be necessary for the transfer of [^{14}C]mannose from GDP-[^{14}C]mannose onto polyprenol phosphate in the *S. coelicolor* membrane. Although the monosaccharide composition of the PstS glycan was not determined, the work by Wehmeier et al. (2009) suggests that mannose is a key constituent thereof.

In the closely related *M. tuberculosis*, the 45/47 kDa secreted antigen MPT 32 was previously shown to be modified with mannose, as well as with α(1,2) linked mannobiose and mannotriose (Dobos et al. 1996). These linkages are also observed in the mannose branches of the lipomannan (LM) and lipoarabinomannan (LAM) core, as well as in the phosphatidyl inositol mannoside (PIM) family of phospholipids in mycobacteria (Hunter and Brennan 1990; Chatterjee et al. 1992). In *M. avium* the capsular glycoprotein SmT was shown to be glycosylated with dihexose (Taylor et al. 2006). Additionally, the secreted antigen MPB83 in *M. bovis* was shown to be modified with mannose, as well as with α(1,3) linked mannobiose and mannotriose (Michell et al. 2003). It is currently unclear whether α(1,3) mannose linkages exist in *M. tuberculosis*, although they have been described in the mannan core of *Mycobacterium chelonae* LAM and LM (Guérardel et al. 2002).

The *S. coelicolor* membrane glycoproteome has been shown to consist of glycoproteins modified with Hex, Hex_2 and Hex_3 glycan moieties (discussed in Chapter 4). However, the nature of the monosaccharide and the linkages involved have yet to be fully investigated. The aim of the work described in this chapter was to structurally characterise the glycans modifying the *S. coelicolor* glycoproteins that were identified in the work described in Chapter 4. The work in this chapter was carried out in collaboration with Dr Jerry Thomas and Rachel Bates, from the University of York Bioscience Technology Facility. All spectra shown in sections 5.1 and 5.2 were generated by Rachel Bates.

5.1 Mass spectrometric analysis of O-linked glycans from *S. coelicolor* glycoproteins

The samples used in these experiments were obtained after the Con A enrichment of *S. coelicolor* membrane proteins, isolated after 20 h, 35 h, 43 h and 60 h of growth as described in Chapter 4, section 4.1. Previous attempts to characterise the glycans from a single glycoprotein, by another researcher in the Smith laboratory, were unsuccessful due to difficulties encountered because of limited sample amounts. Therefore, analyses were carried out on the glycans from the total *S. coelicolor* glycoproteome. A preliminary analysis of the collective glycans from the individual time points was carried out using MALDI-TOF-MS. The mass spectra from each of the four time points were nearly identical, suggesting that there was little difference in the sample compositions. Therefore, the membrane preparations from the four time points were pooled to increase the amount of sample material.

The *S. coelicolor* glycoprotein fractions were run into an SDS-PAGE gel for 10 min. This was carried out in order to separate the glycoproteins from the detergent (n-dodecyl-β-D-maltoside) used to solubilise them. The total O-linked glycans were removed by in-gel, non-reductive β-elimination (as described in Chapter 2), permethylated and subsequently analysed by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry (MALDI-FT-ICR-MS carried out on the solariX XR FTMS 9.4T). Two disaccharide standards, cellobiose (β1,4 linked glucose) and maltose (α1,4 linked glucose) were permethylated and analysed alongside the glycans from the pooled sample. The high resolution and excellent mass accuracy of the solariX XR FTMS 9.4T instrument enables accurate mass determination for the identification of small molecules, such as small carbohydrates. The mass spectrum of the *S. coelicolor* β-eliminated and permethylated glycans showed signals at *m/z* 477.23104 and 681.33104. These signals are consistent with $[M+Na]^+$ species for permethylated Hex_2 ($C_{20}H_{38}O_{11}Na$ – calculated *m/z* 477.23119) and Hex_3 ($C_{29}H_{54}O_{16}Na$ - calculated *m/z* 681.33096) respectively (Figure 5.1.A). Intense signals consistent with the presence of Hex_2 and weaker signals for Hex_3 were also observed in both the cellobiose and maltose standards (Figure 5.1.B and 5.1.C respectively). A low intensity ion at *m/z* 631.40361 was observed in the *S. coelicolor* glycan sample (Figure 5.1.A.I). This signal is consistent with the *m/z* for sodiated permethylated n-dodecyl-β-D-maltoside ($C_{31}H_{60}O_{11}Na$ – calculated *m/z* 631.403335). This finding suggests that despite efforts to remove the detergent by running the *S. coelicolor* glycoprotein fractions into an SDS-PAGE gel for 10 min, some of the glycoside detergent was still present in the sample.

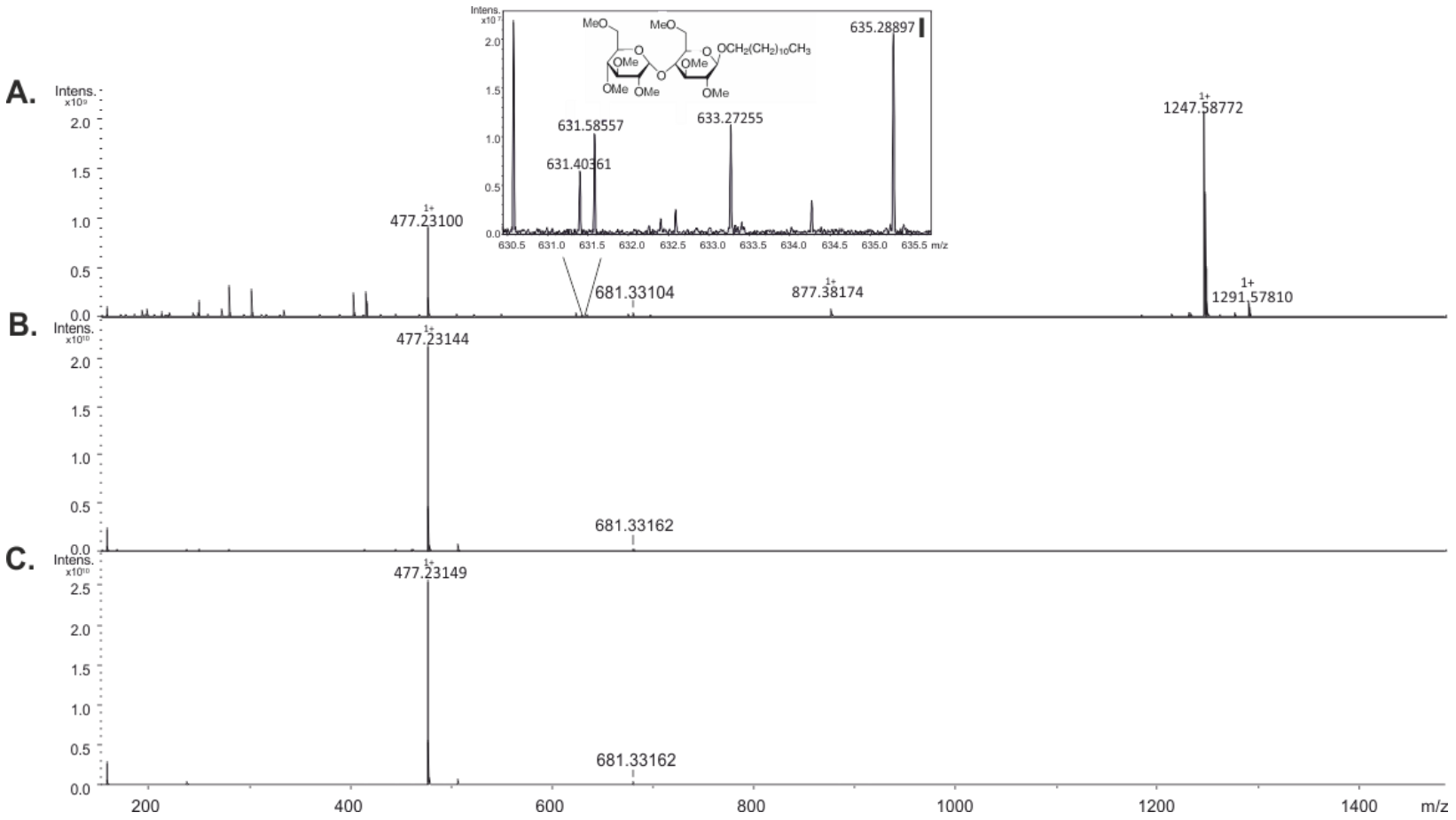


Figure 5.1 MALDI-FT-ICR-MS analysis of the permethylated *S. coelicolor* glycans. The *S. coelicolor* glycans (A) were analysed alongside permethylated cellobiose (B) and maltose (C) standards. For clarity of the ion at *m/z* 631.4036 in the *S. coelicolor* glycan sample (A), a magnified section of the spectrum is shown (I).

A high intensity signal at *m/z* 1247.58772 was also observed in the *S. coelicolor* glycan sample (Figure 5.1.A). In order to investigate the origin of this signal, in-gel β-elimination was carried out on a protein-free section of an SDS-PAGE gel and permethylated alongside a fresh preparation of *S. coelicolor* glycans. After the analysis of both samples by matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) it was found that the signal at *m/z* 1247 was also present in the reagent control, submitted to conditions for in-gel β-elimination of the protein-free SDS-PAGE gel piece (Figure A.16). Product ion analysis of the contaminant ion at *m/z* 1247 in the *S. coelicolor* glycan sample was carried out using MALDI-TOF-MS/MS. The product ion spectrum was dominated by ions differing by 204 *m/z* units, suggesting the presence of an oligosaccharide that was made up of hexoses (Figure 5.2). The *m/z* value of 1247, and the absence of signals indicating a loss of 221 *m/z* units is consistent with a cyclic glucan (Dell et al. 1984), in this case containing six residues.

It is well established that the signal intensities of permethylated carbohydrates in MALDI mass spectra are directly related to the concentration of those carbohydrates in a sample (Wada et al. 2007). The ion at *m/z* 1247.58772 in the FT-ICR spectrum from the *S. coelicolor* glycan sample (Figure 5.1.A) was more than double the intensity of the next most intense ion in the spectrum (*m/z* 477.23100), suggesting that the contaminant was the most abundant carbohydrate in the sample. The ion at *m/z* 1247.58772 is consistent with the *m/z* for a sodiated permethylated Hex_6 cyclic glucan ($C_{54}H_{96}O_{30}Na$ – calculated *m/z* 1247.588420). α-Cyclodextrin is an example of a cyclic oligosaccharide consisting of six glucose units linked via α-1, 4 linkages. It is likely that the Hex_6 cyclic glucan in the *S. coelicolor* glycan sample is α-cyclodextrin that originated from the SafeBlue protein stain (NBS Biologicals Ltd) used to stain the SDS-PAGE gels, prior to carrying out the in-gel β-elimination. While the manufacturer would not confirm this due to "contractual obligations", the use of α-cyclodextrin in Coomassie blue based protein assays has been shown to make it compatible with detergents (Rabilloud 2016).

The presence in the sample of a large amount of what is probably α-cyclodextrin raises the concern that the Hex_2 and Hex_3 species could have arisen from this by hydrolysis. However, there is no evidence, even at very low abundance, of linear Hex_6 (calculated *m/z* 1293.63028), Hex_5 (calculated *m/z* 1089.530501), Hex_4 (calculated *m/z* 885.43072768) (Figure 5.3 A-C) or indeed a hexose monosaccharide (calculated *m/z* 273.13141) that would also be expected products of such hydrolysis, making it very unlikely that the Hex_2 and Hex_3 species derive from the contaminant in this way.

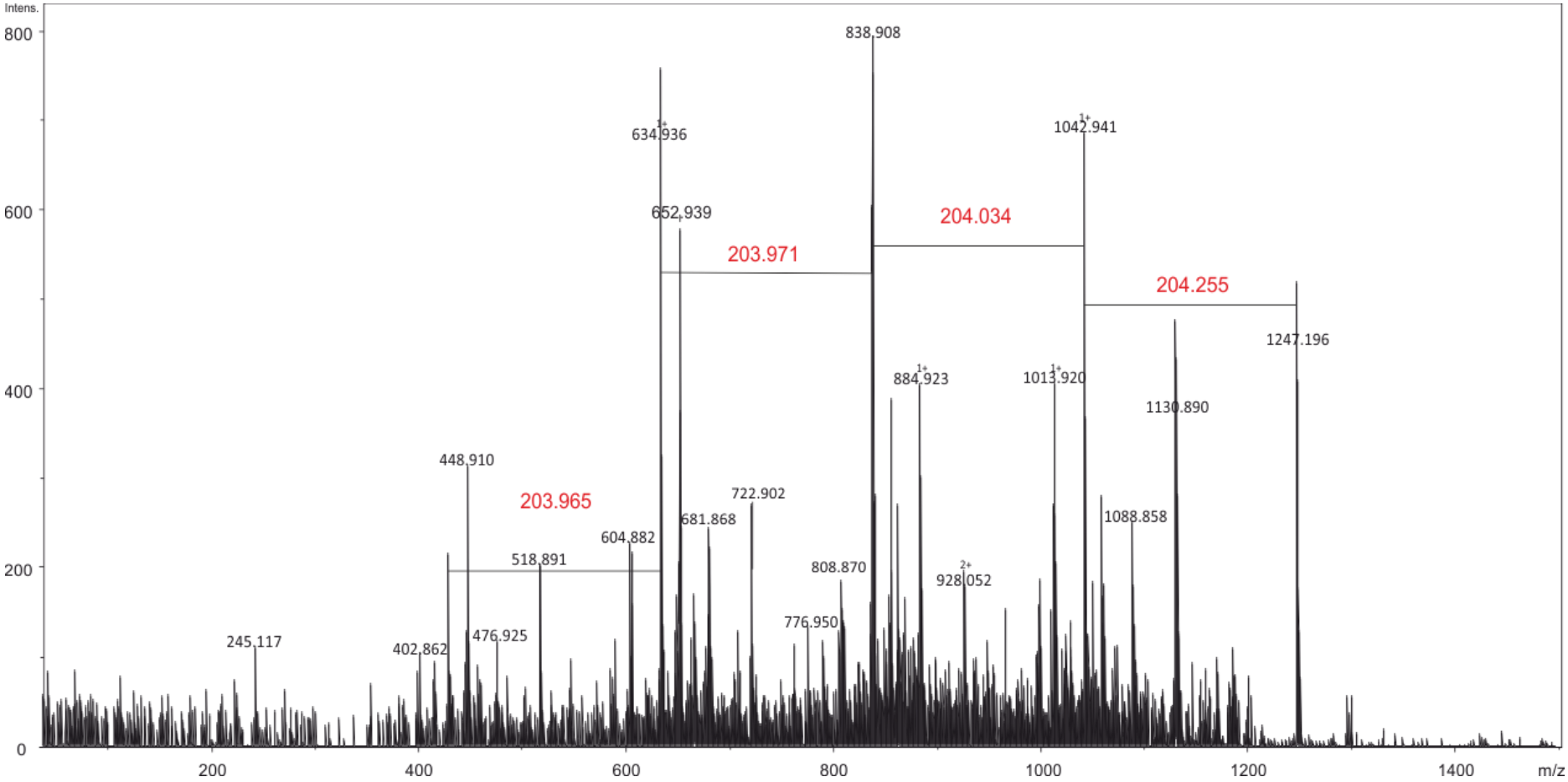


Figure 5.2 The product ion spectrum of the contaminant ion at *m/z* 1247 from the *S. coelicolor* glycan sample, analysed by MALD-TOF-MS/MS.

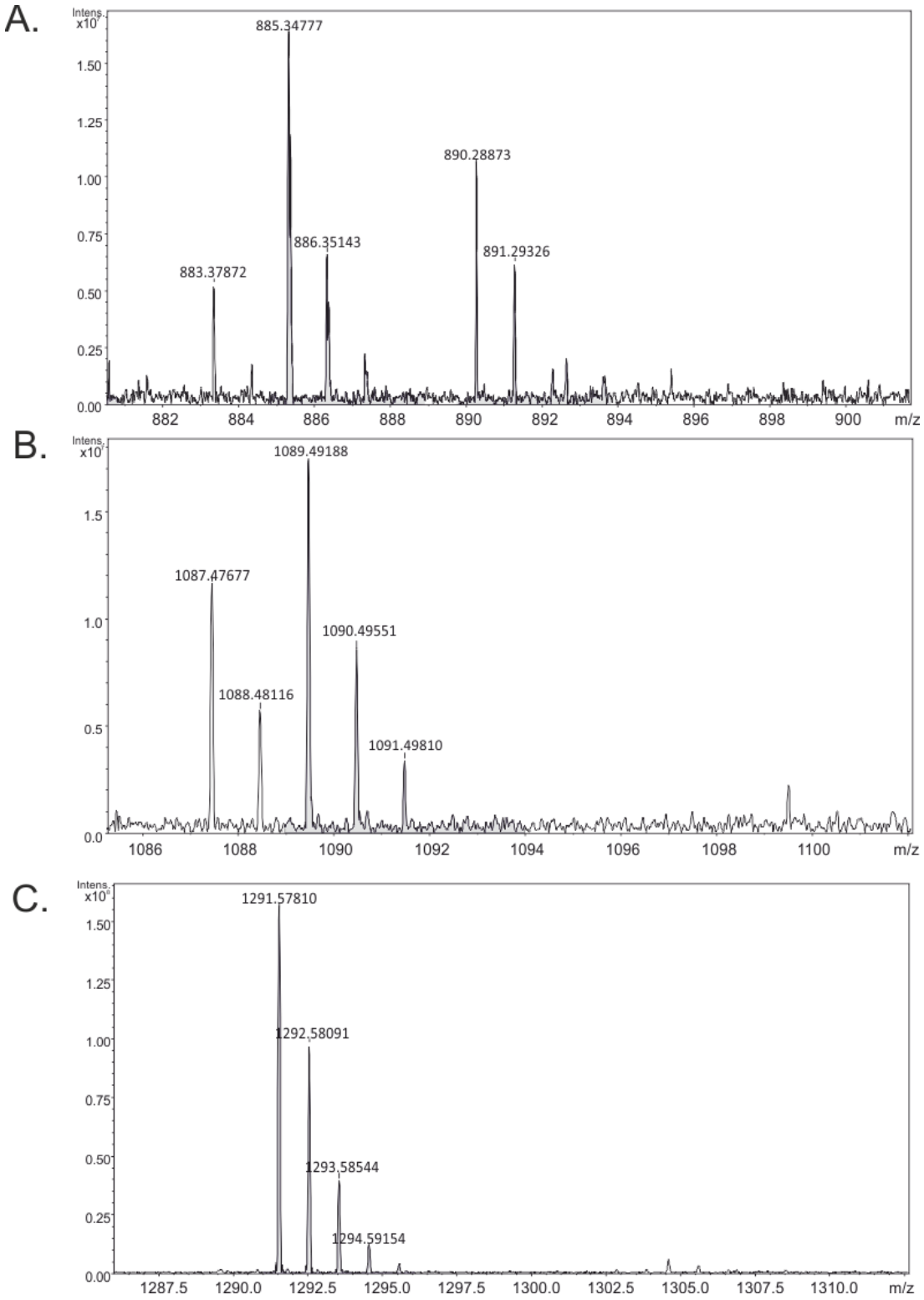


Figure 5.3 The MALDI-FT-ICR-MS spectrum of the *S. coelicolor* glycan sample, confirming the absence of signals corresponding to $[M+Na]^+$ species for hydrolysed and then permethylated cyclic glucan. Magnified sections of the mass spectrum show the m/z regions of 882 – 900 (A), 1086 – 1100 (B) and 1287.5 – 1310 (C).

In order to further investigate the ions corresponding to Hex_2 (*m/z* 477.23119) and Hex_3 (*m/z* 681.33096) in the *S. coelicolor* glycans and the standards, CID of the ions was carried out. Tandem mass spectrometric analysis of the ion at *m/z* 477.23119 from the *S. coelicolor* glycan sample yielded product ions *m/z* 259.11549 and 241.10492 (Figure 5.4.A). These ions are consistent with the *m/z* for the diagnostic sodiated Y_1 ($C_{10}H_{20}O_6Na$ – calculated *m/z* 259.115760) fragment ion and Y_1 exhibiting a neutral loss of H_2O (or a sodiated B_1 ion: $C_{10}H_{18}O_5Na$ – calculated *m/z* 241.105195), respectively. Similarly, sodiated Y_1 and Y_1-H_2O fragment ions were observed in the product ion spectra of the ions at *m/z* 477.231 in the cellobiose and maltose standards (Figure 5.4.B and 5.4.C respectively). Tandem mass spectrometric analysis of the ion at *m/z* 681.33096 from the *S. coelicolor* glycan sample yielded product ions *m/z* 463.21408 and 259.11544 (Figure 5.5.A). These ions are consistent with the *m/z* for the diagnostic sodiated Y_2 ($C_{19}H_{36}O_{11}Na$ – calculated *m/z* 463.215535) and Y_1 ($C_{10}H_{20}O_6Na$ – calculated *m/z* 259.115760) fragment ions, respectively. The sodiated Y_2 and Y_1 fragment ions were observed in the product ion spectra of the ions at *m/z* 681.330 for a low level Hex_3 contaminant in both of the standards, suggesting the presence of cellotriose and maltotriose respectively (Figure 5.5.B and 5.5.C). Additionally, sodiated B_2 (*m/z* 445.205) fragment ions were observed in the product ion spectra of the ions at *m/z* 681.330 in both of the standards. Taken together, these results demonstrate the presence of disaccharides and trisaccharides in the *S. coelicolor* glycans. Additionally, the cellobiose and maltose standards were shown to contain disaccharide as expected, as well as some trisaccharide.

5.2 Methylation and linkage analysis of *S. coelicolor* glycans

After determining the presence of disaccharides and trisaccharides in the *S. coelicolor* glycan sample, methylation analysis was carried out in order to investigate the linkages present in the *S. coelicolor* glycans. Methylation is an important step in the analysis of the structure of a carbohydrate and can be used to determine the linkages present, as illustrated in Figure 5.6 (Carpita and Shea 1989). During this procedure the glycans are methylated, which modifies all of the free hydroxyls to become methoxy groups. The hydroxyls participating in the linkage and the closure of the monosaccharide rings are protected. The partially methylated glycans are then subjected to acid hydrolysis, which breaks the glycosidic linkages and reveals the unmodified hydroxyls. The information regarding the anomeric

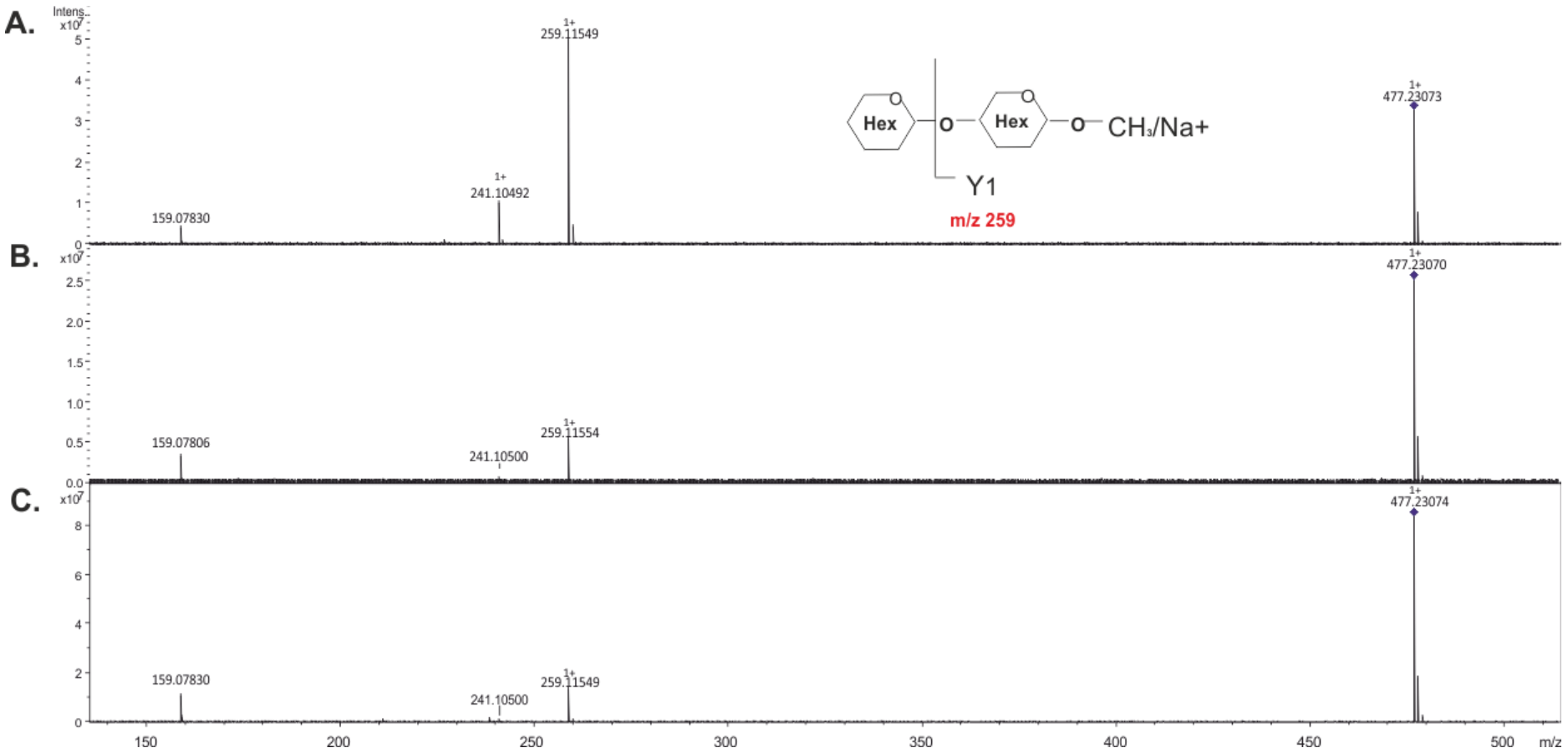


Figure 5.4 MALDI-FT-ICR-MS analysis of the precursor ions at *m/z* 477.231 demonstrate the presence of Hex_2. The precursor ions at m/z 477.231 in the *S. coelicolor* glycans (A), cellobiose (B) and maltose (C) standards are indicated by ◆ .

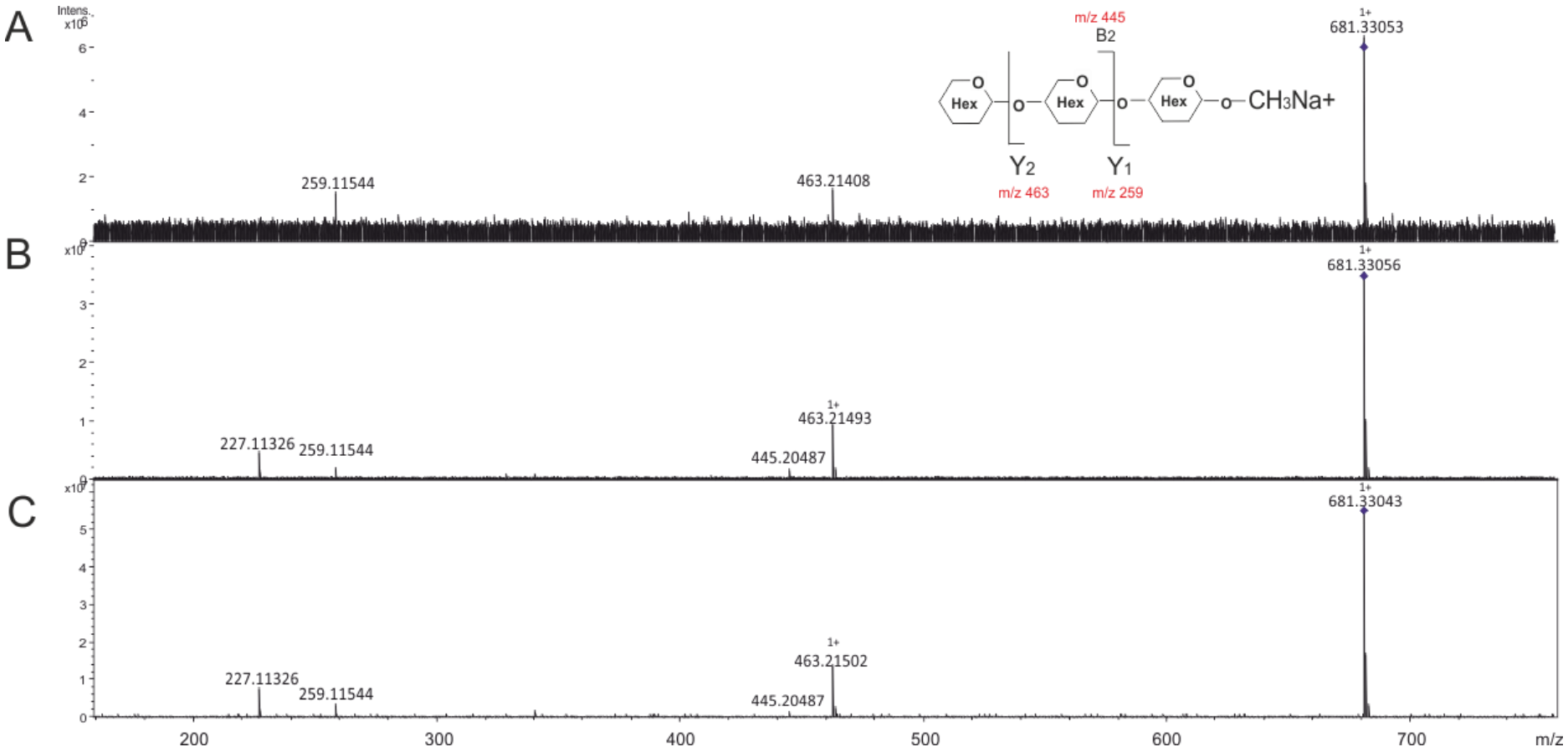


Figure 5.5 MALDI-FT-ICR-MS analysis of the precursor ions at *m/z* 681.330 demonstrate the presence of trisaccharides. The precursor ions at *m/z* 681 in the *S. coelicolor* glycans (A) and the standards (cellotriose (B) and maltotriose (C)) are indicated by ◆ .

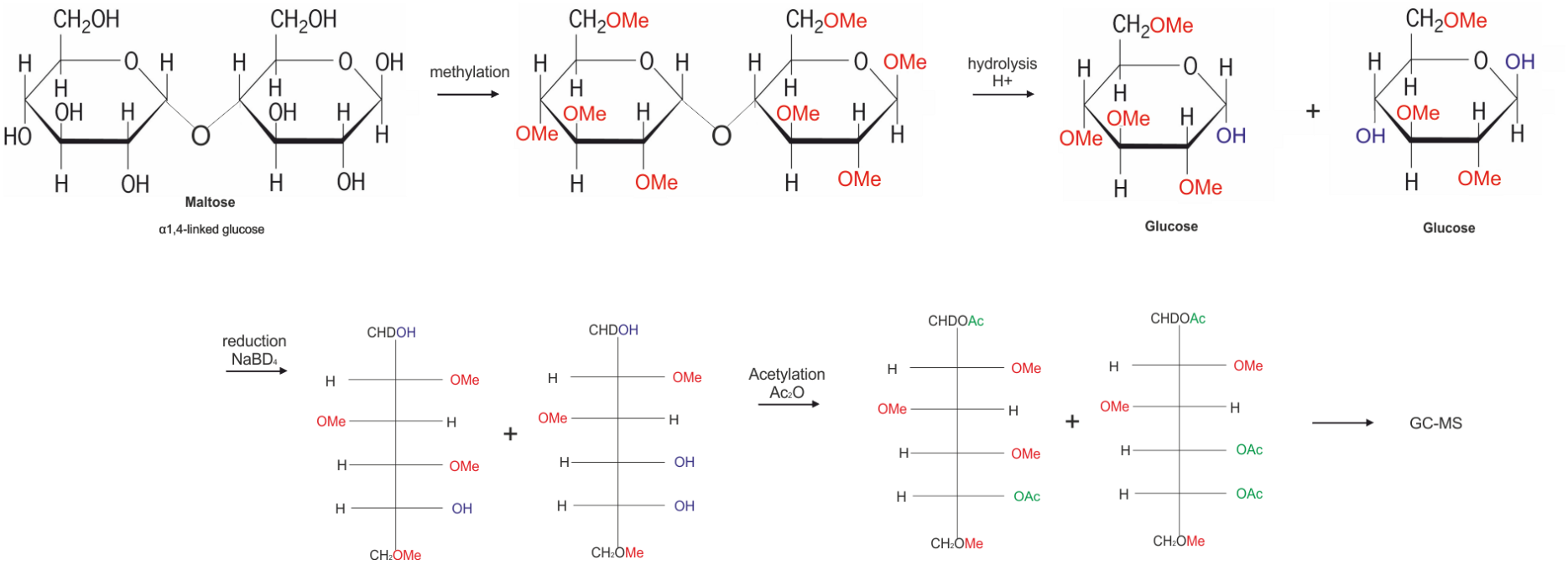


Figure 5.6 Methylation analysis of maltose. Carbohydrates are methylated, resulting in the modification of all of the free hydroxyl groups. After acid hydrolysis, the partially methylated monosaccharides are reduced, acetylated and analysed by GC-MS.

configuration (α/β) is lost during this step. The partially methylated monosaccharides are reduced using $NaBD_4$ which reduces the aldose to form an alditol and simultaneously tags the anomeric carbon atom (C1 in this case) with a deuterium atom. Finally, the partially methylated alditols are acetylated to increase their volatility for GC-MS analysis, which results in the acetylation of the free hydroxyl groups. The acetylation and methylation direct the electron ionisation (EI)-induced fragmentation of the partially methylated alditol acetate (PMAA) which allows the structure of the PMAA and thus the substitution pattern of the original monosaccharide to be determined. When analysed by GC-MS, the partially methylated alditol acetates (PMAAs) fragment in a predictable manner and these fragmentation patterns can be used to deduce the original distribution of methoxylated and acetoxylated carbons on the carbon backbone (Carpita and Shea 1989). Cleavage is favoured between adjacent carbon atoms bearing methoxyl groups, over that between adjacent carbons bearing a methoxyl group and an acetoxyl group (Figure 5.7). Due to the little electronic influence exhibited by methyl ethers on the alditol carbon, if cleavage occurs between two adjacent methoxylated carbons, carbonium ions of either fragment are produced at an equal frequency (Figure 5.7.A). However, cleavage between adjacent carbons one of which is methoxylated and the other acetoxylated results in a cation formed exclusively by the methyoxylated carbon (Figure 5.7.B). This is due to the electron withdrawing function exhibited by the keto group of the acetyl substitution. Therefore, only the fragment incorporating the methoxylated carbon is observed after EI-MS.

An example of the predicted fragmentation of a non-reducing terminal (i.e. 1-substituted) hexose is illustrated in Figure 5.8. Cleavage of the alditol chain usually results in fragments with two, three, four and five carbons (Carpita and Shea 1989). Cleavage is favoured between C2-C3 and C3-C4 resulting in fragments of *m/z* of 118 and 205, and 162 and 161 respectively. In the less likely event of a cleavage between a methoxylated and an acetoxylated carbon such as between C4-C5, the fragment ion at *m/z* 206 is observed. In addition to primary fragmentation, secondary fragments can result from the loss of acetic acid (CH_3COOH; loss of 60 *m/z* units), methanol (CH_3OH; loss of 32 *m/z* units), formaldehyde (CH_2O; loss of 30 *m/z* units) and ketene (CH_2CO; loss of 42 *m/z* units).

The *S. coelicolor* glycans were permethylated, hydrolysed, reduced and acetylated, and the partially methylated alditol acetates (PMAAs) were subsequently analysed by GC-MS. The *S. coelicolor* glycan derived PMAAs were analysed alongside PMAAs of mannose and glucose as standards (provided by the Proteomics Laboratory of the York Bioscience Technology Facility).

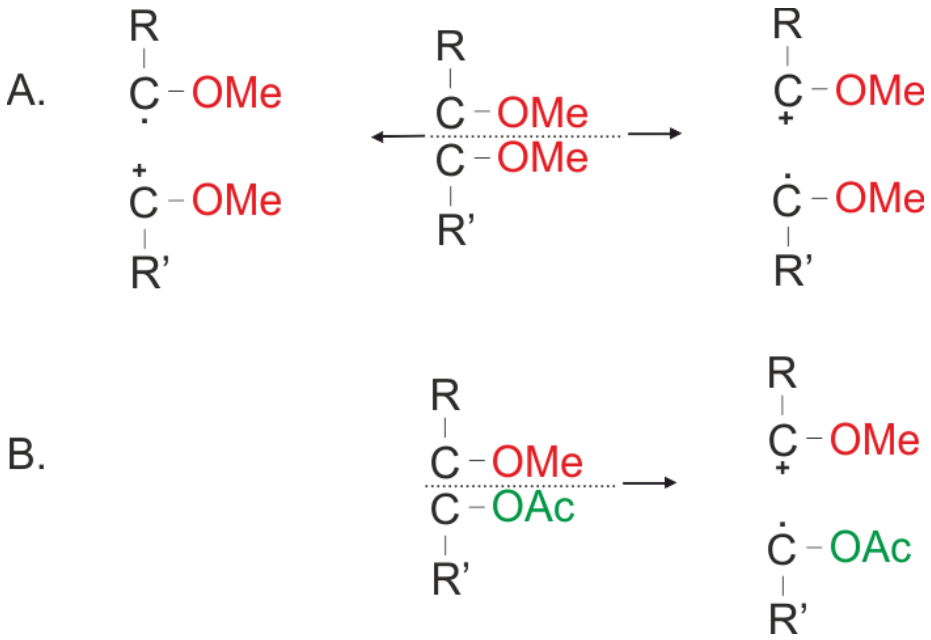


Figure 5.7 Primary fragmentation of PMAAs occurs more favourably between adjacent methoxylated carbons. A. Fragmentation between two methoxylated carbons generates cations of either fragment at an equal frequency. B. After fragmentation between a methoxylated and an acetoxylated carbon, only cations from the fragment bearing the methoxylated carbon are produced.

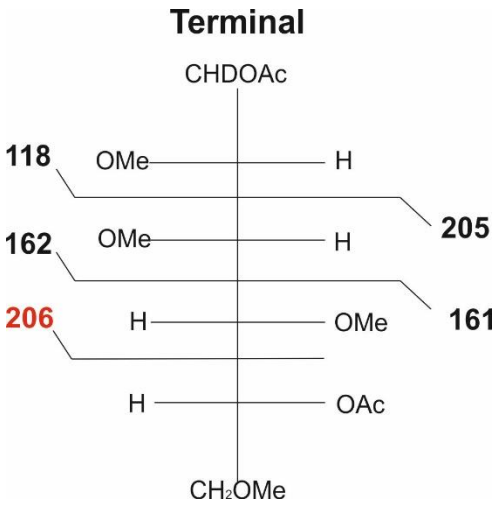


Figure 5.8 The predicted primary fragmentation of a PMAA from a terminal hexose. The most likely fragment ions are predicted to result from the cleavage between two methoxylated carbons, such as the fragments at *m/z* 118, 205, 162 and 161. Fragmentation is also possible between methoxylated and acetoxylated carbons, such as between C4-C5. However, only the fragment bearing the methoxylated carbon is observed in this case, the fragment at *m/z* 206.

The mass spectra enabled the assignment of the resulting PMAAs. The identities of the monosaccharides were determined by comparison of the retention times of component peaks in the sample, with the peaks in the glucose and mannose standard derived PMAAs that were separated under the same conditions. The GC-MS analysis of the PMAAs derived from the permethylated *S. coelicolor* glycans gave peaks corresponding to terminal mannose, 2-substituted mannose, 4-substituted mannose, 4-substituted glucose and terminal glucose, as well as spectra consistent with the respective hexitols (Table 5.1; Figure A.17-A.20).

As an example, the mass spectrum of 2-substituted mannose is shown in Figure 5.9. A peak was observed in the *S. coelicolor* glycan sample with a retention time of 15.585 min, that corresponded to the retention time (15.585 min) of the 2-substituted mannose peak in the PMAA mannose standard. The mass spectrum of the peak in the mannose standard was consistent with a 2-substituted hexitol, since diagnostic product ions at *m/z* 190 and 161 resulting from the cleavage between C3-C4 were observed (Figure 5.9.A). Additionally, secondary fragment ions were also observed at *m/z* 130 and 129, resulting from the loss of acetic acid and methanol respectively. Similarly, product ions diagnostic of the 2-substituted hexitol (*m/z* 190, 160, 130 ,129) were observed in the mass spectrum of the peak in the *S. coelicolor* glycan sample (Figure 5.9.B). Therefore, the peak in the PMAA data from the sample was identified as 2-substituted mannose.

The observation of 4-substituted glucose can be explained by contamination of the sample with α-cyclodextrin, an oligosaccharide composed of α-1,4 linked glucose units. The presence of n-dodecyl-β-D-maltoside which is composed of maltose (α-1,4 linked glucose) with an alkyl aglycone could also be a source of the 4-substituted glucose, as well as the terminal glucose. After acid hydrolysis, the glycosidic bond linking the two glucose units, as well as that bonding the alkyl chain to the disaccharide would be cleaved, resulting in terminal and 4-substituted glucose derivatives. The chromatogram peak intensity of the 4-substituted glucose in the *S. coelicolor* glycan sample was more than three times that of terminal glucose and more than forty times that of terminal, 2-substituted and 4-substituted mannose (Figure A.21). This is consistent with the observation that the α-cyclodextrin contaminant was the most abundant carbohydrate in the *S. coelicolor* glycan sample (Figure 5.1).

Taken together, these results suggest that the *S. coelicolor* glycans are composed of terminal, 2-substituted and 4-substituted mannose.

Table 5.1 GC-MS analysis of the PMAAs derived from the permethylated glycans, released by non-reductive β-elimination of *S. coelicolor* glycoproteins. The GC-MS chromatograms for each peak are shown in Figure A.21.

Monosaccharide[1]	Retention time (min)[2]	Fragment ions of the PMAA (*m/z*)
2-substituted mannose	15.585	190, 161, 130, 129
4-substituted mannose	15.751	233, 118, 162, 173
Terminal mannose	13.650	102, 118, 129, 145, 161, 162, 205
Terminal glucose	13.556	102, 118, 129, 145, 161, 162, 205
4-substituted glucose	15.839	233, 118, 162, 173

[1] The monosaccharide residue was identified by diagnostic fragment ions observed in the mass spectrum of the PMAAs (Figures A.17-A.20) and by co-elution with mannose or glucose standards from the column (Figure A.21).

[2] Retention time of the PMAAs separated by GC.

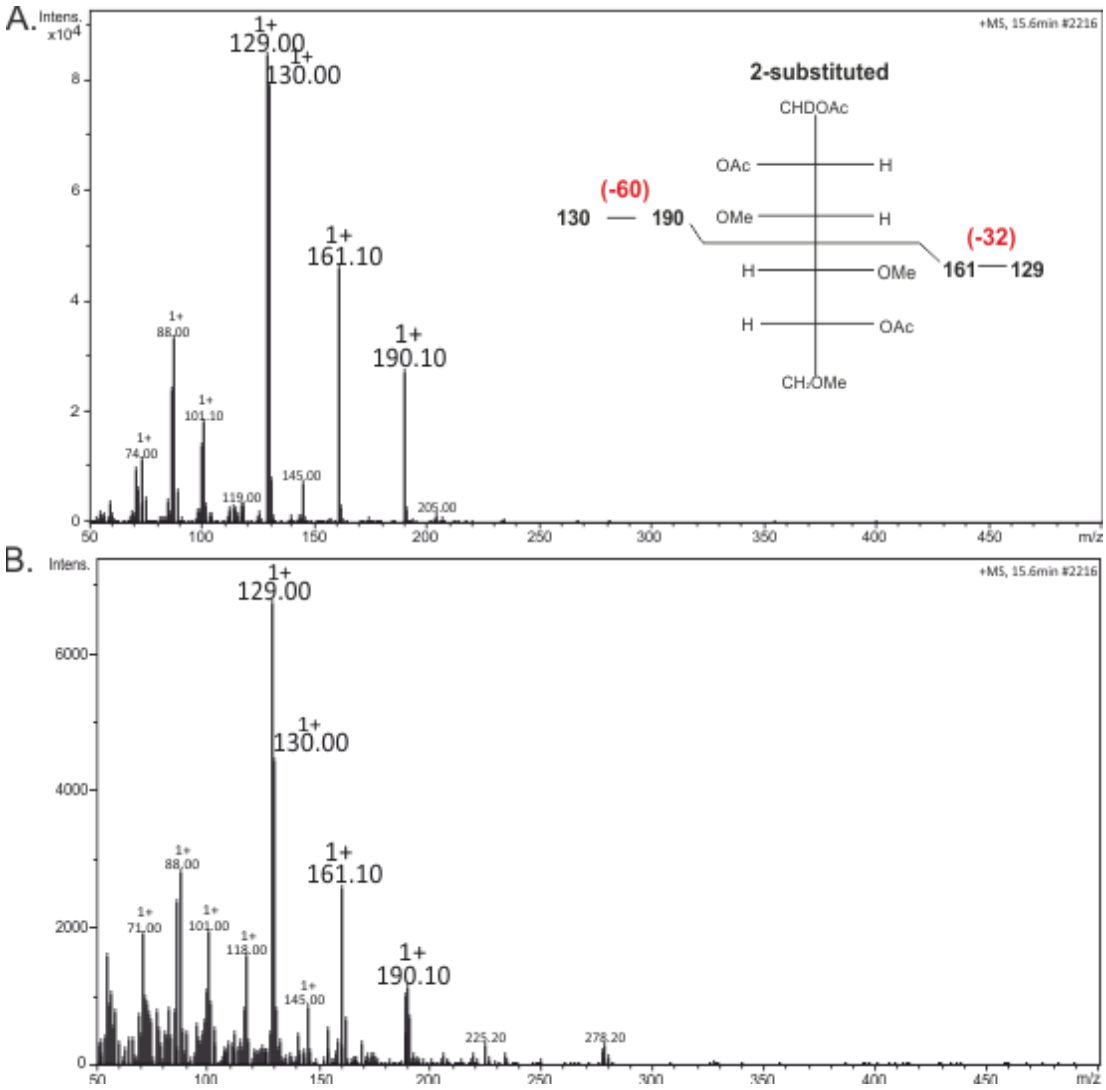


Figure 5.9 GC mass spectrum determining 2-substituted mannose in the partially methylated mannose standard (A) and the *S. coelicolor* glycans (B). Diagnostic fragment ions are *m/z* 190, 161, 130 and 129 are indicated.

5.3 Analysis of disaccharides and trisaccharides in the *S. coelicolor* glycans

After determining that the *S. coelicolor* glycans were composed of terminal, 2-substituted and 4-substituted mannose, an analysis of the di/tri saccharide alditols by GC-MS was carried out. This experiment was carried out to investigate the arrangement of the linkages identified in the analysis of PMAAs (Table 5.1) in the disaccharide and trisaccharide. The *S. coelicolor* glycans were reduced using $NaBD_4$, which reduces the aldose to form an alditol and simultaneously tags the anomeric carbon atom (C1) with a deuterium atom (Figure 5.10). The di/trisaccharides were methylated which modifies all of the free hydroxyls to methoxy groups. The permethylated di/trisaccharide alditols were then analysed by GC-MS. Disaccharides maltose (α-1,4 linked glucose) and α3 mannobiose (Manα(1->3)Man) were prepared and analysed alongside the *S. coelicolor* glycans as standards.

In the case of a disaccharide, the molecular mass of the monosaccharide units and the position of the glycosidic linkage can be determined based on the GC mass spectra (Kärkkäinen 1970). The primary fragmentation patterns that are expected from the (1->2), (1->3), (1->4) and (1->6) linked permethylated disaccharide alditols are exemplified in Figure 5.11. In the case of hexose disaccharides, ions at *m/z* 219 (B_1) and 236 (Z_1) are expected to form through the cleavage of the hexose unit from the molecule and from the cleavage to generate an ion containing the alditol unit respectively. Secondary and tertiary fragments of these ions at *m/z* 187 (219 - 32), 155 (187 – 32), 204 (236 – 32) and 172 (204 – 32) resulting from a loss of one and two molecules of methanol are also expected. Fragmentation within the alditol enables the assignment of the glycosidic linkage.

In the case of a trisaccharide it is possible to determine the molecular mass of the monosaccharide units and the glycosidic linkage to the alditol unit, based on the GC mass spectrum (Kärkkäinen 1971). The primary fragmentation patterns expected from permethylated trisaccharide alditols with (1->2), (1->3), (1->4) and (1->6) linkages to the alditol are exemplified in Figure 5.12. In the case of unbranched trisaccharides composed of hexoses, ions formed through the cleavage of the hexose unit and of the alditol unit from the molecule (at *m/z* 219 (B_1), 236 (Z_1), 423 (B_2) and 440 (Z_2)) are expected, as well as secondary and tertiary fragments resulting from loss of methanol. As in the case of the disaccharide alditols, fragmentation of the alditol backbone enables the assignment of the glycosidic linkage to the residue.

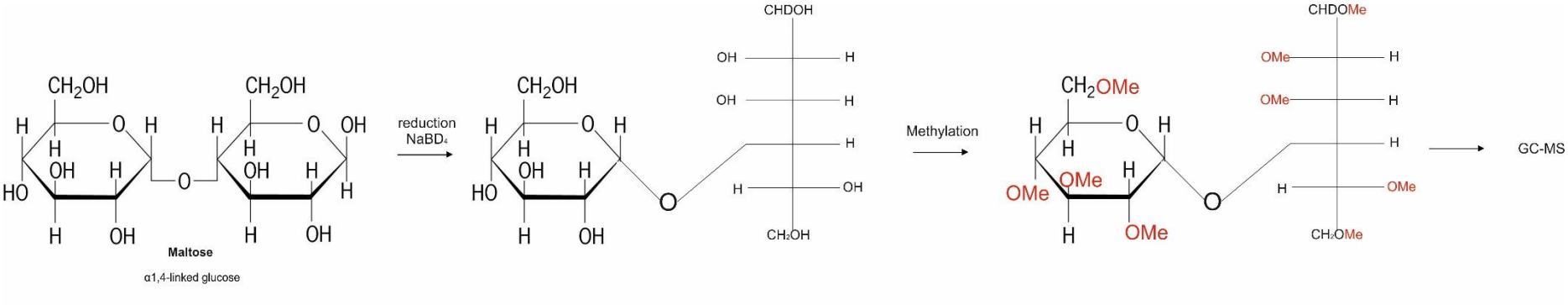


Figure 5.10 Analysis of methylated di/trisaccharide alditols exemplified using maltose. Carbohydrates are reduced using $NaBD_4$, methylated and analysed by GC-MS.

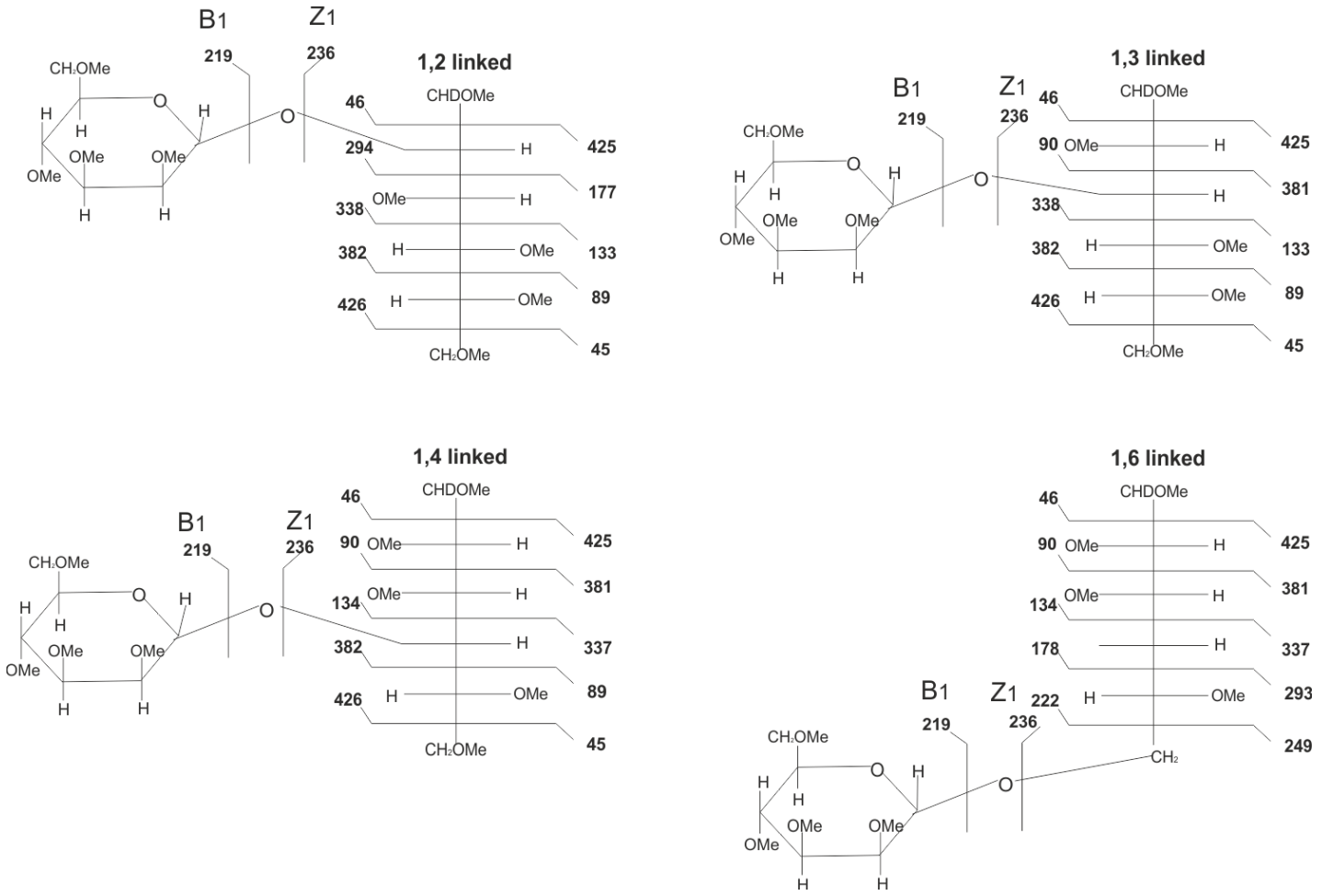


Figure 5.11 The predicted primary fragmentation of permethylated dihexose alditols with (1->2), (1->3), (1->4) and (1->6) linkages.

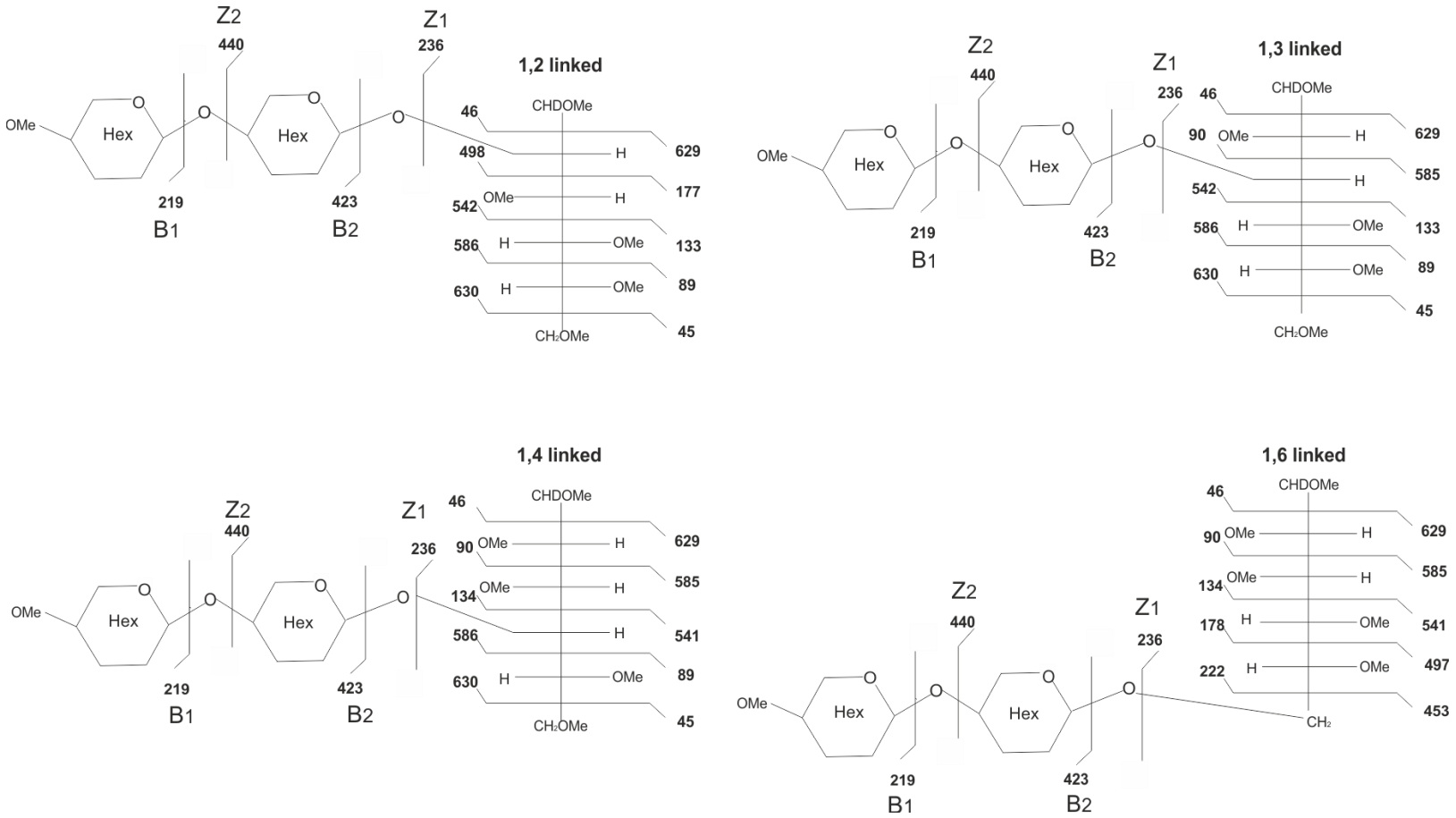


Figure 5.12 The predicted primary fragmentation of permethylated, unbranched trisaccharide alditols with (1->2), (1->3), (1->4) and (1->6) linkages respectively.

A. Standards

The relative retention times of the maltose and α3 mannobiose (standards) derived methylated disaccharide alditols were 9.900 min and 9.875 min respectively, and the partial mass spectra data are summarised in Table 5.2. The fragment ion intensities are shown as a percentage of that of the base peak at *m/z* 88, for each individual spectrum. The ions at *m/z* 219 and 236 formed as a result of the cleavage of the hexose unit and the alditol unit from the molecule respectively (Figure 5.11), were observed in the spectra of both of the standards. These ions can undergo secondary and tertiary fragmentations by the loss of one and two molecules of methanol, resulting in ions at *m/z* 187 (219 - 32), 155 (187 – 32), 204 (236 – 32) and 172 (204 – 32) respectively. The secondary and tertiary fragmentation ions were all observed in spectra of the standards. Taken together, these ions are indicative of hexose containing disaccharides.

The absence of ions at *m/z* 177 and 178 in the spectra of both standards rules out the possibility of (1->2) or (1->6) linked disaccharides respectively. The deuteration was necessary to distinguish between the (1->3) (α3 mannobiose) and (1->4) (maltose) linked disaccharide. The (1->4) linked disaccharide is identified by the presence of *m/z* 134 arising from the cleavage between the C3 and C4 bonds in the alditol moiety (Figure 5.11; Table 5.2), which is absent from the spectrum of the (1->3) disaccharide. Similarly, the presence of the ion at *m/z* 133 in the spectrum of the α3 mannobiose standard along with the absence of the ion at *m/z* 134 identifies the (1->3) linked disaccharide. Due to the mass cut-off of the instrument, ions at *m/z* 45 consistent with the cleavage between C5 and C6 ((1->2), (1->3) and (1->4) linked disaccharides) and at *m/z* 46 consistent with the cleavage between C1 and C2 in the alditol moiety ((1->2), (1->3), (1->4) and (1->6) linked disaccharides) could not be observed. The low intensity of the higher molecular mass fragment ions at *m/z* 425, 426, 381 and 382 meant that these ions were not particularly useful as diagnostic fragment ions in the linkage assignment. The ions at *m/z* 89 and 90 were not useful in distinguishing between (1->3) and (1->4) linked disaccharides due to the presence of the isotope peaks of the prominent ion at *m/z* 88. Therefore, the glycosidic linkage assignment is based on the product ions at *m/z* 177, 178, 133 and 134.

B. Sample

A peak in the *S. coelicolor* glycan sample with a retention time of 9.890 min was observed. The similarity in the retention time of the peak in the sample compared to the peaks in the

maltose and α3 mannobiose standards, suggested that the sample peak was a disaccharide. The partial mass spectra data of the unknown methylated disaccharide alditol in the *S. coelicolor* glycan sample is shown in Table 5.2. The fragment ion intensities are shown as a percentage of that of the base peak at *m/z* 101. The ions at *m/z* 219 and 236, as well as the secondary and tertiary fragment ions resulting from a neutral loss of methanol at *m/z* 187, 155, 204 and 172 respectively, are consistent with hexose residues. The ion at *m/z* 134 was more than double the intensity of the ion at *m/z* 133, suggesting that the disaccharide could be either (1->4) or (1->6) linked. However, the low abundance of the ion at *m/z* 178 compared to the ion at *m/z* 177 rules out the possibility of a (1->6) linked disaccharide. These results suggest the presence of a 4-linked disaccharide alditol made up of hexose monosaccharides. A second peak with a retention time of 12.985 min was observed in the *S. coelicolor* glycan sample, where the mass spectrum contained product ions that appeared to be consistent with the fragmentation of a permethylated trisaccharide alditol. The partial mass spectra data from the peak is shown in Table 5.2. The fragment ion intensities are shown as a percentage of the base peak at *m/z* 88. The later retention time observed (12.985 min) compared to the disaccharide standards (~ 9.9 min) suggests that the peak is not a disaccharide. The ions at *m/z* 219 and 236, as well as the secondary and tertiary fragment ions resulting from neutral loss of methanol at *m/z* 187, 155, 204 and 172 are consistent with the presence of hexose residues. The ion observed at *m/z* 440 is consistent with a Z_2 fragment ion of a permethylated trisaccharide alditol (Figure 5.12). The B_2 fragment ion at *m/z* 423 was observed in very low abundance. The higher abundance of the ion at *m/z* 133 compared to the ion at *m/z* 134 suggests that the linkage is either (1->2) or (1->3). The higher abundance of the ion at *m/z* 177 compared to the ion at *m/z* 178 suggests that the linkage is not (1->6) but could be (1->2). The fragment ion at *m/z* 89 was more than six times the intensity of the ion at *m/z* 90, which is consistent with a (1->2) linkage. Additionally, higher molecular mass ions consistent with a (1->2) linkage at *m/z* 498, 542 and 586 were observed in low abundance. No ions were observed at *m/z* 629 or 630. These results suggest the presence of a (1->2) substitution of the alditol residue in the trisaccharide.

Taken together, these results provide evidence for a (1->4) linked disaccharide in the *S. coelicolor* glycan sample. Additionally, there is preliminary evidence of a trisaccharide with a (1->2) linkage to the alditol moiety. However, without an appropriate trisaccharide standard it was not possible to determine the expected intensities of the diagnostic fragment ions of a permethylated trisaccharide alditol. In addition, a (1->2) linked di/trisaccharide standard

Table 5.2 Partial mass spectra data from the methylated disaccharide and trisaccharide alditols analysed by GC-MS. Ions pertinent to the structural interpretations of the glycans are shown. Intensities are shown as a percentage of the base peak.

m/z	Sample (RT: 9.890 min)	Maltose (RT: 9.900 min)	3α Mannobiose (RT: 9.875 min)	Sample (RT: 12.985 min)
88	88.1	100.0	100.0	100.0
89	33.5	53.1	26.2	29.6
90	13.3	21.1	13.1	4.7
101	100.0	100.0	94.2	69.0
133	3.1	3.2	5.6	8.8
134	6.6	10.4	0.8	5.0
145	9.0	12.3	5.4	13.9
146	1.8	2.2	2.0	4.4
155	18.0	25.7	7.4	12.2
172	10.3	18.0	15.5	10.1
177	1.7	0.3	0.2	3.1
178	1.2	0.6	0.1	1.2
187	64.3	99.5	33.1	24.2
204	1.6	1.6	0.6	1.2
219	9.5	12.6	19.6	6.4
222	0.8	0.4	0.6	1.3
236	30.6	50.7	28.2	59.4
249	1.4	1.3	0.8	1.1
293	0.7	0.6	0.3	0.4
294	0.5	0.1	0.1	0.4
296	2.7	3.4	3.7	0.7
337	0.1	0.1	-	0.4
338	0.1	0.1	0.1	0.2
359	-	-	-	0.4
376	-	-	-	0.3
381	0.2	0.3	0.0	0.4
382	0.2	0.3	0.1	0.4
391	-	-	-	0.4
408	-	-	-	0.3
423	-	-	-	0.4
425	0.2	0.3	0.0	0.3
426	0.2	0.4	0.0	0.3
440	-	-	-	4.4

m/z	**Sample (RT: 9.890 min)**	**Maltose (RT: 9.900 min)**	**3α Mannobiose (RT: 9.875 min)**	**Sample (RT: 12.985 min)**
498	-	-	-	0.1
541	-	-	-	
542	-	-	-	0.1
585	-	-	-	
586	-	-	-	0.2
629	-	-	-	-
630	-	-	-	-

could help to determine the lower molecular mass fragmentation pattern of the (1->2) linked sugar.

5.4 Discussion

In order to characterise the glycans modifying the *S. coelicolor* glycoproteins, the glycans were analysed by mass spectrometry. Using MALDI-FT-ICR-MS, the *S. coelicolor* glycans were shown to contain Hex_2 and Hex_3. There was significant contamination of the *S. coelicolor* glycan sample with a Hex_6 cyclic glucan, probably α-cyclodextrin most likely to have come from the SafeBlue protein stain used to stain the SDS-PAGE gels. There was no evidence of partially hydrolysed and permethylated forms of α-cyclodextrin, suggesting that the Hex_2 and Hex_3 observed in the *S. coelicolor* glycan sample were unlikely to be α-cyclodextrin derived. Additionally, despite efforts to remove the detergent n-dodecyl-β-D-maltoside, some low level contamination was observed in the *S. coelicolor* glycan sample.

The presence of Hex_2 and Hex_3 in the *S. coelicolor* glycans is consistent with observations made in the glycoproteomics characterisation of the *S. coelicolor* membrane proteome, where glycopeptides were shown to be modified with Hex, Hex_2 and Hex_3 glycan moieties (Chapter 4). Additionally, the previously characterised *S. coelicolor* glycoprotein PstS (SCO4142) was shown to be modified with a trihexose (Wehmeier et al. 2009). The carbohydrate methylation and linkage analysis of the *S. coelicolor* glycans resulted in the identification of terminal, 4-substituted and 2-substituted mannose, as well as terminal and 4-substituted glucose. The main source of the 4-substituted glucose is most likely to be the α-cyclodextrin, however the n-dodecyl-β-D-maltoside could also have contributed to this. The most likely source of the terminal glucose is the n-dodecyl-β-D-maltoside. The identification of 2-substituted mannose in the *S. coelicolor* glycans was not surprising considering the similarity to the pathway in *M. tuberculosis*, where the glycoprotein MPT32 was previously shown to be modified with both α (1->2) mannobiose and α (1->2), (1->2) mannotriose (Dobos et al. 1996). The identification of 4-substituted mannose in the *S. coelicolor* glycans however, was surprising since there is no record in the literature of (1->4) linked mannose in Actinobacterial glycoproteins. The configuration (α/β) of the linked mannose monosaccharides identified in this study was not determined, since the non-reductive β-elimination removes all O-linked glycans non-specifically. Additionally, the anomeric configuration of the monosaccharides identified in the linkage analysis is lost after hydrolysis.

The methylation analysis of the di/trisaccharides in this study provides preliminary evidence of a trisaccharide with a (1->2) linkage, which is consistent with observation of 2-substituted mannose in the *S. coelicolor* glycans on linkage analysis. Additionally, the observation of a disaccharide with a (1->4) linkage would be consistent with the observation of 4-substituted mannose in the *S. coelicolor* glycans on linkage analysis. Taken together these results could suggest that the glycans modifying the *S. coelicolor* glycoproteins are composed of trisaccharides that contain (1->2) linked mannose, and disaccharides that contain (1->4) mannose. Since only the linkage to the alditol in the permethylated trisaccharide alditol was characterised, it is possible that the trisaccharides are composed of both (1->2) and (1->4) linked mannose chains. However, this remains to be experimentally validated. Due to contamination of the *S. coelicolor* glycan sample with α-1, 4 linked glucose containing carbohydrates (α-cyclodextrin and n-dodecyl-β-D-maltoside), the origin of the disaccharide with a (1->4) linkage in the *S. coelicolor* glycan sample could not be conclusively determined. To gain more conclusive evidence that the (1->4) linked disaccharide is *S. coelicolor* glycoprotein derived, these experiments should be repeated with measures taken to eliminate the carbohydrate contaminants. SDS-PAGE gels could be stained with Coomassie blue R-250 stain without added cyclodextrins (He 2011). Additionally, the solubilisation of *S. coelicolor* membrane proteins could be carried out with a non-glucoside detergent or its removal could be monitored mass spectrometrically before beta-elimination is carried out.

Due to the experimental approach used in this study, where the glycans eliminated from the total *S. coelicolor* membrane glycoproteome fractions were analysed, it was not possible to determine which glycoproteins the respective di/trisaccharides originated from. However, the previous characterisation of the *S. coelicolor* membrane glycoproteome in this work (Chapter 4) has resulted the identification of 37 new glycoproteins, which could serve as targets for any future validation of the work reported in this chapter, on a single glycoprotein. The use of commercially available glycosidases that are specific for certain glycosidic linkages (e.g. α1,2 mannosidase) could be used alongside the β-elimination for the removal of glycans, whilst enabling the determination of the anomeric configuration of the glycosidic linkages. Additionally, the Glyconeer oligosaccharide synthesiser based in the University of York Department of Chemistry, could be used to synthesise an appropriate disaccharide and trisaccharide standard.

Chapter 6 - Characterisation of knockout mutants in genes encoding glycoproteins required for cell wall biosynthesis.

Chapter 6 – Characterisation of knockout mutants in genes encoding glycoproteins required for cell wall biosynthesis.

Among the diverse physiological roles of protein O-glycosylation in nature, the importance of glycosylation for maintaining growth and cell wall integrity has been widely reported. In fungi, the disruption of PMT family proteins has been shown to result in changes in growth, cell morphology, and cell wall rigidity and integrity (Gentzsch and Tanner 1996; Prill et al. 2005; Mouyna et al. 2010; Guo et al. 2016). In *M. smegmatis*, *pmt* mutants displayed a reduced tolerance to cell wall stress induced by SDS treatment, suggesting changes in the cell wall composition (Liu et al. 2013a). Additionally, Ppm1 in *M. tuberculosis* has been shown to be required for the synthesis of the cell wall glycoconjugates, lipomannan and lipoarabinomannan (Gurcha et al. 2002). Recently, Howlett et al. (2016) demonstrated that mutants defective in proteins required for the protein O-glycosylation pathway in *S. coelicolor* (discussed in detail in Chapter 1), were hypersusceptible to cell-wall targeting antibiotics. They hypothesised that protein O-glycosylation in *S. coelicolor* is required for the optimal function of membrane and periplasmic proteins, including cell wall biosynthetic enzymes.

During my characterisation of the *S. coelicolor* membrane glycoproteome (Chapter 4) a number of glycoproteins with predicted functions in the cell wall were identified. Included in these were two putative peptidoglycan biosynthetic enzymes, SCO4847 and SCO4934.

SCO4847 is a D-Ala-D-Ala carboxypeptidase and predicted low molecular mass penicillin binding protein (LMM PBP) (Sauvage et al. 2008). The prediction of a C-terminal transmembrane domain in SCO4847 (Chapter 4, Figure 4.9) is consistent with previous observations that many LMM PBPs are membrane associated (Pratt 2008). D-Ala-D-Ala carboxypeptidases recognise the terminal D-Ala-D-Ala of stem pentapeptides or the already crosslinked analogues (Pratt 2008). The reaction of the enzyme with the peptide stem results in the formation of an acyl enzyme intermediate, which then undergoes hydrolysis to cleave the terminal D-Ala residue and thereby limiting crosslinking (Figure 6.1).

SCO4934 is a putative L, D transpeptidase. L, D transpeptidases were first reported in *Enterococcus faecium* but have since been discovered in other bacteria, including *E. coli, Bacillus subtilis* and *M. tuberculosis* (Mainardi et al. 2002; Magnet et al. 2007; Bramkamp 2010; Sanders et al. 2014). As discussed in Chapter 1, L, D transpeptidases catalyse the crosslinking between the third amino acids (commonly meso diaminopimelic acid (meso-

DAP)) of tetrapeptide stems (known as 3->3 crosslinking). This is in contrast to the more widely studied crosslinking between the fourth and third amino acids respectively (normally meso-DAP and D-alanine) of peptidoglycan stem pentapeptides (4->3 crosslinking) that is carried out by D-Ala-D-Ala transpeptidases (Vollmer et al. 2008). Both 3->3 crosslinking and 4->3 crosslinking have been observed in *S. coelicolor* (Hugonnet et al. 2014).

The aim of the work in this chapter was to understand the roles of the cell wall active glycoproteins SCO4934 and SCO4847 in *S. coelicolor.* By the characterisation of *sco4934* and *sco4847* knockout mutants, I aim to investigate whether loss of these glycoproteins is likely to contribute to the antibiotic hypersensitivity phenotype observed previously in the *S. coelicolor* glycosylation deficient *pmt* mutant, such as DT1025 (discussed in Chapter 1). The *S. coelicolor* parent strain, J1929, will be used for generating the gene knockouts. I expect the mutations might result in one of several possible phenotypes:

1. A phenotype similar to that of the parent strain. This phenotype could be indicative of protein functional redundancy and that loss of this glycoprotein does not contribute to the antibiotic susceptible phenotype in the *pmt*$^{-}$ mutant.
2. A phenotype resembling that of DT1025 (*pmt*). Loss of this glycoprotein probably contributes to the antibiotic hypersensitivity phenotype in the *pmt* mutant.
3. A phenotype with increased antibiotic susceptibility to that of DT1025. This phenotype could indicate that both the glycosylated and the unglycosylated isoforms of the protein have a role in the intrinsic antibiotic resistance we observe in the parent strain.

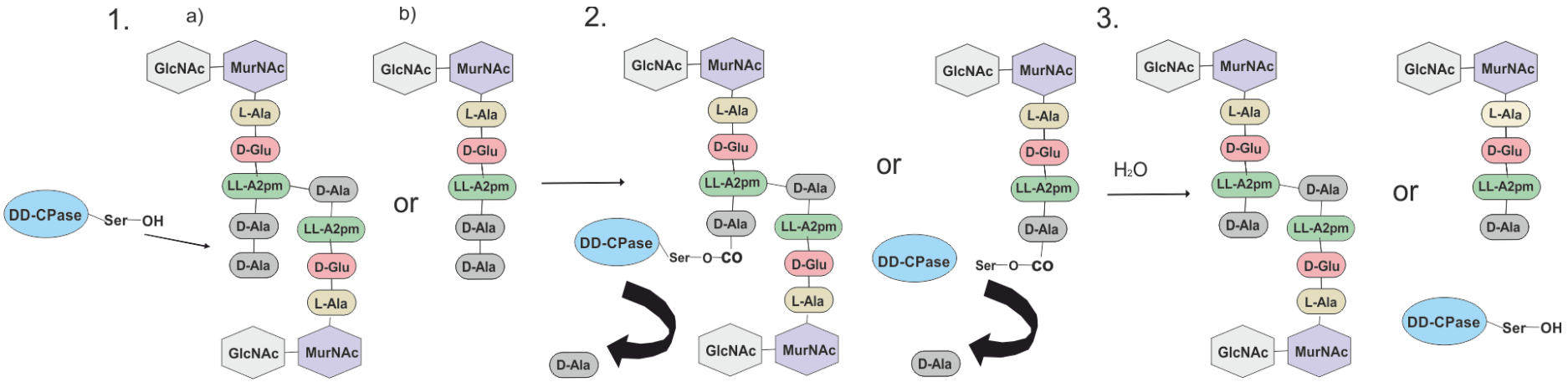


Figure 6.1 Schematic summary of the function of D-Ala-D-Ala carboxypeptidases in peptidoglycan crosslinking. D-Ala-D-Ala carboxypeptidases recognise the terminal D-Ala-D-Ala of stem pentapeptides (b) or the already crosslinked analogues (a) (1). The reaction of the enzyme with the peptide stem results in the formation of an acyl enzyme intermediate (2), which then undergoes hydrolysis to cleave the terminal D-Ala residue (3).

6.1 Transposon insertion mutagenesis of genes encoding glycoproteins SCO4847, SCO4934 and SCO5204.

In order to investigate the functions of the putative cell-wall biosynthetic enzymes SCO4934 and SCO4847, transposon insertion mutant strains were constructed. A transposon insertion mutant in *sco5204* was also constructed. *sco5204* encodes a putative integral membrane protein of unknown function, previously identified as a glycoprotein after the CID mass spectrometry analysis of Con A enriched membrane proteins (Chapter 4, Table 4.1).

The mutants were constructed by taking advantage of the *S. coelicolor* transposon insertion cosmid library available from Dr Paul Dyson (University of Swansea). The positions of the transposon insertions in *sco5204, sco4934* and *sco4847* (indicated in Figure 6.2) were identified using the StrepDB (http://strepdb.streptomyces.org.uk). Conserved domains in each of the three protein sequences were predicted using the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and transposon insertions positioned upstream of these were selected for mutant construction. The cosmids with the transposon insertions were isolated from the strains sent by Meirwyn Evans (University of Swansea). Restriction digestion of the cosmids with EcoRI confirmed that the cosmids contained Tn5062 at the correct site (Figure 6.3, Table 6.1). The cosmids were introduced into *E. coli* ET12567(pUZ8002) and then into *S. coelicolor* J1929 as described in Materials and Methods.

The apramycin resistant, kanamycin sensitive exconjugates from the 2St3B6.1.G02-J1929, 2St3B6.1.G06-J1929 (transposon insertions in *sco5204*) and 2SCK31.2.F11-J1929 (transposon insertion in *sco4934*) conjugations were screened by Southern blot analysis to verify the genomic integration of the cosmids (Figure 6.4 – 6.5). Genomic DNA isolated from the 2St3B6.1.G02:J1929 and 2St3B6.1.G06:J1929 exconjugates was digested with XhoI. Similarly, the genomic DNA isolated from the 2SCK31.2.F11-J1929 exconjugates was double digested with ScaI and SacI. The digested DNA was separated in a 1 % agarose gel and blotted onto a Zeta probe membrane (Bio-Rad) as described in the Materials and Methods. The probes were prepared by amplifying a region upstream of the *sco5204* and *sco4934* genes, and downstream of the XhoI (-214 bp) (Figure 6.4.A) and SacI (- 147 bp) (Figure 6.5.A) restriction sites respectively, using PCR.

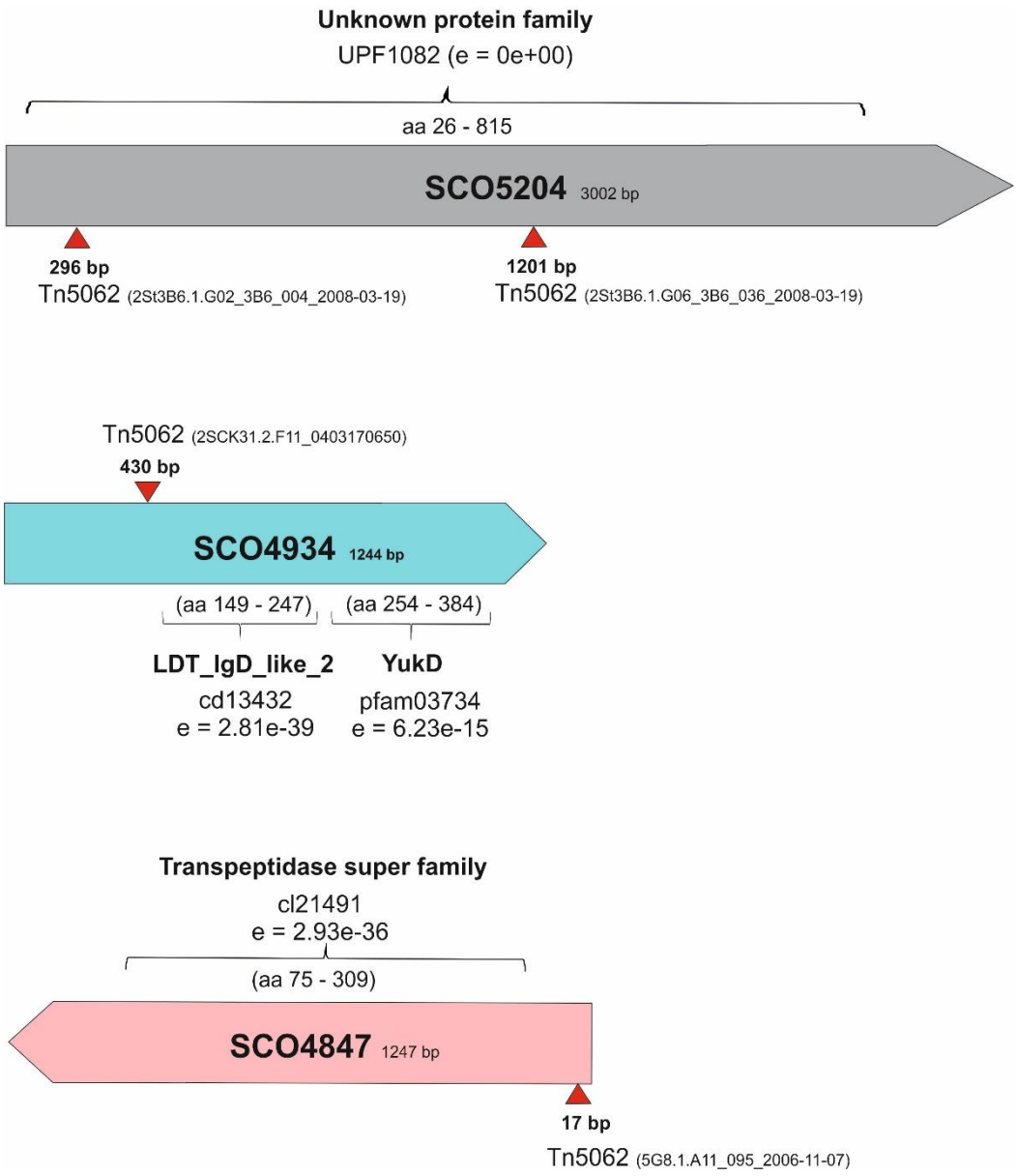


Figure 6.2 Conserved domains and the positions of the transposon insertions in *sco5204*, *sco4934* and *sco4847*. SCO5204 is predicted to belong to an uncharacterised family of integral membrane proteins (UPF1082). SCO4934 has a predicted lgD domain of actinobacterial L, D transpeptidases (LDT_lgD_like_2) as well as the catalytic YukD domain of L, D transpeptidases. SCO4847 has a predicted penicillin binding protein transpeptidase domain (cl21491). The expect (e) scores of each predicted domain generated by the CDD are indicated. Positions of the transposon insertions in the cosmids selected for mutagenesis are indicated by red arrow heads.

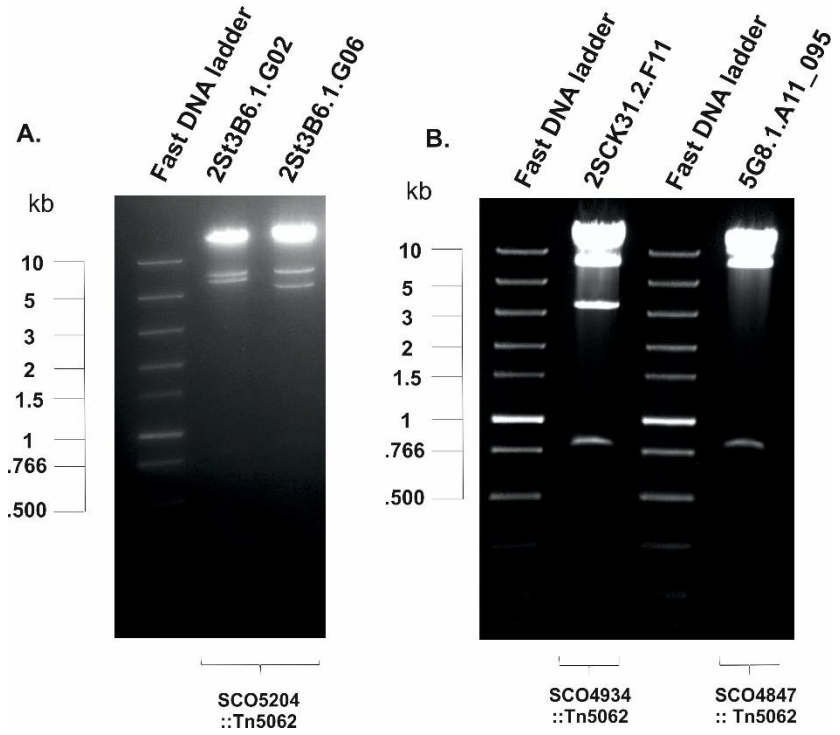


Figure 6.3 Restriction digests of the transposon insertion cosmids with EcoRI. The digested cosmids were separated in a 1 % agarose gel. Cosmids containing a transposon insertion in *sco5204* (2St3B6.1.G02, 2St3B6.1.G06) are indicated in panel A. Cosmids containing a transposon insertion in *sco4934* (2SCK31.2.F11) and *sco4847* (5G8.1.A11) respectively are indicated in panel B. The genotypes of the cosmids are indicated at the bottom of each panel.

Table 6.1 Expected fragment sizes for the transposon insertion containing cosmids after restriction digestion with EcoRI.

Gene	Transposon cosmid	Expected fragments sizes (bp) after EcoRI digest
sco5204	2St3B6.1.G02	38839, 6792, 6791, 5942, 782, 1
sco5204	2St3B6.1.G06	39744, 6792, 6791, 5037, 782, 1
sco4934	2SCK31.2.F11	24662, 13432, 6791, 3172, 782, 1
sco4847	5G8.1.A11	16696, 12103, 12058, 6791, 782, 1

A signal at ~ 2405 bp was observed in all of the 2St3B6.1.G02:J1929 exconjugates screened, while the wild type signal at 3035 bp was absent (Figure 6.4.C). These results indicate the presence of double cross-overs, thus confirming the successful generation of *sco5204* mutants. None of the 2St3B6.1.G06:J1929 exconjugates were confirmed as double cross-overs, as it was difficult to distinguish between the wild type signal at 3035 bp and the mutant signal at 3310 bp.

A signal at ~ 1901 bp was observed in all of the 2SCK31.2.F11-J1929 exconjugates screened, while the wild type signal at 987 bp was absent (Figure 6.5.C). These results indicate the presence of double cross-overs, thus confirming the successful generation of *sco4934* mutants.

Attempts to confirm the *sco4847* mutants using Southern blotting were unsuccessful, therefore the genomic DNA from the apramycin resistant, kanamycin sensitive exconjugates from the 5G8.1.A11-J1929 conjugation was screened using PCR (Figure 6.6). PCR primers were designed upstream and downstream of the transposon insertion (Figure 6.6.A). Two of the five exconjugates screened contained a PCR product at 4017 bp while the wild type PCR product at 575 bp was absent. These results are consistent with the presence of double cross-overs, thus confirming the presence *sco4847* mutants (Figure 6.6.B). The new knockout strains and genotypes are summarised in Table 6.2.

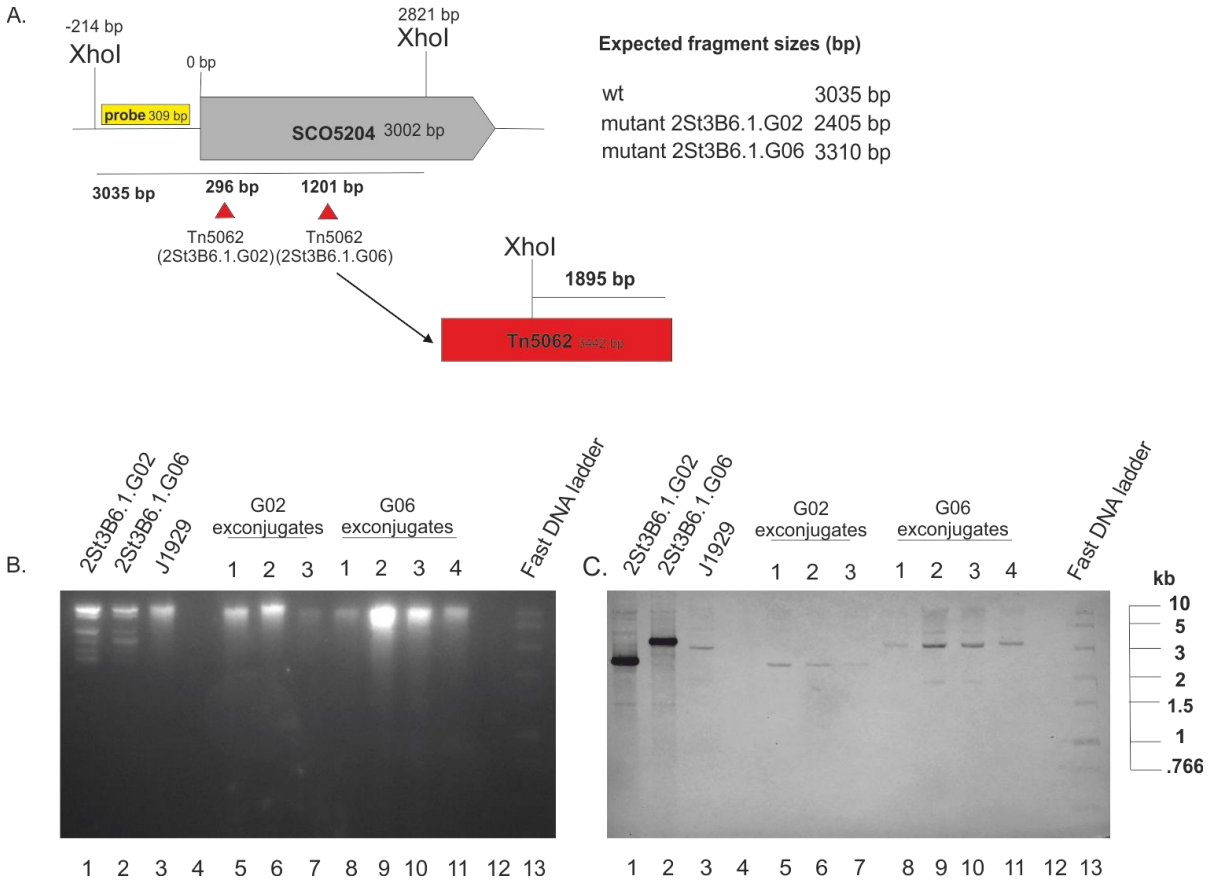


Figure 6.4 Southern blot analysis of the putative transposon mutants in *sco5204*. A. A schematic showing the genomic positions of the XhoI restriction sites relative in *sco5204* and the position of the transposon insertions. The expected fragment sizes are indicated. B. Ethidium bromide stained agarose gel showing the digested and separated genomic DNA prior to alkaline blotting. C. Southern blot of 2St3B6.1.G02-J1929 (G02) 2St3B6.1.G06-J1929 (G06) exconjugates probed with a DIG labelled probe, using the DIG High Prime DNA labelling and detection kit (Roche Life Science). J1929 genomic DNA served as a negative control. 2St3B6.1.G02 and 2St3B6.1.G06 cosmid DNA served as the positive controls.

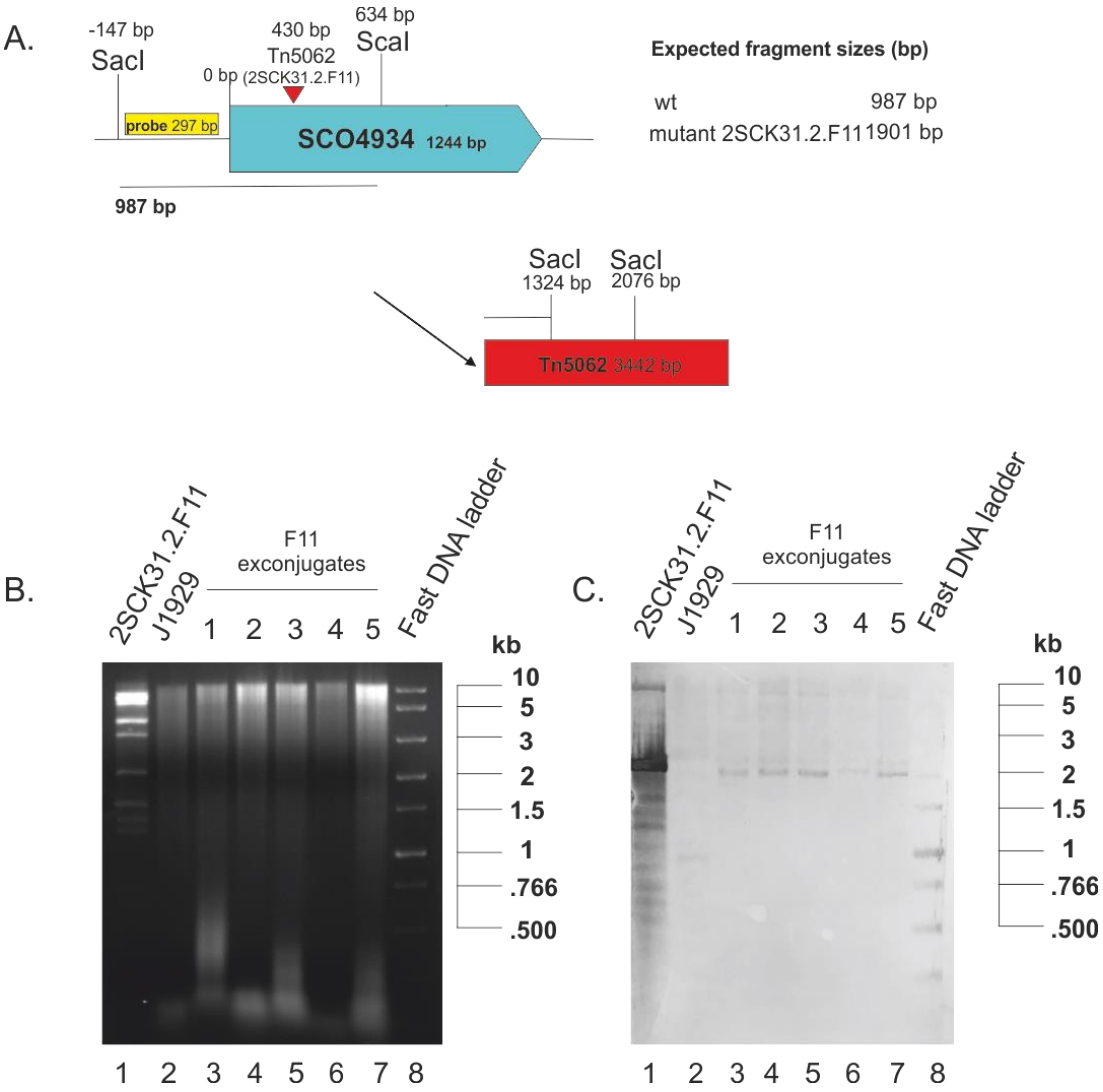


Figure 6.5 Southern blot analysis of the putative transposon mutants in *sco4934*. A. A schematic showing the genomic positions of the ScaI and SacI restriction sites relative to *sco4934* and the position of the transposon insertions. The expected fragment sizes are indicated. B. Ethidium bromide stained agarose gel showing the digested and separated genomic DNA prior to alkaline blotting. C. Southern blot of 2SCK31.2.F11-J1929 (F11) exconjugates probed with a biotin labelled probe, using the Biotin chromogenic detection kit (Thermo Fisher Scientific). J1929 genomic DNA served as a negative control. 2SCK31.2.F11 cosmid DNA served as a positive control.

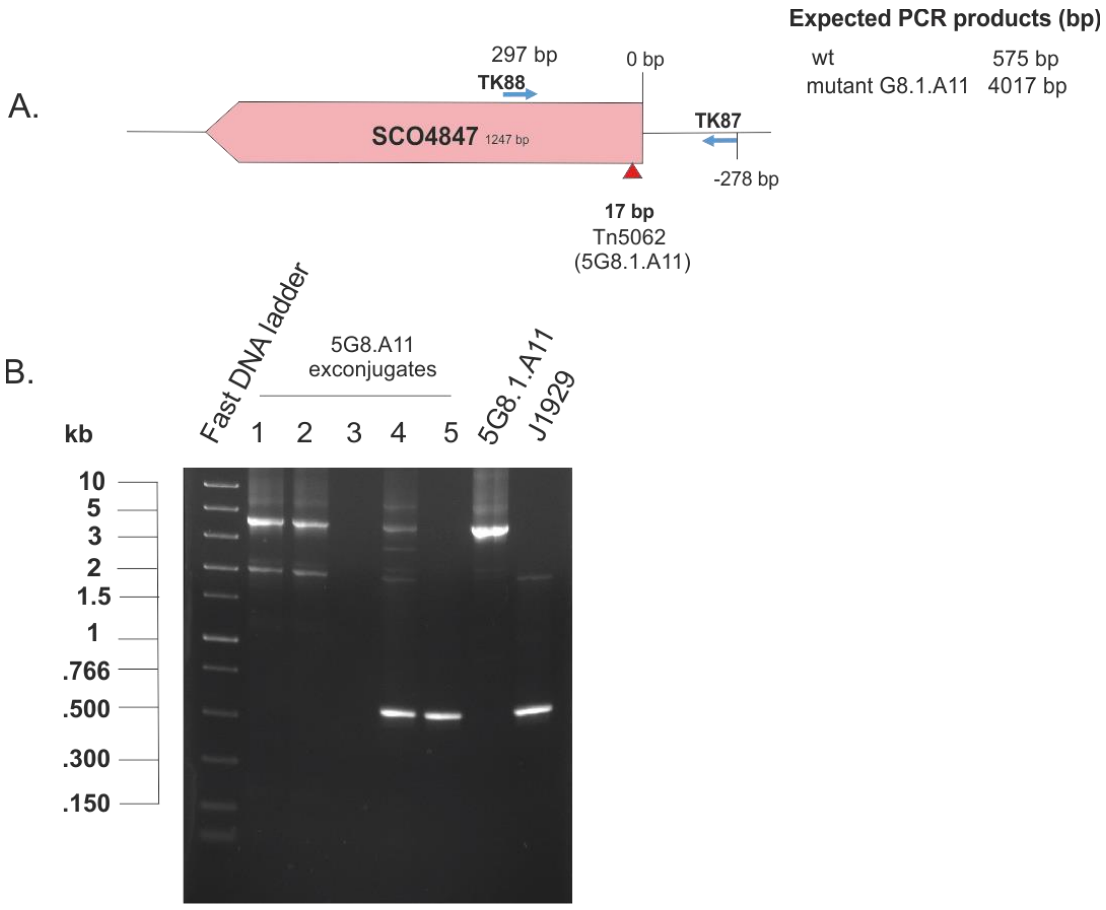


Figure 6.6 PCR analysis of the putative transposon mutants in *sco4847*. A. A schematic showing the positions of the primers used to confirm the presence of double crossovers and the expected sizes of the PCR products. B. Ethidium bromide stained agarose gel showing the PCR products generated after the amplification of genomic DNA from 5G8.1.A11-J1929 exconjugates using primers TK87 and TK88. J1929 genomic DNA served as a negative control. 2SCK31.2.F11 served as a positive control.

Table 6.2 New strains generated by transposon insertion mutagenesis of *sco5204*, *sco4847* and *sco4934* respectively.

Strain	Gene knockout	Transposon cosmid
TK005	*sco5204*	2St3B6.1.G02
TK006	*sco4847*	5G8.1.A11
TK008	*sco4934*	2SCK31.2.F11

In order to investigate whether the glycoprotein deficient mutants displayed changes in colony morphology as displayed previously in the DT3017 (*ppm1*) strain, TK005 (*sco5204*), TK006 (*sco4847*) and TK008 (*sco4934*) spores were plated onto DNA and grown for 2 days (Figure 6.7). J1929 spores served as a negative control. DT1025 (*pmt*) and DT3017 (*ppm1*) spores were grown for comparison. No changes in colony morphology compared to J1929 were observed in any of the glycoprotein deficient mutants. The ϕC31cΔ25 phage sensitivity of the glycoprotein deficient mutants was tested on DNA (Figure 6.8). Large clear plaques bearing resemblance to those observed on J1929 were observed on TK005 (*sco5204*), TK006 (*sco4847*) and TK008 (*sco4934*). These results indicate that the ϕC31cΔ25 phage receptor is present in all of the knockout mutants, suggesting that neither SCO5204, SCO4934 nor SCO4847 is the ϕC31cΔ25 phage receptor in *S. coelicolor*.

To investigate whether the glycoprotein deficient mutants displayed changes in sporulation they were streaked onto MSA and grown for 7 days (Figure 6.9). All three mutant strains sporulated efficiently and displayed no changes in pigment production compared to J1929. However, after the mutants were plated onto SMMS medium both TK006 (*sco4847*) and TK008 (*sco4934*) displayed a slight increase in antibiotic production compared to the parent strain J1929 (Figure 6.10). No increase in antibiotic production was observed in the TK005 (*sco5204*) strain. These results suggest that the TK006 (*sco4847*) and TK008 (*sco4934*) strains could be stressed, since the early onset of pigment production is a generic stress response in *S. coelicolor*.

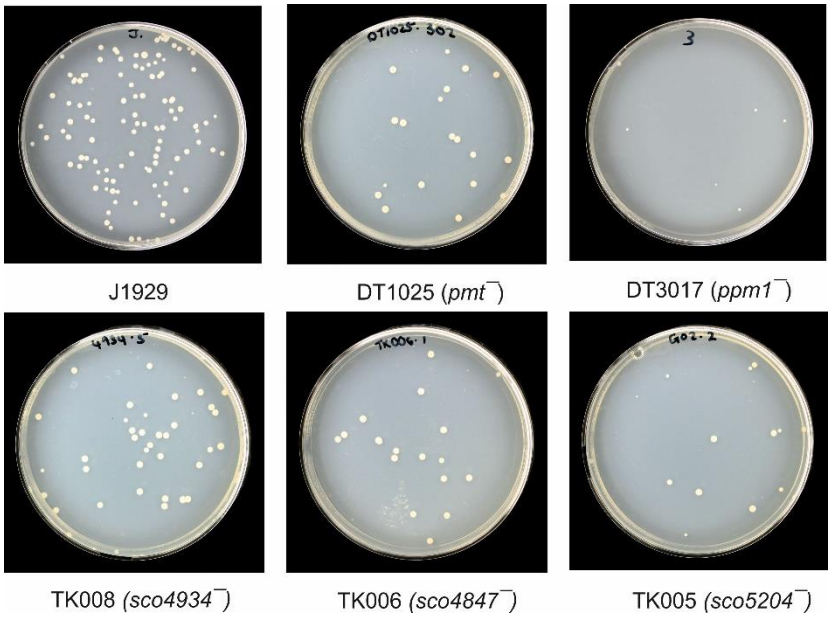


Figure 6.7 The colony morphology of the glycoprotein deficient mutants on DNA. Images are representative of at least two biological replicates and two technical replicates.

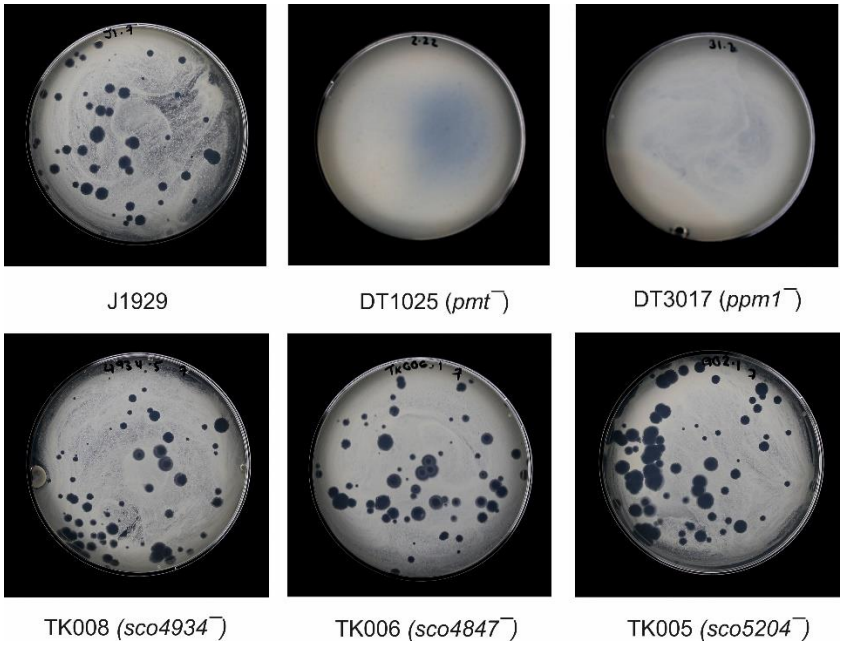


Figure 6.8 The ϕC31cΔ25 phage sensitivity of the glycoprotein deficient mutants on DNA. Images are representative of two biological replicates repeated at least once. ϕC31cΔ25 (~ 1 x 10^3 pfu) was inoculated with J1929, TK008, TK006 and TK005 whereas ~ 1 x 10^8 pfu of ϕC31cΔ25 was inoculated with DT1025 and DT3017. Images are representative of at least two biological replicates and two technical replicates.

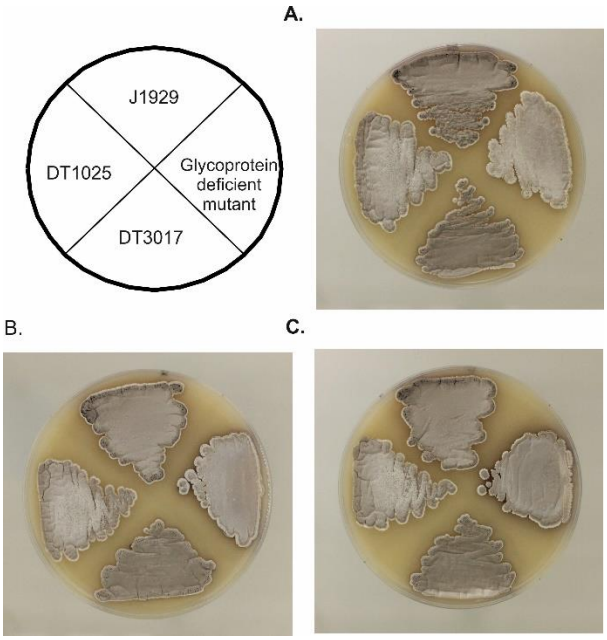


Figure 6.9 Sporulation of the glycoprotein deficient mutants MSA. A. TK005 (*sco5204*). B. TK008 (*sco4934*). C. TK006 (*sco4847*). Images are representative of at least two biological replicates and two technical replicates.

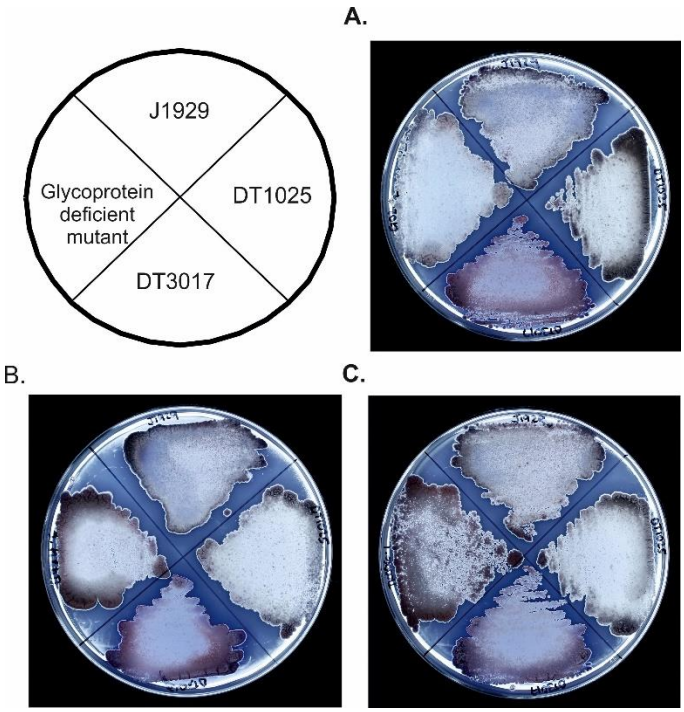


Figure 6.10 Antibiotic production of the glycoprotein deficient mutants SMMS. A. TK005 (*sco5204*). B. TK008 (*sco4934*). C. TK006 (*sco4847*). Approximately 10^6 spores were inoculated. Images are representative of at least two biological replicates and two technical replicates.

The antibiotic sensitivity of *S. coelicolor* TK005 (*sco5204*), TK006 (*sco4847*) and TK008 (*sco4934*) was measured on DNA using disc diffusion assays (Figure 6.11). *S. coelicolor* J1929 served as a negative control. DT1025 (*pmt*) and DT3017 (*ppm1*) served as a comparison for the severity of the antibiotic hypersensitivity phenotypes. The TK005 (*sco5204*), TK006 (*sco4847*) and TK008 (*sco4934*) strains displayed increased susceptibilities to imipenem, meropenem, ampicillin and penicillin (i.e. β-lactam antibiotics) compared to the parent strain J1929 (Figure 6.11.A-C). Additionally, the TK006 (*sco4847*) strain displayed a slight increase in sensitivity to vancomycin (Figure 6.11.B) and the TK005 (*sco5204*) strain displayed a decrease in sensitivity to rifampicin (Figure 6.11.A) compared to the parent strain J1929. The β-lactam hypersensitivity phenotype observed in the TK006 (*sco4847*) and TK008 (*sco4934*) strains was more severe than the hypersensitivity displayed by the DT1025 (*pmt*) strain. In contrast, the hypersensitivity phenotype displayed by the TK005 (*sco5204*) strain was similar to that of the DT1025 (*pmt*) strain. These results show that not all knockout mutants of glycoproteins in *S. coelicolor* display the same antibiotic hypersensitivity profiles. The increased sensitivity of the TK006 (*sco4847*) and TK008 (*sco4934*) strains to β-lactams compared to the DT1025 (*pmt*) strain suggests that non-glycosylated SCO4847 and SCO4934 in the DT1025 (*pmt*) strain have some activity. Overall, with the exception of rifampicin which targets RNA polymerase, the majority of antibiotics with increased efficacy against the glycoprotein deficient mutant strains target different stages in peptidoglycan biosynthesis.

Due to the predicted functions of SCO4847 and SCO4934 in cell wall biosynthesis, the sensitivity of the TK006 (*sco4847*) and TK008 (*sco4934*) strains to lysozyme was assayed. Lysozyme is a glycosyl hydrolase that cleaves the β 1,4 linkage between adjacent N -acetylmuramic acid and N-acetylglucosamine units in peptidoglycan (Salton and Ghuysen 1960). *S. coelicolor* TK006 (*sco4847*) and TK008 (*sco4934*) spores were serial diluted, plated onto Difco nutrient agar with 0.25 mg/mL of lysozyme and grown for 60 h. (Figure 6.12). Spores were plated onto Difco nutrient agar without lysozyme as a negative control for lysozyme treatment. J1929 served as a negative control for lysozyme hypersensitivity. The TK006 (*sco4847*) strain was dramatically more sensitive to the lysozyme treatment than both the parent J1929 and DT1025 (*pmt*) (Figure 6.12.A). The lysozyme sensitivity of TK008 (*sco4934*) however, appeared to similar to that of the parent strain J1929 (Figure 6.12.B).

Taken together, these results suggest that SCO4934 and SCO4847 are both required for maintaining cell wall integrity. However, the lack of lysozyme hypersensitivity observed in the TK008 (*sco4934*) strain could suggest that there is some functional redundancy in its role in cell wall biosynthesis.

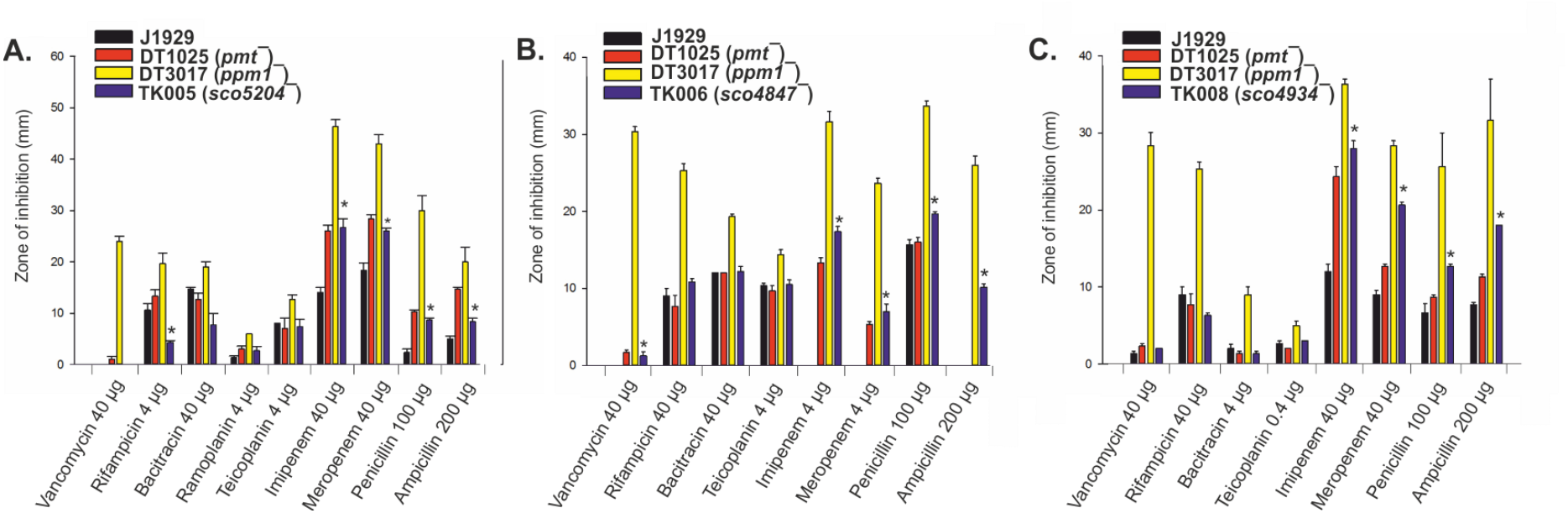


Figure 6.11 Antibiotic sensitivity of the glycoprotein deficient mutants TK005 (*sco5204*) (A), TK006 (*sco4847*) (B) and TK008 (*sco4934*) (C). Shown are the diameters of growth inhibition zones from disc diffusion assays for each mutant against the parent strain J1929, and the glycosylation deficient strains DT1025 (*pmt*) and DT3017 (*ppm1*). Bars represent the mean of three biological replicates except for TK006, where the bars indicate the mean of two biological replicates tested three times each. Error bars indicate SEM. * indicates $p < 0.05$ that the observed difference between the glycoprotein deficient strains (TK005, TK006 and TK008 respectively) and J1929 has occurred by chance. The full data set is in Table A.4 – Table A.6.

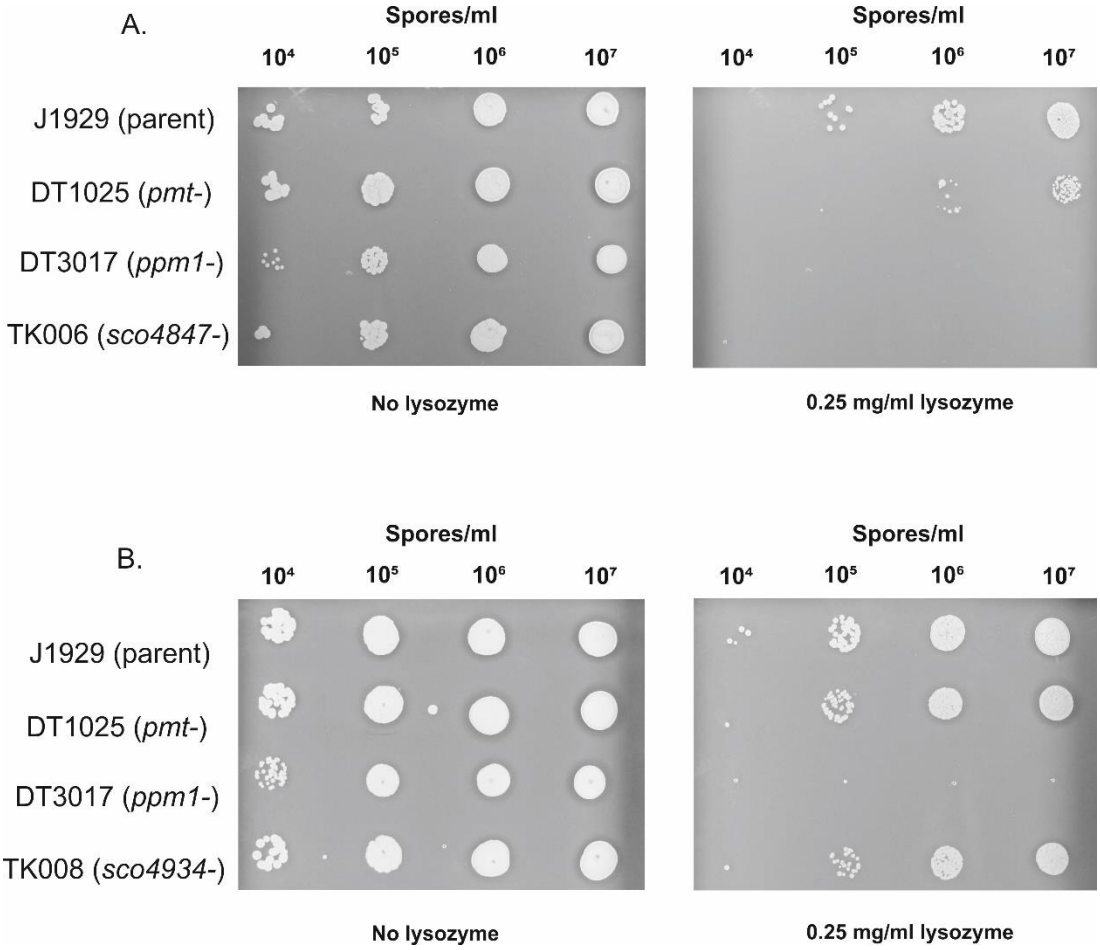


Figure 6.12 Lysozyme sensitivity of TK006 (*sco4847*) (A) and TK008 (*sco4934*) (B) compared to the parent strain J1929, DT1025 (*pmt*) and DT3017 (*ppm1*). Spores were adjusted to 10^8 spores/mL and a ten-fold serial dilution was carried out to get 10^4 spores/mL. 5 µL of each spore stock was plated onto DNA without lysozyme (left-hand panel) and with 0.25 mg/mL of lysozyme (right-hand panel). Images are representative of two biological replicates tested twice each (A) or three biological replicates tested once (B).

Alternatively, it is possible that if the majority of the peptidoglycan in *S. coelicolor* undergoes 4->3 crosslinking rather than the 3->3 crosslinking catalysed by L, D transpeptidases, the loss of SCO4934 may have less of an impact on peptidoglycan biosynthesis overall. SCO5204 may also be required for cell wall integrity, but its role is unclear.

6.2 Complementation of TK005 (*sco5204*), TK006 (*sco4847*) and TK008 (*sco4934*)

To confirm that the changes in the phenotypes observed in the TK005 (*sco5204*), TK006 (*sco4847*) and TK008 (*sco4934*) strains were directly linked to the disruption of *sco5204*, *sco4847* and *sco4934* respectively, it was decided to attempt to complement the mutants. The wild type copies of *sco5204*, *sco4847* and *sco4934* respectively were amplified by PCR and cloned into the NdeI restriction site of the shuttle vector pIJ10257 (as described in Materials and Methods). The new constructs were confirmed by restriction digests with BamHI and EcoRI to confirm the presence of the insert (Figure 6.13). The new constructs and expected digest fragments are summarised in Table 6.3. All fragment sizes in the restriction digests were as expected. Sequencing of the plasmids confirmed the presence of the correct insert. The constructs were introduced into *E. coli* ET12567(pUZ8002) and then into the *S. coelicolor* glycoprotein deficient mutants as described in the Materials and Methods. The empty pIJ10257 vector was introduced into TK005 (*sco5204*), TK006 (*sco4847*) and TK008 (*sco4934*) to serve as a negative control for the complementation.

The shuttle vector pIJ10257 enables the constitutive expression of genes under the *ermEp** promoter (Hong et al. 2005). The plasmid encodes the phiBT1 phage integrase and *attP* site which enables the site specific integration of the vector into the phiBT1 *attB* site in the *S. coelicolor* genome (Gregory et al. 2003). Genomic DNA was isolated from the exconjugates and the presence of the integrated plasmids were confirmed by PCR. To confirm the presence of the empty pIJ10257 vector in the TK005 (*sco5204*), TK006 (*sco4847*) and TK008 (*sco4934*) strains respectively, PCR primers were designed in the *S. coelicolor* genome upstream of *sco4848* and in the pIJ10257 vector upstream of the *ermE*p* site (Figure 6.14.A). To confirm the presence of pTAK29 in TK005 (*sco5204*) PCR primers were designed on the vector backbone and in the *sco5204* gene (Figure 6.14.B). To confirm the presence of pTAK30 and pTAK32 in TK006 (*sco4847*) and TK008 (*sco4934*) respectively, PCR primers were designed on the vector backbone flanking the *sco4847* and *sco4934* genes respectively (Figure 6.14.C-D).

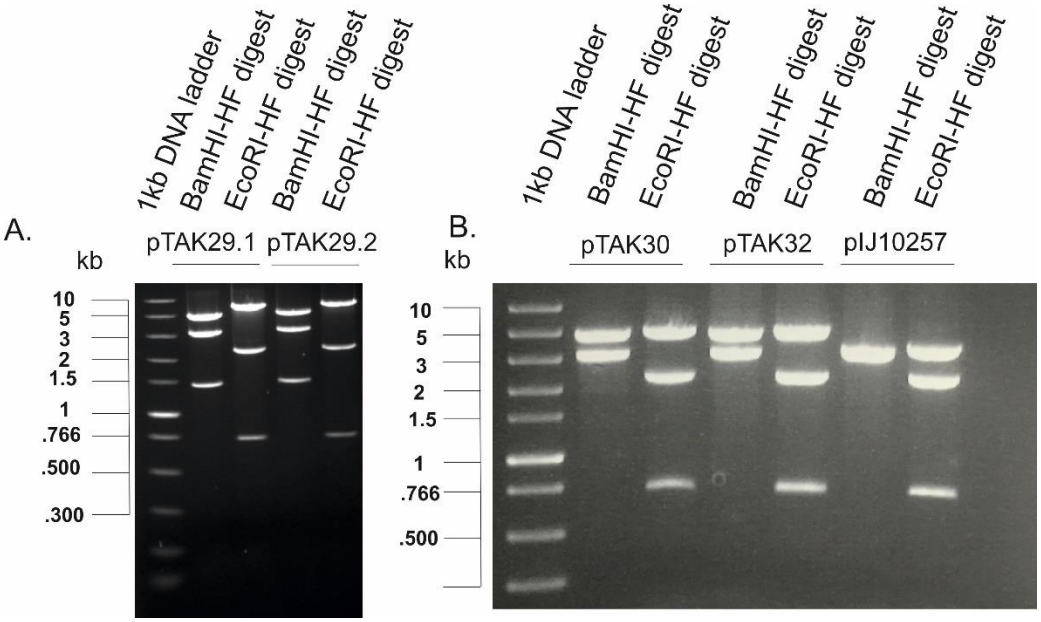


Figure 6.13 Restriction digests to confirm the correct construction of pTAK29, pTAK30 and pTAK32. A. Digestion of pTAK29 (pIJ10257 with *sco5204*) constructs with BamHI-HF and EcoRI-HF. B. Digestion of pTAK30 (pIJ10257 with *sco4847*), pTAK32 (pIJ10257 with *sco4934*) and pIJ10257 constructs with BamHI-HF and EcoRI-HF.

Table 6.3 New constructs generated by cloning *sco5204* (pTAK29), *sco4847* (pTAK30) and *sco4934* (pTAK32) respectively into pIJ10257.

Name	Construct	Function	BamHI-HF fragments (bp)	EcoRI-HF fragments (bp)
pIJ10257	Shuttle vector; Integrates into phiBT1 attachment site (attB); constitutive *ermEp** promoter in front of the MCS; Hygromycin^R .	Cloning vector in *E. coli* and integrating vector in *S. coelicolor*	3280, 3144	2466, 3246, 712
pTAK29	*sco5204* in pIJ10257	Complementation of *sco5204* mutants (TK005)	4896, 1393, 3144	2466, 6255, 712
pTAK30	*sco4847* in pIJ10257	Complementation of *sco4847* mutants (TK006)	4604, 3144	2466, 4570, 712
pTAK32	*sco4934* in pIJ10257	Complementation of *sco4934* mutants (TK008)	4531, 3144	2466, 4497, 712

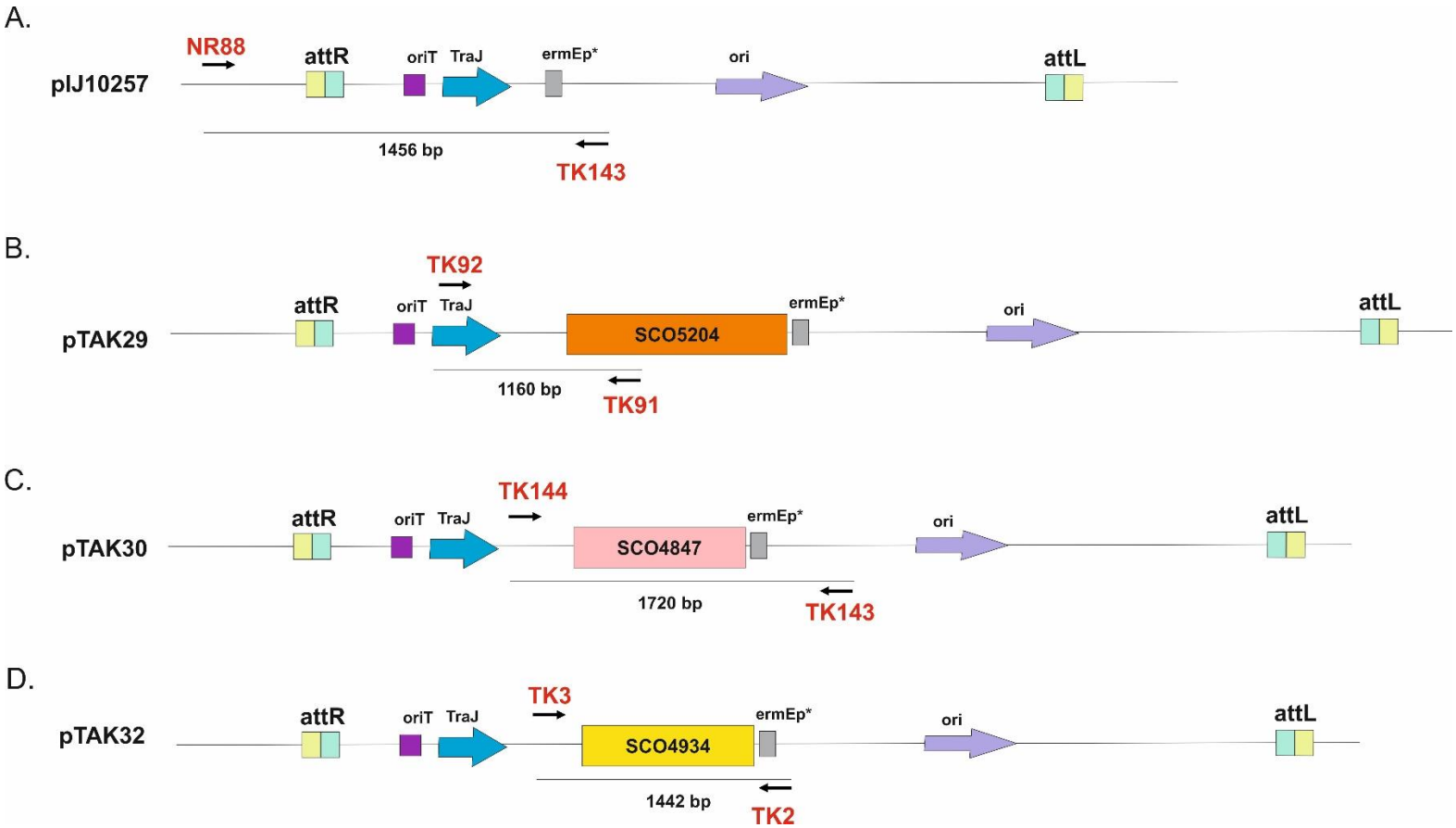


Figure 6.14 PCR primer design to confirm the presence of the integrated pIJ10257 (A), pTAK29 (B), pTAK30 (C) and pTAK32 (D) constructs. OriT: origin of transfer; TraJ: oriT recognising protein; Ori: origin of replication; ermEp*: ermE promoter.

For each PCR the negative control was J1929 genomic DNA and a positive control was the plasmid that was introduced into each respective strain. The PCRs confirmed that all of the pTAK29:TK005, pTAK32:TK008, pIJ10257:TK005, pIJ10257:TK006 and pIJ10257:TK008 exconjugates screened contained the integrated plasmids. However only four of the pTAK30:TK006 exconjugates (Figure 6.15.B – lanes 2,3,8 and 9) were positive for the integrated vector. The new strains, PCR primers and expected sizes of the PCR products are summarised in Table 6.4.

To investigate if the increase in antibiotic production observed previously in the TK006 (*sco4847*) and TK008 (*sco4934*) strains (Figure 6.10) was complemented in TK013 (TK006: pTAK30) and TK010 (TK008:pTAK32) respectively, the strains were plated onto SMMS and grown for 7 days (Figure 6.16). The slight increase in antibiotic production was still observed in TK013 (Figure 6.16.A) and TK010 (Figure 6.16.B). Complementation of the lysozyme sensitivity phenotype observed previously in TK006 (*sco4847*) (Figure 6.12.A) was observed after the reintroduction of wild type *sco4847* in the TK013 (TK006: pTAK30) strain (Figure 6.17.A). No reduction in lysozyme sensitivity was observed in the empty vector control strain TK016 (TK006:pIJ10257) as expected.

Partial complementation of the β-lactam hypersensitivity observed previously in the TK006 (*sco4847*) (Figure 6.11.B) and TK008 (*sco4934*) (Figure 6.11.C) strains was observed upon re-introduction of the wild type *sco4847* and *sco4934* in TK013 (TK006: pTAK30) (Figure 6.18.A) and TK010 (TK008:pTAK32) (Figure 6.18.B) respectively. Additionally, a slight reduction in vancomycin sensitivity was observed in TK013 (TK006: pTAK30). The β-lactam hypersensitivity observed previously in TK005 (*sco5204*) (Figure 6.11.A) was not complemented in TK012 (TK005: pTAK29) (Figure 6.18.C). However, the constitutive expression of *sco5204* in TK012 (TK005: pTAK29) appeared to result in increased sensitivity to imipenem, penicillin and ampicillin. Taken together these results suggest that the phenotypes observed as a result of knocking out *sco4847* and *sco4934* in TK006 (*sco4847*) and TK008 (*sco4934*) respectively, are directly linked to the disruption of the respective genes. Additionally, the slightly exacerbated phenotypes observed in TK012 (TK005: pTAK29) might suggest that constitutive expression of *sco5204* is damaging in *S. coelicolor*.

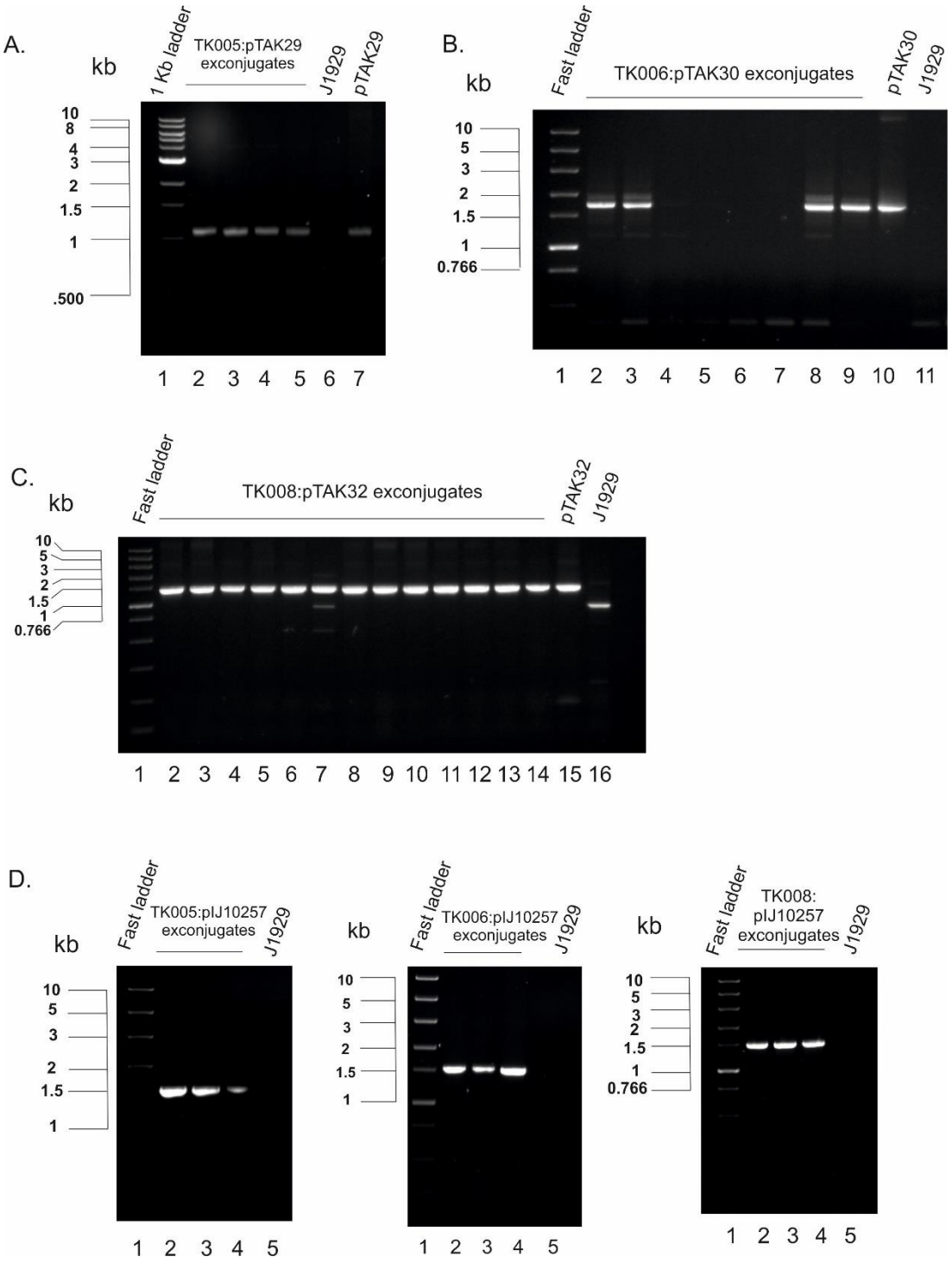


Figure 6.15 PCRs to confirm the genomic integration of the complementation constructs. Panels indicate: A. pTAK29 integration into TK005. B. pTAK30 integration into TK006. C. pTAK32 integration into TK008. D. pIJ10257 integration into TK005, TK006 and TK008 respectively.

Table 6.4 New *S. coelicolor* strains generated by complementation of the glycoprotein deficient mutants and the expected sizes of the PCR products.

***S. coelicolor* strain**	**Genotype**	**Description**	**PCR primers**	**Expected PCR products (bp)**
J1929	N/A	Parent strain	NR88, TK143	no product
			TK91, TK92	no product
			TK143, TK144	no product
			TK2, TK3	no product
TK010	TK008: pTAK32	Complementation of *sco4934* knockout	TK2, TK3	1442
TK012	TK005: pTAK29	Complementation of *sco5204* knockout	TK91, TK92	1160
TK013	TK006: pTAK30	Complementation of *sco4847* knockout	TK143, TK144	1720
TK015	TK008: pIJ12057	Plasmid only controls	NR88, TK143	1456
TK016	TK006: pIJ12057		NR88, TK143	1456
TK017	TK005: pIJ12057		NR88, TK143	1456

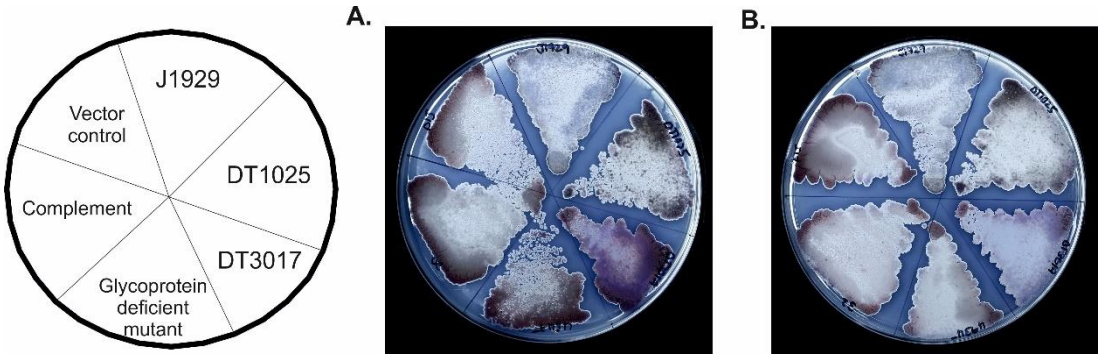


Figure 6.16 Antibiotic production of the complement strains TK013 (TK006:pTAK30) (A) and TK010 (TK008:pTAK32) (B) respectively. A. TK006 (*sco4847*), TK013 (TK006: pTAK30) and TK016 (TK006:pIJ10257). B. TK008 (*sco4934*), TK012 (TK008:pTAK32) and TK015 (TK008:pIJ10257). Approximately 10^6 spores were inoculated. Images are representative of at least two biological replicates and two technical replicates.

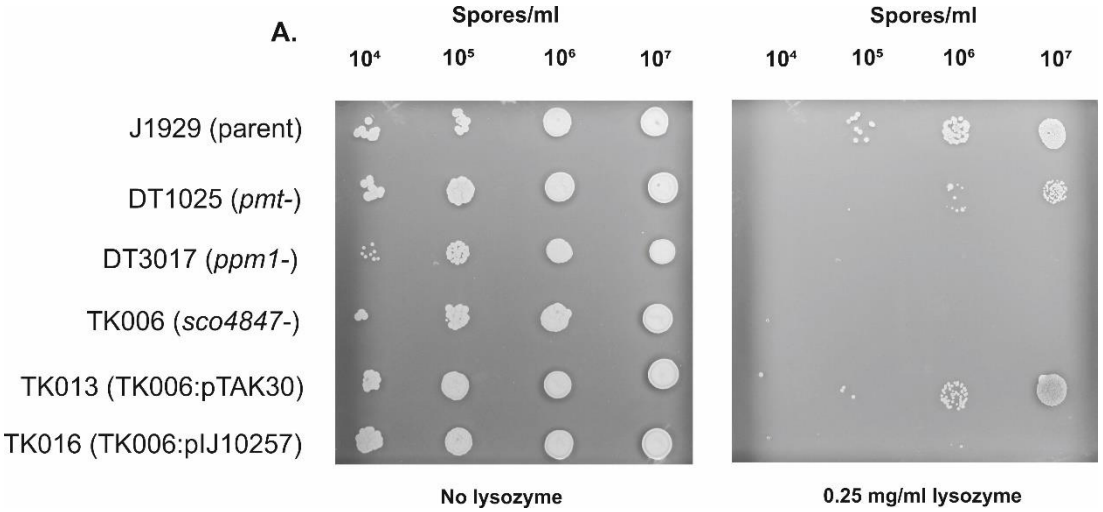


Figure 6.17 Lysozyme sensitivity of the complement strain TK013 (TK006:pTAK30). Spores were adjusted to 10^8 spores/mL and a ten-fold serial dilution was carried out to get 10^4 spores/mL. 5 µL of each spore stock was plated onto DNA without lysozyme (left-hand panel) and with 0.25 mg/mL of lysozyme (right-hand panel). Images are representative of two biological replicates and two technical replicates. These images are the same plates as in Figure 6.12.A.

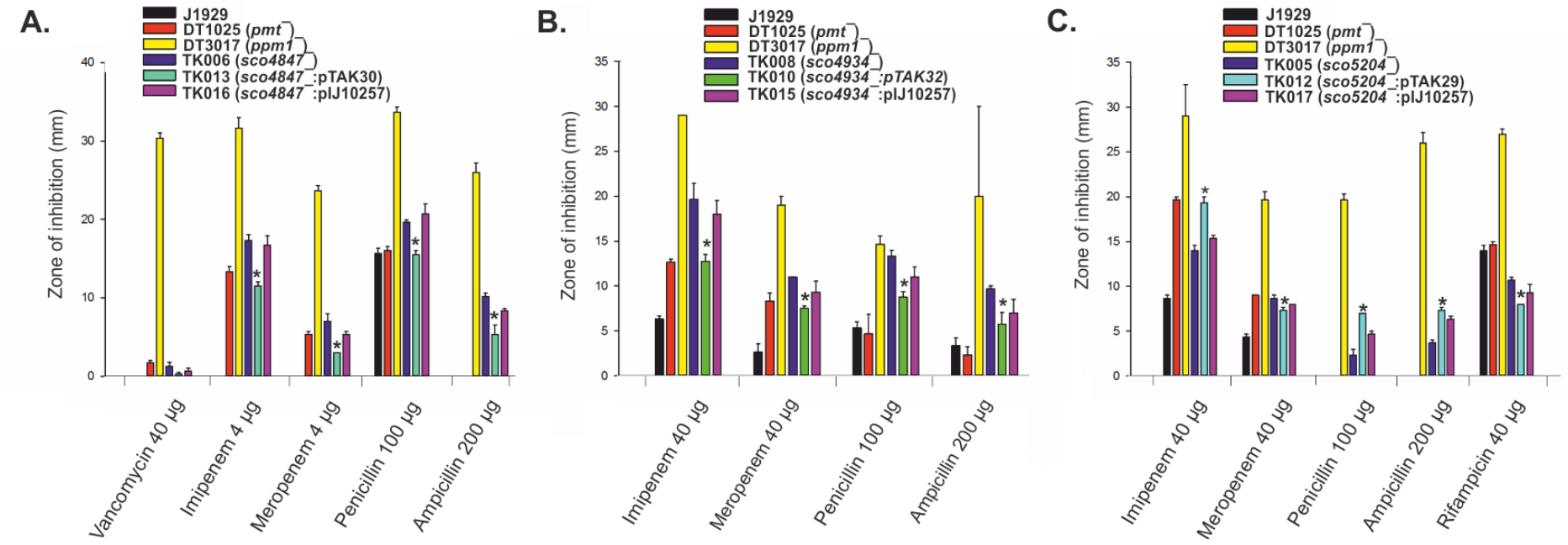


Figure 6.18 Antibiotic sensitivity of the complement strains TK013 (TK006:pTAK30) (A), TK010 (TK008:pTAK32) (B) and TK012 (TK005:pTAK29) (C). Shown are the diameters of growth inhibition zones from disc diffusion assays for each complement strain against the mutant, the parent strain J1929, and the glycosylation deficient strains DT1025 (*pmt*) and DT3017 (*ppm1*). Mutant strains with the integrated empty vector pIJ10257 served as negative controls for the complementations. Bars represent the mean of at least three biological replicates except for TK006, where the bars indicate the mean of two biological replicates tested three times each. Error bars indicate SEM. * indicates $p < 0.05$ that the observed difference between the glycoprotein deficient mutant and the complement strain has occurred by chance. The full data set is in Table A.5, A.7 and A.8. Data shown for J1929, DT1025, DT3017 and TK006 in panel A is the same as for Figure 6.11.B.

In a further attempt to rescue the β-lactam hypersensitivity phenotype in TK005 (*sco5204*) it was decided to complement the strains with *sco5204* under the control of its native promoter. *sco5204*, along with an extended upstream sequence (251 bp) to include potential regulatory elements, was amplified from J1929 genomic DNA by PCR and cloned into the HindIII site in the shuttle vector pMS82 (as described in Materials and Methods). pMS82 is an integrating vector similar to pIJ10257 without a promoter sequence (Gregory et al. 2003). The new construct, pTAK28 was confirmed by restriction digestion with PstI (expected fragments: 802 bp and 8793 bp) and BamHI (expected fragments: 1292 bp, 1904 bp and 6399 bp) (Figure 6.19.A) and sequenced to confirm the presence of the correct insert. The construct was introduced into *E. coli* ET12567(pUZ8002) and then into *S. coelicolor* TK005 (*sco5204*), as described in the Materials and Methods. The empty pMS82 vector was introduced into TK005 (*sco5204*) to serve as a negative control for the complementation.

Genomic DNA was isolated from the exconjugates and the presence of the integrated plasmids were confirmed by PCR (Figure 6.19 B-C). To confirm the presence of the empty pMS82 vector in the TK005 (*sco5204*), PCR primers were designed in the *S. coelicolor* genome upstream of *sco4848* and in the pMS82 vector upstream of the origin of replication (Figure 6.19.D). To confirm the presence of pTAK28 in TK005 (*sco5204*) PCR primers were designed on the vector backbone and in the *sco5204* gene (Figure 6.19.E). The PCR products observed from both of the TK005:pTAK28 and TK005:pMS82 exconjugates were slightly smaller than the expected 3022 bp (Figure 6.19.B) and 3261 bp (Figure 6.19.C) respectively. However, the PCR product amplified from the positive control pTAK28 was similarly smaller than expected suggesting that the TK005:pTAK28 exconjugates do contain the integrated pTAK28 vector (Figure 6.19.B). The new strains were named TK009 (TK005:pTAK28) and TK018 (TK005: pMS82) respectively.

No complementation of the β-lactam hypersensitivity phenotype observed previously in the TK005 (*sco5204*) (Figure 6.11.A) was observed in the TK009 (TK005:pTAK28) strains, however ampicillin sensitivity was increased (Figure 6.20). Additionally, no change in rifampicin sensitivity was observed in the TK009 (TK005:pTAK28) strains. These results suggest that the changes in antibiotic sensitivity in TK005 (*sco5204*) compared to the parent strain J1929 may not be a direct result of a loss of *sco5204*.

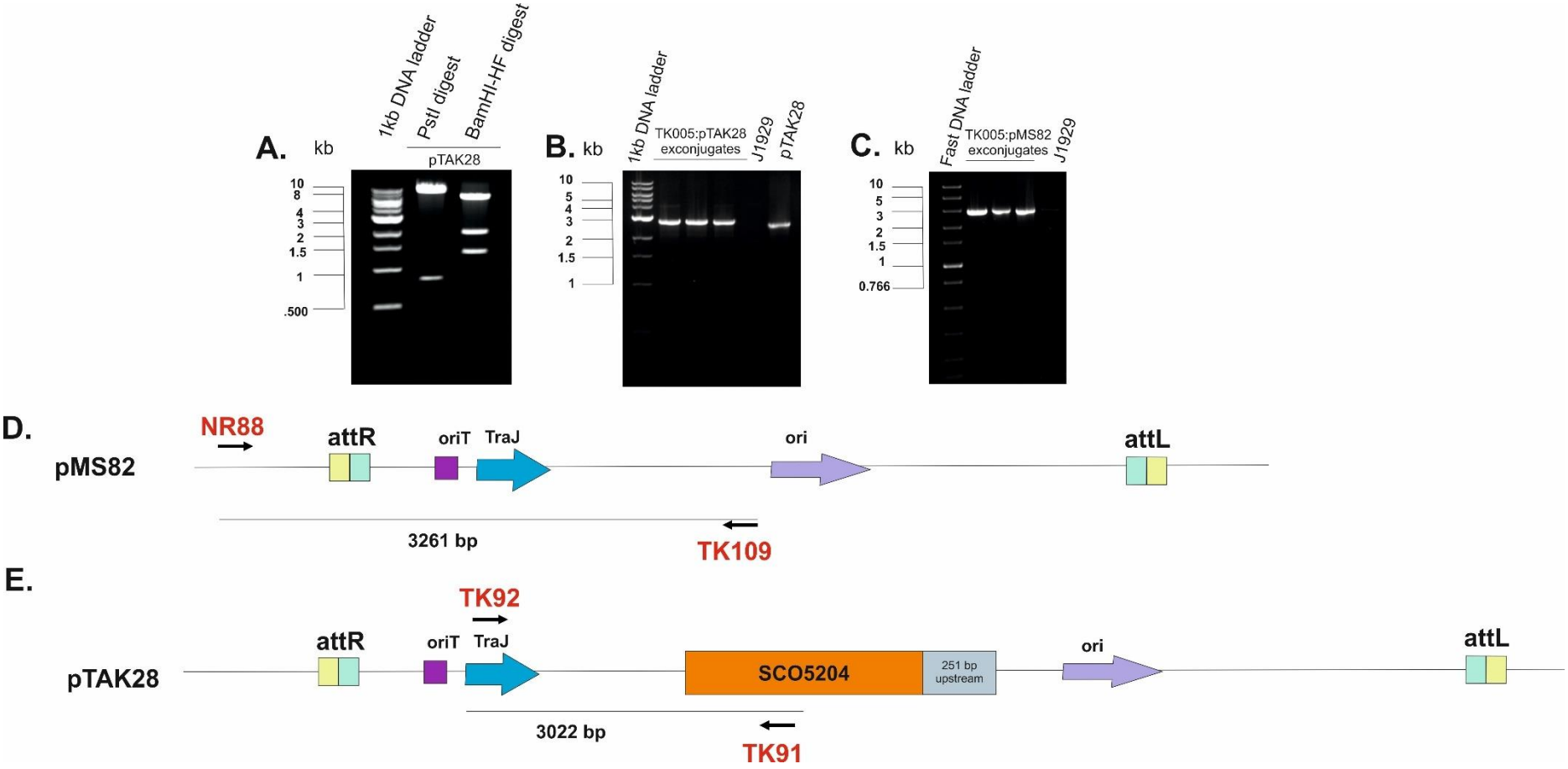


Figure 6.19 Restriction digest analysis of pTAK28 and confirmation of vector integration into TK005 (*sco5204*). A. Restriction digest analysis of pTAK28 with PstI and BamHI-HF. B. PCR confirming pTAK28 in TK005:pTAK28 exconjugates. C. PCR confirming pMS82 in TK005:pMS82 exconjugates. D-E. PCR primer design to confirm pMS82 (D) and pTAK28 (E) in the exconjugates. OriT: origin of transfer; TraJ: oriT recognising protein; Ori: origin of replication; ermEp*: ermE promoter.

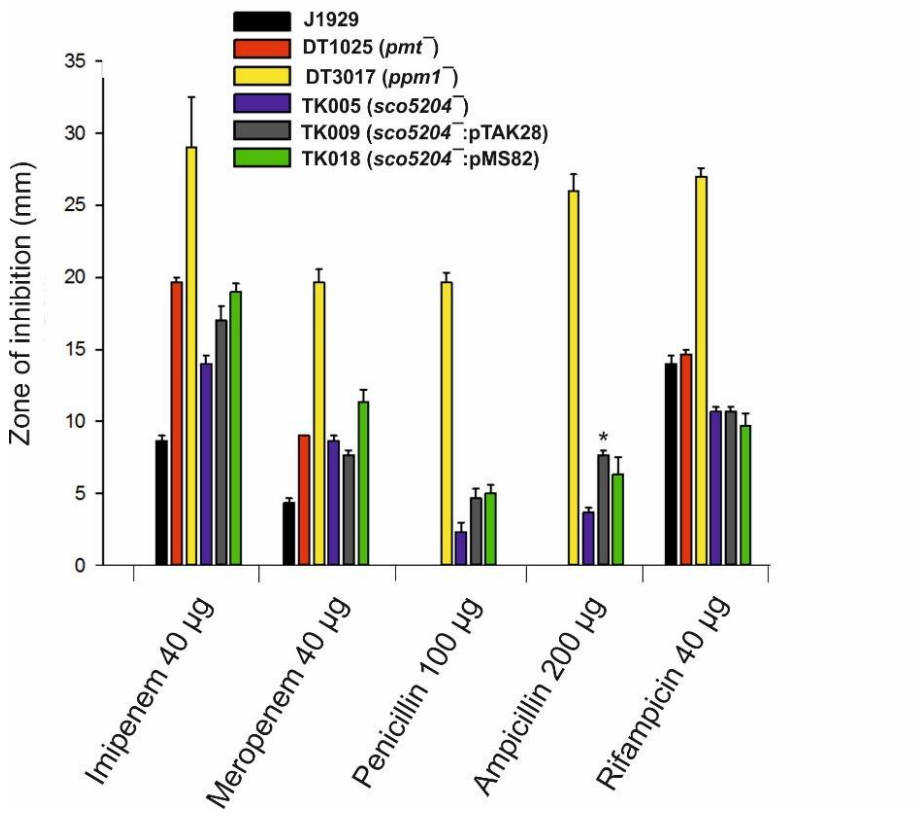


Figure 6.20 Antibiotic sensitivity of the complement strain TK009 (TK005:pTAK28). Shown are the diameters of growth inhibition zones from disc diffusion assays for the complement strain against the mutant, the parent strain J1929, and the glycosylation deficient strains DT1025 (*pmt*) and DT3017 (*ppm1*). The mutants strain with the integrated empty vector pIJ10257 served as negative control for the complementation. Bars represent the mean of at least three biological replicates and error bars indicate SEM. * indicates $p < 0.05$ that the observed difference between the glycoprotein deficient mutant and the complement strain has occurred by chance. The full data set is in Table A.8. Data for J1929, DT1025, DT3017 and TK005 is the same as shown in Figure 6.18.C.

6.3 Investigating the importance of the glycosylated amino acid in the SCO4934 glycopeptide TSQAEVDEAAAK

After confirming the complementation of the β-lactam hypersensitivity phenotype in the *sco4934* mutant (Figure 6.18) an investigation into whether changes in the glycosylated amino acid residue could affect the ability to complement the phenotype, was carried out.

The SCO4934 glycopeptide, N-40-TSQAEVDEAAAK-51-C identified in the glycoproteomics analysis of the *S. coelicolor* membrane proteome (Chapter 4, Table 4.1) was shown to be modified with up to three hexose residues. The glycosylated residue was not identified, however two possible glycosylated residues reside within the glycopeptide, T40 and S41. The positions of the conserved domains identified in SCO4934 that are associated with L, D transpeptidase function (Figure 6.2) (LDT_lgD_like_2: aa 149 – 247; YukD: aa 254 – 384) suggest that the glycosylated residue/s in the glycopeptide N-40-TSQAEVDEAAAK-51-C are not involved in the enzyme's catalytic activity. Therefore, we hypothesised that if the glycosylated residue/s in SCO4934 were mutated and the mutagenised genes introduced into TK008 (*sco4934*), a loss of the complementation of the β-lactam hypersensitivity could suggest that the glycosylation of that residue is required for the function of the protein.

sco4934 was amplified from pTAK32 and subcloned into the small vector (~ 3 Kb) pGEM7. A restriction digest of the new plasmid, pTAK47 with ScaI (expected fragments: 2477 bp, 1816 bp) and a double digest with BamHI and EcoRI (expected fragments: 1280 bp, 2963 bp) confirmed the presence of the insert (Figure 6.21). Using pTAK47 as a template, PCR based site directed mutagenesis (described in Materials and methods) was carried out to create amino acid substitutions of either T40, S41 or both T40 and S41 in SCO4934 to either alanine or valine respectively, as summarised in Figure 6.22. Alanine was selected due to its small and chemically inert functional group. Valine was selected since the substitution from threonine to valine via an exchange of a methyl group to a hydroxyl group, was the most conservative change in mass. The new constructs were sequenced (Figure A.22) to confirm the correct mutation and are summarised in Table 6.5.A. Mutated *sco4934* was amplified from pTAK48 – pTAK53 respectively and cloned into the NdeI site of pIJ10257. The constructs were confirmed by restriction enzyme analysis, and are summarised in Table 6.5.B. The constructs were introduced into *E. coli* ET12567(pUZ8002) and then into *S. coelicolor* TK008 (*sco4934*).

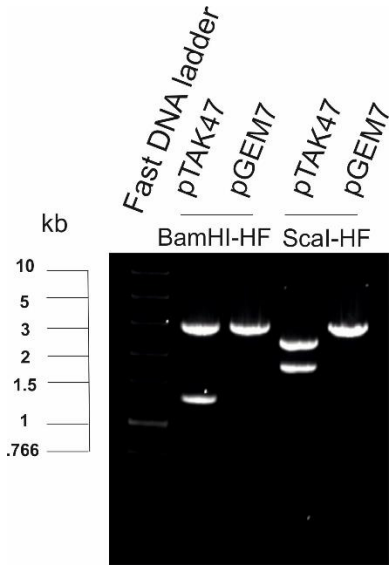


Figure 6.21 Restriction digest analysis of pTAK47 with BamHI-HF and ScaI-HF.

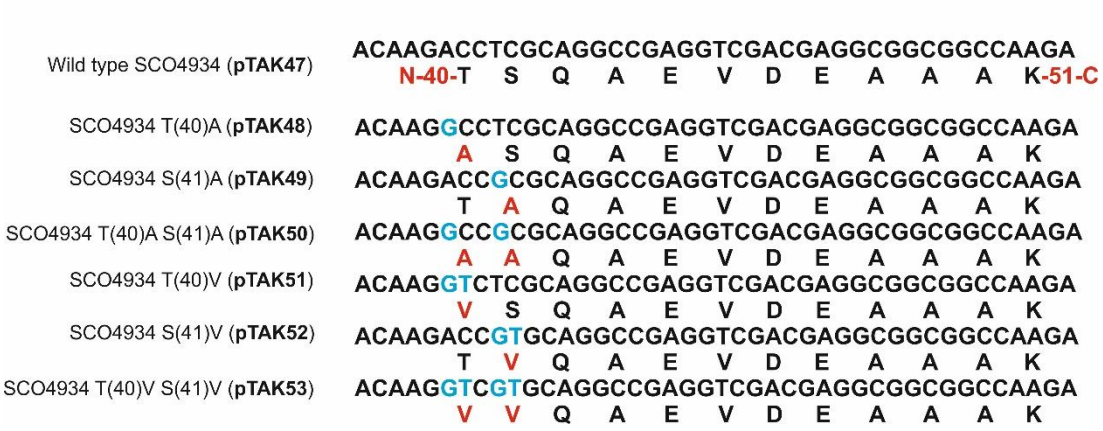


Figure 6.22 Site directed mutagenesis of T40 and S41 amino acid residues of SCO4934 positioned in the glycopeptide N-40-TSQAEVDEAAAK-51-C. A segment of the nucleotide sequence encoding *sco4934* (and mutants) is shown, indicating the translation of the glycopeptide N-40-TSQAEVDEAAAK-51-C. Amino acid substitutions are indicated in red. Nucleotide base changes are indicated in blue.

Table 6.5 New constructs generated after site-directed mutagenesis of *sco4934* (A) and cloning the mutagenised *sco4934* in pIJ10257 for conjugation into *S. coelicolor*.

A.

Construct	Insert	Vector	Mutation
pTAK47	*sco4934*	pGEM7	none (wild type)
pTAK48	*sco4934*	pGEM7	T(40)A
pTAK49	*sco4934*	pGEM7	S(41)A
pTAK50	*sco4934*	pGEM7	T(40)A, S(41)A
pTAK51	*sco4934*	pGEM7	T(40)V
pTAK52	*sco4934*	pGEM7	S(41)V
pTAK53	*sco4934*	pGEM7	T(40)V, S(41)V

B.

Construct	Insert	Vector	Mutation
pTAK48.b	*sco4934*	pIJ10257	T(40)A
pTAK49.b	*sco4934*	pIJ10257	S(41)A
pTAK50.b	*sco4934*	pIJ10257	T(40)A, S(41)A
pTAK51.b	*sco4934*	pIJ10257	T(40)V
pTAK52.b	*sco4934*	pIJ10257	S(41)V
pTAK53.b	*sco4934*	pIJ10257	T(40)V, S(41)V

Table 6.6 New *S. coelicolor* strains generated by complementation of TK008 (*sco4934*) with constructs with mutated *sco4934*.

S. coelicolor strain	Genotype	Description
TK020	TK008: pTAK48.b	*sco4934* T(40)A
TK021	TK008: pTAK49.b	*sco4934* S(41)A
TK022	TK008: pTAK50.b	*sco4934* T(40)A, S(41)A
TK023	TK008: pTAK51.b	*sco4934* T(40)V
TK024	TK008: pTAK52.b	*sco4934* S(41)V
TK025	TK008: pTAK53.b	*sco4934* T(40)V, S(41)V

Genomic DNA was isolated from the exconjugates and screened by PCR to confirm the presence of the integrated vector. The new *S. coelicolor* strains are summarised in Table 6.6. The sensitivity of the new strains to imipenem was assayed using disc diffusion assays (Figure 6.23). The complementation of the imipenem hypersensitivity phenotype observed in TK010 (*sco4934*:pTAK32) appeared to be lost, at least to some degree, in all of the strains complemented with the mutated *sco4934* variants, except in TK021 (*sco4934*:pTAK49). This was most obvious in the inhibition zones at 10 µg of imipenem (Figure 6.23.B). These results suggest that the mutation S(41)A does not affect the function of SCO4934, possibly indicating that S41 is not glycosylated. In contrast, the partial loss of complementation observed after the mutations of the T40 to both alanine and valine (TK020 and TK023 respectively), as well as after the mutation of both T40 and S41 to either valine or alanine (TK022 and TK025), could be consistent with the glycosylation of T40. The mutation S(41)V (TK024) however, appeared to result in a partial loss of complementation as well. A possible explanation for this could be that the valine in position 41 interferes with the glycosylation of the T40 residue, due to its slightly bulkier side chain. Taken together, these results could suggest that T40 is the most likely glycosylated residue in the glycopeptide N-40-TSQAEVDEAAAK-51-C. However, since the differences in the complementation phenotypes were extremely subtle, these results would require further validation.

6.4 Discussion

The discovery that mutants defective in protein O-glycosylation in *S. coelicolor* displayed increased sensitivity to cell-wall active antibiotics (discussed in Chapter 1) made by Howlett et al. (2016), led to the hypothesis that the *S. coelicolor* glycoproteome contained proteins required for cell wall biosynthesis. After the identification of several predicted cell-wall active glycoproteins in this work (Chapter 4), it was decided that the roles of the *S. coelicolor* glycoproteins SCO4847 (a putative D-Ala-D-Ala carboxypeptidase) and SCO4934 (a putative L, D transpeptidase) should be investigated. Mutants in both *sco4934* and *sco4847* were successfully generated by transposon insertion mutagenesis and validated. The mutants displayed no changes in colony morphology suggesting the loss of SCO4934 and SCO4847, had no significant impact on the growth the mutants on DNA. Similarly, no change in ϕC31cΔ25 phage sensitivity was observed in the glycoprotein deficient mutants compared to the parent strain, J1929.

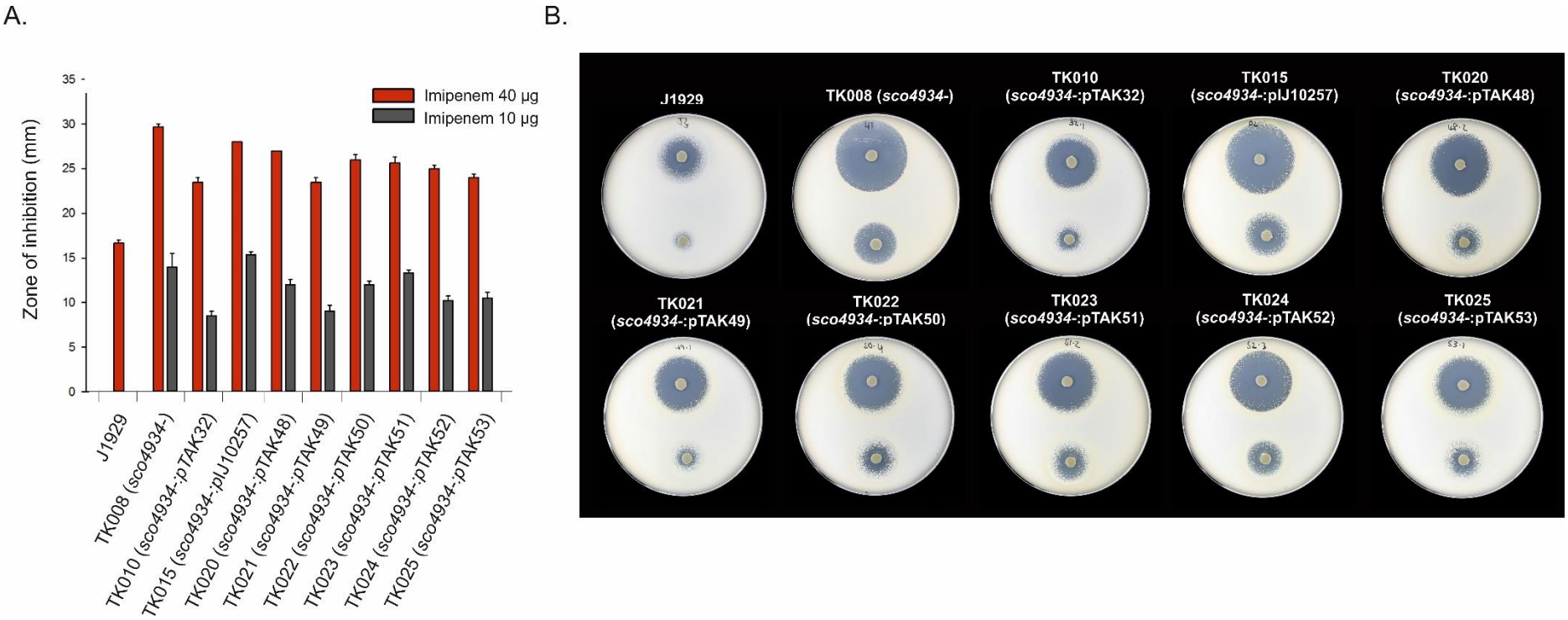


Figure 6.23 Imipenem sensitivity of the strains generated by the complementation of the *sco4934* strain with mutant variants of *sco4934*. A. Shown are the diameters of growth inhibition zones from disc diffusion assays for the complement strains against the *sco4934* mutant (TK008) and the parent strain J1929. The mutant strain with the integrated empty vector pIJ10257 served as negative control for the complementation. Bars represent the mean of at least three biological replicates and error bars indicate SEM. The full data set is in Table A.9. B. Plates showing imipenem sensitivity of the SCO4934 mutants complemented with mutant variants of *sco4934.* All replicates are shown in Figure A.1.

This suggests that neither of the glycoproteins SCO4934 and SCO4847 is the ϕC31cΔ25 phage receptor, which has been suggested to be a cell-wall exposed glycoprotein in *S. coelicolor* (Cowlishaw and Smith 2001). Additionally, the glycoprotein deficient mutants displayed no changes in sporulation showing that the glycoproteins are not required for sporulation in *S. coelicolor*.

The increase in susceptibility to β-lactam antibiotics, which target cell-wall biosynthesis, in the glycoprotein deficient mutants suggests that both proteins are required for cell wall biosynthesis. The antibiotic susceptibility phenotype was rescued partially upon the re-introduction of the wild type copies of *sco4934* and *sco4847*. The antibiotic susceptibility phenotype of both mutants was greater than that observed in the DT1025 (*pmt*) strain, suggesting the non-glycosylated SCO4934 and SCO4847 isoforms in DT1025 (*pmt*) have some activity. The loss of SCO4934 and SCO4847 respectively, in the glycoprotein deficient mutants could affect specific aspects of cell wall biosynthesis, such as the ability to form (3->3) crosslinks in peptidoglycan (*sco4934* mutant), which makes them more susceptible to antibiotics. The subsequent treatment of the mutants with antibiotics which non-specifically target proteins required for peptidoglycan crosslinking, could further the impact on cell wall biosynthesis, leading to an eventual loss of cell wall integrity and cell lysis.

The *sco4847* mutant displayed a significant increase in sensitivity to lysozyme, which is further evidence that this protein is required for maintaining the cell-wall in *S. coelicolor*. The phenotype was rescued after the introduction of the wild type copy of *sco4847*. *Streptomyces sp.* are known to produce large numbers of penicillin binding proteins, most likely as a resistance mechanism due their synthesis of β-lactam antibiotics (Ogawara 2015). In the StrepDB, the *S. coelicolor* genome alone has 10 genes that are annotated with D-Ala-D-Ala carboxypeptidase function (*sco0830, sco3408, sco3811, sco4439, sco4847, sco5660, sco6131, sco7050, sco7148* and *sco7607*). The serious phenotypes observed in the *sco4847* deficient mutant in this work might suggest that some D-Ala-D-Ala carboxypeptidases in *S. coelicolor*, such as SCO4847 have specific roles in cell wall biosynthesis or may be required during specific growth stages. The SCO4847 glycopeptides identified previously by mass spectrometry (Chapter 4) were isolated after 20 h and 35 h of growth. This could suggest that SCO4847 is required for actively growing cells. However, this is purely speculative and requires further validation.

The *sco4934* mutant displayed no changes in sensitivity to lysozyme, suggesting that a loss of this protein alone is unlikely to directly affect cell wall biosynthesis. This could indicate

that the majority of the peptidoglycan in *S. coelicolor* is (4->3) crosslinked. While (3->3) crosslinking in *S. coelicolor* has been identified, the ratio of (3->3) to (4->3) crosslinking has not been investigated (Hugonnet et al. 2014). Another explanation could be that other proteins compensate for the loss of SCO4934 in the *sco4934* mutant. A BLAST search of the SCO4934 protein sequence against the StrepDB revealed at least three other putative L, D transpeptidases in the *S. coelicolor* genome (SCO3194, SCO5458 and SCO5457).

Despite the increase in sensitivity to β-lactam antibiotics observed in the *sco5204* glycoprotein deficient mutant, the inability to complement the mutant phenotype by the introduction of the wild type copy of *sco5204*, suggests that the phenotype is not the direct result of a loss of SCO5204. The distance between the stop codon of *sco5204* and the start codon of the downstream gene *sco5205* is 392 bp, suggesting that these genes are not translationally coupled. The start codon of the gene immediately upstream of *sco5204*, *sco5203* is positioned 36 bp upstream of *sco5204* on the -1 strand. However, the position of the transposon insertion in the *sco5204* mutant was 296 bp into the *sco5204* gene and therefore unlikely to have disrupted any regulatory elements of *sco5203*. It is therefore unclear whether the knockout of *sco5204* resulted in the antibiotic sensitivity observed in the TK005 (*sco5204*-) strain.

The complete inability to complement the Tn5062 mutation of *sco5204* is an odd case. The complemented phenotypes of both *sco4847* and *sco4934* mutants were at best partial. In other work, the use of pIJ10257 has been problem free in complementation studies in *Streptomyces* (Hong et al. 2005; Howlett et al. 2016). Therefore, it is possible that there could be an unknown effect, such as gene context, that is common amongst the three genes, that has yet to be investigated.

Chapter 7 - General Discussion

Chapter 7 – General Discussion

7.1 The importance of protein O-glycosylation in *S. coelicolor*.

The disruption of PMT family proteins in several fungal species has demonstrated that protein O-mannosylation is essential and that it is required for maintaining cell wall structure and stability, septum formation, hyphal formation and conidiation (Gentzsch and Tanner 1996; Willer et al. 2005; Mouyna et al. 2010). Additionally, the inactivation of Pmts in *S. cerevisiae* using rhodanine-3-acetic acid derivatives resulted in the transcriptional activation of the cell wall integrity (CWI) pathway, a cellular response that includes the activation of several genes that have been implicated in cell wall assembly (Levin 2005; Arroyo et al. 2011).

In Actinobacteria, protein O-glycosylation is thought to be similarly important for maintaining cell wall integrity. In *M. smegmatis*, *pmt* mutants displayed a reduced tolerance to cell wall stress induced by SDS treatment, suggesting changes in the cell wall (Liu et al. 2013a). In *M. tuberculosis*, the culture filtrate glycoproteome is thought to consist of more than 40 glycoproteins, including a putative glycosyl hydrolase (Rv1096) and a beta-lactamase BlaC (Rv2068c) (Gonzalez-Zamorano et al. 2009; Smith et al. 2014). In *S. coelicolor*, mutations in *pmt* and *ppm1* confer pleiotropic phenotypes with increased sensitivity to antibiotics that target cell wall biosynthesis (Howlett et al. 2016). The antibiotic sensitivity phenotype in the *pmt* strains was less extreme than in the *ppm1* strains. However the *pmt* strains still displayed considerable sensitivity to β-lactam antibiotics and vancomycin. It was hypothesised that protein O-glycosylation in *S. coelicolor* was required for the correct functioning of periplasmic or membrane enzymes that are required for cell wall biosynthesis. The observation that the *S. coelicolor* glycoproteome (Chapter 4) contains several glycoproteins that are thought to be required for cell wall biosynthesis, supports this hypothesis. The phenotype of null mutants that ablate genes encoding two of these glycoproteins, a putative D-Ala-D-Ala carboxypeptidase (SCO4847) and an L, D transpeptidase (SCO4934) suggests that they are required for cell wall biosynthesis in *S. coelicolor* and that glycosylation could be affecting one or more of the enzymes' properties. Since the *sco4934* and *sco4847* mutants displayed increased antibiotic susceptibilities compared to the glycosylation deficient *pmt* strain, it is likely that the non-glycosylated isoforms of the proteins have some activity in the *pmt* mutant. As in *M. tuberculosis* (Gonzalez-Zamorano et al. 2009; Smith et al. 2014) *S. coelicolor* glycosylates a large number

of proteins with a wide range of biological functions, including solute binding, polysaccharide hydrolases, ABC transporters and cell wall biosynthesis.

The role of the glycan on each glycoprotein is still unknown but could conceivably affect protein activity, stability, folding or localisation. Protein glycosylation has been shown to have roles in maintaining protein solubility, as well as in aiding protein folding and protecting proteins from protease degradation (Olden et al. 1982; Harty et al. 2001). Protein *N*-glycosylation was recently shown to be required to protect *Campylobacter jejuni* surface proteins from protease degradation in the chicken gut (Alemka et al. 2013). Streptomycetes require large numbers of extracellular enzymes such as proteases and hydrolases to enable survival in the soil on complex, and often insoluble organic polymers. Morphological development in *Streptomyces* is thought to be regulated by trypsin-like proteases that are inhibited by protease inhibitors until late in the developmental growth cycle (Kim and Lee 1995; Kim et al. 2008). These secreted proteases are thought to be required for the digestion of the substrate mycelium to be used as a food source when nutrients are limited (Kang et al. 1995). It is possible that protein O-glycosylation in *S. coelicolor* is required for the protection of proteins against protease degradation. In the glycosylation deficient *pmt* strain, proteins could be exposed to protease degradation, potentially leading to reduced enzyme levels, aberrant stoichiometries and increased rates of protein turnover. This could led to depleted levels of some peptidoglycan biosynthetic enzymes or even a limitation in cell wall precursors, leading to a greater susceptibility to antibiotics that target the cell wall. This hypothesis could be tested by isolating one of the *S. coelicolor* glycoproteins identified in this study and its non-glycosylated isoform (from the *pmt* mutant) and evaluating the susceptibility of the proteins to protease degradation.

7.2 Intracellular glycosylation in *S. coelicolor*.

Since the early work carried out in protoplasts, it has generally been accepted that protein O-mannosylation in yeast is initiated on proteins as they are secreted into the ER (Larriba et al. 1976). This idea is supported by the fact that many fungal secreted and cell wall proteins have been found to be O-mannosylated (Strahl-Bolsinger et al. 1999; De Groot et al. 2005). Additionally, the activity of Pmt1 in *S. cerevisiae* (ScPmt1p) was shown to require two ER-lumen exposed loops (loop 1 and 5) (Figure 7.1 A), further suggesting that Pmt-mediated O-mannosylation occurs on secreted proteins in the ER lumen (Girrbach et al. 2000; Lommel et

al. 2011). Analogously in mycobacteria, protein O-mannosylation was shown to be coupled to protein translocation via the SEC pathway and secreted glycoproteins have been characterised in several actinomycetes, including *M. tuberculosis*, *Corynebacterium glutamicum* (*C. glutamicum*) and *S. coelicolor* (Dobos et al. 1995; Dobos et al. 1996; Michell et al. 2003; Mahne et al. 2006; Wehmeier et al. 2009; Smith et al. 2014; VanderVen et al. 2005) .

While 33 of the *S. coelicolor* glycoproteins identified in this study are consistent with the dogma that protein O-glycosylation in Actinobacteria is coupled to protein secretion (discussed in Chapter 4), five of the glycoproteins identified here had no predicted transmembrane domains or secretory signals, suggesting that they could be intracellular proteins (Chapter 4, Table 4.6). While it is probable that these intracellular proteins may be associated with the inner membrane, these findings could suggest that glycosylation may also possible on the cytoplasmic face of the membrane in *S. coelicolor*. Additionally, some glycopeptides belonging to membrane glycoproteins identified in this study (SCO2963, SCO4141, SCO4256, SCO4968 and SCO4548), are located in regions predicted to be on the cytoplasmic face of the membrane.

Membrane proteins are generally targeted to the SEC machinery as a ribosome nascent chain complex for integration into the membrane (Driessen and Nouwen 2008). They are recognised by a signal recognition particle (SRP) that binds directly to hydrophobic transmembrane segments of the nascent polypeptide chain and are targeted to the SecYEG protein conducting channel, in a cotranslational manner. If protein O-glycosylation in *S. coelicolor* is coupled to protein translocation via the SEC machinery, as has been described in mycobacteria (VanderVen et al. 2005), it is possible that during membrane protein translocation and insertion, loops destined for the cytoplasm come into contact with the glycosylation machinery before they are correctly localised on the cytoplasmic face of the membrane. Another possibility to consider is that the membrane protein topology could have been incorrectly predicted by the TMHMM algorithm. The authors of the TMHMM Server v 2.0 found that inverse topology predictions occurred approximately 7 % of the time in a test data set of 160 proteins and that this was one of the most common mistakes made by the program, when predicting the topology of a protein with a single transmembrane domain (Krogh et al. 2001). Three out of the five membrane proteins identified in this study, as having a glycopeptide located in regions predicted in the cytoplasm have only a single predicted transmembrane domain (SCO2963, SCO4256, SCO4968).

The O-glycosylation of intracellular proteins in *S. coelicolor* could suggest that protein O-glycosylation is possible on the cytoplasmic face of the membrane. The enzymatic activity of ScPmt1p was previously shown to require two arginine residues located within transmembrane domains 1 and 2 (Figure 7.1.A), an aspartic acid and glutamic acid residue (DE motif) in the ER facing loop 1, and a large ER facing loop 5 (Girrbach et al. 2000). The DE motif forms part of the enzyme's catalytic site and is crucial for acceptor peptide binding, suggesting that substrate binding occurs in the ER lumen (Lommel et al. 2011). Similarly, the DE motif was shown to be required for the activity of Pmt in *M. tuberculosis* (VanderVen et al. 2005). The Arg residues and the DE motif are conserved in Pmt from *M. tuberculosis* and *S. coelicolor* (Figure 7.1.D) and are located in similar topological positions to those in ScPmt1p (Figure 7.1.B-C). These findings suggest that similarly to ScPmt1p, loop 1 in Pmt from Actinomycetes is required for acceptor peptide binding and that this process occurs in the periplasm. It is therefore unclear how cytoplasmic proteins might be glycosylated by Pmt in *S. coelicolor*. One possibility is that another glycosyltransferase has a role in the glycosylation of proteins on the cytoplasmic face of the membrane in *S. coelicolor*. The putative integral membrane protein, SCO4023 was identified previously as a *pmt* homologue in *S. coelicolor* (Varghese 2008). A CDD database search found that the protein has a PMT_2 superfamily domain (e = 7.78e-30) that belongs to dolichol phosphate mannose-protein mannosyltransferases (CDD database). This protein has thirteen predicted transmembrane domains, suggesting that it might be a part of the GT-C glycosyltransferase family like Pmt. *sco4023* mutants in *S. coelicolor* were previously shown to display no changes in growth or phiC31 phage sensitivity compared to the wild type J1929 strain, suggesting that SCO4023 was not required for the glycosylation of the extracellular phiC31 phage receptor (Varghese 2008).

However, the presence of a *pmt* homologue in *S. coelicolor* would not explain the absence of Con A-HRP reactivity in the membrane and soluble protein fractions isolated from the *S. coelicolor pmt* (DT1025) mutants (Chapter 3, Figure 3.9). Another possibility is that a cytoplasmic facing domain in the *S. coelicolor* Pmt can also catalyse glycosyltransferase activity. The cytoplasmic N-terminal domain of *S. coelicolor* Pmt contains a DE motif and is extended by 55 amino acids in comparison to its *M. tuberculosis* homologue (Figure 7.1.D). This might suggest that it has a function that is unique to *S. coelicolor*.

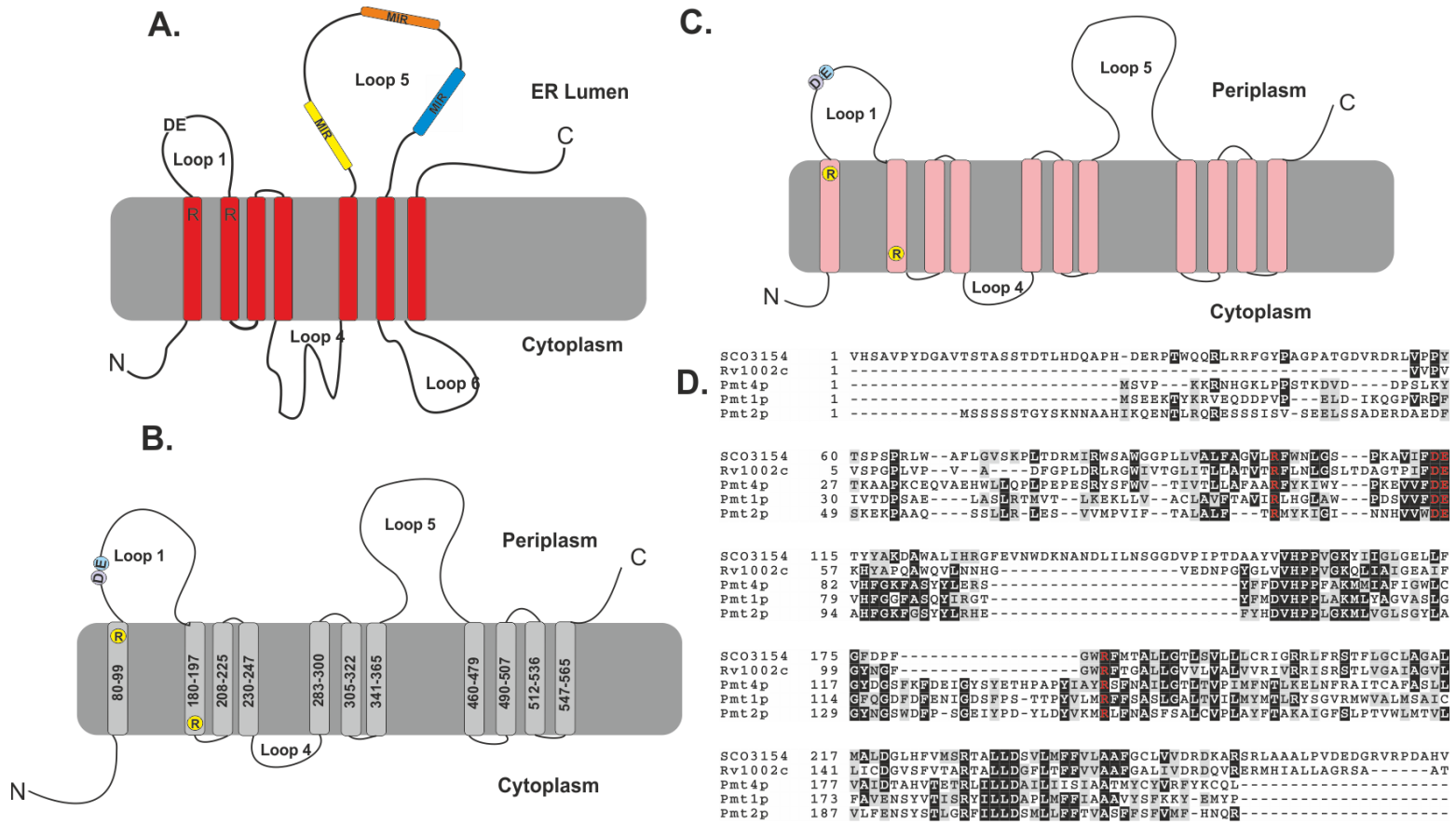


Figure 7.1 Models of Pmts from *S. cerevisiae* (A), *M. tuberculosis* (B) and *S. coelicolor* (C). A. ScPmt1p based on the work of Girrbach et al. (2000). B. A model of Mt-Pmt as described by VanderVen et al. (2005). C. A model of *S. coelicolor* Pmt (SCO3154). The membrane topology was predicted using the TMPred algorithm. Amino acid residues predicted in the transmembrane domains are indicated. The conserved Arg residues (R) and DE motif are indicated. D. Amino acid sequence alignment of Pmt1p, Pmt2p and Pmt4p from *S. cerevisiae*, as well as Pmts from *M. tuberculosis* (Rv1002c) and *S. coelicolor* (SCO3154). Conserved Arg residues and the DE motif are highlighted in red.

An in depth characterisation of the various domains of the *S. coelicolor* Pmt, similar to the work carried out by Girrbach et al. (2000) on ScPmt1p in yeast, could help to test this hypothesis.

7.3 Glycosylation of folded proteins in *S. coelicolor*

In addition to the O-mannosylation of unfolded eukaryotic proteins as they are translocated into the ER lumen, Pmt mediated glycosylation of misfolded proteins after they have been translocated into the ER has been demonstrated (Harty et al. 2001). These findings led to the proposal that surface exposed O-mannosylation acceptor sites within misfolded proteins are modified after prolonged ER residence, as a protein quality control mechanism. In mycobacteria, protein O-glycosylation was shown to be coupled to protein translocation via the SEC pathway, suggesting that protein O-glycosylation occurs on unfolded proteins (VanderVen et al. 2005). In contrast, a number of the *S. coelicolor* glycoproteins identified in this study (Chapter 4, Table 4.3) are predicted to contain TAT pathway signal peptides. The TAT protein transport system functions to secrete folded proteins across the cytoplasmic membrane and to insert some integral membrane proteins into the membrane (Berks et al. 2003). The pathway is well characterised in *S. coelicolor* and it is known to translocate large numbers of lipoproteins (Widdick et al. 2006; Thompson et al. 2010). SCO4934, a glycoprotein and predicted TAT substrate identified in this study, was experimentally verified as a TAT substrate by Thompson et al. (2010) after it was shown to be absent from *S. coelicolor ΔtatC* strains. The translocation of glycoproteins via the TAT pathway in *S. coelicolor* suggests that glycosylation is also possible on folded glycoproteins. Considering that Pmt mediated protein O-mannosylation of misfolded proteins occurs in yeast (Harty et al. 2001), it is plausible that the glycosylation of folded proteins in *S. coelicolor* is Pmt mediated. One could hypothesise that the glycosylation occurs on surface exposed regions of the protein or in flexible loops that link secondary structure elements.

7.4 NetOGlyc for the prediction of O-glycosylation sites in Actinobacteria.

While a consensus sequence for protein *N*-glycosylation has been defined in eukaryotes (N-X-S/T) and prokaryotes (D/E-X-N-X-S/T), little is known about the consensus sequence required for protein O-glycosylation. The NetOGlyc algorithm was developed for the

prediction of eukaryotic, mucin type O-glycosylation sites and is based on the amino acid sequence context (Hansen et al. 1998). The glycosylation of mucin-like proteins occurs on serine and threonine residues, in serine, threonine and proline rich sequences (Varki et al. 2009). Herrmann et al. (2000) found that the NetOGlyc algorithm correctly predicted 8 out of 11 O-glycosylation sites in mycobacterial glycoproteins. Additionally, they found that the amino acids flanking the glycosylation sites were important, and that they often contained alanine and proline.

While no definitive O-glycosylation consensus sequence in *S. coelicolor* was identified in this study, a higher propensity for alanine, proline and glycine was observed in sequences surrounding the O-glycosylation sites (Chapter 4, Figure 4.11). However, out of the eighteen O-glycosylation sites identified in thirteen of the *S. coelicolor* glycoproteins, only 27 % were correctly identified by the NetOGlc algorithm. A recent characterisation of the culture filtrate glycoproteome in *M. tuberculosis* led to the identification of thirteen glycoproteins, in which eight had validated glycosylation sites (Smith et al. 2014). Interestingly, around 76 % of the glycosylation sites were correctly predicted by the NetOGlyc algorithm. These findings suggest that while the NetOGlyc algorithm might perform well when predicting O-glycosylation in mycobacterial glycoproteins, it is not particularly accurate for the prediction of *S. coelicolor* O-glycosylation sites. This observation is supported by the fact that NetOGlyc predicted O-glycosylation sites in the putative *S. coelicolor* glycoprotein SCO4471 could not be experimentally verified by mass spectrometry (Chapter 3, Table 3.1). However, it is worth noting that many of these mycobacterial O-glycosylation sites (10 out of 17) were clustered along the sequence. Mucin type O-glycosylation sites were often found to be clustered in the NetOGlyc algorithm training dataset (Hansen et al. 1998). Therefore, it could be possible that if O-glycosylation sites in Actinobacterial glycoproteins are clustered in serine/threonine rich sequences, they are more likely to be identified by the NetOGlyc algorithm. Based on these findings, I propose that there is a real requirement for an algorithm that can predict protein O-glycosylation sites in Actinobacterial glycoproteins. This would require the further identification of novel Actinobacterial glycoproteins and the definitive characterisation of more O-glycosylation sites. The characterisation of the *S. coelicolor* glycoproteome presented in this study would undoubtedly make a significant contribution to this.

7.5 Glycan chain elongation in *S. coelicolor*.

In yeast, while protein O-mannosylation is initiated by Pmt in the ER, the glycans can undergo further extension in the Golgi apparatus (Loibl and Strahl 2013). In *S. cerevisiae,* the glycan is extended by α-1,2 mannosyltransferases of the KTR family to form α-1,2-linked mannobiose and mannotriose (Lussier et al. 1999). Then, an MNN1-family α-1,3 mannosyltransferase can further extend the glycan by the addition of α-1,3-linked mannose to form an oligosaccharide that generally contains five mannose residues. Similarly, in *M. tuberculosis* Pmt is thought to catalyse the first step of protein O-mannosylation, while extension is thought to be carried out by an additional glycosyltransferase, PimE (VanderVen et al. 2005; Liu et al. 2013a). In *M. tuberculosis*, the 45/47 kDa secreted antigen (Apa) was shown to be modified with α1,2 linked mannobiose and mannotriose glycans (Dobos et al. 1996). PimE, a putative PPM-dependent α1,2-mannosyltransferase was shown to be required for the elongation of the glycan moiety on the glycoprotein FasC in *M. smegmatis* (Liu et al. 2013a). *pimE* mutants in *M. smegmatis* could be complemented with *pimE* (Rv1159) from *M. tuberculosis*, suggesting that glycan chain elongation by PimE also occurs in *M. tuberculosis*.

I have shown that 2-linked, 4-linked and terminal mannose are components of the glycans that modify *S. coelicolor* glycoproteins (Chapter 5). Additionally, there was preliminary evidence of an O-linked trisaccharide, where the reducing residue was substituted via a 1,2 linkage, and a disaccharide with a 1,4 linkage. These findings suggest that the glycans on *S. coelicolor* glycoproteins could be different to those present in *M. tuberculosis,* based on the linkages present. In *M. bovis*, glycans modifying the glycoprotein MBP83 are composed of α1,3-linked mannobiose and mannose, suggesting that there is even variability in the glycan linkages between mycobacteria spp. It is possible that the glycans on *S. coelicolor* glycoproteins are extended sequentially, in a fashion that is similar to *M. tuberculosis*. A *pimE* homologue in *S. coelicolor* (SCO2335) has been identified. *S. coelicolor sco2335* mutants displayed a slight increase in sensitivity to vancomycin, ampicillin, rifampicin and imipenem when compared to the wildtype strain J1929. However no change in phiC31 phage sensitivity was observed (R. Howlett. Unpublished). These findings could suggest that a lack of extended glycans on *S. coelicolor* glycoproteins might not drastically affect glycoprotein function. Alternately, there may be other glycosyltransferases involved. According to the CAZy database, there are more than 50 proteins identified as glycosyltransferases in the *S. coelicolor* genome, many of which are genes that have not yet been characterised. It would be interesting to identify the glycosyltransferases responsible for the formation of 1,2 and

1,4 linkages in *S. coelicolor* glycoproteins, as well as to understand the factors that influence the formation of certain glycosidic linkages over others and essentially whether these contribute to the overall role of the glycan.

7.6 A revised model for protein O-glycosylation for in *S. coelicolor*

Based on the work reported in this thesis, I present a revised model for protein O-glycosylation in *S. coelicolor* (Figure 7.2). I have demonstrated that *S. coelicolor* has a glycoproteome consisting of glycoproteins with a variety of functions, including cell wall biosynthesis, transport and solute binding. Some of the glycoproteins identified here are predicted to be translocated via the TAT pathway, suggesting that protein O-glycosylation in *S. coelicolor* is possible on folded proteins. I have hypothesised that this process is Pmt mediated, although this remains to be experimentally validated. I have presented preliminary evidence that *S. coelicolor* glycoproteins are modified glycans that are composed of 2-linked, 4-linked and terminal mannose and I therefore propose that there may be other glycosyltransferases required for glycan extension. The *S. coelicolor* homologue (SCO2335) of *pimE* in *M. tuberculosis* could be involved in glycan chain extension. However, this remains to be experimentally validated. Additionally, I have shown that some putative intracellular proteins are glycosylated and I have hypothesised that this could be Pmt mediated. However, a full characterisation of *S. coelicolor* Pmt is required to understand the roles of the different functional domains of the protein and to further investigate its role, if any in the glycosylation of proteins on the cytoplasmic face of the membrane.

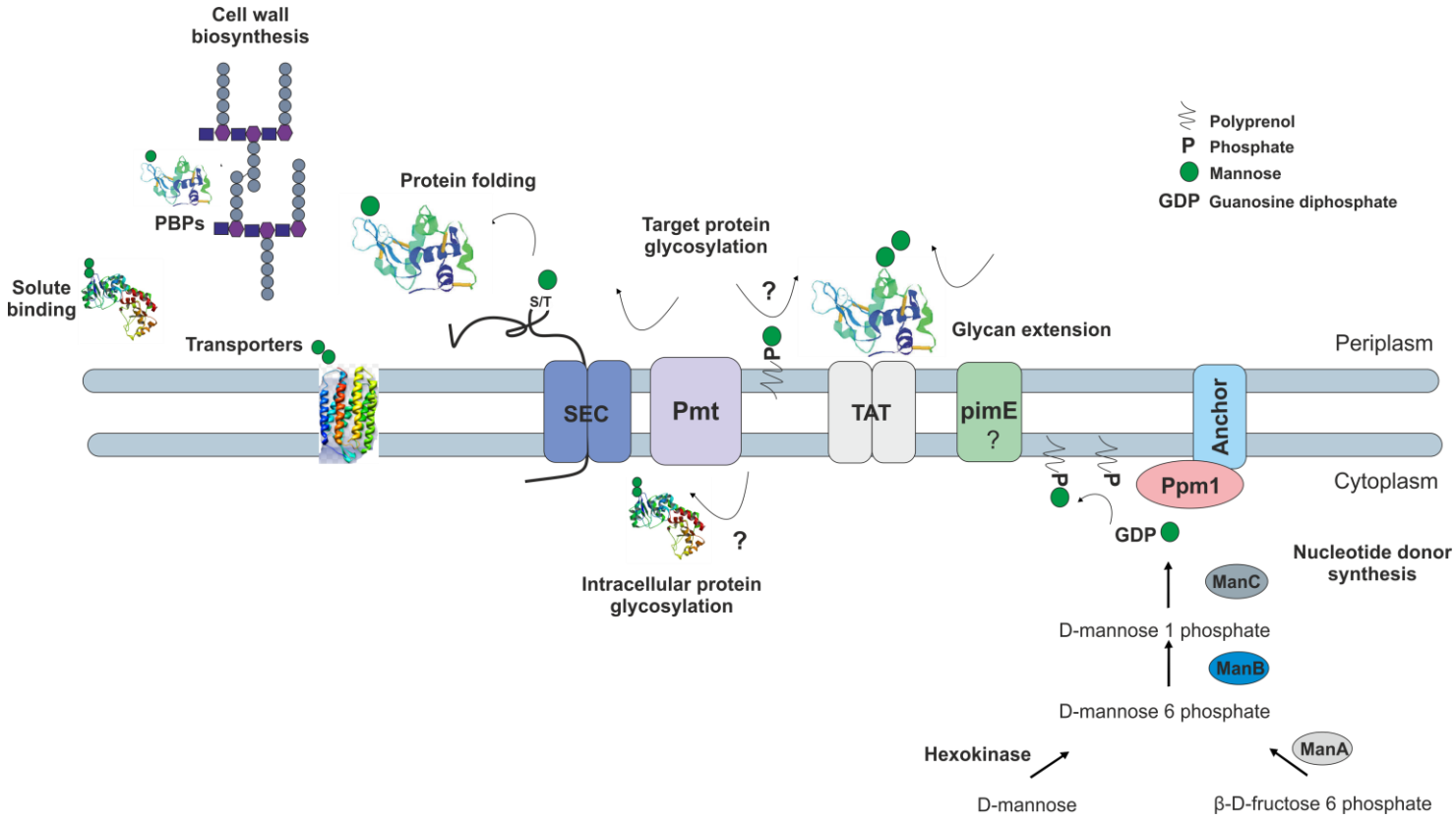


Figure 7.2 A revised model of protein O-glycosylation in *S. coelicolor*.

Appendix

Appendix

A.1 List of primers

Table A.1 Primers used for PCR and sequencing

Primer	Sequence	Description
TK2	GTGCGGGGAGGATCTGAC	forward primer pIJ10257
TK3	CCAACGTCATCTCGTTCTCC	reverse primer pIJ10257
F_pGEM7-test:	TGCAAGGCGATTAAGTTGGG	pGEM7 sequencing primer
R_pGEM7-seq	GCGTTGGATGCATAGCTTGA	pGEM7 sequencing primer
NR88	CGTGCGGCGATACTGAAA C	binds upstream of the phiBT1 attB site in *S. coelicolor*
TK33	GGGAACCGAGTAAACGATGA	Forward primer - for the generation of probe for southern blotting to confirm *sco5204* mutant
TK34	ACGAGTAGTGGACCGACCTG	Reverse primer - for the generation of probe for southern blotting to confirm *sco5204* mutant
TK82	CGGTCGTAGCGTTGAAAGTT	Forward primer - for the generation of probe for southern blotting to confirm *sco4934* mutant
TK83	GAGTTGTTGATGGAGGCGTT	Reverse primer - for the generation of probe for southern blotting to confirm *sco4934* mutant
TK85	CCTAGGATCCAAGCTTCC GGAGACGTACTCGTCGAT	Forward primer - cloning *sco5204* into pMS82
TK86	GCATAGATCTAAGCT TTCAGCCGGTGTCCGCTCC	Reverse primer - cloning *sco5204* into pMS82
TK87	CCGTCCCCAATACGAAGGAG	Forward primer - confirming *sco4847* mutants by PCR
TK88	CTCCCAGTGCGCGTTGT	Reverse primer - confirming *sco4847* mutants by PCR
TK91	CAAGGTCCTGGTGAGCTACG	Forward primer - confirm *S. coelicolor* genome integration of pTAK29
TK92	GCTACAAAATCACGGGCGTC	Reverse primer - confirm *S. coelicolor* genome integration of pTAK29
TK93	ACAGGAGGCCCCATC TGATGCCGGACCGCGGCGGA	Forward primer - cloning *sco5204* into pIJ10257
TK94	ACTCGAGATCTCATA TGTCAGCCGGTGTCCGCTCC	Reverse primer - cloning *sco5204* into pIJ10257
TK97	ACAGGAGGCCCCATAT GGTGCCCGCTCCCAAGAAG	Forward primer - cloning *sco4847* into pIJ10257
TK98	ACTCGAGATCTCATA TGGGCAGCAAGGCGCAGGAA	Reverse primer - cloning *sco4847* into pIJ10257
TK101	ACAGGAGGCCCCATATGA TGACGGACGGTAAGCGG	Forward primer - cloning *sco4934* into pIJ10257
TK102	ACTCGAGATCTCATATGT CAGACCGCCGAACCCGC	Reverse primer - cloning *sco4934* into pIJ10257
TK109	GCCGTTGATCGTGCTATGA	Forward primer pMS82

Primer	Sequence	Description
TK110	CGCCAGTGGTATTTATGTC	Reverse primer pMS82
TK116	GGGGGTGGGGACAAG GCCTCGCAGGCCGAGG	Forward primer - SDM of *sco4934* T(40)A
TK117	GGAGGTGCCGGCGTCGCT	Reverse primer - SDM of *sco4934* T(40)A
TK143	ACGTCCATGCGAGTGTCC	forward primer - to confirm integration of pTAK30 into the *S. coelicolor* genome
TK144	CCAAACGGCATTGAGCGTC	Reverse primer - to confirm integration of pTAK30 into the *S. coelicolor* genome
TK160	GGAATTCGGTACCCCAT GACGGACGGTAAGCGG	Forward primer - cloning *sco4934* into pGEM7
TK161	ATCGATTTCGAACCCTC AGACCGCCGAACCCGC	Reverse primer - cloning *sco4934* into pGEM7
TK203	CTCCGGGGGTGGGGAC AAGGCCGCGCAGGCCGAGGTC GAC	Forward primer - SDM of *sco4934* T(40)A, S(41)A
TK204	GTGCCGGCGTCGCTACTG	Reverse primer - SDM of *sco4934* T(40)A, S(41)A
TK205	CCTCCGGGGGTGGGGA CAAGGTCTCGCAGGCCGAGGTC G	Forward primer - SDM of *sco4934* T(40)V
TK206	TGCCGGCGTCGCTACTGGA	Reverse primer - SDM of *sco4934* T(40)V
TK207	CCGGGGGTGGGGA CAAGACCGTGCAGGCCGAGGTC G	Forward primer - SDM of *sco4934* S(41)V
TK208	AGGTGCCGGCGTCGCTACT	Reverse primer - SDM of *sco4934* S(41)V and S(41)A
TK209	CCTCCGGGGGTGGGGA CAAGGTCGTGCAGGCCGAGGTC GAC	Forward primer - SDM of *sco4934* T(40)V, S(41)V
TK210	TGCCGGCGTCGCTACTGG	Reverse primer - SDM of *sco4934* T(40)V, S(41)V
TK211	CCGGGGGTGGGGACAA GACCGCGCAGGCCGAGGTCG	Forward primer - SDM of *sco4934* S(41)A

A.2 Antibiotic susceptibility disc diffusion assays

Table A.2 Antibiotic susceptibility of *S. coelicolor* J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) on DNA assayed by disc diffusion. Raw data for three biological replicates including the zone of growth inhibition (mm), mean (AVERAGE), standard error of the mean (SEM) and p-value of a two-sample t-test assuming unequal variances between samples.

		J1929					DT1025					Vs J1929	DT3017					Vs J1929
		Zone of inhibition (mm)					Zone of inhibition (mm)						Zone of inhibition (mm)					
Antibiotic	conc (µg/disc)	1	2	3	AVERAGE	SEM	1	2	3	AVERAGE	SEM	P-value	1	2	3	AVERAGE	SEM	P-value
Ampicillin	200	0	0	0	0.0	0.00	10	9	12	10.3	0.88	0.0072	34	32	33	33.0	0.58	0.0003
Ampicillin	20	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	29	27	25	27.0	1.15	0.0018
Ampicillin	2	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	23	21	22	22.0	0.58	0.0007
Ampicillin	0.2	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	17	12	14	14.3	1.45	0.0101
Teicoplanin	40	24	25	25	24.7	0.33	25	26	25	25.3	0.33	0.2302	29	29	29	29.0	0.00	0.0059
Teicoplanin	4	22	23	25	23.3	0.88	23	24	23	23.3	0.33	1.0000	28	26	26	26.7	0.67	0.0432
Teicoplanin	0.4	20	22	21	21.0	0.58	21	21	20	20.7	0.33	0.6495	25	25	21	23.7	1.33	0.1730
Teicoplanin	0.04	17	18	18	17.7	0.33	17	18	17	17.3	0.33	0.5185	23	26	21	23.3	1.45	0.0535
Rifampicin	40	20	22	23	21.7	0.88	24	25	25	24.7	0.33	0.0621	37	36	35	36.0	0.58	0.0004
Rifampicin	4	15	18	15	16.0	1.00	20	20	19	19.7	0.33	0.0551	31	30	31	30.7	0.33	0.0022
Rifampicin	0.4	8	11	12	10.3	1.20	13	14	14	13.7	0.33	0.1000	27	25	27	26.3	0.67	0.0011
Rifampicin	0.04	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	22	21	23	22.0	0.58	0.0007
Bacitracin	40	17	18	19	18.0	0.58	19	18	18	18.3	0.33	0.6495	23	25	25	24.3	0.67	0.0022
Bacitracin	4	14	15	14	14.3	0.33	14	15	14	14.3	0.33	1.0000	21	20	20	20.3	0.33	0.0002
Bacitracin	0.4	11	10	11	10.7	0.33	10	10	12	10.7	0.67	1.0000	15	16	16	15.7	0.33	0.0004
Bacitracin	0.04	11	8	8	9.0	1.00	8	8	9	8.3	0.33	0.5813	12	12	12	12.0	0.00	0.0955
Imipenem	40	16	16	17	16.3	0.33	24	25	25	24.7	0.33	0.0001	42	40	42	41.3	0.67	0.0001
Imipenem	4	11	12	13	12.0	0.58	18	22	20	20.0	1.15	0.0090	36	34	36	35.3	0.67	0.0000
Imipenem	0.4	0	0	0	0.0	0.00	12	13	13	12.7	0.33	0.0007	28	26	26	26.7	0.67	0.0006
Imipenem	0.04	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	21	21	20	20.7	0.33	0.0003
Meropenem	40	12	13	11	12.0	0.58	17	17	17	17.0	0.00	0.0131	32	31	31	31.3	0.33	0.0001
Meropenem	4	0	0	0	0.0	0.00	11	10	10	10.3	0.33	0.0010	26	25	25	25.3	0.33	0.0002
Meropenem	0.4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	22	22	20	21.3	0.67	0.0010
Meropenem	0.04	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	14	16	15	15.0	0.58	0.0015
Penicillin	100	6	7	7	6.7	0.33	12	13	9	11.3	1.20	0.0516	30	30	29	29.7	0.33	0.0000
Penicillin	10	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	24	23	23	23.3	0.33	0.0002
Penicillin	1	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	17	14	16	15.7	0.88	0.0032
Penicillin	0.1	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Vancomycin	40	0	0	0	0.0	0.00	7	6	6	6.3	0.33	0.0028	32	31	31	31.3	0.33	0.0001
Vancomycin	4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	30	29	28	29.0	0.58	0.0004
Vancomycin	0.4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	27	26	26	26.3	0.33	0.0002
Vancomycin	0.04	0	0	0	0.0	0.00	0	0	0	0.00	0.00	-	25	24	24	24.3	0.33	0.0002

Table A.3 Antibiotic susceptibility of *S. coelicolor* J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) on F134 agar assayed by disc diffusion. Raw data for three biological replicates including the zone of growth inhibition (mm), mean (AVERAGE), standard error of the mean (SEM) and p-value of a two-sample t-test assuming unequal variances between samples.

		J1929					DT1025					Vs J1929	DT3017					Vs J1929
		Zone of inhibition (mm)					Zone of inhibition (mm)						Zone of inhibition (mm)					
Antibiotic	conc (µg/disc)	1	2	3	AVERAGE	SEM	1	2	3	AVERAGE	SEM	P-value	1	2	3	AVERAGE	SEM	P-value
Ampicillin	200	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Ampicillin	20	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Ampicillin	2	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Ampicillin	0.2	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Teicoplanin	40	40	40	40	40.0	0.00	35	37	37	36.3	0.67	0.0315	35	32	35	34.0	1.00	0.0267
Teicoplanin	4	36	34	35	35.0	0.58	35	32	32	33.0	1.00	0.1759	32	31	31	31.3	0.33	0.0100
Teicoplanin	0.4	33	31	32	32.0	0.58	30	30	27	29.0	1.00	0.0754	30	27	27	28.0	1.00	0.0366
Teicoplanin	0.04	30	28	30	29.3	0.67	27	25	25	25.7	0.67	0.0177	25	25	25	25.0	0.00	0.0229
Rifampicin	40	42	46	40	42.7	1.76	33	32	33	32.7	0.33	0.0263	33	31	32	32.0	0.58	0.0183
Rifampicin	4	38	40	37	38.3	0.88	32	28	28	29.3	1.33	0.0074	28	28	28	28.0	0.00	0.0072
Rifampicin	0.4	34	32	32	32.7	0.67	25	23	25	24.3	0.67	0.0009	23	23	21	22.3	0.67	0.0004
Rifampicin	0.04	28	28	29	28.3	0.33	20	18	18	18.7	0.67	0.0011	20	18	16	18.0	1.15	0.0081
Bacitracin	40	17	22	21	20.0	1.53	18	17	17	17.3	0.33	0.2193	18	17	17	17.3	0.33	0.2193
Bacitracin	4	12	15	16	14.3	1.20	15	13	13	13.7	0.67	0.6596	13	12	13	12.7	0.33	0.2980
Bacitracin	0.4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Bacitracin	0.04	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Imipenem	40	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Imipenem	4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Imipenem	0.4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Imipenem	0.04	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Meropenem	40	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Meropenem	4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Meropenem	0.4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Meropenem	0.04	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Penicillin	100	23	21	25	23.0	1.15	25	21	21	22.3	1.33	0.7250	30	29	29	29.3	0.33	0.0244
Penicillin	10	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	17	15	20	17.3	1.45	0.0070
Penicillin	1	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Penicillin	0.1	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Vancomycin	40	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	15	15	14	14.7	0.33	0.0005
Vancomycin	4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	10	13	11	11.3	0.88	0.0060
Vancomycin	0.4	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-
Vancomycin	0.04	0	0	0	0.0	0.00	0	0	0	0.0	0.00	-	0	0	0	0.0	0.00	-

Table A.4 Antibiotic susceptibility of *S. coelicolor* TK005 (*sco5204*) compared to the parent strain J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) on DNA assayed by disc diffusion. Raw data is for three biological replicates including the zone of growth inhibition (mm), mean (AVERAGE), standard error of the mean (SEM) and p-value of a two-sample t-test assuming unequal variances between samples.

		J1929					DT1025				
		Zone of inhibition (mm)			AVERAGE	SEM	Zone of inhibition (mm)			AVERAGE	SEM
Antibiotic	conc (µg/disc)	1	2	3			1	2	3		
Vancomycin	40	0	0	0	0.0	0.0	2	1	0	1.0	0.6
Rifampicin	40	25	23	26	24.7	0.9	24	29	27	26.7	1.5
Rifampicin	4	13	9	10	10.7	1.2	11	15	14	13.3	1.2
Bacitracin	40	15	14	15	14.7	0.3	15	11	12	12.7	1.2
Bacitracin	4	3	2	2	2.3	0.3	0	4	5	3.0	1.5
Ramoplanin	40	8	11	8	9.0	1.0	9	9	8	8.7	0.3
Ramoplanin	4	2	1	1	1.3	0.3	2	3	4	3.0	0.6
Teicoplanin	40	17	14	18	16.3	1.2	16	14	16	15.3	0.7
Teicoplanin	4	8	8	8	8.0	0.0	8	3	10	7.0	2.1
Imipenem	40	12	15	15	14.0	1.0	26	28	24	26.0	1.2
Imipenem	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0
Imipenem	0.4	0	0	0	0.0	0.0	0	0	0	0.0	0.0
Meropenem	40	16	18	21	18.3	1.5	27	28	30	28.3	0.9
Meropenem	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0
Penicillin	100	3	3	1	2.3	0.7	10	10	11	10.3	0.3
Ampicillin	200	4	5	6	5.0	0.6	15	14	15	14.7	0.3

		DT3017					TK005 (*sco5204-*)					Vs J1929
		Zone of inhibition (mm)			AVERAGE	SEM	Zone of inhibition (mm)			AVERAGE	SEM	
Antibiotic	conc (µg/disc)	1	2	3			1	2	3			P-value
Vancomycin	40	26	23	23	24.0	1.0	0	0	0	0.0	0.0	-
Rifampicin	40	30	30	29	29.7	0.3	22	22	20	21.3	0.7	0.04
Rifampicin	4	23	20	16	19.7	2.0	4	4	5	4.3	0.3	0.03
Bacitracin	40	20	20	17	19.0	1.0	4	12	7	7.7	2.3	0.09
Bacitracin	4	10	9	13	10.7	1.2	0	1	1	0.7	0.3	0.02
Ramoplanin	40	12	14	13	13.0	0.6	8	9	8	8.3	0.3	0.58
Ramoplanin	4	6	6	6	6.0	0.0	1	3	4	2.7	0.9	0.27
Teicoplanin	40	23	22	22	22.3	0.3	13	17	16	15.3	1.2	0.59
Teicoplanin	4	14	11	13	12.7	0.9	7	5	10	7.3	1.5	0.69
Imipenem	40	45	45	49	46.3	1.3	25	30	25	26.7	1.7	0.01
Imipenem	4	25	29	31	28.3	1.8	0	0	0	0.0	0.0	-
Imipenem	0.4	5	9	7	7.0	1.2	0	0	0	0.0	0.0	-
Meropenem	40	46	40	43	43.0	1.7	26	27	25	26.0	0.6	0.02
Meropenem	4	25	32	23	26.7	2.7	0	0	0	0.0	0.0	-
Penicillin	100	35	25	30	30.0	2.9	9	8	9	8.7	0.3	0.00
Ampicillin	200	15	20	25	20.0	2.9	7	9	9	8.3	0.7	0.02

Table A.5 Antibiotic susceptibility of *S. coelicolor* TK006 (*sco4847*), TK013 (*sco4847*: pTAK30) and TK016 (*sco4847*: pIJ10257), compared to the parent strain J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) on DNA assayed by disc diffusion. Raw data is for at least three biological replicates, except for TK006 where the mean of 2 biological replicates tested three times is shown. Includes the zone of growth inhibition (mm), mean (AVERAGE), standard error of the mean (SEM) and p-value of a two-sample t-test assuming unequal variances between samples.

		J1929					DT1025					DT3017				
		Zone of inhibition (mm)			AVERAGE	SEM	Zone of inhibition (mm)			AVERAGE	SEM	Zone of inhibition (mm)			AVERAGE	SEM
Antibiotic	conc (µg/disc)	1	2	3			1	2	3			1	2	3		
Vancomycin	40	0	0	0	0.0	0.0	1	2	2	1.7	0.3	31	31	29	30.3	0.7
Vancomycin	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	29	28	27	28.0	0.6
Vancomycin	0.4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	25	25	26	25.3	0.3
Vancomycin	0.04	0	0	0	0.0	0.0	0	0	0	0.0	0.0	23	23	23	23.0	0.0
Rifampicin	40	8	8	11	9.0	1.0	5	8	10	7.7	1.5	27	25	24	25.3	0.9
Rifampicin	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	7	6	5	6.0	0.6
Bacitracin	40	12	12	12	12.0	0.0	12	12	12	12.0	0.0	19	20	19	19.3	0.3
Bacitracin	4	0.5	2	2	1.5	0.5	1	0.5	2	1.2	0.4	6	6	5	5.7	0.3
Teicoplanin	40	18	18	20	18.7	0.7	20	18	20	19.3	0.7	22	24	23	23.0	0.6
Teicoplanin	4	10	10	11	10.3	0.3	9	9	11	9.7	0.7	13	15	15	14.3	0.7
Teicoplanin	0.4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	4	3	5	4.0	0.6
Imipenem	40	7	8	5	6.7	0.9	20	21	21	20.7	0.3	41	37	41	39.7	1.3
Imipenem	4	0	0	0	0.0	0.0	12	14	14	13.3	0.7	29	33	33	31.7	1.3
Imipenem	0.4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	25	27	27	26.3	0.7
Imipenem	0.04	0	0	0	0.0	0.0	0	0	0	0.0	0.0	19	23	15	19.0	2.3
Meropenem	40	5	7	5	5.7	0.7	12	11	12	11.7	0.3	30	33	29	30.7	1.2
Meropenem	4	0	0	0	0.0	0.0	5	6	5	5.3	0.3	25	23	23	23.7	0.7
Meropenem	0.4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	10	12	15	12.3	1.5
Penicillin	100	15	15	17	15.7	0.7	15	17	16	16.0	0.6	35	33	33	33.7	0.7
Penicillin	10	0	2	0	0.7	0.7	3	7	7	5.7	1.3	29	23	25	25.7	1.8
Penicillin	1	0	0	0	0.0	0.0	0	0	0	0.0	0.0	19	19	19	19.0	0.0
Penicillin	0.1	0	0	0	0.0	0.0	0	0	0	0.0	0.0	10	7	4	7.0	1.7
Ampicillin	200	0	0	0	0.0	0.0	0	0	0	0.0	0.0	24	26	28	26.0	1.2

Appendix

		TK006 (sco4847-)								vs J1929	TK013 (*sco4847-*: pTAK30)						vs TK006
		Zone of inhibition (mm)						AVERAGE	SEM	p-value	Zone of inhibition (mm)				AVERAGE	SEM	p-value
Antibiotic	conc (µg/disc)	1	2	3	4	5	6				1	2	3	4			
Vancomycin	40	0	0.5	1	2	2	2	1.3	0.5	0.02	0	0	1	0	0.3	0.3	0.05
Vancomycin	4	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Vancomycin	0.4	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Vancomycin	0.04	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Rifampicin	40	11	10	10	12	11	11	10.8	0.4	0.20	7	4	7	-	6.0	1.0	0.03
Rifampicin	4	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Bacitracin	40	13	12	13	12	10	13	12.2	0.7	0.74	12	12	12	-	12.0	0.0	0.74
Bacitracin	4	2	2	1	2	1	2	1.7	0.3	0.78	1	1	1	-	1.0	0.0	0.03
Teicoplanin	40	19	19	20	19	19	19	19.2	0.2	0.53	18	16	18	-	17.3	0.7	0.10
Teicoplanin	4	9	10	12	10	11	11	10.5	0.6	0.77	10	9	10	-	9.7	0.3	0.17
Teicoplanin	0.4	0	0	0	0	0	0	0.0	0.0	-	0	0	0	-	0.0	0.0	-
Imipenem	40	27	27	27	25	28	28	27.0	0.6	0.00	18	17	19	17	17.8	0.5	0.00
Imipenem	4	17	16	16	18	18	19	17.3	0.7	0.00	12	10	12	12	11.5	0.5	0.00
Imipenem	0.4	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Imipenem	0.04	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Meropenem	40	16	18	15	17	17	17	16.7	0.6	0.00	12	10	12	10	11.0	0.6	0.00
Meropenem	4	5	7	5	8	8	9	7.0	1.0	0.00	3	3	3	3	3.0	0.0	0.00
Meropenem	0.4	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Penicillin	100	20	20	20	19	19	20	19.7	0.3	0.02	15	15	17	15	15.5	0.5	0.00
Penicillin	10	8	10	6	10	9	12	9.2	1.2	0.00	3	0	8	8	4.8	2.0	0.11
Penicillin	1	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Penicillin	0.1	0	0	0	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-
Ampicillin	200	10	11	9	11	10	10	10.2	0.4	0.00	6	3	7	-	5.3	1.2	0.05

		TK016 (sco4847-:pIJ10257)					vs TK006
		Zone of inhibition (mm)			AVERAGE	SEM	p-value
Antibiotic	conc (µg/disc)	1	2	3			
Vancomycin	40	0	1	1	0.7	0.3	0.28
Vancomycin	4	0	0	0	0.0	0.0	-
Vancomycin	0.4	0	0	0	0.0	0.0	-
Vancomycin	0.04	0	0	0	0.0	0.0	-
Rifampicin	40	6	3	7	5.3	1.2	0.04
Rifampicin	4	0	0	0	0.0	0.0	-
Bacitracin	40	12	13	11	12.0	0.6	0.83
Bacitracin	4	2	1	0.5	1.2	0.4	0.38
Teicoplanin	40	18	18	18	18.0	0.0	0.00
Teicoplanin	4	11	9	11	10.3	0.7	0.84
Teicoplanin	0.4	0	0	0	0.0	0.0	-
Imipenem	40	27	28	30	28.3	0.9	0.27
Imipenem	4	19	15	16	16.7	1.2	0.65
Imipenem	0.4	0	0	0	0.0	0.0	-
Imipenem	0.04	0	0	0	0.0	0.0	-
Meropenem	40	16	15	16	15.7	0.3	0.11
Meropenem	4	5	6	5	5.3	0.3	0.07
Meropenem	0.4	0	0	0	0.0	0.0	-
Penicillin	100	22	18	22	20.7	1.3	0.53
Penicillin	10	8	9	9	8.7	0.3	0.60
Penicillin	1	0	0	0	0.0	0.0	-
Penicillin	0.1	0	0	0	0.0	0.0	-
Ampicillin	200	9	8	8	8.3	0.3	0.01

Table A.6 Antibiotic susceptibility of *S. coelicolor* TK008 (*sco4934*) compared to the parent strain J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) on DNA assayed by disc diffusion. Raw data is for at least three biological replicates including the zone of growth inhibition (mm), mean (AVERAGE), standard error of the mean (SEM) and p-value of a two-sample t-test assuming unequal variances between samples.

		J1929					DT1025				
		Zone of inhibition (mm)			Average	SEM	Zone of inhibition (mm)			Average	SEM
Antibiotic	conc (µg/disc)	1	2	3			1	2	3		
Vancomycin	40	1	1	2	1.3	0.3	2	3	2	2.3	0.3
Vancomycin	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0
Vancomycin	0.4	0	0	0	0.0	0.0	0	0	0	0.0	0.0
Vancomycin	0.04	0	0	0	0.0	0.0	0	0	0	0.0	0.0
Rifampicin	40	8	8	11	9.0	1.0	5	8	10	7.7	1.5
Rifampicin	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0
Bacitracin	40	10	10	10	10.0	0.0	13	12	10	11.7	0.9
Bacitracin	4	3	1	2	2.0	0.6	1	1	2	1.3	0.3
Teicoplanin	40	19	21	20	20.0	0.6	19	19	16	18.0	1.0
Teicoplanin	4	12	11	12	11.7	0.3	10	11	10	10.3	0.3
Teicoplanin	0.4	3	3	2	2.7	0.3	2	2	2	2.0	0.0
Imipenem	40	10	13	13	12.0	1.0	23	23	27	24.3	1.3
Imipenem	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0
Meropenem	40	8	9	10	9.0	0.6	13	13	12	12.7	0.3
Penicillin	100	5	6	9	6.7	1.2	9	8	9	8.7	0.3
Ampicillin	200	8	7	8	7.7	0.3	12	11	11	11.3	0.3
Ampicillin	20	0	0	0	0.0	0.0	0	0	0	0.0	0.0

		DT3017					TK008 (*sco4934-*)					vs J1929
		Zone of inhibition (mm)			Average	SEM	Zone of inhibition (mm)			Average	SEM	P-VALUE
Antibiotic	conc (µg/disc)	1	2	3			1	2	3			
Vancomycin	40	31	29	25	28.3	1.8	2	2	2	2.0	0.0	0.18
Vancomycin	4	23	23	20	22.0	1.0	0	0	0	0.0	0.0	-
Vancomycin	0.4	14	13	11	12.7	0.9	0	0	0	0.0	0.0	-
Vancomycin	0.04	1	1	3	1.7	0.7	0	0	0	0.0	0.0	-
Rifampicin	40	27	25	24	25.3	0.9	6	6	7	6.3	0.3	0.10
Rifampicin	4	7	6	5	6.0	0.6	0	0	0	0.0	0.0	-
Bacitracin	40	22	22	18	20.7	1.3	10	10	9	9.7	0.3	0.42
Bacitracin	4	10	10	7	9.0	1.0	2	1	1	1.3	0.3	0.39
Teicoplanin	40	24	25	22	23.7	0.9	20	21	19	20.0	0.6	1.00
Teicoplanin	4	15	15	13	14.3	0.7	12	11	12	11.7	0.3	1.00
Teicoplanin	0.4	6	5	4	5.0	0.6	3	3	3	3.0	0.0	0.42
Imipenem	40	37	37	35	36.3	0.7	26	29	29	28.0	1.0	0.00
Imipenem	4	10	8	12	10.0	1.2	4	7	6	5.7	0.9	0.02
Meropenem	40	29	29	27	28.3	0.7	21	21	20	20.7	0.3	0.00
Penicillin	100	31	29	17	25.7	4.4	13	13	12	12.7	0.3	0.03
Ampicillin	200	37	37	21	31.7	5.3	18	18	18	18.0	0.0	0.00
Ampicillin	20	16	19	0	11.7	5.9	0	0	0	0.0	0.0	-

Table A.7 β-lactam antibiotic susceptibility of *S. coelicolor* TK008 (*sco4934*), TK010 (*sco4934*: pTAK32) and TK015 (*sco4934*: pIJ10257), compared to the parent strain J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) on DNA assayed by disc diffusion. Raw data is for at least three biological replicates, including the zone of growth inhibition (mm), mean (AVERAGE), standard error of the mean (SEM) and p-value of a two-sample t-test assuming unequal variances between samples.

		J1929					DT1025					DT3017			
		Zone of inhibition (mm)					Zone of inhibition (mm)					Zone of inhibition (mm)			
Antibiotic	conc (µg/disc)	1	2	3	Mean	SEM	1	2	3	Mean	SEM	1	2	3	Mean
Imipenem	40	12	12	14	12.7	0.7	21	20	18	19.7	0.9	37	37	37	37.0
Imipenem	4	6	6	7	6.3	0.3	12	13	13	12.7	0.3	29	29	29	29.0
Imipenem	0.4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	18	16	18	17.3
Meropenem	40	8	8	9	8.3	0.3	15	14	15	14.7	0.3	29	27	26	27.3
Meropenem	4	3	1	4	2.7	0.9	7	8	10	8.3	0.9	20	17	20	19.0
Penicillin	100	6	4	6	5.3	0.7	3	2	9	4.7	2.2	13	16	15	14.7
Ampicillin	200	2	3	5	3.3	0.9	2	1	4	2.3	0.9	0	29	31	20.0
Ampicillin	20	0	0	0	0.0	0.0	0	0	0	0.0	0.0	0	19	23	14.0
Ampicillin	2	0	0	0	0.0	0.0	0	0	0	0.0	0.0	0	11	15	8.7

			TK008 (*sco4934*-)						TK010 (*sco4934* -: pTAK32)								TK015 (*sco4934* -: pIJ10257)					
			Zone of inhibition (mm)					vs J1929	Zone of inhibition (mm)						vs TK008	Zone of inhibition (mm)					vs TK008
Antibiotic	conc (µg/disc)	SEM	1	2	3	Mean	SEM	p-value	1	2	3	4	Mean	SEM	p-value	1	2	3	Mean	SEM	p-value
Imipenem	40	0.0	30	29	29	29.3	0.3	0.00	22	25	19	20	21.5	1.3	0.01	27	25	24	25.3	0.9	0.03
Imipenem	4	0.0	23	17	19	19.7	1.8	0.01	12	15	12	12	12.8	0.8	0.04	21	17	16	18.0	1.5	0.52
Imipenem	0.4	0.7	13	10	11	11.3	0.9	0.01	10	7	8	7	8.0	0.7	0.04	10	11	6	9.0	1.5	0.27
Meropenem	40	0.9	19	19	20	19.3	0.3	0.00	14	14	13	13	13.5	0.3	0.00	17	17	16	16.7	0.3	0.00
Meropenem	4	1.0	11	11	11	11.0	0.0	0.01	7	7	8	8	7.5	0.3	0.00	10	11	7	9.3	1.2	0.30
Penicillin	100	0.9	12	14	14	13.3	0.7	0.00	7	9	10	9	8.8	0.6	0.00	13	11	9	11.0	1.2	0.17
Ampicillin	200	10.0	10	9	10	9.7	0.3	0.01	3	4	8	7	5.5	1.2	0.04	10	5	6	7.0	1.5	0.22
Ampicillin	20	7.1	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-	0	0	0	0.0	0.0	-
Ampicillin	2	4.5	0	0	0	0.0	0.0	-	0	0	0	0	0.0	0.0	-	0	0	0	0.0	0.0	-

Table A.8 Antibiotic susceptibility of *S. coelicolor* TK005 *(sco5205*), TK009 (*sco5204*: pTAK28), TK018 (*sco5204*: pMS82), TK012 (*sco5204*: pTAK29) and TK017 (*sco5204*: pIJ10257) compared to the parent strain J1929, DT1025 (*pmt*) and DT3017 (*ppm1*) on DNA assayed by disc diffusion. Raw data is for three biological replicates, including the zone of growth inhibition (mm), mean (AVERAGE), standard error of the mean (SEM) and p-value of a two-sample t-test assuming unequal variances between samples.

		J1929					DT1025					DT3017				
		Zone of inhibition (mm)					Zone of inhibition (mm)					Zone of inhibition (mm)				
Antibiotic	conc (µg/disc)	1	2	3	Average	SEM	1	2	3	Average	SEM	1	2	3	Average	SEM
Imipenem	40	8	9	9	8.7	0.3	20	20	19	19.7	0.3	32	22	33	29.0	3.5
Imipenem	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	13	15	13	13.7	0.7
Meropenem	40	4	4	5	4.3	0.3	9	9	9	9.0	0.0	18	20	21	19.7	0.9
Penicillin	100	0	0	0	0.0	0.0	0	0	0	0.0	0.0	21	19	19	19.7	0.7
Ampicillin	200	0	0	0	0.0	0.0	0	0	0	0.0	0.0	24	26	28	26.0	1.2
Rifampicin	40	13	14	15	14.0	0.6	15	14	15	14.7	0.3	26	28	27	27.0	0.6
Rifampicin	4	0	0	0	0.0	0.0	0	0	0	0.0	0.0	10	10	8	9.3	0.7

		TK005 (*sco5204* -)						TK009 (*sco5204*- : pTAK28)						TK018 (*sco5204* -: pMS82)					
		Zone of inhibition (mm)					vs J1929	Zone of inhibition (mm)					vs TK005	Zone of inhibition (mm)					vs TK005
Antibiotic	conc (µg/disc)	1	2	3	Average	SEM	p-value	1	2	3	Average	SEM	p-value	1	2	3	Average	SEM	p-value
Imipenem	40	15	13	14	14.0	0.6	0.00	16	19	16	17.0	1.0	0.08	19	18	20	19.0	0.6	0.00
Imipenem	4	0	0	0	0.0	0.0	-	0	0	0	0.0	0.0	-	0	0	0	0.0	0.0	-
Meropenem	40	9	8	9	8.7	0.3	0.00	8	8	7	7.7	0.3	0.10	11	13	10	11.3	0.9	0.08
Penicillin	100	3	1	3	2.3	0.7	0.07	4	6	4	4.7	0.7	0.07	4	6	5	5.0	0.6	0.04
Ampicillin	200	4	3	4	3.7	0.3	0.01	8	8	7	7.7	0.3	0.00	4	8	7	6.3	1.2	0.15
Rifampicin	40	11	10	11	10.7	0.3	0.01	11	10	11	10.7	0.3	-	10	11	8	9.7	0.9	0.38
Rifampicin	4	0	0	0	0	0	-	0	0	0	0.0	0.0	-	0	0	0	0.0	0.0	-

		TK012 (*sco5204*- : pTAK29)						TK017 (*sco5204* -: pIJ10257)					
		Zone of inhibition (mm)					vs TK005	Zone of inhibition (mm)					vs TK005
Antibiotic	conc (µg/disc)	1	2	3	Average	SEM	p-value	1	2	3	Average	SEM	p-value
Imipenem	40	20	20	18	19.3	0.7	0.00	15	15	16	15.3	0.3	0.13
Imipenem	4	0	0	0	0.0	0.0	-	0	0	0	0.0	0.0	-
Meropenem	40	8	7	7	7.3	0.3	0.05	8	8	8	8.0	0.0	0.18
Penicillin	100	7	7	7	7.0	0.0	0.02	5	4	5	4.7	0.3	0.05
Ampicillin	200	8	7	7	7.3	0.3	0.01	6	7	6	6.3	0.3	0.02
Rifampicin	40	8	8	8	8.0	0.0	0.02	8	9	11	9.3	0.9	0.27
Rifampicin	4	0	0	0	0.0	0.0	-	0	0	0	0.0	0.0	-

Table A.9 Imipenem sensitivity of the strains generated by the complementation of the *sco4934* knockout strain with mutant variants of *sco4934* assayed by disc diffusion assays. Raw data is for at least three biological replicates, including the zone of growth inhibition (mm), mean (MEAN), standard error of the mean (SEM) and p-value of a two-sample t-test assuming unequal variances between samples.

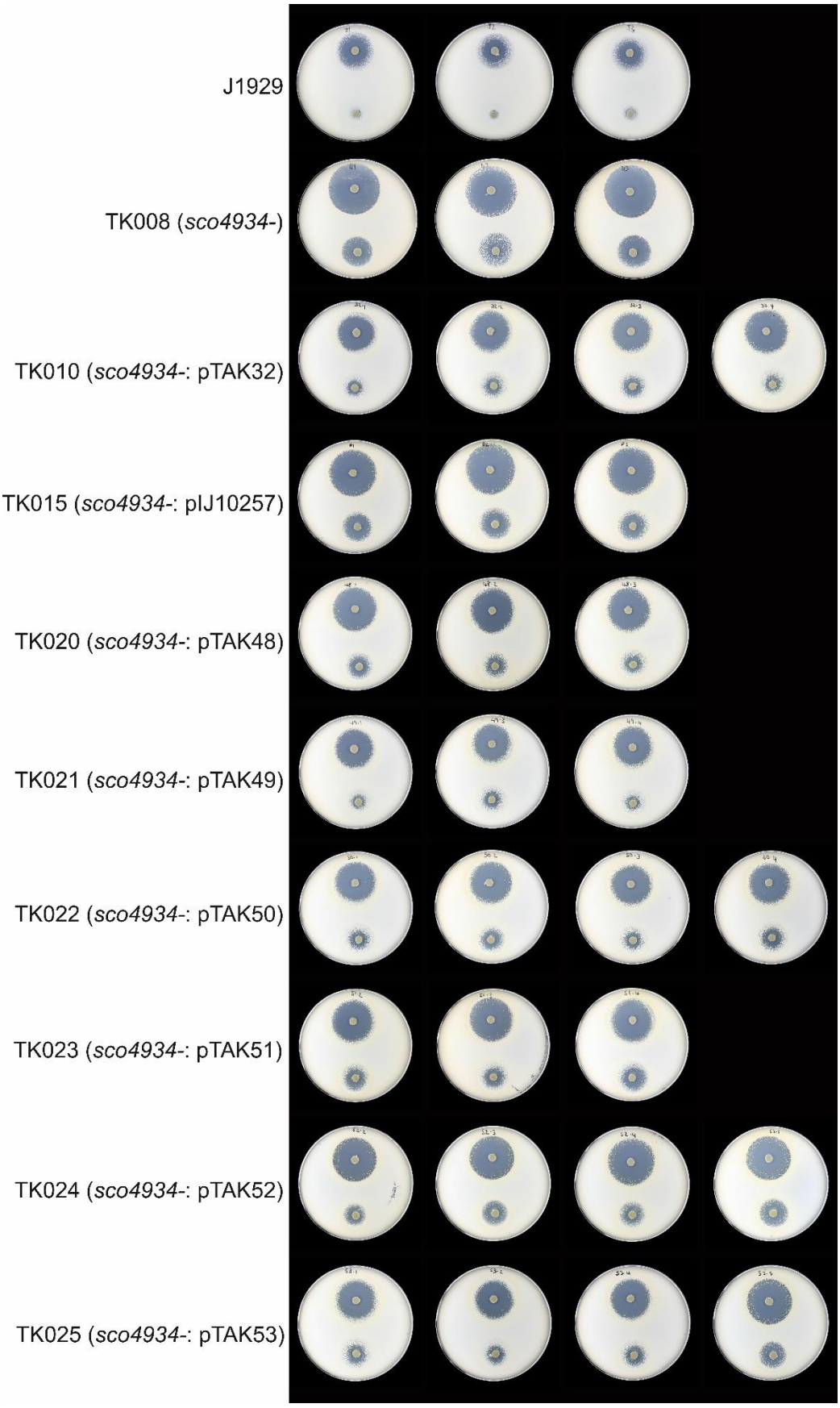
		J1929					TK008 (*sco4934-*)					vs J1929	TK010 (*sco4934-* :pTAK32)						vs TK008
Antibiotic	conc (µg/disc)	1	2	3	MEAN	SEM	1	2	3	MEAN	SEM	p-value	1	2	3	4	MEAN	SEM	p-value
Imipenem	40	12	11	12	11.67	0.33	25	25	24	24.67	0.33	0.000	18	18	18	20	18.50	0.58	0.000
Imipenem	4	0	0	0	0.00	0.00	11	10	6	9.00	1.53	0.028	3	3	5	3	3.50	0.58	0.057

		TK015 (*sco4934-* :pIJ10257)					vs TK008	TK020 (*sco4934-* :pTAK48)					vs TK010	TK021 (*sco4934-* :pTAK49)					
Antibiotic	conc (µg/disc)	1	2	3	MEAN	SEM	p-value	1	2	3	MEAN	SEM	p-value	1	2	3	4	MEAN	SEM
Imipenem	40	23	23	23	23.00	0.00	0.038	22	22	22	22.00	0.00	0.006	18	18	18	20	18.50	0.58
Imipenem	4	10	11	10	10.33	0.33	0.477	8	6	7	7.00	0.58	0.008	3	4	3	6	4.00	0.82

		vs TK010	TK022 (*sco4934-* :pTAK50)						vs TK010	TK023 (*sco4934-* :pTAK51)					vs TK010	TK024 (*sco4934-* :pTAK52)						vs TK010
Antibiotic	conc (µg/disc)	p-value	1	2	3	4	MEAN	SEM	p-value	1	2	3	MEAN	SEM	p-value	1	2	3	4	MEAN	SEM	p-value
Imipenem	40	1.000	22	20	20	22	21.00	0.67	0.017	22	20	20	20.67	0.67	0.060	19	20	20	21	19.67	0.33	0.061
Imipenem	4	0.587	8	7	6	7	7.00	0.47	0.002	8	8	9	8.33	0.33	0.001	6	4	6	5	5.33	0.67	0.045

		TK025 (*sco4934-* :pTAK53)						vs TK010
Antibiotic	conc (µg/disc)	1	2	3	4	MEAN	SEM	p-value
Imipenem	40	19	18	19	20	19.00	0.47	0.469
Imipenem	4	5	4	6	7	5.50	0.75	0.052

Figure A.1 Imipenem sensitivity of the strains generated by the complementation of the *sco4934* mutant with mutant variants of *sco4934*.



A.3 Identification of putative glycoproteins and deglycosylation analysis

Table A.10 Protein identification summary of the ~ 45 kDa band enriched from the J1929 culture filtrate (a) and membrane (b).

a)

	gi|21222858 Mass: 46976 Score: 441 Matches: 8(8) Sequences: 8(8)						
	hypothetical protein SCO4471 [Streptomyces coelicolor A3(2)]						
Peptide	**Observed *m/z***	**Expected MW**	**Calculated MW**	**ppm**	**Miss**	**Score**	**Expect**
R.YDASSPGGR.Q	909.3867	908.3794	908.3988	-21.3	0	43	5.70E-03
R.TPCLLGQDR.L	1073.533	1072.5257	1072.5335	-7.31	0	52	1.80E-03
R.LYDPPNNPR.G	1085.5369	1084.5296	1084.5301	-0.5	0	49	2.90E-03
K.APAYNYPGWR.E	1194.5698	1193.5625	1193.5618	0.65	0	67	3.80E-05
R.GASDIGYLTDEHIK.A	1518.73	1517.7227	1517.7362	-8.88	0	105	7.50E-09
F.SWDGAGEVGNGLFPR.F	1561.7313	1560.724	1560.7321	-5.17	0	111	1.80E-09
R.FLDLAQEHGAHMTFF.L	1763.7933	1762.7861	1762.8137	-15.7	0	104	7.10E-09
R.QFTDWLDVQKPEVLAK.L	1916.9784	1915.9711	1916.0044	-17.4	0	140	1.80E-12

b)

	gi|21222858 Mass: 47147 Score: 477 Matches: 6(6) Sequences: 4(4)						
	hypothetical protein SCO4471 [Streptomyces coelicolor A3(2)]						
Peptide	**Observed *m/z***	**Expected MW**	**Calculated MW**	**ppm**	**Miss**	**Score**	**Expect**
K.APAYNYPGWR.E	1194.5898	1193.5826	1193.5618	17.4	0	67	3.90E-06
K.APAYNYPGWR.E	1210.5829	1209.5756	1209.5567	15.6	0	-37	3.70E-03
R.TVEVGQRPAGGWNAFLK.E	1829.9786	1828.9714	1828.9584	7.08	0	95	6.80E-09
R.TVEVGQRPAGGWNAFLK.E	1845.9647	1844.9574	1844.9533	2.23	0	-46	5.80E-04
R.QFTDWLDVQKPEVLAK.L	1917.0106	1916.0033	1916.0044	-0.53	0	127	3.80E-12
R.TNSGWTDLPSLPFDYDKELVGGR.T	2567.2017	2566.1944	2566.234	-15.4	1	188	3.60E-18

Table A.11 MASCOT search results summary of trypsin and Asp-N digested SCO4471 analysed by LC-MS. Fixed modifications indicated in the MASCOT search: Carbamidomethyl (C). Variable modifications indicated in the MASCOT search: Oxidation (M), Ser->Diamino-propanoate (S), Thr->Diamino-butyrate (T).

Observed *m/z*	Expected MW	Calculated MW	ppm	Peptide e-value	Peptide sequence
632.3151	1262.6156	1262.6143	1.03	4.20E-02	K.APAKPIGDGSTSY.T
517.2666	1548.778	1548.7784	-0.26	2.30E-06	K.APAKPIGDGSTSYTGK.Q
775.3974	1548.7802	1548.7784	1.21	1.60E-06	K.APAKPIGDGSTSYTGK.Q
517.2704	1548.7893	1548.7784	7.07	2.00E-02	K.APAKPIGDGSTSYTGK.Q
775.4043	1548.794	1548.7784	10.1	4.60E-03	K.APAKPIGDGSTSYTGK.Q
517.2721	1548.7945	1548.7784	10.4	1.80E-06	K.APAKPIGDGSTSYTGK.Q
691.353	1380.6915	1380.6885	2.17	1.30E-05	P.AKPIGDGSTSYTGK.Q
691.3542	1380.6939	1380.6885	3.94	4.90E-03	P.AKPIGDGSTSYTGK.Q
461.2388	1380.6946	1380.6885	4.39	2.10E-03	P.AKPIGDGSTSYTGK.Q
461.2404	1380.6992	1380.6885	7.77	6.80E-03	P.AKPIGDGSTSYTGK.Q
591.7857	1181.5569	1181.5564	0.35	1.20E-02	K.PIGDGSTSYTGK.Q
969.4952	1936.9759	1936.9935	-9.07	5.90E-03	P.DAPVPLEPGQTPPQFVVF.S
969.5079	1937.0013	1936.9935	4.04	7.90E-05	P.DAPVPLEPGQTPPQFVVF.S
969.5109	1937.0072	1936.9935	7.06	1.10E-02	P.DAPVPLEPGQTPPQFVVF.S
686.8054	1371.5963	1371.5878	6.22	3.60E-02	L.DLAQEHGAHMTF.F + Oxidation (M)
752.3353	1502.6561	1502.6613	-3.43	5.40E-05	L.DLAQEHGAHMTFF.L
760.335	1518.6554	1518.6562	-0.53	3.90E-03	L.DLAQEHGAHMTFF.L + Oxidation (M)
760.3384	1518.6623	1518.6562	4.05	1.10E-02	L.DLAQEHGAHMTFF.L + Oxidation (M)
683.8837	1365.7528	1365.7544	-1.17	3.20E-03	F.FLSGLYLLPESK.K

Observed *m/z*	Expected MW	Calculated MW	ppm	Peptide e-value	Peptide sequence
683.8856	1365.7566	1365.7544	1.6	1.90E-02	F.FLSGLYLLPESK.K
610.3508	1218.6871	1218.686	0.94	1.80E-04	F.LSGLYLLPESK.K
610.3522	1218.6898	1218.686	3.15	7.40E-05	F.LSGLYLLPESK.K
610.3532	1218.6919	1218.686	4.85	1.20E-03	F.LSGLYLLPESK.K
553.8087	1105.6029	1105.6019	0.89	4.90E-04	L.SGLYLLPESK.K
553.8099	1105.6052	1105.6019	2.98	4.10E-05	L.SGLYLLPESK.K
510.2933	1018.572	1018.5699	2.09	3.90E-04	S.GLYLLPESK.K
510.2935	1018.5724	1018.5699	2.51	7.80E-03	S.GLYLLPESK.K
759.876	1517.7374	1517.7362	0.81	3.00E-04	R.GASDIGYLTDEHIK.A
759.8782	1517.7418	1517.7362	3.7	1.00E-03	R.GASDIGYLTDEHIK.A
506.9219	1517.7438	1517.7362	5.02	1.10E-02	R.GASDIGYLTDEHIK.A
695.8464	1389.6783	1389.6776	0.52	9.30E-03	A.SDIGYLTDEHIK.A
652.3275	1302.6405	1302.6456	-3.91	7.60E-03	S.DIGYLTDEHIK.A
652.3293	1302.6441	1302.6456	-1.1	1.70E-04	S.DIGYLTDEHIK.A
387.7295	773.4444	773.4395	6.24	2.10E-02	K.ATLTNVR.R
506.7271	1011.4396	1011.441	-1.4	7.90E-04	L.DGHEIGTHF.N
576.2925	1150.5704	1150.5618	7.44	2.30E-03	F.DYDKELVGGR.T
576.2928	1150.571	1150.5618	7.98	3.30E-02	F.DYDKELVGGR.T
455.2057	908.3969	908.3988	-2.11	7.60E-04	R.YDASSPGGR.Q
455.2071	908.3996	908.3988	0.85	8.80E-04	R.YDASSPGGR.Q
455.208	908.4015	908.3988	2.93	1.20E-03	R.YDASSPGGR.Q
455.2082	908.4018	908.3988	3.26	1.40E-03	R.YDASSPGGR.Q

Observed *m/z*	**Expected MW**	**Calculated MW**	**ppm**	**Peptide e-value**	**Peptide sequence**
455.2085	908.4024	908.3988	3.94	1.10E-02	R.YDASSPGGR.Q
455.2085	908.4024	908.3988	3.94	9.20E-05	R.YDASSPGGR.Q
455.2095	908.4044	908.3988	6.15	4.80E-05	R.YDASSPGGR.Q
455.2096	908.4047	908.3988	6.49	2.20E-03	R.YDASSPGGR.Q
455.2117	908.4088	908.3988	11	3.30E-03	R.YDASSPGGR.Q
455.2117	908.4089	908.3988	11.1	1.60E-02	R.YDASSPGGR.Q
681.3679	1360.7213	1360.7139	5.4	8.70E-04	W.DLPLQQIPFPGH.S
681.3709	1360.7271	1360.7139	9.7	1.30E-04	W.DLPLQQIPFPGH.S
798.4073	1594.8001	1594.8144	-8.93	1.50E-02	W.DLPLQQIPFPGHSF.E
798.4163	1594.818	1594.8144	2.24	3.10E-06	W.DLPLQQIPFPGHSF.E
798.4169	1594.8193	1594.8144	3.08	1.00E-03	W.DLPLQQIPFPGHSF.E
1086.0393	2170.0641	2170.0769	-5.91	5.20E-04	W.DLPLQQIPFPGHSFEVLSM.D + Oxidation (M)
1086.0513	2170.088	2170.0769	5.11	1.50E-03	W.DLPLQQIPFPGHSFEVLSM.D + Oxidation (M)
597.7886	1193.5627	1193.5618	0.8	2.70E-02	K.APAYNYPGWR.E
597.791	1193.5675	1193.5618	4.79	9.60E-04	K.APAYNYPGWR.E
513.7422	1025.4698	1025.4719	-2.02	7.90E-03	P.AYNYPGWR.E
393.2128	784.411	784.4119	-1.2	8.20E-03	K.AYISGFK.R
393.2131	784.4117	784.4119	-0.27	2.80E-03	K.AYISGFK.R
393.2155	784.4164	784.4119	5.73	2.40E-04	K.AYISGFK.R
393.2182	784.4218	784.4119	12.6	1.00E-04	K.AYISGFK.R
563.8278	1125.641	1125.6394	1.43	1.40E-02	L.DVQKPEVLAK.L
563.829	1125.6435	1125.6394	3.71	1.70E-04	L.DVQKPEVLAK.L

Observed *m/z*	**Expected MW**	**Calculated MW**	**ppm**	**Peptide e-value**	**Peptide sequence**
376.2223	1125.645	1125.6394	5.05	3.10E-02	L.DVQKPEVLAK.L
563.83	1125.6455	1125.6394	5.44	1.10E-03	L.DVQKPEVLAK.L
376.223	1125.6471	1125.6394	6.84	1.20E-02	L.DVQKPEVLAK.L

Table A.12 Bioinformatic predictions of candidate glycoproteins enriched by Con A affinity chromatography. Candidate glycoproteins were searched through the bioinformatics tools indicated in order to classify the proteins. Searches were carried out on the 12/07/2016.

SCO number	LipoP 1.0 [1]	LipoP 1.0 score	SignalP 4.1	SignalP 4.0 D-score [2]	TMHMM 2.0 [3]	Number of predicted transmembrane helices	PRED_LIPO [4]	TatP 1.0 [5]	TATP.1.0 D-score	Overall verdict
SCO5204	-	-	N	0.177	Y	7	membrane	N	0.282	membrane protein
SCO4856	-	-	N	0.175	N	-	cytoplasmic	N	0.206	cytoplasmic protein
SCO4471	SpII	21.718	Y	0.612	Y	1	lipoprotein	Y	0.610	lipoprotein
SCO4142	SpII	26.798	Y	0.661	N	-	lipoprotein	N	0.322	lipoprotein
SCO6009	SpII	18.197	Y	0.671	N	-	lipoprotein	N	0.217	lipoprotein
SCO1796	-	-	N	0.161	Y	1	membrane	N	0.053	membrane protein
SCO5776	SpI	21.851	Y	0.820	N	-	lipoprotein	N	0.173	lipoprotein

[1] LipoP 1.0 software produces predictions of lipoproteins (http://www.cbs.dtu.dk/services/LipoP/). SpI denotes SEC signal peptide; SpII denotes lipoprotein

[2] SignalP 4.1 software predicts the presence of a signal peptide (http://www.cbs.dtu.dk/services/SignalP/). D-score is a score used to discriminate signal peptides from non-signal peptides. Scores > 0.450 indicate a signal peptide.

[3] TMHMM 2.0 software predicts the presence of transmembrane helices in a protein sequence (http://www.cbs.dtu.dk/services/TMHMM/).

[4] PRED_LIPO enables the prediction of lipoproteins in Gram positive bacteria (http://www.compgen.org/tools/PRED-LIPO).

[5] TatP 1.0 predicts the presence of twin arginine (TAT) signal peptides. D-sore > 0.36 predicts the presence of a TAT pathway signal

A.4 Glycoproteomics

Table A.13 A full list of the high confidence glycopeptides identified by mass spectrometry using CID fragmentation.

SCO Number	Accession	Expect score[1]	Peptide sequence	# Hex[2]	Site allocation	Time point	Precursor ion *m/z*	Retention time (min)	Charge
SCO0996	gi|499337797	3.00E-05	ATAPSAEGFPVTIDNCGVK	3	-	20	1210.56	75.7	2
SCO4739	gi|499340123	1.10E-08	TEQSASAGGAEESAPAGK	3	-	20	1067.46	25.7	2
SCO4739	gi|499340123	4.10E-06	TEQSASAGGAEESAPAGK	4	-	20	1148.48	25.1	2
SCO4739	gi|499340123	2.50E-04	TEQSASAGGAEESAPAGK	6	-	20	874.023	24.3	3
SCO4739	gi|499340123	3.40E-04	TEQSASAGGAEESAPAGK	9	-	20	1036.08	23.6	3
SCO4739	gi|499340123	1.10E-03	TEQSASAGGAEESAPAGK	8	-	20	982.061	23.7	3
SCO4739	gi|499340123	1.80E-03	TEQSASAGGAEESAPAGK	8	-	20	982.06	23.8	3
SCO4739	gi|499340123	2.20E-03	TEQSASAGGAEESAPAGK	8	-	20	982.061	23.9	3
SCO4739	gi|499340123	2.50E-03	TEQSASAGGAEESAPAGK	5	-	20	1229.5	24.8	2
SCO4739	gi|499340123	1.20E-02	TEQSASAGGAEESAPAGK	7	-	20	928.04	24.1	3
SCO4739	gi|499340123	2.90E-02	TEQSASAGGAEESAPAGK	7	-	20	928.045	24.3	3
SCO4739	gi|499340123	3.40E-02	TEQSASAGGAEESAPAGK	8	-	20	982.063	23.8	3
SCO4905	gi|490071893	1.60E-08	ATPGLPAQVFLLCGSSLVAVDR	3	-	20	919.788	121.6	3
SCO4905	gi|490071893	1.90E-04	ATPGLPAQVFLLCGSSLVAVDR	2	-	20	865.783	122.5	3
SCO4934	gi|499340251	2.30E-02	TSQAEVDEAAAK	3	-	20	853.378	27.9	2
SCO5736	gi|490071116	5.00E-04	EGDTGSPEVQVALLSR	1	-	20	910.45	79.3	2
SCO0472	gi|21219012	2.10E-04	GGGSTPSATPAASVQDPLVATFDGGLYILDGK	9	-	35	1507.67	136.5	3
SCO0472	gi|21219012	1.00E-02	GGGSTPSATPAASVQDPLVATFDGGLYILDGK	9	-	35	1507.67	136.5	3
SCO1714	gi|499338228	1.90E-02	TVTEPAADR	3	-	35	723.318	25.9	2

SCO Number	Accession	Expect score[1]	Peptide sequence	# Hex[2]	Site allocation	Time point	Precursor ion *m/z*	Retention time (min)	Charge
SCO3540	gi|21221959	5.50E-05	ATPAELSPYYEQK	2	-	35	910.927	53.2	2
SCO4739	gi|499340123	4.50E-03	TEQSASAGGAEESAPAGK	8	-	35	982.063	23.3	3
SCO4739	gi|499340123	8.70E-03	TEQSASAGGAEESAPAGK	9	-	35	1036.08	23.1	3
SCO4739	gi|499340123	3.10E-02	TEQSASAGGAEESAPAGK	8	-	35	982.063	23.4	3
SCO4847	gi|21223223	8.30E-04	SATAASPSAEASGEAGGTGK	9	-	20	1055.76	59.5	3
SCO4847	gi|21223223	3.00E-04	SATAASPSAEASGEAGGTGK	9	-	35	1055.76	59.5	3
SCO4934	gi|499340251	1.90E-04	TSQAEVDEAAAK	2	-	35	772.341	27.4	2
SCO4934	gi|499340251	2.40E-04	TSQAEVDEAAAK	3	-	35	853.368	27	2
SCO3357	gi|21221786	1.30E-02	DEGPAHADAVGGAGSASPAPAAK	6	-	43	992.761	35.9	3
SCO3540	gi|21221959	1.90E-02	ATPAELSPYYEQK	2	T2	43	910.921	56.4	2
SCO4934	gi|499340251	3.60E-03	TSQAEVDEAAAK	3	-	43	853.373	30.1	2
SCO5204	gi|490071596	1.20E-02	QVQSQFNSEQDIAESIR	1	-	43	1071	73.2	2
SCO6558	gi|499341298	1.20E-02	IPDITLER	1	T5	43	559.797	87.9	2
SCO6558	gi|499341298	4.60E-02	IPDITLER	1	T5	43	559.798	85.6	2
SCO4141	gi|499339752	4.00E-02	TPQPPATEDTRPGR	1	-	60	562.28	31.2	3
SCO4739	gi|499340123	4.10E-03	TEQSASAGGAEESAPAGK	8	-	60	982.063	23.8	3
SCO4905	gi|490071893	9.10E-03	ATEVPTDYGPAPSR	3	-	60	973.935	46.7	2
SCO4934	gi|499340251	1.90E-04	TSQAEVDEAAAK	2	-	60	772.35	27.7	2
SCO4934	gi|499340251	2.40E-02	TSQAEVDEAAAK	3	-	60	853.375	27.4	2
SCO5115	gi|1532204	8.90E-03	AVDGLSFDLER	1	S6	60	692.336	103.6	2
SCO5818	gi|499340768	4.10E-02	SPHAARLAALVTK	3	S1, T12	60	910.982	128.8	2

[1] Expect value assigned to the glycopeptide match in the MASCOT search

[2] Number of hexose residues on the glycopeptide.

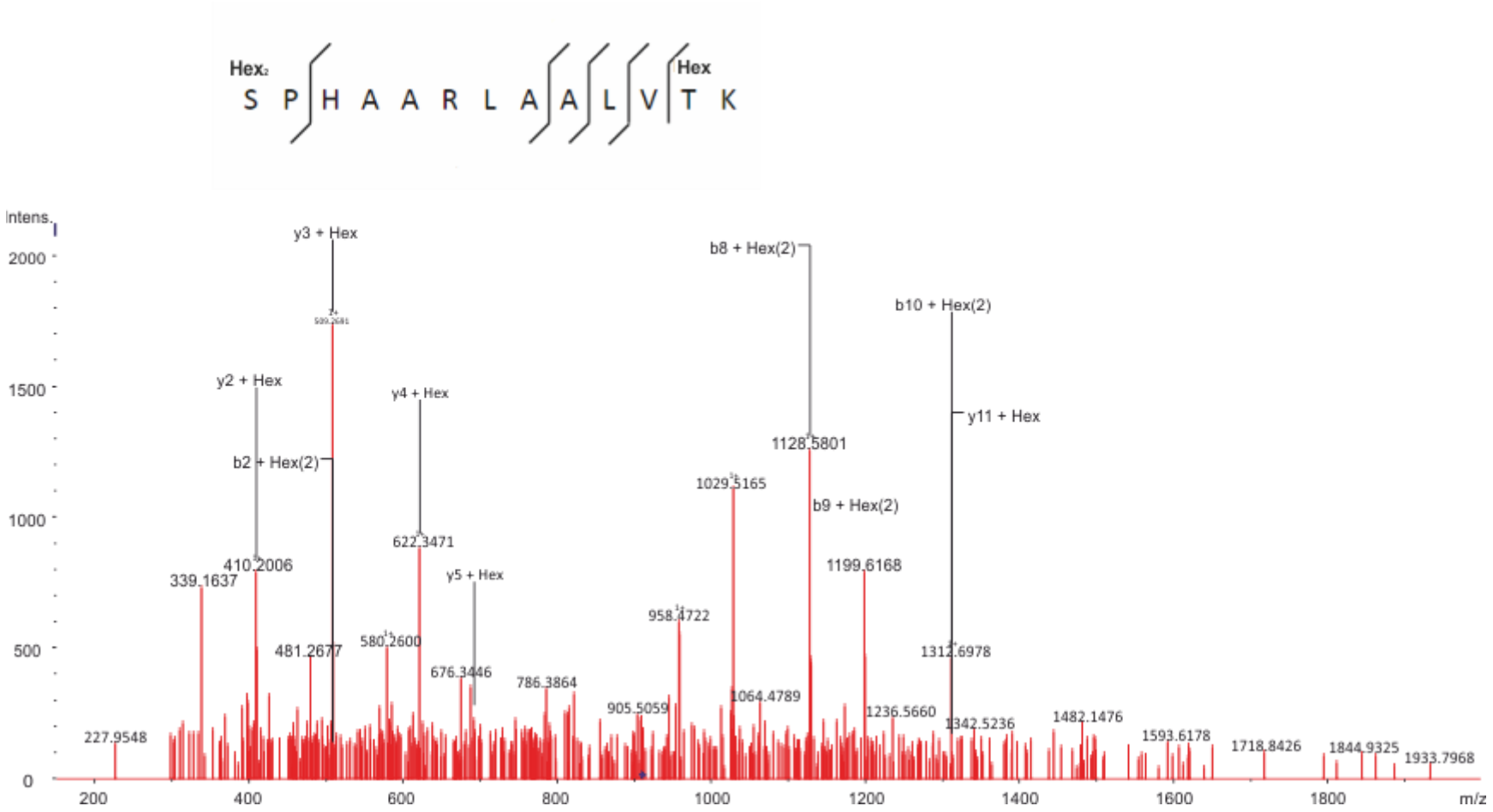


Figure A.2 CID spectrum of the SCO5815 tryptic peptide SPHAARLAALVTK (aa 228 – 240) indicates Hex_2 on Ser228 and Hex on Thr239. Precursor *m/z* 910.981 charge = 2+; retention time = 128.8 min; e-value = 4.10E-02. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. The precursor ion is indicated by a ◆ . (++) Indicates 2+ ion series. (0) Indicates ions that have lost H_2O. (*) Indicates ions that have lost NH_3

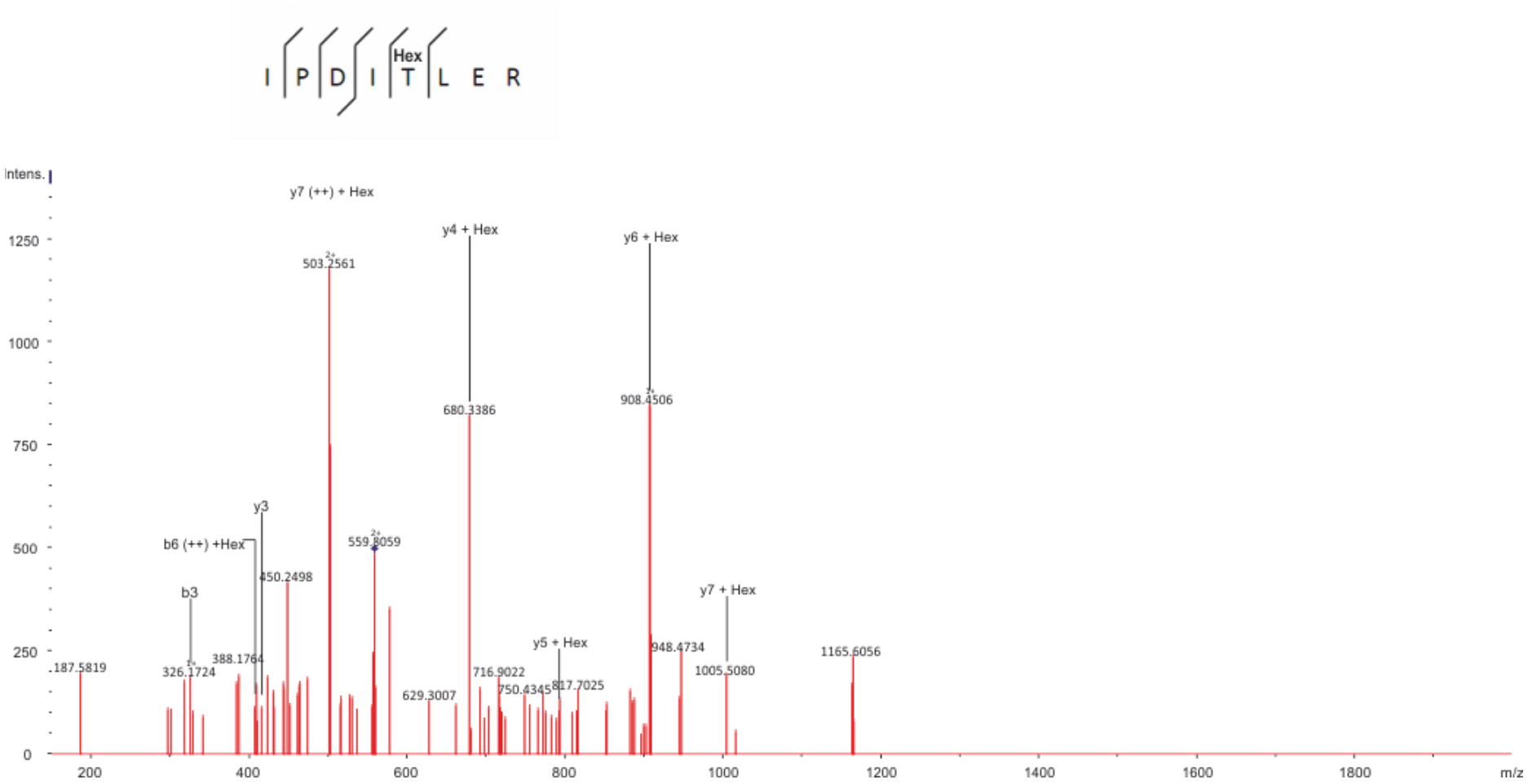


Figure A.3 CID spectrum of the SCO6558 tryptic peptide IPDITLER (aa 100 – 107) indicates monohexosylation on Thr104. Precursor *m/z* 559.797; charge = 2+; retention time = 87.9 min; e-value = 1.20E-02. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. The precursor ion is indicated by a ◆ (++) Indicates 2+ ion series. (0) Indicates ions that have lost H_2O. (*) Indicates ions that have lost NH_3.

Table A.14 A full list of the high confidence glycopeptides identified by mass spectrometry in the HCD_IT, ETD_IT and ETD_OT acquisitions.

SCO Number	Accession[1]	Expect score[2]	Peptide sequence	# Hex[3]	Site allocation	MD-score[4]	Method	Precursor ion *m/z*	Charge	Scan
SCO0996	gi|21219513	8.20E-04	ATAPSAEGFPVTIDNCGVK	2	-	-	HCD_IT	753.3517	3	46383
SCO0996	gi|21219513	2.10E-02	ATAPSAEGFPVTIDNCGVK	3	-	-	HCD_IT	1210.55	2	45042
SCO2035	gi|21220516	6.60E-03	DDGSESAGPVVAPSGAQGK	2	-	-	HCD_IT	1026.959	2	14239
SCO2096	gi|21220576	5.90E-03	KLDACPNESAVAVPVTGDDGPK	3	-	-	HCD_IT	909.4205	3	26945
SCO2156	gi|21220633	1.50E-03	EGTFLGKCAELCGVDHSR	1	-	-	HCD_IT	733.3296	3	25242
SCO2838	gi|21221288	2.00E-03	AAGAGITQQPK	2	T7	Only possible site	ETD_IT	683.3412	2	2576
SCO2963	gi|21221408	6.50E-03	GRGSSDADR	1	-	-	ETD_IT	541.7381	2	8834
SCO3044	gi|21221487	3.20E-05	GDAGQPSDEPAADSEIGVLVQNATR	3	-	-	HCD_IT	995.119	3	63502
SCO3044	gi|21221487	6.40E-03	GDAGQPSDEPAADSEIGVLVQNATR	3	-	-	HCD_IT	995.1174	3	63661
SCO3046	gi|32141189	3.10E-06	VAKPTPNAAGQTPLNILVIGSDAR	2	T5	32	ETD_OT	909.8197	3	Sum of 2 scans in range 22872 to 22874
SCO3046	gi|32141189	8.60E-05	VAKPTPNAAGQTPLNILVIGSDAR	2	-	-	HCD_IT	909.8185	3	56534
SCO3184	gi|32141194	1.10E-05	ATVETAAPDRGDGYGVALR	1	-	4.1	ETD_OT	694.3423	3	9131
SCO3184	gi|32141194	2.60E-04	ATVETAAPDRGDGYGVALR	1	-	8.4	ETD_IT	694.3432	3	10892

SCO Number	Accession[1]	Expect score[2]	Peptide sequence	# Hex[3]	Site allocation	MD-score[4]	Method	Precursor ion *m/z*	Charge	Scan
SCO3184	gi|32141194	1.60E-03	ATVETAAPDRGDGYGVALR	1	-	2.1	ETD_OT	694.3431	3	Sum of 2 scans in range 10601 to 10603
SCO3184	gi|32141194	1.40E-03	KATVETAAPDRGDGYGVALR	1	-	-	HCD_IT	737.0407	3	Sum of scans in range 16301 to 16323
SCO3184	gi|32141194	4.50E-02	KATVETAAPDRGDGYGVALR	1	-	-	HCD_IT	737.0419	3	16390
SCO3353	gi|21221782	1.60E-05	KPSAPECGTPPAGSAK	2	T9	35.2	ETD_OT	626.9599	3	2027
SCO3353	gi|21221782	5.50E-04	KPSAPECGTPPAGSAK	2	-	-	HCD_IT	939.9368	2	4189
SCO3353	gi|21221782	7.10E-04	KPSAPECGTPPAGSAK	3	-	-	HCD_IT	680.9775	3	4050
SCO3353	gi|21221782	7.80E-04	KPSAPECGTPPAGSAK	2	T9	28.2	ETD_IT	626.9593	3	2303
SCO3353	gi|21221782	1.50E-03	KPSAPECGTPPAGSAK	3	T9	15.5	ETD_OT	680.9777	3	1965
SCO3353	gi|21221782	1.60E-03	KPSAPECGTPPAGSAK	3	T9	23.5	ETD_IT	680.9773	3	2241
SCO3357	gi|21221786	2.90E-02	ASPSKAPDRVDAVR	6	-	-	ETD_IT	814.373	3	2155
SCO3357	gi|21221786	7.60E-04	DEGPAHADAVGGAGSASPAPAAK	6	S15 S17	Only possible sites	ETD_OT	992.7591	3	Sum of 2 scans in range 2970 to 2972

SCO Number	Accesion[1]	Expect score[2]	Peptide sequence	# Hex[3]	Site allocation	MD-score[4]	Method	Precursor ion *m/z*	Charge	Scan
SCO4130	gi|21222529	6.50E-05	TSATAPSGTRPVQSGFAHDAQ GAQSAAANYAVALGSDGMFDK	2	-	-	HCD_IT	1109.511	4	58511
SCO4130	gi|21222529	3.60E-04	TSATAPSGTRPVQSGFAHDAQ GAQSAAANYAVALGSDGMFDK	2	-	4.7	HCD_IT	1109.514	4	Sum of 2 scans in range 23637to 23639
SCO4130	gi|21222529	9.80E-03	TSATAPSGTRPVQSGFAHDAQ GAQSAAANYAVALGSDGMFDK	2	-	-	HCD_IT	1113.511	4	Sum of 2 scans in range 21239 to 21241
SCO4141	gi|21222539	2.10E-04	TPQPPATEDTRPGR	1	T1	27.8	ETD_OT	562.2769	3	Sum of 2 scans in range 2571 to 2584
SCO4141	gi|21222539	4.10E-03	TPQPPATEDTRPGR	1	T1	13	ETD_OT	562.2767	3	2323
SCO4141	gi|21222539	6.10E-03	TPQPPATEDTRPGR	1	T1	21.7	ETD_IT	562.2783	3	Sum of 2 scans in range 2611 to 2665
SCO4141	gi|21222539	3.70E-02	TPQPPATEDTRPGR	1	-	-	HCD_IT	842.9104	2	5650
SCO4142	gi|21222540	7.80E-04	ADTLPATKSFLNYMASED GQGLLADAGYAPMPTEIITK	1	-	-	HCD_IT	1388.343	3	91188
SCO4142	gi|21222540	4.80E-02	CDDAKGQLQASGSSAQK	1	-	-	HCD_IT	956.9326	2	4208

SCO Number	Accesion[1]	Expect score[2]	Peptide sequence	# Hex[3]	Site allocation	MD-score[4]	Method	Precursor ion *m/z*	Charge	Scan
SCO4142	gi|21222540	8.10E-03	DGIKTVDVK	1	T5	Only possible site	ETD_OT	379.5407	3	3598
SCO4142	gi|21222540	3.60E-05	GGQSAQGSSGLAGQV KQTPGAISYFELSYAK	1	-	-	HCD_IT	1083.868	3	58627
SCO4142	gi|21222540	2.80E-03	QTPGAISYFELSYAKDGIK	1	S12	24.8	ETD_IT	750.7115	3	24203
SCO4142	gi|21222540	1.90E-03	TAAAEPVKATVENATAAIGAAK	1	-	-	HCD_IT	739.7253	3	42536
SCO4142	gi|21222540	1.70E-02	TAAAEPVKATVENATAAIGAAK	1	-	0.8	ETD_IT	739.7249	3	Sum of 2 scans in range 18332 to 18361
SCO4142	gi|21222540	1.40E-02	VCKDGQAIDLPMVGGPIAV GFNVTGVDSLVLDAPTMAK	1	-	-	HCD_IT	1345.343	3	92043
SCO4256	gi|21222651	1.80E-02	GGGGGGGGESKKPKPPVR	3	S10	Only possible site	ETD_OT	527.7643	4	Sum of 2 scans in range 1711 to 1712
SCO4307	gi|21222700	6.00E-03	LIYAGAGTAGR	1	T8	Only possible site	ETD_IT	404.5446	3	6751
SCO4548	gi|32141241	4.20E-04	TTSSSSSTAPSAPSAPR	1	-	-	HCD_IT	877.4079	2	6204
SCO4885	gi|21223231	8.10E-07	SDQAPEPGFADSPYITVTFR	1	-	-	HCD_IT	1180.552	2	71923
SCO4905	gi|21223279	7.20E-05	ATPGLPAQVFLLCGSSLVAVDR	2	-	-	HCD_IT	865.7785	3	88724

SCO Number	Accesion[1]	Expect score[2]	Peptide sequence	# Hex[3]	Site allocation	MD-score[4]	Method	Precursor ion *m/z*	Charge	Scan
SCO4905	gi|21223279	2.40E-02	ATPGLPAQVFLLCGSSLVAVDR	3	-	-	HCD_IT	919.7954	3	88232
SCO4968	gi|21223341	9.10E-06	VDFKEPAEQDASAGPEAKPQR	1	S12	Only possible site	ETD_OT	811.3916	3	Sum of 2 scans in range 6230 to 6232
SCO4968	gi|21223341	3.20E-02	VDFKEPAEQDASAGPEAKPQR	1	S12	Only possible site	HCD_IT	811.3906	3	13570
SCO5646	gi|21223997	5.90E-03	AILTKDNPQGDVFFGVDNTLLSR	1	-	-	HCD_IT	894.7915	3	72287
SCO5751	gi|21224097	4.40E-04	KPADPKPEPSDSAIAAAPADKVTVK	6	S10 S12	31.8	ETD_OT	869.6701	4	Sum of 4 scans in range 4856 to 4896
SCO5751	gi|21224097	2.90E-03	KPADPKPEPSDSAIAAAPADKVTVK	6	S10 S12	28	ETD_OT	695.9383	5	4335
SCO5776	gi|21224122	1.30E-03	SEKVDFAGPYLLAHQDVLIR	1	S1	Only possible site	HCD_IT	609.072	4	60208
SCO7218	gi|21225495	6.60E-04	ASSGGHYPVTVENCGEK	3	-	7.8	ETD_OT	759.9905	3	Sum of 2 scans in range 3702 to 3705

SCO Number	Accesion[1]	Expect score[2]	Peptide sequence	# Hex[3]	Site allocation	MD-score[4]	Method	Precursor ion *m/z*	Charge	Scan
SCO7218	gi|21225495	2.70E-02	ASSGGHYPVTVENCGEK	3	-	-	ETD_OT	759.9905	3	3332
SCO7218	gi|21225495	4.70E-03	ASSGGHYPVTVENCGEKLTFEK	3	-	6.1	ETD_IT	724.83	4	11845

[1] GenBank accession number

[2] Expect value assigned to the glycopeptide match in the MASCOT search

[3] Number of hexose residues on the glycopeptide.

[4] The difference between the MASCOT ion scores for the two best alternative modification sites in the same peptide, assigned by the database search. The MD-score for HCD was not considered to be reliable and therefore is not reported. Sites within glycopeptides containing only one possible glycosylated residue were automatically assigned.

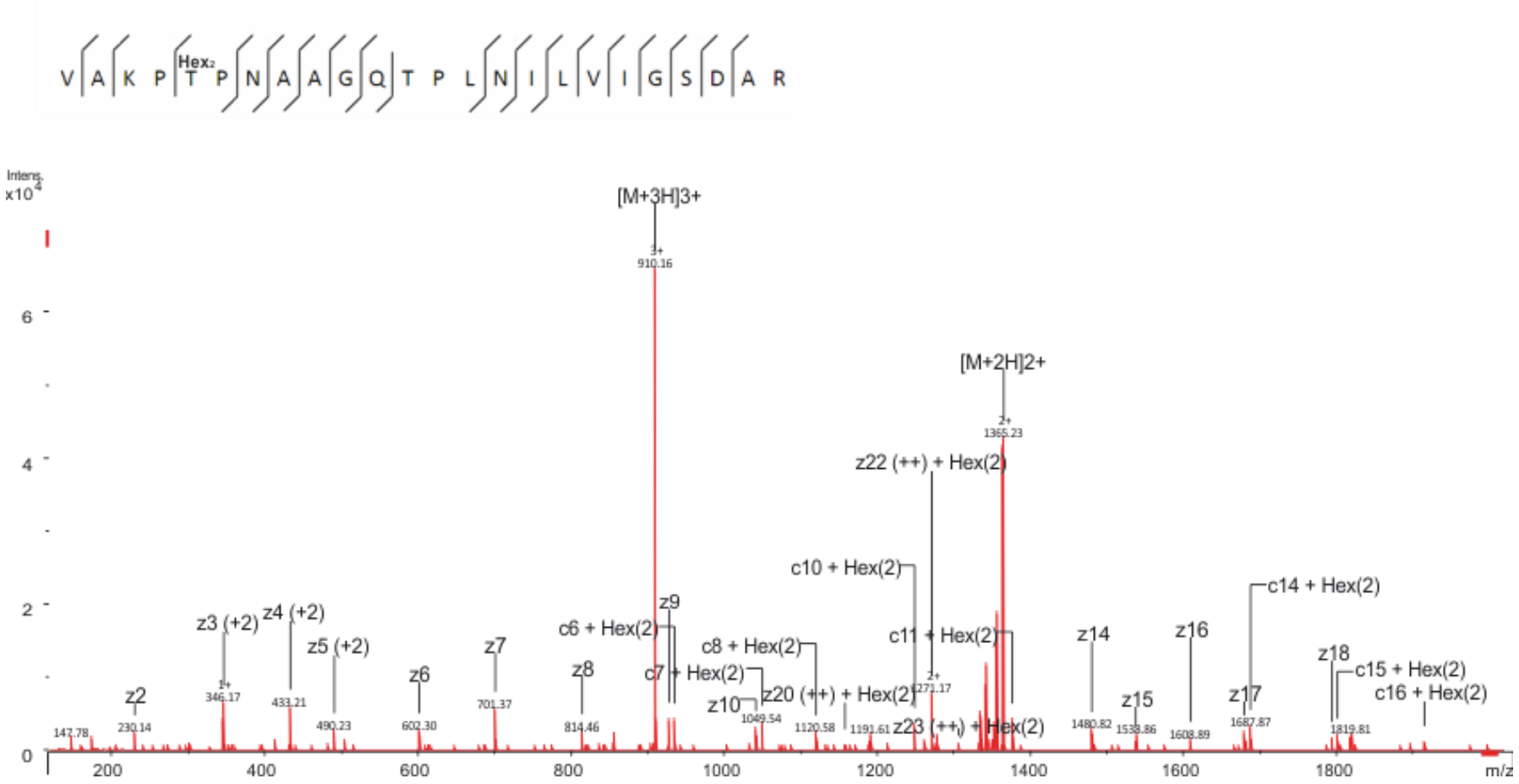


Figure A.4 ETD spectrum of the SCO3046 tryptic peptide VAKPTPNAAGQTPLNILVIGSDAR (aa 43 – 66) indicates Hex_2 on Thr47. Precursor *m/z* 909.819; charge = 3+; retention time = 89.4 min; scan = 22874; e-value = 3.10E-06; method = HCD_IT, ETD_OT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

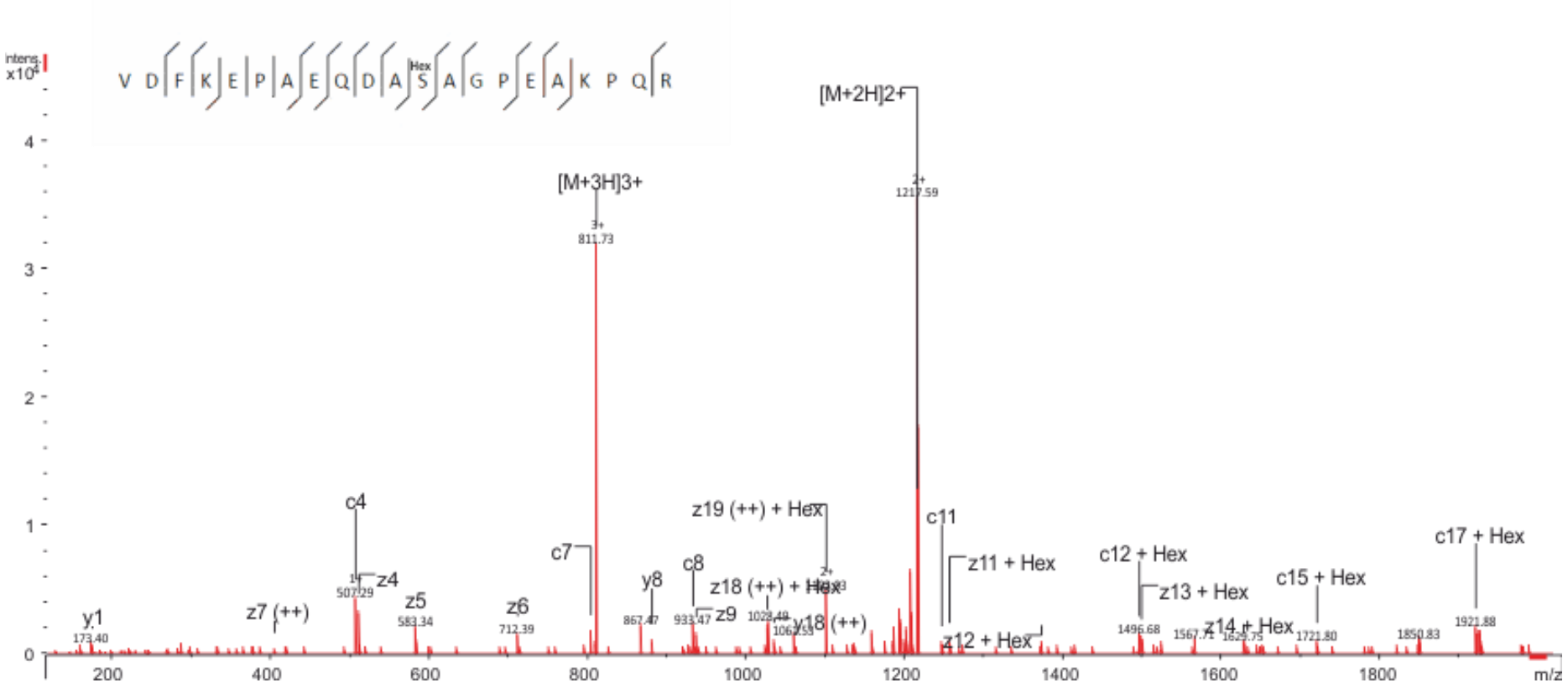


Figure A.5 ETD spectrum of the SCO4938 tryptic peptide VDFKEPAEQDASAGPEAKPQR (aa 54 – 74) indicates Hex on Ser65. Precursor *m/z* 811.391; charge = 3+; retention time = 31.4 min; scan = 6232; e-value = 9.10E-06; method = ETD_OT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

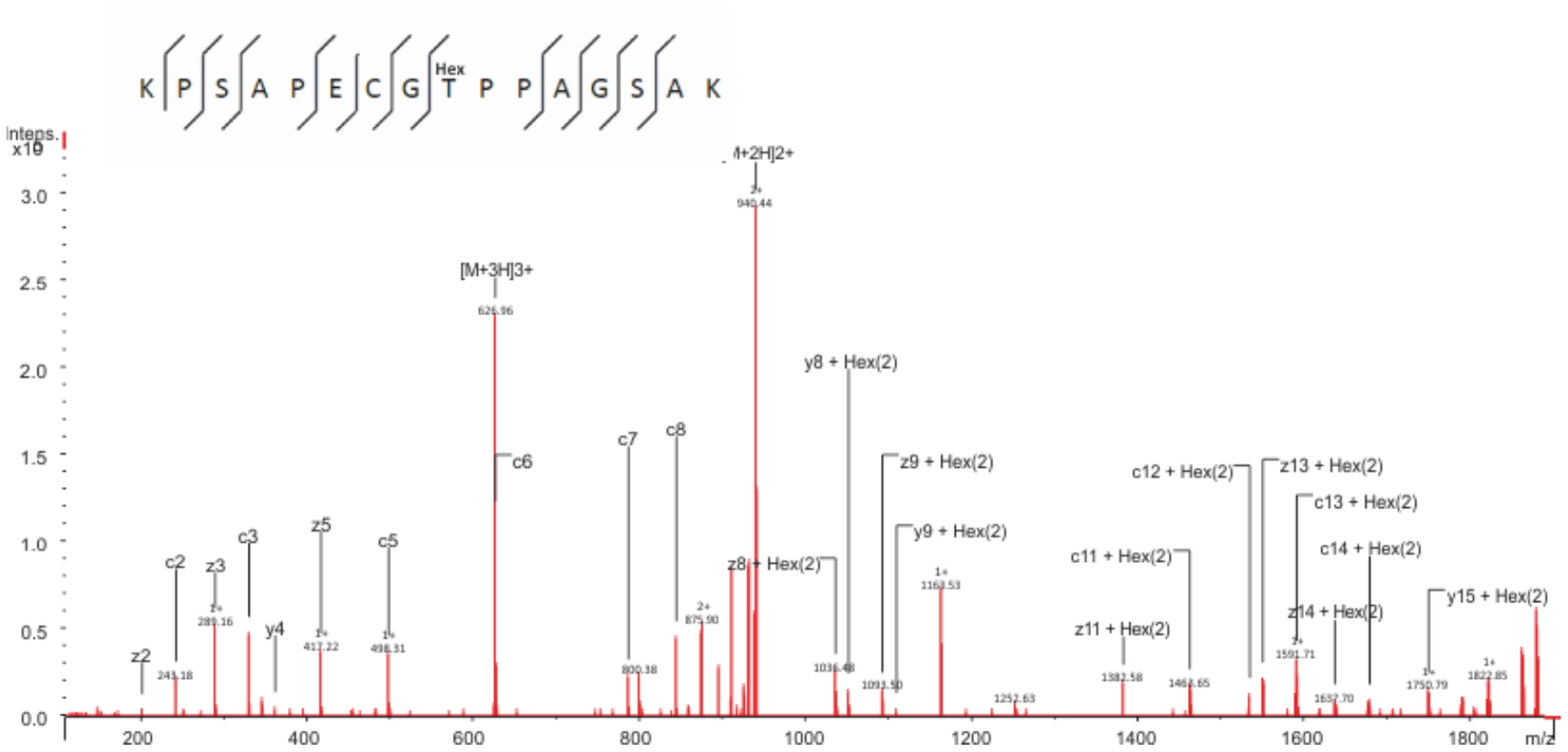


Figure A.6 ETD spectrum of the SCO3353 tryptic peptide KPSAPECGTPPAGSAK (aa 86 – 101) indicates Hex_2 on Thr94. Precursor *m/z* 626.959; charge = 3+; retention time = 17.5 min; scan = 2027; e-value = 1.60E-05; method = HCT_IT, ETD_IT, ETD_OT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues.

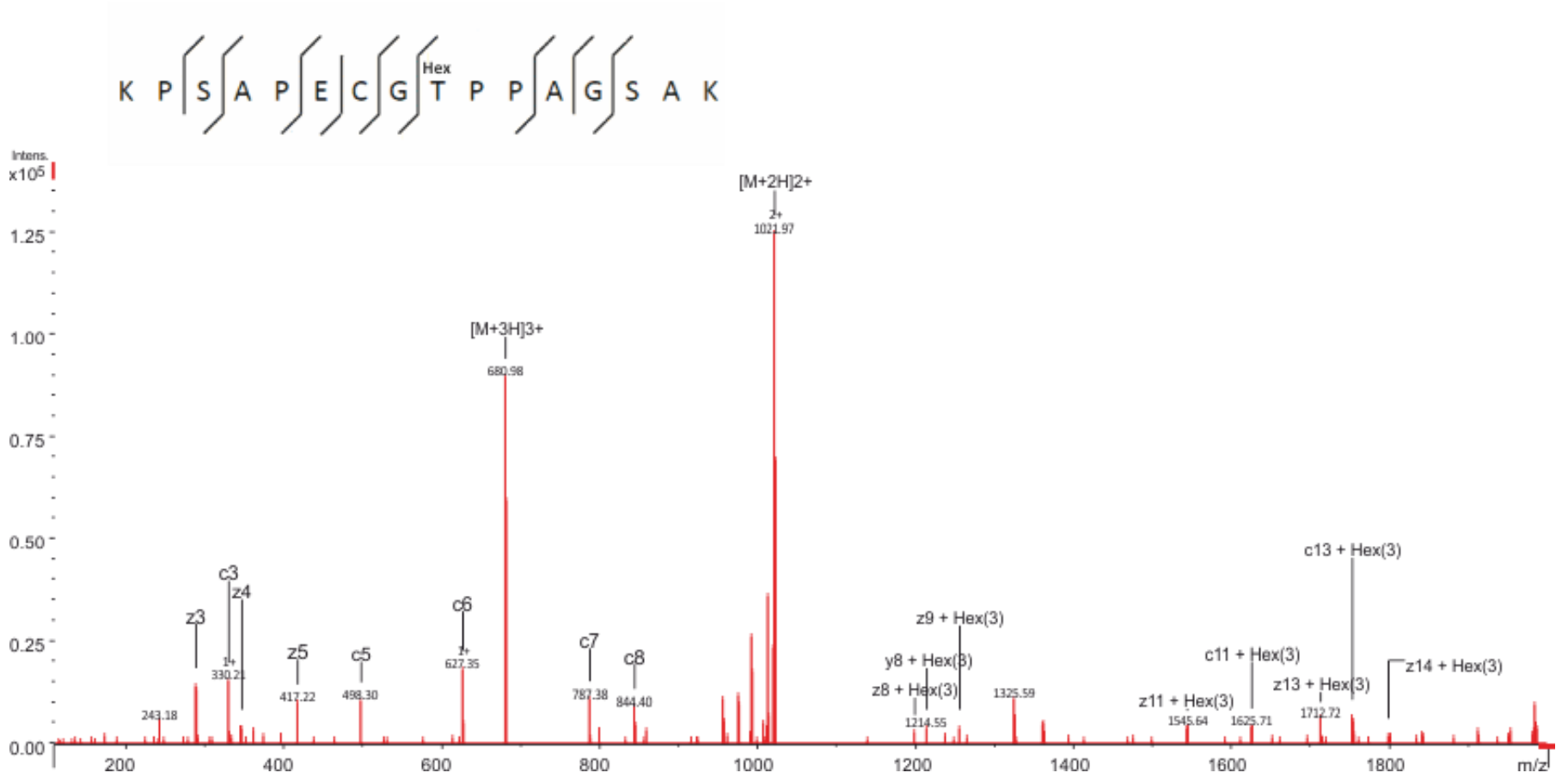


Figure A.7 ETD spectrum of the SCO3353 tryptic peptide KPSAPECGTPPAGSAK (aa 86 – 101) indicates Hex_3 on Thr94. Precursor *m/z* 680.977; charge = 3+; 4retention time = 17.3 min; scan = 4050; e-value = 7.10E-04; method = HCT_IT, ETD_IT, ETD_OT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues.

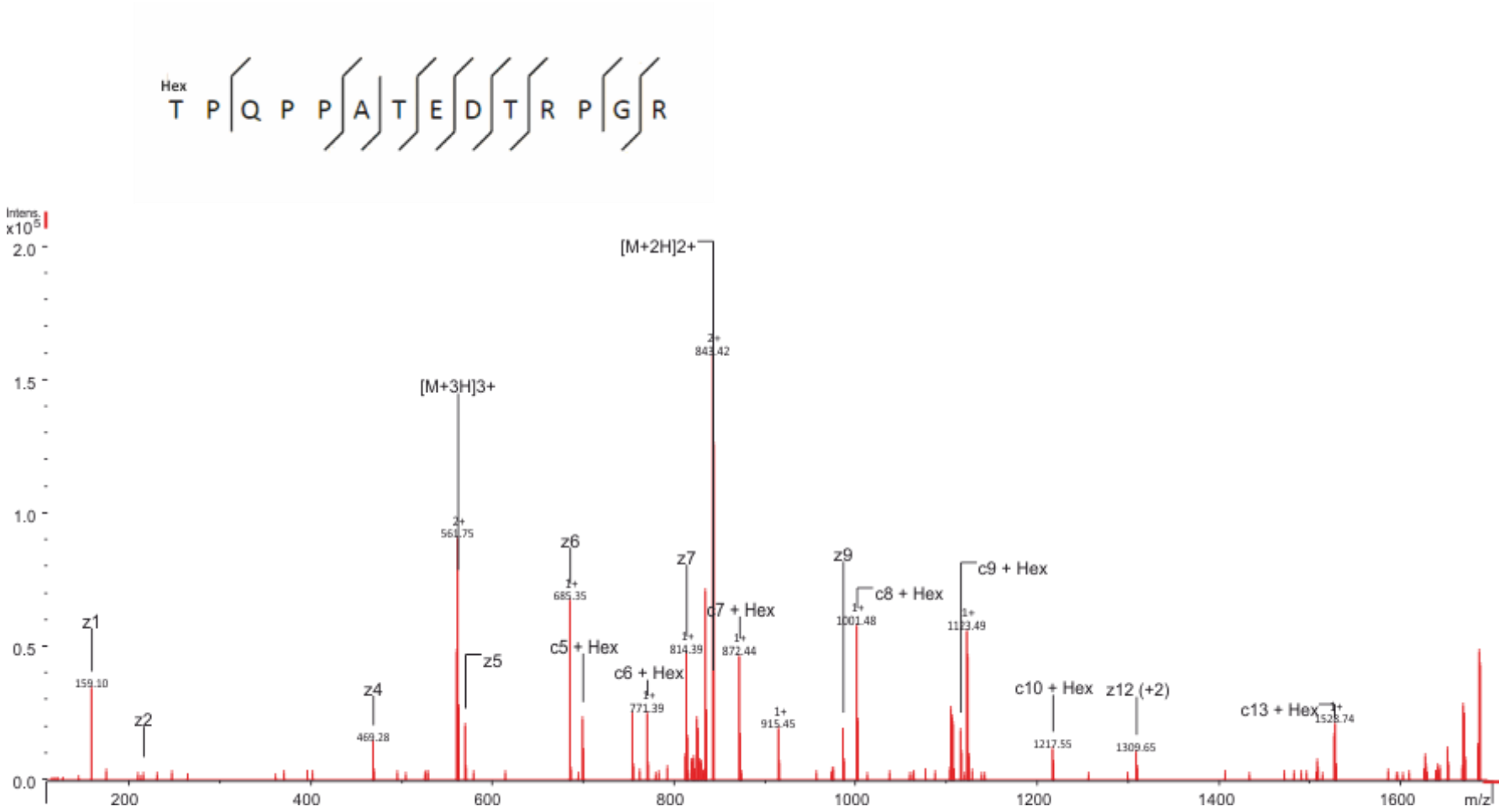


Figure A.8 ETD spectrum of the SCO4141 tryptic peptide TPQPPATEDTRPGR (aa 15 – 28) indicates Hex on Thr15. Precursor *m/z* 562.276 charge = 3+; retention time = 18.7 min; scan = 2571; e-value = 2.10E-04; method = HCD_IT, ETD_OT, ETD_IT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

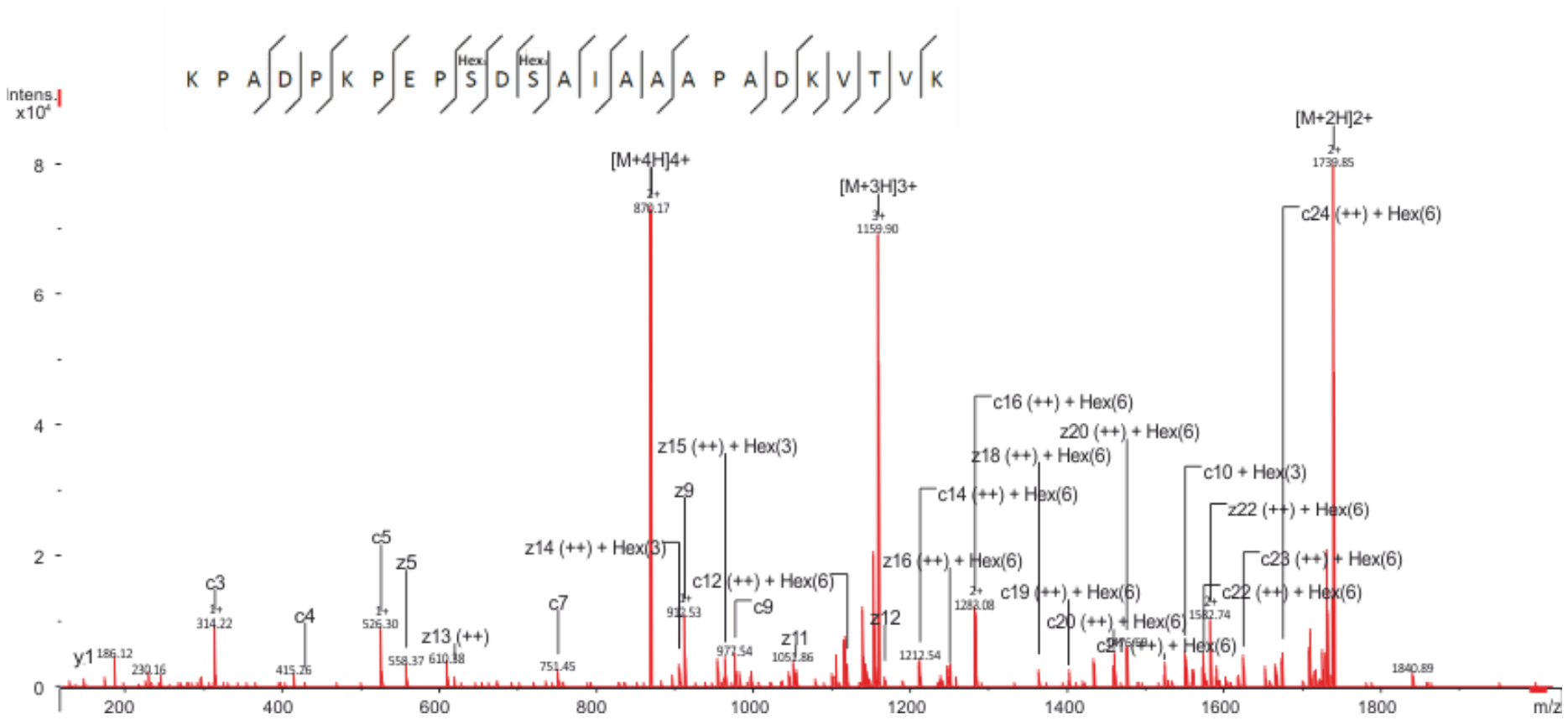


Figure A.9 ETD spectrum of the SCO5751 tryptic peptide KPADPKPEPSDSAIAAAPADKVTVK (aa 184 – 207) indicates a trihexose on both Ser193 and Ser195. Precursor *m/z* 695.938 charge = 5+; retention time = 26.8 min; scan = 4858; e-value = 4.40E-04; method = ETD_OT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

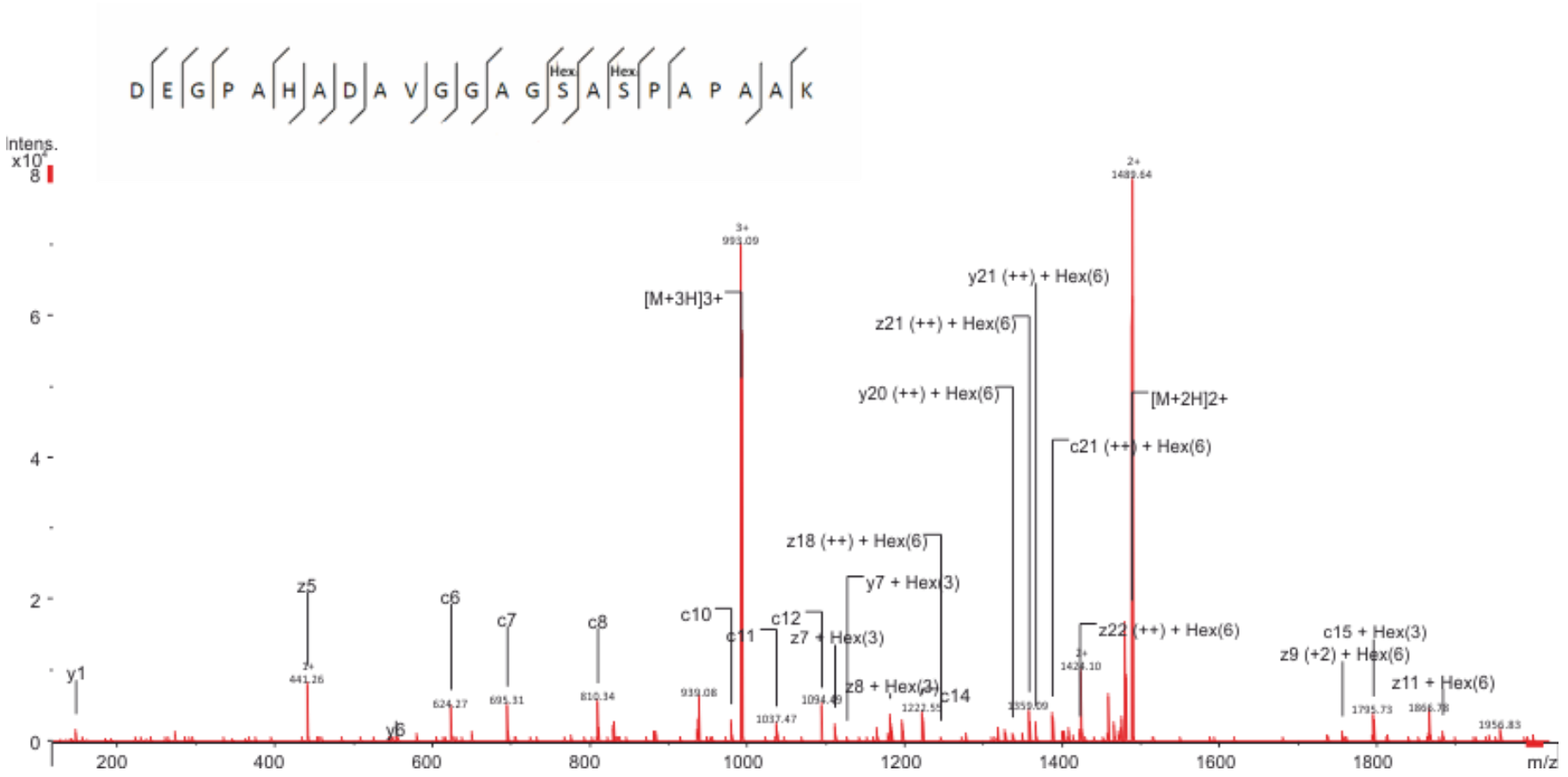


Figure A.10 ETD spectrum of the SCO3357 tryptic peptide DEGPAHADAVGGAGSASPAPAAK (aa 23 – 45) indicates a trihexose on both Ser37 and Ser39. Precursor *m/z* 938.741 charge = 3+; retention time = 20.4 min; scan = 2972; e-value = 7.60E-04; method = ETD_OT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

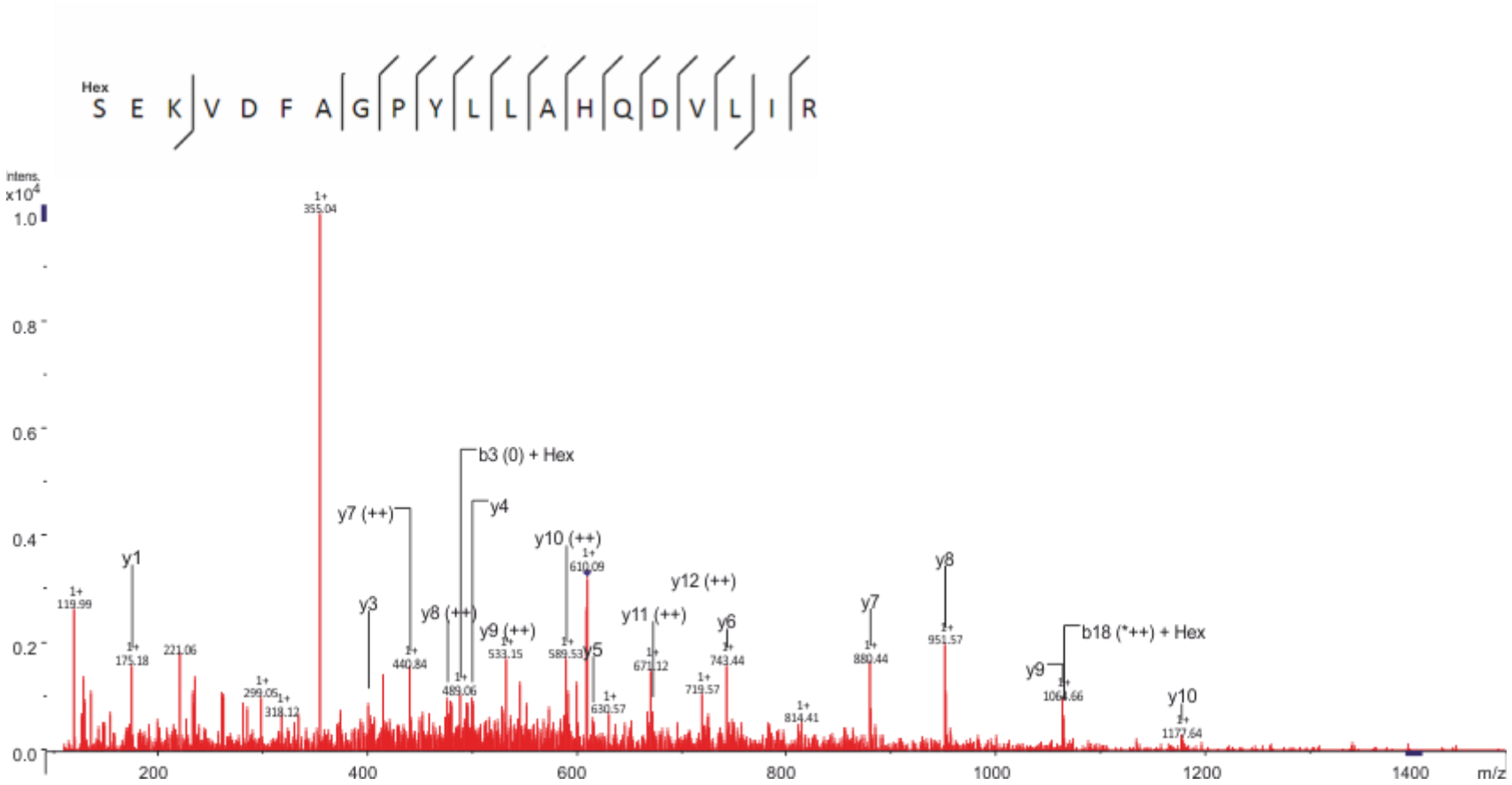


Figure A.11 HCD spectrum of the SCO5776 tryptic peptide SEKVDFAGPYLLAHQDVLIR (aa 114 – 133) indicates a hexose on Ser114. Precursor *m/z* 609.072 charge = 4+; retention time = 94.1 min; scan = 60208; e-value = 1.30E-03; method = HCD_IT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. The precursor ion is indicated by a ◆ (++) Indicates 2+ ion series. (0) Indicates ions that have lost H_2O. (*) Indicates ions that have lost NH_3

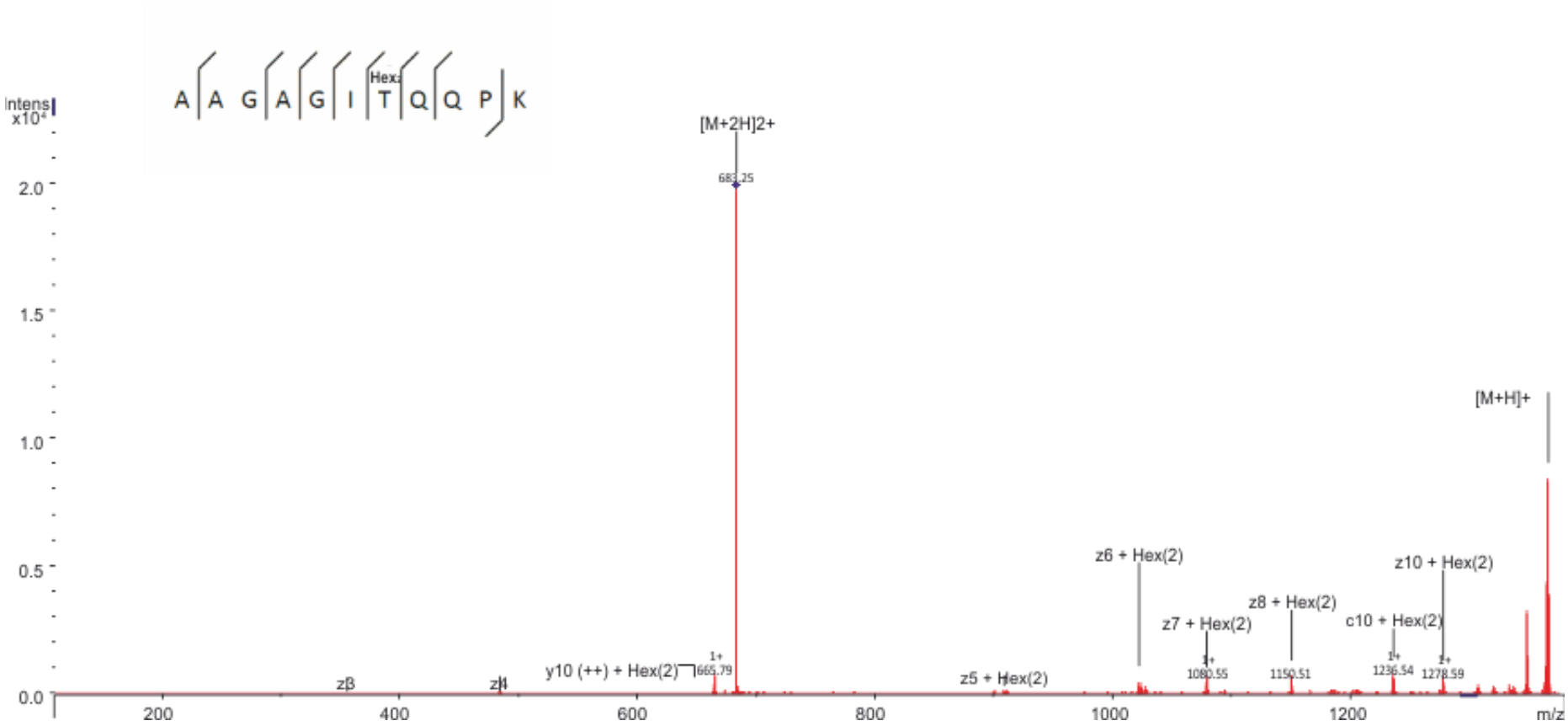


Figure A.12 ETD spectrum of the SCO2838 tryptic peptide AAGAGITQQPK (aa 32 – 42) indicates Hex_2 on Thr38. Precursor *m/z* 683.3412 charge = 2+; retention time = 18.5 min; scan = 2576; e-value = 2.00E-03; method = ETD_IT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. The precursor ion is indicated by a ◆ (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

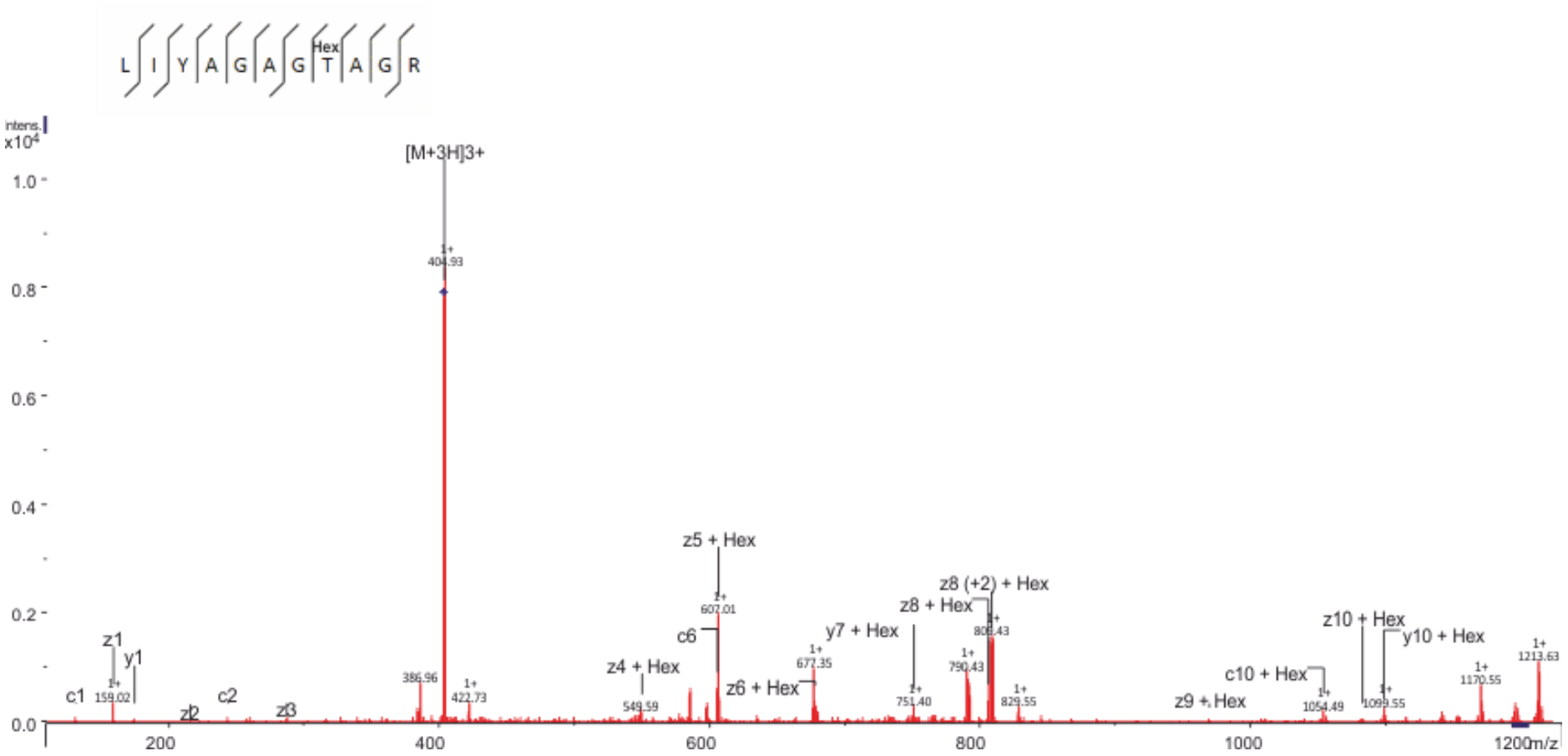


Figure A.13 ETD spectrum of the SCO4307 tryptic peptide LIYAGAGTAGR (aa 76 – 86) indicates a hexose on Thr83. Precursor *m/z* 404.5446 charge = 3+; retention time = 32.5 min; scan = 6751; e-value = 6.00E-04; method = ETD_IT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. The precursor ion is indicated by a ◆ (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

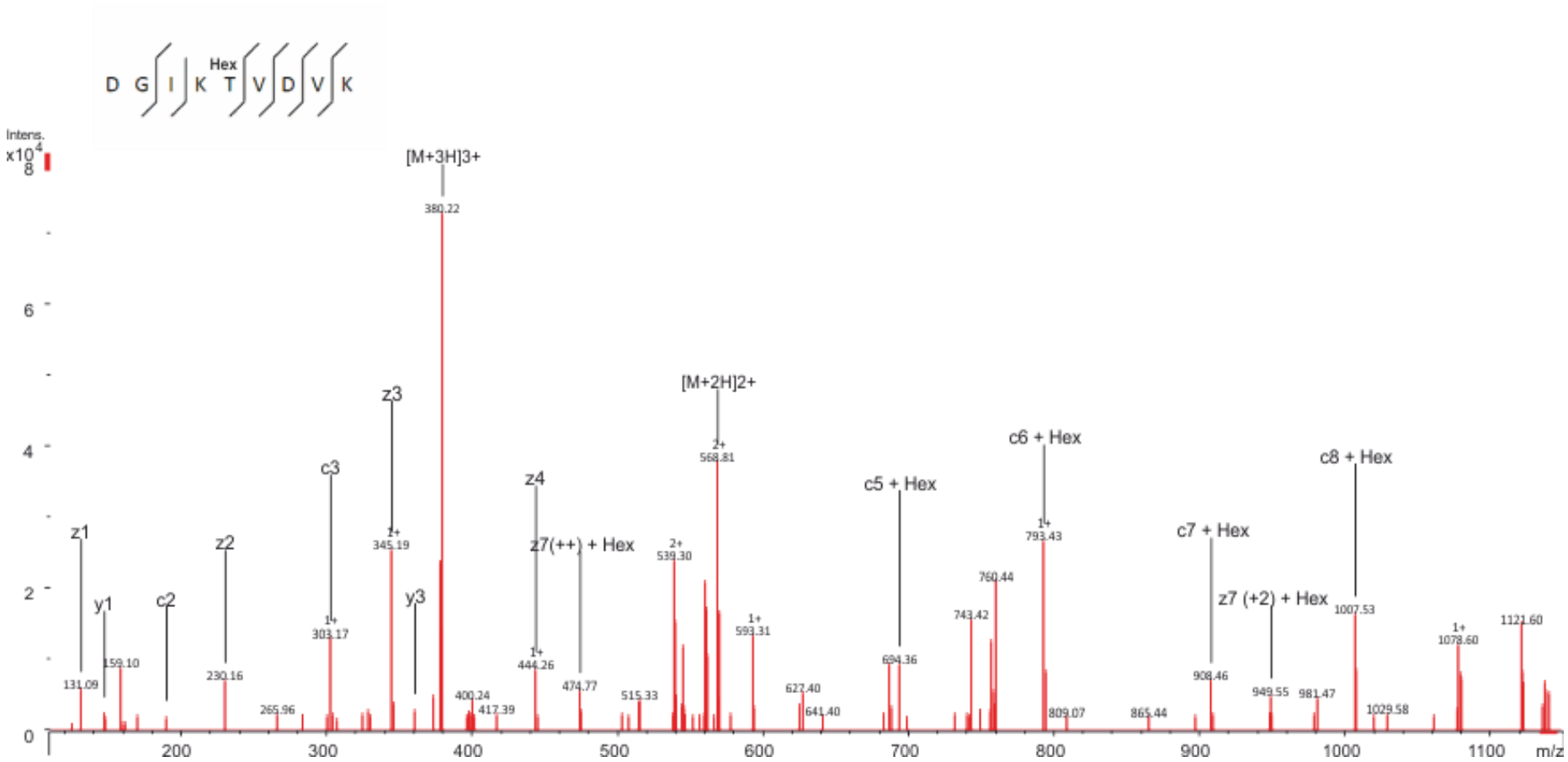


Figure A.14 ETD spectrum of the SCO4142 tryptic peptide DGIKTVDVK (aa 255 – 263) indicates a hexose on Thr259. Precursor *m/z* 379.5407 charge = 3+; retention time = 23.9 min; scan = 3598; e-value = 8.10E-03; method = ETD_OT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

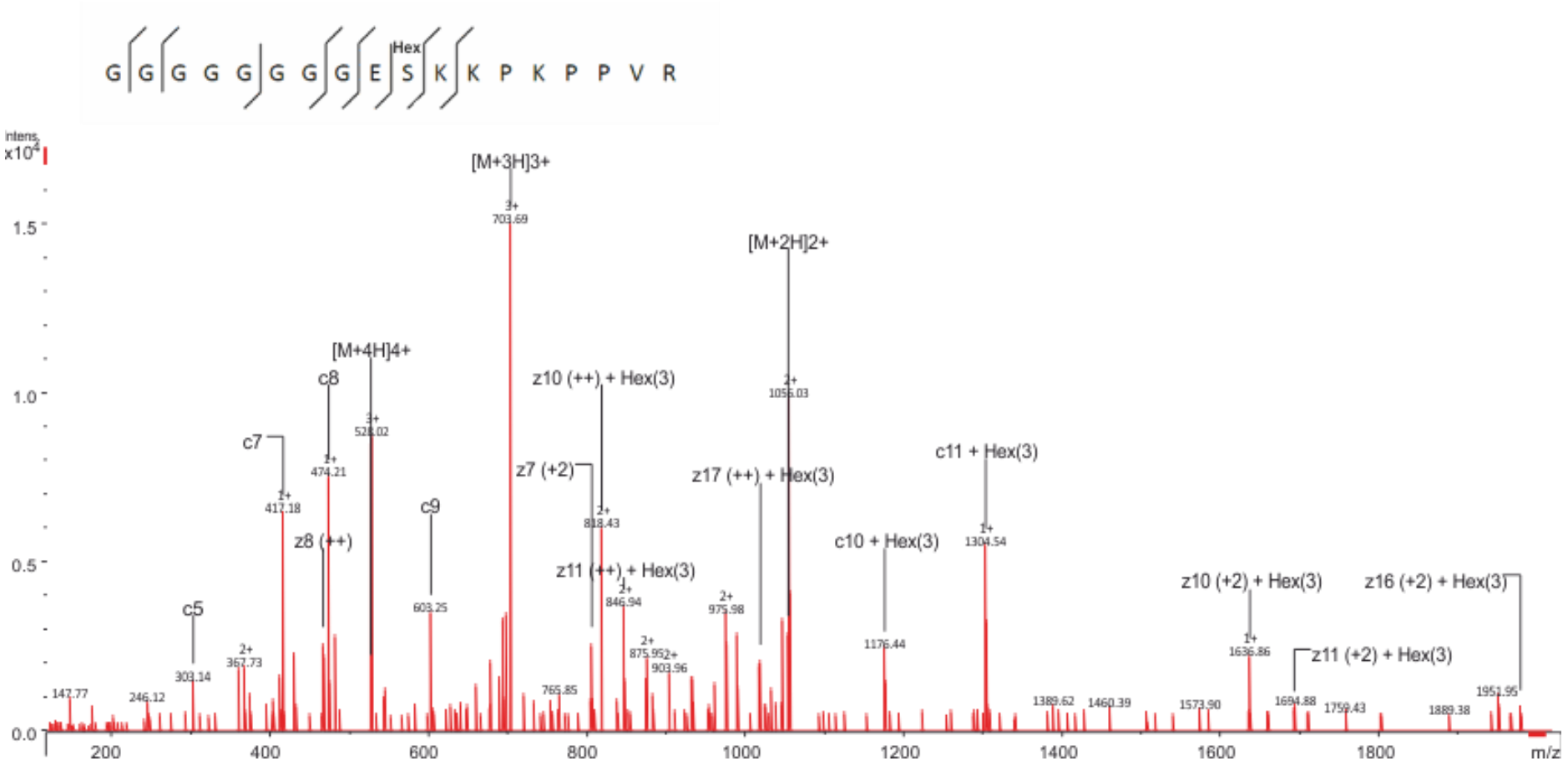


Figure A.15 ETD spectrum of the SCO4356 tryptic peptide GGGGGGGGESKKPKPPVR (aa 308 – 325) indicates a trihexose on Ser317. Precursor *m/z* 527.7643 charge = 4+; retention time = 15.5 min; scan = 1712; e-value = 8.10E-02; method = ETD_OT. Product ions that contain the hexose are indicated by + Hex(n), where n = number of hexose residues. (++) Indicates 2+ ion series. (+2) Indicates z+2 ions.

A.5 Glycomics

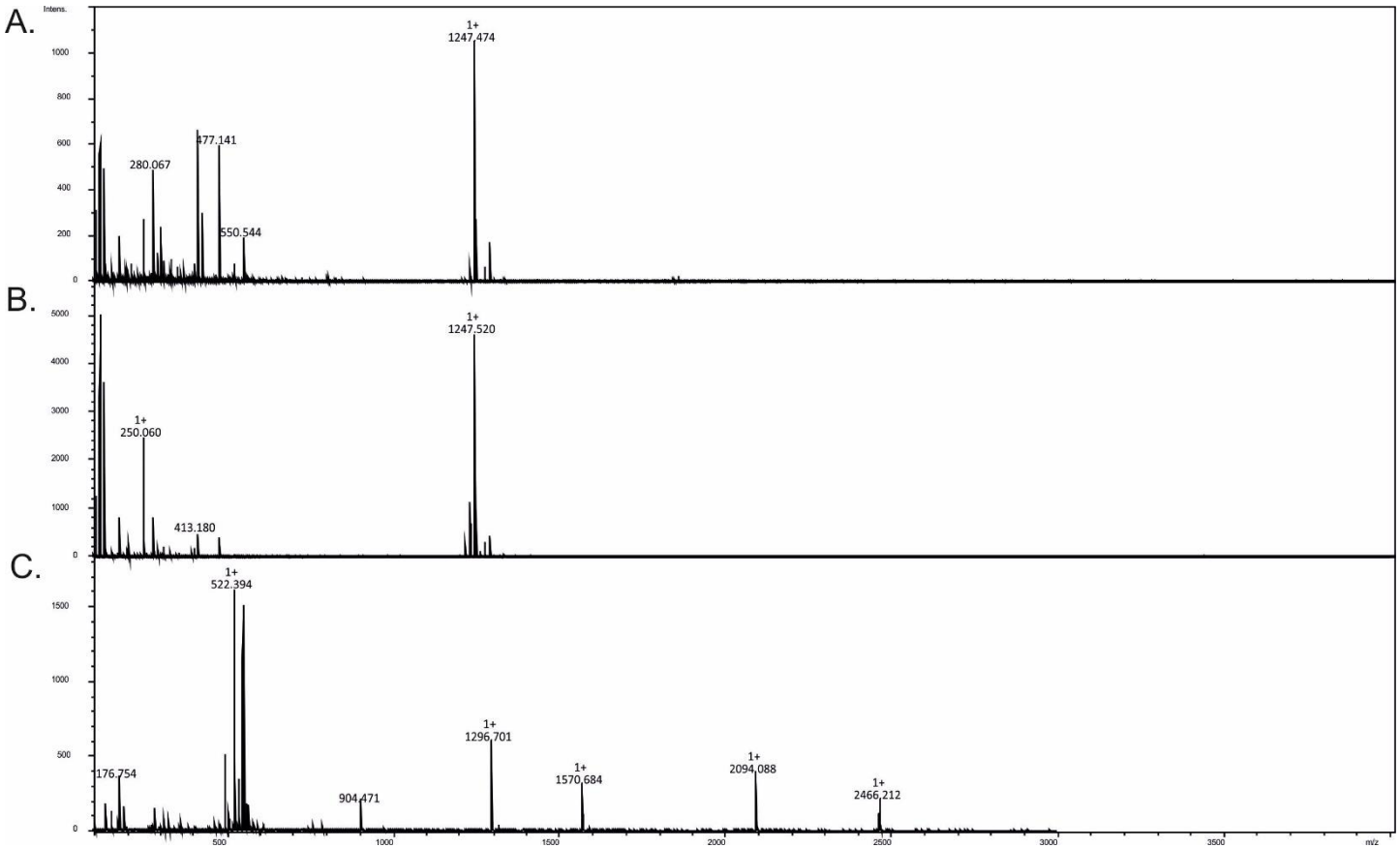


Figure A.16 MALDI-TOF-MS spectrum of permethylated *S. coelicolor* glycans alongside a mock glycan preparation from a blank SDS-PAGE gel. *S. coelicolor* glycans (A) were prepared alongside an in-gel β-elimination of a protein free section of SDS-PAGE gel (B). Panel C indicates a matrix blank. This analysis was carried out by Rachel Bates at the University of York Bioscience Technology Facility.

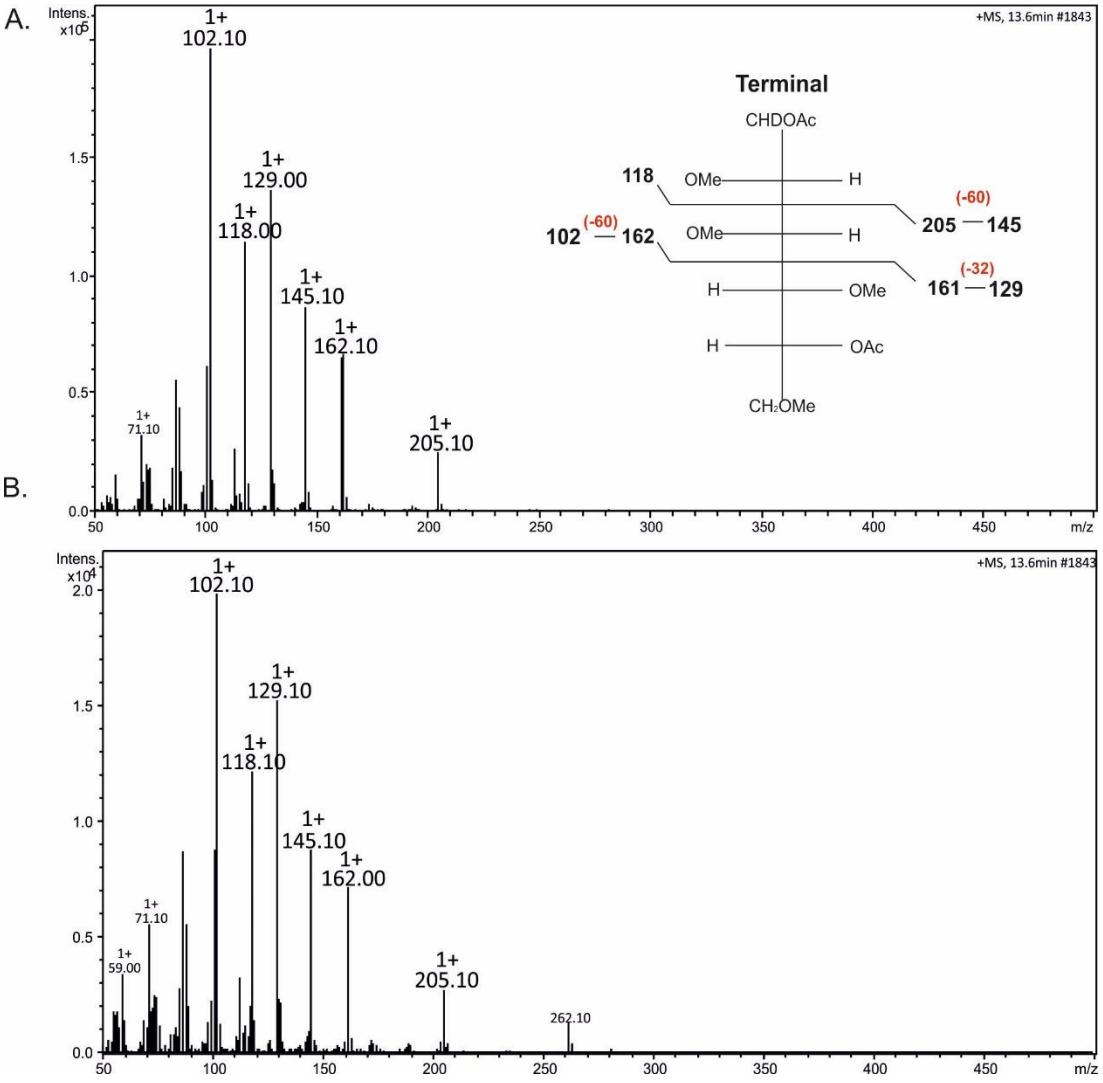


Figure A.17 GC mass spectrum confirming terminal mannose in the mannose standard (A) and the *S. coelicolor* glycans (B). Diagnostic fragment ions of *m/z* 118, 162, 205, 161, 129, 145 and 102 are indicated.

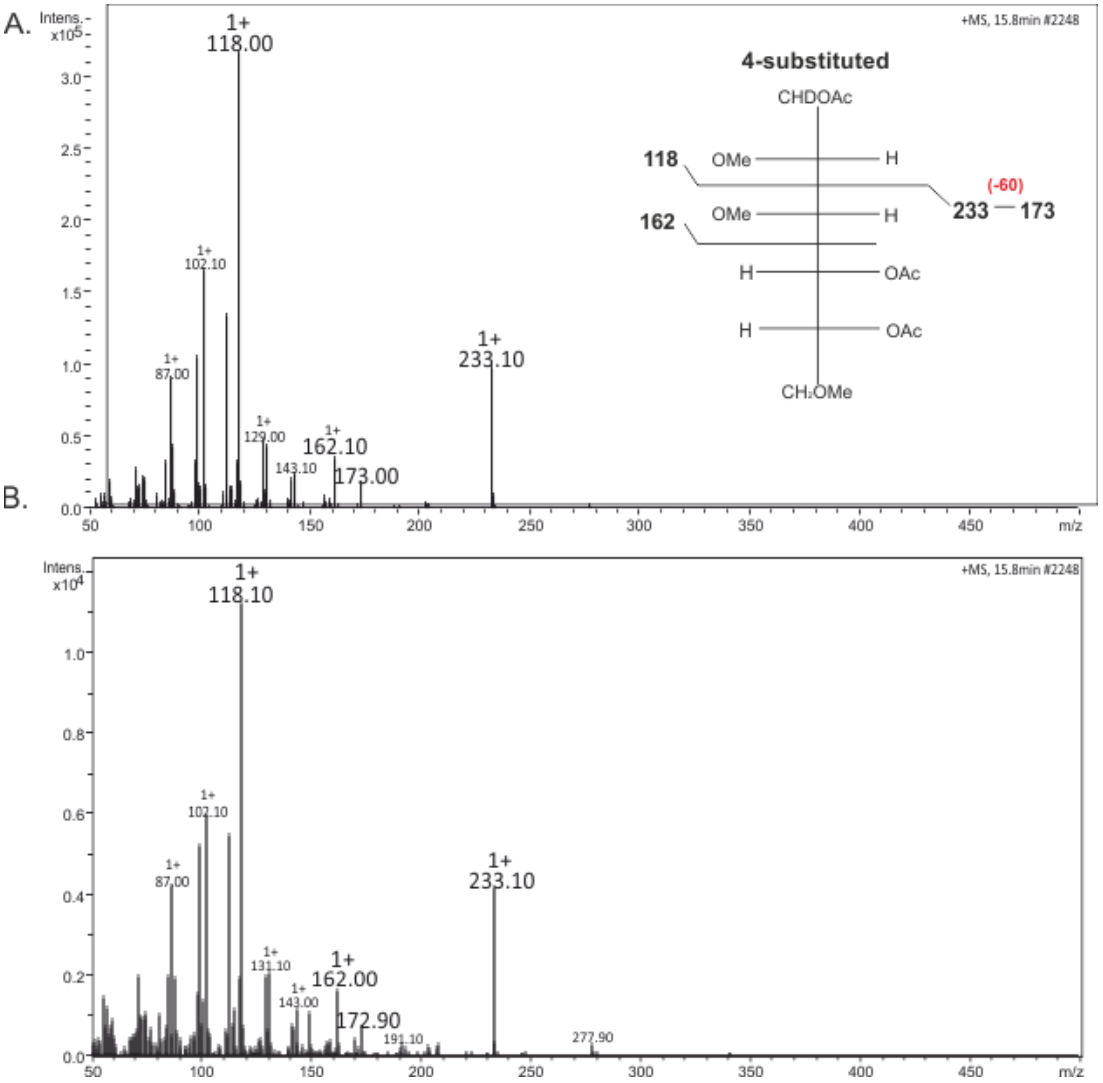


Figure A.18 GC mass spectrum confirming 4-substituted mannose in the mannose standard (A) and the *S. coelicolor* glycans (B). Diagnostic fragment ions of *m/z* 118, 162, 233 and 173 are indicated.

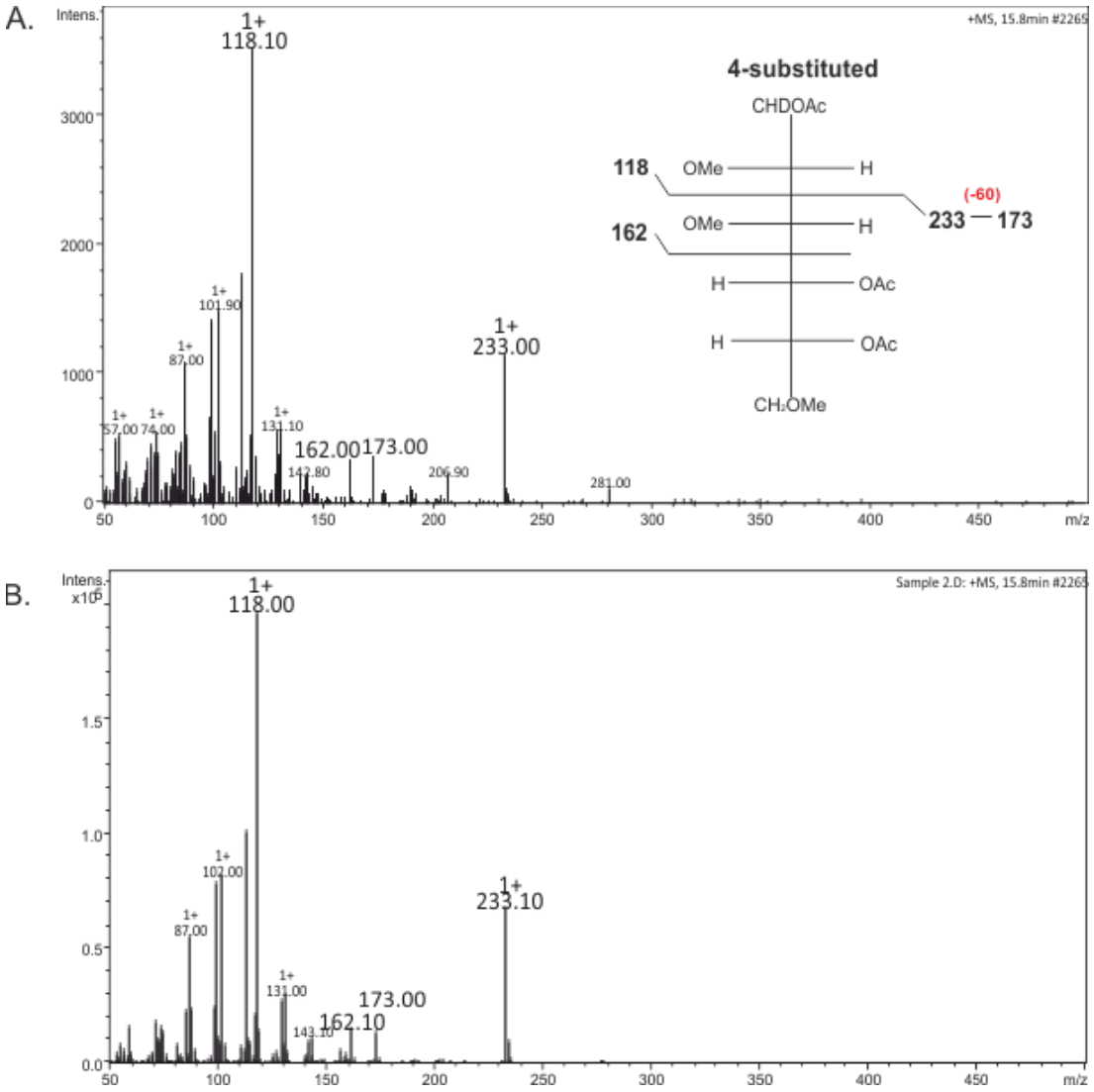


Figure A.19 GC mass spectrum confirming 4-substituted glucose in the glucose standard (A) and the *S. coelicolor* glycans (B). Diagnostic fragment ions of *m/z* 118, 162, 233 and 173 are indicated.

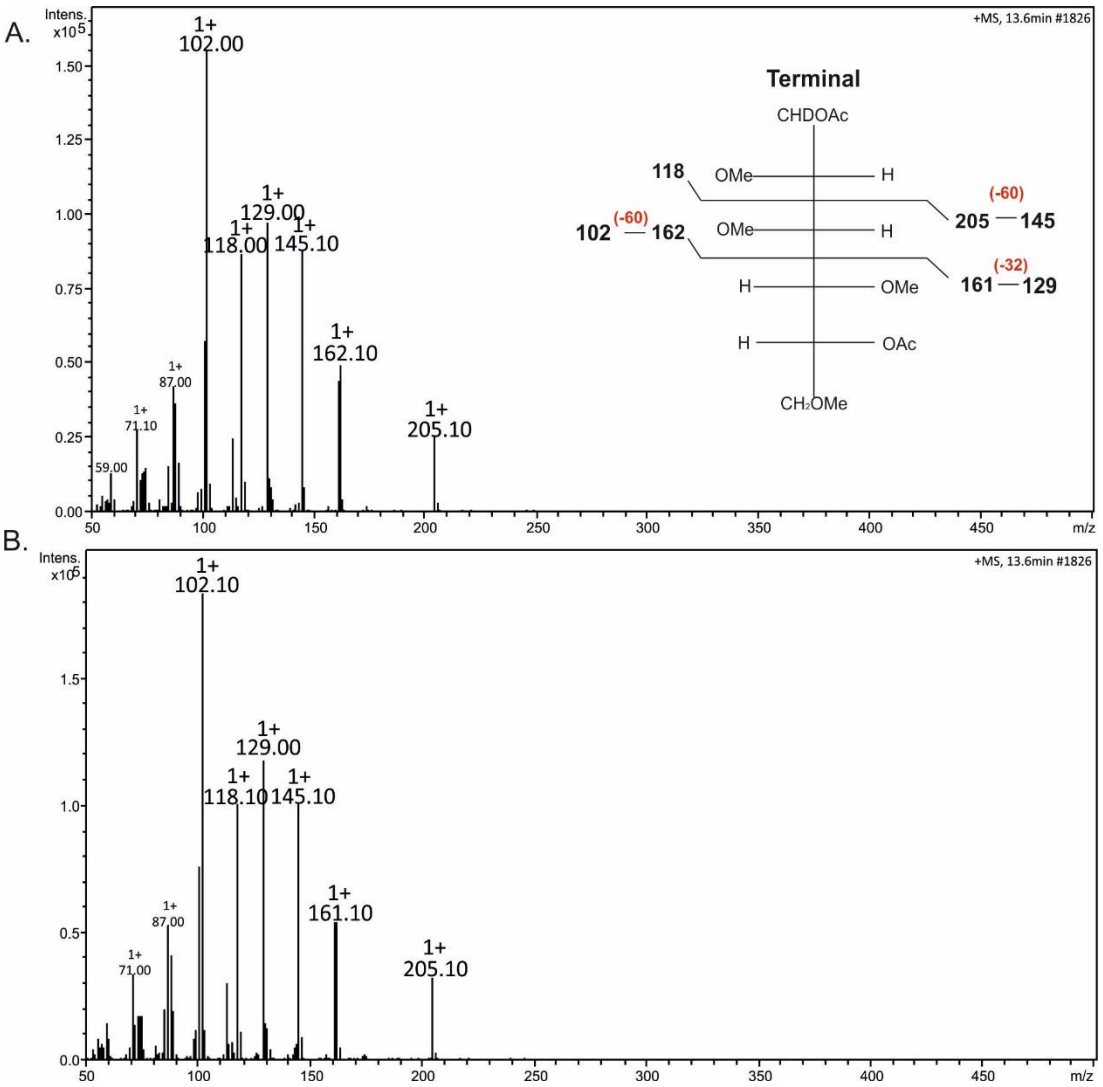


Figure A.20 GC mass spectrum confirming terminal glucose in the glucose standard (A) and the *S. coelicolor* glycans (B). Diagnostic fragment ions of *m/z* 118, 162, 205, 161, 129, 145 and 102 are indicated.

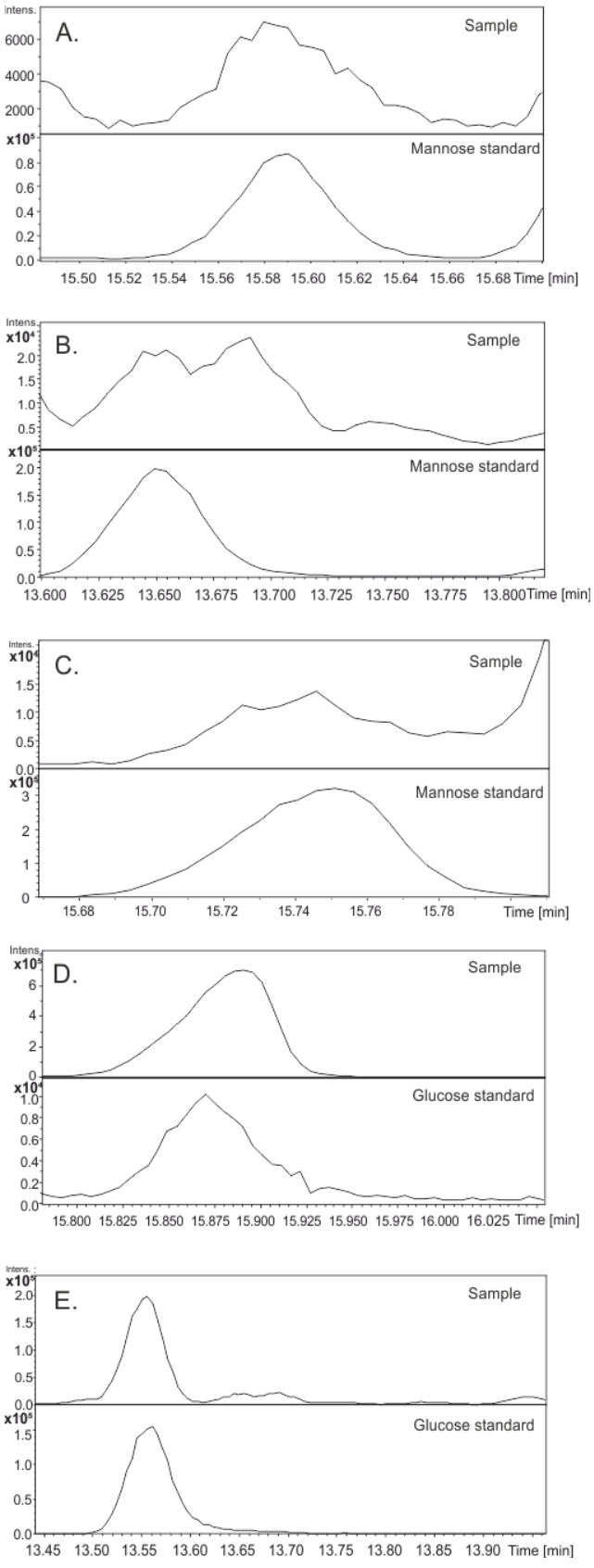
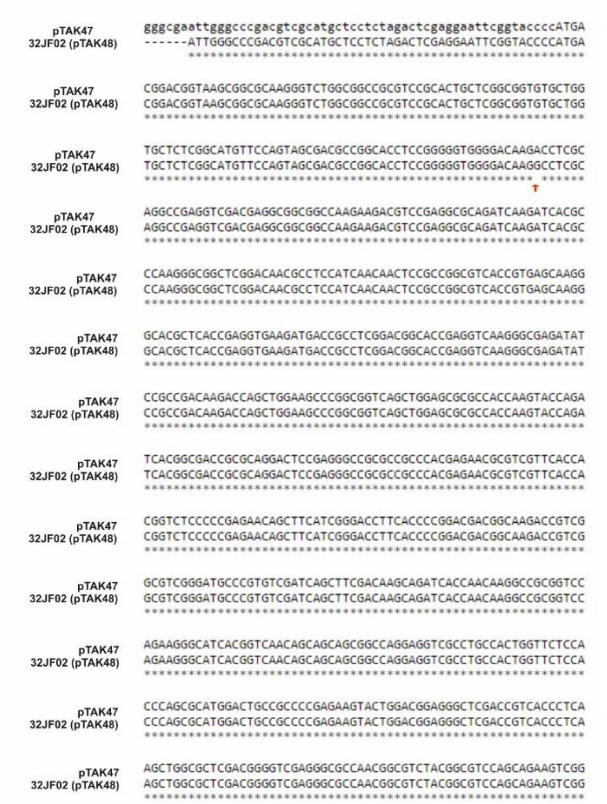


Figure A.21 GC chromatograms showing the co-elution of peaks in the *S. coelicolor* glycan sample with peaks in the PMAAs mannose and glucose standards. Panels show peaks corresponding to: A. 2-substituted mannose. B. terminal mannose. C. 4-substituted mannose. D. 4-substituted glucose. E. Terminal glucose.

A.6 Sequencing data

Figure A. 22 Sequencing of the mutated *sco4934* genes in pTAK48 – pTAK53 to confirm the mutations introduced by site directed mutagenesis. Red arrows indicate the positions of the mutations.




pTAK47
56II23 (pTAK50)
CGGACGGTAAGCGGCGCAAGGGTCTGGCGGCCGCGTCCGCACTGCTCGGCGGTGTGCTGG
--GACGGTAAGCGGCGCAAGGGTCTGGCGGCCGCGTCCGCACTGCTCGGCGGTGTGCTGG
TGCTCTCGGCATGTTCCAGTAGCGACGCCGGCACCTCCGGGGGGTGGGGACAAGACCTCGC
TGCTCTCGGCATGTTCCAGTAGCGACGCCGGCACCTCCGGGGGGTGGGGACAAGGCCGCGC
AGGCCGAGGTCGACGAGGCGGCGGCCAAGAAGACGTCCGAGGCGCAGATCAAGATCACGC
AGGCCGAGGTCGACGAGGCGGCGGCCAAGAAGACGTCCGAGGCGCAGATCAAGATCACGC
CCAAGGGCGGCTCGGACAACGCCTCCATCAACAACTCCGCCGGCGTCACCGTGAGCAAGG
CCAAGGGCGGCTCGGACAACGCCTCCATCAACAACTCCGCCGGCGTCACCGTGAGCAAGG
GCACGCTCACCGAGGTGAAGATGACCGCCTCGGACGGCACCGAGGTCAAGGGCGAGATAT
GCACGCTCACCGAGGTGAAGATGACCGCCTCGGACGGCACCGAGGTCAAGGGCGAGATAT
CCGCCGACAAGACCAGCTGGAAGCCCGGCGGTCAGCTGGAGCGCGCCACCAAGTACCAGA
CCGCCGACAAGACCAGCTGGAAGCCCGGCGGTCAGCTGGAGCGCGCCACCAAGTACCAGA
TCACGGCGACCGCGCAGGACTCCGAGGGCCGCGCCGCCCACGAGAACGCGTCGTTCACCA
TCACGGCGACCGCGCAGGACTCCGAGGGCCGCGCCGCCCACGAGAACGCGTCGTTCACCA
CGGTCTCCCCCGAGAACAGCTTCATCGGGACCTTCACCCCGGACGACGGCAAGACCGTCG
CGGTCTCCCCCGAGAACAGCTTCATCGGGACCTTCACCCCGGACGACGGCAAGACCGTCG
GCGTCGGGATGCCCGTGTCGATCAGCTTCGACAAGCAGATCACCAACAAGGCCGCGGTCC
GCGTC----------


pTAK47
24GC15 (pTAK51)
gggcgaattgggcccgacgtcgcatgctcctctagactcgaggaattcggtaccccATGA
----------GCCCGACGTCGCATGCTCCTCTAGACTCGAGGAATTCGGTACCCCATGA
CGGACGGTAAGCGGCGCAAGGGTCTGGCGGCCGCGTCCGCACTGCTCGGCGGTGTGCTGG
TGCTCTCGGCATGTTCCAGTAGCGACGCCGGCACCTCCGGGGGGTGGGGACAAGACCTCGC
AGGCCGAGGTCGACGAGGCGGCGGCCAAGAAGACGTCCGAGGCGCAGATCAAGATCACGC
CCAAGGGCGGCTCGGACAACGCCTCCATCAACAACTCCGCCGGCGTCACCGTGAGCAAGG
GCACGCTCACCGAGGTGAAGATGACCGCCTCGGACGGCACCGAGGTCAAGGGCGAGATAT
CCGCCGACAAGACCAGCTGGAAGCCCGGCGGTCAGCTGGAGCGCGCCACCAAGTACCAGA
TCACGGCGACCGCGCAGGACTCCGAGGGCCGCGCCGCCCACGAGAACGCGTCGTTCACCA
CGGTCTCCCCCGAGAACAGCTTCATCGGGACCTTCACCCCGGACGACGGCAAGACCGTCG
GCGTCGGGATGCCCGTGTCGATCAGCTTCGACAAGCAGATCACCAACAAGGCCGCGGTCC

```
pTAK47          gggcgaattgggcccgacgtcgcatgctcctctagactcgaggaattcggtaccccATGA
24GC19 (pTAK52) ---------------------------------AGACTCGAGGAATTCGGTACCCCATGA
                                                  ****************************

pTAK47          CGGACGGTAAGCGGCGCAAGGGTCTGGCGGCCGCGTCCGCACTGCTCGGCGGTGTGCTGG
24GC19 (pTAK52) CGGACGGTAAGCGGCGCAAGGGTCTGGCGGCCGCGTCCGCACTGCTCGGCGGTGTGCTGG
                ************************************************************

pTAK47          TGCTCTCGGCATGTTCCAGTAGCGACGCCGGCACCTCCGGGGGGTGGGGACAAGACCTCGC
24GC19 (pTAK52) TGCTCTCGGCATGTTCCAGTAGCGACGCCGGCACCTCCGGGGGGTGGGGACAAGACCGTGC
                ************************************************************ **

pTAK47          AGGCCGAGGTCGACGAGGCGGCGGCCAAGAAGACGTCCGAGGCGCAGATCAAGATCACGC
24GC19 (pTAK52) AGGCCGAGGTCGACGAGGCGGCGGCCAAGAAGACGTCCGAGGCGCAGATCAAGATCACGC
                ************************************************************

pTAK47          CCAAGGGCGGCTCGGACAACGCCTCCATCAACAACTCCGCCGGCGTCACCGTGAGCAAGG
24GC19 (pTAK52) CCAAGGGCGGCTCGGACAACGCCTCCATCAACAACTCCGCCGGCGTCACCGTGAGCAAGG
                ************************************************************

pTAK47          GCACGCTCACCGAGGTGAAGATGACCGCCTCGGACGGCACCGAGGTCAAGGGCGAGATAT
24GC19 (pTAK52) GCACGCTCACCGAGGTGAAGATGACCGCCTCGGACGGCACCGAGGTCAAGGGCGAGATAT
                ************************************************************

pTAK47          CCGCCGACAAGACCAGCTGGAAGCCCGGCGGTCAGCTGGAGCGCGCCACCAAGTACCAGA
24GC19 (pTAK52) CCGCCGACAAGACCAGCTGGAAGCCCGGCGGTCAGCTGGAGCGCGCCACCAAGTACCAGA
                ************************************************************

pTAK47          TCACGGCGACCGCGCAGGACTCCGAGGGCCGCGCCGCCCACGAGAACGCGTCGTTCACCA
24GC19 (pTAK52) TCACGGCGACCGCGCAGGACTCCGAGGGCCGCGCCGCCCACGAGAACGCGTCGTTCACCA
                ************************************************************

pTAK47          CGGTCTCCCCCGAGAACAGCTTCATCGGGACCTTCACCCCGGACGACGGCAAGACCGTCG
24GC19 (pTAK52) CGGTCTCCCCCGAGAACAGCTTCATCGGGACCTTCACCCCGGACGACGGCAAGACCGTCG
                ************************************************************

pTAK47          GCGTCGGGATGCCCGTGTCGATCAGCTTCGACAAGCAGATCACCAACAAGGCCGCGGTCC
24GC19 (pTAK52) GCGTCGGGATGCCCGTGTCGATCAGCTTCGACAAGCAGATCACCAACAAGGCCGCGGTCC
                ************************************************************

pTAK47          AGAAGGGCATCACGGTCAACAGCAGCAGCGGCCAGGAGGTCGCCTGCCACTGGTTCTCCA
24GC19 (pTAK52) AGAA------------------------------------------------------------
                ****
```

```
pTAK47          gggcgaattgggcccgacgtcgcatgctcctctagactcgaggaattcggtaccccATGA
24GC17 (pTAK53) ---------------------------------AGACTCGAGGAATTCGGTACCCCATGA
                                                  ****************************

pTAK47          CGGACGGTAAGCGGCGCAAGGGTCTGGCGGCCGCGTCCGCACTGCTCGGCGGTGTGCTGG
24GC17 (pTAK53) CGGACGGTAAGCGGCGCAAGGGTCTGGCGGCCGCGTCCGCACTGCTCGGCGGTGTGCTGG
                ************************************************************

pTAK47          TGCTCTCGGCATGTTCCAGTAGCGACGCCGGCACCTCCGGGGGGTGGGGACAAGACCTCGC
24GC17 (pTAK53) TGCTCTCGGCATGTTCCAGTAGCGACGCCGGCACCTCCGGGGGGTGGGGACAAGGTCGTGC
                ************************************************************ * **

pTAK47          AGGCCGAGGTCGACGAGGCGGCGGCCAAGAAGACGTCCGAGGCGCAGATCAAGATCACGC
24GC17 (pTAK53) AGGCCGAGGTCGACGAGGCGGCGGCCAAGAAGACGTCCGAGGCGCAGATCAAGATCACGC
                ************************************************************

pTAK47          CCAAGGGCGGCTCGGACAACGCCTCCATCAACAACTCCGCCGGCGTCACCGTGAGCAAGG
24GC17 (pTAK53) CCAAGGGCGGCTCGGACAACGCCTCCATCAACAACTCCGCCGGCGTCACCGTGAGCAAGG
                ************************************************************

pTAK47          GCACGCTCACCGAGGTGAAGATGACCGCCTCGGACGGCACCGAGGTCAAGGGCGAGATAT
24GC17 (pTAK53) GCACGCTCACCGAGGTGAAGATGACCGCCTCGGACGGCACCGAGGTCAAGGGCGAGATAT
                ************************************************************

pTAK47          CCGCCGACAAGACCAGCTGGAAGCCCGGCGGTCAGCTGGAGCGCGCCACCAAGTACCAGA
24GC17 (pTAK53) CCGCCGACAAGACCAGCTGGAAGCCCGGCGGTCAGCTGGAGCGCGCCACCAAGTACCAGA
                ************************************************************

pTAK47          TCACGGCGACCGCGCAGGACTCCGAGGGCCGCGCCGCCCACGAGAACGCGTCGTTCACCA
24GC17 (pTAK53) TCACGGCGACCGCGCAGGACTCCGAGGGCCGCGCCGCCCACGAGAACGCGTCGTTCACCA
                ************************************************************

pTAK47          CGGTCTCCCCCGAGAACAGCTTCATCGGGACCTTCACCCCGGACGACGGCAAGACCGTCG
24GC17 (pTAK53) CGGTCTCCCCCGAGAACAGCTTCATCGGGACCTTCACCCCGGACGACGGCAAGACCGTCG
                ************************************************************

pTAK47          GCGTCGGGATGCCCGTGTCGATCAGCTTCGACAAGCAGATCACCAACAAGGCCGCGGTCC
24GC17 (pTAK53) GCGTCGGGATGCCCGTGTC---------------------------------------------
                ******************
```

Abbreviations

Abbreviation	Description
~	approximately
°C	degree Celsius
µg	microgram
µL	microliter
Asn	asparagine
CID	collision induced dissociation
Con A	Concanavalin A
Con A-HRP	Concanavalin A from *Canavalia ensiformis* (Jack bean) peroxidase conjugate
CV	column volume
ddH_2O	double distilled water
DNA	Deoxyribonucleic acid
DNB	Difco nutrient broth
Dol	Dolichol
ER	endoplasmic reticulum
ETD	electron transfer dissociation
eV	electron volt
x g	times gravity
g	gram
GalNAc	N-acetylgalactosamine
GDP	Guanosine diphosphate
GlcNAc	N-acetylglucosamine
h	hour
Hex	hexose
kPsi	kilopound per square inch, Pressure
LB	Luria-Burtani mediuma
LC-ESI-MS/MS	Liquid chromatography coupled to electrospray ionisation tandem mass spectrometry
LC-MS	Liquid chromatography coupled to mass spectrometry
MALDI-FT-ICR-MS	Matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance tandem mass spectrometry
MALDI-TOF-MS	Matrix-assisted laser desorption/ionization time of flight mass spectrometry
Man	mannose
MeOH	methanol
mg	milligram
min	minute
mL	millilitre

Abbreviation	Description
MSA	mannitol soya flour agar
MurNAc	N-acetylmuramic acid
MW	molecular weight
NEB	New England Biolabs
N-linked	asparagine linked
nm	nanometre
O-linked	serine/threonine linked
OST	oligosaccharyl transferase
PCR	polymerase chain reaction
pmt	protein O-mannosyl transferase (gene)
Pmt	protein O-mannosyl transferase (protein)
PP	polyprenol phosphate
ppm1	polyprenol phosphate mannose synthase (gene)
Ppm1	polyprenol phosphate mannose synthase (protein)
rpm	revolutions per minute
RT	room temperature
s	second
SDS-PAGE	Sodium dodecyl sulfate Polyacrylamide gel electrophoresis
Ser	serine
TFA	trifluoroacetic acid
Thr	threonine
UV	Ultraviolet

References

Aas, F., Å. Vik, J. Vedde, M. Koomey, and W. Egge-Jacobsen. 2007. '*Neisseria gonorrhoeae* O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure', *Mol Microbiol*, 65: 607-24.

Alaimo, C., I. Catrein, L. Morf, C. Marolda, N. Callewaert, M. Valvano, M. F. Feldman, and M. Aebi. 2006. 'Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides', *Embo j*, 25: 967-76.

Alemka, A., H. Nothaft, J. Zheng, and C. Szymanski. 2013. 'N-glycosylation of *Campylobacter jejuni* surface proteins promotes bacterial fitness', *Infect Immun*, 81: 1674-82.

Anttonen, K. P. 2010. 'Protein glycosylation in actinomycetes', University of Aberdeen.

Arroyo, J., J. Hutzler, C. Bermejo, E. Ragni, J. García-Cantalejo, P. Botías, H. Piberger, A. Schott, A. Sanz, and S. Strahl. 2011. 'Functional and genomic analyses of blocked protein O-mannosylation in baker's yeast', *Mol Microbiol*, 79: 1529-46.

Bagos, P. G., K. Tsirigos, T. Liakopoulos, and S. Hamodrakas. 2008. 'Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model', *J Proteome Res*, 7: 5082-93.

Barna, J. C. J., and D. H. Williams. 1984. 'The structure and mode of action of glycopeptide antibiotics of the vancomycin group', *Annual Reviews in Microbiology*, 38: 339-57.

Barresi, R., and K. Campbell. 2006. 'Dystroglycan: from biosynthesis to pathogenesis of human disease', *Journal of cell science*, 119: 199-207.

Barreteau, H., A. Kovač, A. Boniface, M. Sova, S. Gobec, and D. Blanot. 2008. 'Cytoplasmic steps of peptidoglycan biosynthesis', *FEMS microbiology reviews*, 32: 168-207.

Baulard, A. R., S. S. Gurcha, J. Engohang-Ndong, K. Gouffi, C. Locht, and G. S. Besra. 2003. 'In vivo interaction between the polyprenol phosphate mannose synthase Ppm1 and the integral membrane protein Ppm2 from *Mycobacterium smegmatis* revealed by a bacterial two-hybrid system', *J Biol Chem*, 278: 2242-8.

Bedford, D. J., C. Laity, and M. J. Buttner. 1995. 'Two genes involved in the phase-variable phi C31 resistance mechanism of *Streptomyces coelicolor* A3 (2)', *J Bacteriol*, 177: 4681-89.

Bendtsen, J. D., H. Nielsen, D. Widdick, T. Palmer, and S. Brunak. 2005. 'Prediction of twin-arginine signal peptides', *BMC Bioinformatics*, 6: 1-9.

Bentley, S., K. Chater, A. Cerdeno-Tarraga, G. L. Challis, N. Thomson, K. James, D. Harris, M. Quail, H Kieser, and D. Harper. 2002. 'Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2)', *Nature*, 417: 141-47.

Berks, B. C., T. Palmer, and F. Sargent. 2003. 'The Tat protein translocation pathway and its role in microbial physiology', *Adv Microb Physiol*, 47.

Bouhss, A., A. Trunkfield, T. Bugg, and D. Mengin-Lecreulx. 2008. 'The biosynthesis of peptidoglycan lipid-linked intermediates', *FEMS microbiology reviews*, 32: 208-33.

Bramkamp, M. 2010. 'The putative Bacillus subtilis L,D-transpeptidase YciB is a lipoprotein that localizes to the cell poles in a divisome-dependent manner', *Arch Microbiol*, 192: 57-68.

Brockhausen, I. 1999. 'Pathways of O-glycan biosynthesis in cancer cells', *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1473: 67-95.

Brockhausen, I., J. Yang, J. Burchell, C. Whitehouse, and J. Taylor-Papadimitriou. 1995. 'Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells', *European Journal of Biochemistry*, 233: 607-17.

Brodbelt, J. S. 2015. 'Ion activation methods for peptides and proteins', *Analytical chemistry*, 88: 30-51.

Carpita, N., and E. Shea. 1989. 'Linkage structure of carbohydrates by gas chromatography-mass spectrometry (GC-MS) of partially methylated alditol acetates', *Analysis of Carbohydrates by GLC and MS*: 157-216.

Castric, P., F. Cassels, and R. Carlson. 2001. 'Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan', *Journal of Biological Chemistry*, 276: 26479-85.

Chan, Y. G., H. K. Kim, O. Schneewind, and D. Missiakas. 2014. 'The capsular polysaccharide of *Staphylococcus aureus* is attached to peptidoglycan by the LytR-CpsA-Psr (LCP) family of enzymes', *Journal of Biological Chemistry*, 289: 15680-90.

Chatterjee, D., K. Lowell, B. Rivoire, M. R. McNeil, and P. J. Brennan. 1992. 'Lipoarabinomannan of *Mycobacterium tuberculosis*. Capping with mannosyl residues in some strains', *J Biol Chem*, 267: 6234-9.

Colussi, P., C. Taron, J. Mack, and P. Orlean. 1997. 'Human and *Saccharomyces cerevisiae* dolichol phosphate mannose synthases represent two classes of the enzyme, but both function in *Schizosaccharomyces pombe*', *Proceedings of the National Academy of Sciences*, 94: 7873-78.

Córdova-Dávalos, L., C. Espitia, G. González-Cerón, R. Arreguín-Espinosa, G. Soberón-Chávez, and L. Servín-González. 2014. 'Lipoprotein N-acyl transferase (Lnt1) is dispensable

for protein O-mannosylation by *Streptomyces coelicolor*', *FEMS microbiology letters*, 350: 72-82.

Cowlishaw, D. A., and M. C. Smith. 2001. 'Glycosylation of a *Streptomyces coelicolor* A3(2) cell envelope protein is required for infection by bacteriophage phi C31', *Mol Microbiol*, 41: 601-10.

Cowlishaw, D. A., and M. C. M. Smith. 2002. 'A gene encoding a homologue of dolichol phosphate-β-D-mannose synthase is required for infection of *Streptomyces coelicolor* A3 (2) by phage ϕC31', *J Bacteriol*, 184: 6081-83.

Crooks, G. E., G. Hon, J. Chandonia, and S. E. Brenner. 2004. 'WebLogo: a sequence logo generator', *Genome research*, 14: 1188-90.

Cummings, R. D. 1994. 'Use of lectins in analysis of glycoconjugates', *Methods in enzymology*, 230: 66-86.

De Groot, P., A. Ram, and F. Klis. 2005. 'Features and functions of covalently linked proteins in fungal cell walls', *Fungal Genetics and Biology*, 42: 657-75.

Dell, A., A. Galadari, F. Sastre, and P. Hitchen. 2011. 'Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes', *International journal of microbiology*, 2010.

Dell, A., J. Oates, C. Lugowski, E. Romanowska, L. Kenne, and B. Lindberg. 1984. 'The enterobacterial common-antigen, a cyclic polysaccharide', *Carbohydrate Research*, 133: 95-104.

Dobos, K. M., K. H. Khoo, K. M. Swiderek, P. J. Brennan, and J. T. Belisle. 1996. 'Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*', *J Bacteriol*, 178: 2498-506.

Dobos, K. M., K. Swiderek, K. H. Khoo, P. J. Brennan, and J. T. Belisle. 1995. 'Evidence for glycosylation sites on the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*', *Infect Immun*, 63: 2846-53.

Doering, T., W. Masterson, G. Hart, and P. Englund. 1990. 'Biosynthesis of glycosyl phosphatidylinositol membrane anchors', *J Biol Chem*, 265: 611-14.

Driessen, Arnold JM, and Nico Nouwen. 2008. 'Protein translocation across the bacterial cytoplasmic membrane', *Annu. Rev. Biochem.*, 77: 643-67.

Dubée, V., M. Arthur, H. Fief, S. Triboulet, J. L. Mainardi, L. Gutmann, M. Sollogoub, L. B. Rice, M. Ethève-Quelquejeu, and J. Hugonnet. 2012a. 'Kinetic analysis of *Enterococcus faecium* L, D-transpeptidase inactivation by carbapenems', *Antimicrob Agents Chemother*, 56: 3409-12.

Dubée, V., S. Triboulet, J. L. Mainardi, M. Ethève-Quelquejeu, L.. Gutmann, A. Marie, L. Dubost, J. Hugonnet, and M. Arthur. 2012b. 'Inactivation of *Mycobacterium tuberculosis* L, D-transpeptidase LdtMt1 by carbapenems and cephalosporins', *Antimicrob Agents Chemother*, 56: 4189-95.

Eberhardt, A. l., C. N. Hoyland, D. Vollmer, S. Bisle, R. M. Cleverley, O. Johnsborg, L. S. Håvarstein, R. J. Lewis, and W. Vollmer. 2012. 'Attachment of capsular polysaccharide to the cell wall in *Streptococcus pneumoniae*', *Microbial Drug Resistance*, 18: 240-55.

Eichler, J. 2013. 'Extreme sweetness: protein glycosylation in archaea', *Nature Reviews Microbiology*, 11: 151-56.

Elorza, M., G. Larriba, J. Villanueva, and R. Sentandreu. 1977. 'Biosynthesis of the yeast cell wall: Selective assays and regulation of some mannosyl transferase activities', *Antonie van Leeuwenhoek*, 43: 129-42.

Espitia, C., and R. Mancilla. 1989. 'Identification, isolation and partial characterization of *Mycobacterium tuberculosis* glycoprotein antigens', *Clinical and experimental immunology*, 77: 378.

Fang, X., K. Tiyanont, Y. Zhang, J. Wanner, D. Boger, and S. Walker. 2006. 'The mechanism of action of ramoplanin and enduracidin', *Molecular BioSystems*, 2: 69-76.

Fifis, T., C. Costopoulos, A. Radford, A. Bacic, and P. Wood. 1991. 'Purification and characterization of major antigens from a *Mycobacterium bovis* culture filtrate', *Infect Immun*, 59: 800-07.

Flärdh, K., and M. J. Buttner. 2009. 'Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium', *Nature Reviews Microbiology*, 7: 36-49.

Garbe, T., D. Harris, M. Vordermeier, R. Lathigra, J. Ivanyi, and D. Young. 1993. 'Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation', *Infect Immun*, 61: 260-67.

Gentzsch, M., and W. Tanner. 1996. 'The PMT gene family: protein O-glycosylation in *Saccharomyces cerevisiae* is vital', *Embo j*, 15: 5752.

Geyer, H., and R. Geyer. 2006. 'Strategies for analysis of glycoprotein glycosylation', *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1764: 1853-69.

Girrbach, V., T. Zeller, M.e Priesmeier, and S. Strahl-Bolsinger. 2000. 'Structure-function analysis of the dolichyl phosphate-mannose: protein O-mannosyltransferase ScPmt1p', *Journal of Biological Chemistry*, 275: 19288-96.

Gonzalez-Zamorano, M., G. Mendoza-Hernandez, W. Xolalpa, C. Parada, A. J. Vallecillo, F. Bigi, and C. Espitia. 2009. '*Mycobacterium tuberculosis* glycoproteomics based on ConA-lectin affinity capture of mannosylated proteins', *J Proteome Res*, 8: 721-33.

Goon, S., J. Kelly, S. Logan, C. Ewing, and P. Guerry. 2003. 'Pseudaminic acid, the major modification on *Campylobacter flagellin*, is synthesized via the Cj1293 gene', *Mol Microbiol*, 50: 659-71.

Gregory, M. A., R. Till, and M. C. Smith. 2003. 'Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors', *J Bacteriol*, 185: 5320-3.

Guérardel, Y., E. Maes, E. Elass, Y. Leroy, P. Timmerman, G. S. Besra, C. Locht, G. Strecker, and L. Kremer. 2002. 'Structural Study of Lipomannan and Lipoarabinomannan from *Mycobacterium chelonae* presence of unusual components with α1, 3-mannopyrannose side chains', *Journal of Biological Chemistry*, 277: 30635-48.

Guo, M., L. Tan, X. Nie, X. Zhu, Y. Pan, and Z. Gao. 2016. 'The Pmt2p-Mediated Protein O-Mannosylation Is Required for Morphogenesis, Adhesive Properties, Cell Wall Integrity and Full Virulence of *Magnaporthe oryzae*', *Frontiers in microbiology*, 7.

Gurcha, S. S., A. Baulard, L. Kremer, C. Locht, D. Moody, W. Muhlecker, C. Costello, D. C. Crick, P. J. Brennan, and G. S. Besra. 2002. 'Ppm1, a novel polyprenol monophosphomannose synthase from *Mycobacterium tuberculosis*', *Biochem J*, 365: 441-50.

Gust, B., S. O'Rourke, N. Bird, T. Kieser, and K. Chater. 2003. 'Recombineering in *Streptomyces coelicolor*', *Norwich: The John Innes Foundation*.

Han, I., and J. Kudlow. 1997. 'Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility', *Molecular and cellular biology*, 17: 2550-58.

Hansen, Jan E, Ole Lund, Niels Tolstrup, Andrew A Gooley, Keith L Williams, and Søren Brunak. 1998. 'NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility', *Glycoconjugate journal*, 15: 115-30.

Harrison, J., G. Lloyd, M. Joe, T. L. Lowary, E. Reynolds, H. Walters-Morgan, A. Bhatt, A. Lovering, G. S. Besra, and L. J. Alderwick. 2016. 'Lcp1 Is a Phosphotransferase Responsible for Ligating Arabinogalactan to Peptidoglycan in *Mycobacterium tuberculosis*', *MBio*, 7.

Hart, G. 1997. 'Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins', *Annual review of biochemistry*, 66: 315-35.

Harty, C., S. Strahl, and K. Römisch. 2001. 'O-mannosylation protects mutant alpha-factor precursor from endoplasmic reticulum-associated degradation', *Molecular biology of the cell*, 12: 1093-101.

Haselbeck, A. 1989. 'Purification of GDP mannose: dolichyl-phosphate O-β-D-mannosyltransferase from *Saccharomyces cerevisiae*', *European Journal of Biochemistry*, 181: 663-68.

Haselbeck, A., and W. Tanner. 1983. 'O-Glycosylation in *Saccharomyces cerevisiae* is initiated at the endoplasmic reticulum', *FEBS letters*, 158: 335-38.

He, F. 2011. 'Coomassie Blue Staining', *Bio-protocol*, Bio101: e78. DOI: 10.21769/BioProtoc.78.

Hegge, F., P. Hitchen, F. Aas, H. Kristiansen, C. Løvold, W. Egge-Jacobsen, M. Panico, W. Leong, V. Bull, and M. Virji. 2004. 'Unique modifications with phosphocholine and phosphoethanolamine define alternate antigenic forms of *Neisseria gonorrhoeae* type IV pili', *Proc Natl Acad Sci U S A*, 101: 10798-803.

Heifetz, A., R. Keenan, and A. Elbein. 1979. 'Mechanism of action of tunicamycin on the UDP-GlcNAc: dolichyl-phosphate GlcNAc-1-phosphate transferase', *Biochemistry*, 18: 2186-92.

Herrmann, J. L., R. Delahay, A. Gallagher, B. Robertson, and D. Young. 2000. 'Analysis of post-translational modification of mycobacterial proteins using a cassette expression system', *FEBS letters*, 473: 358-62.

Herrmann, J. L., P. O'Gaora, A. Gallagher, J. E. Thole, and D. B. Young. 1996. 'Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*', *Embo j*, 15: 3547-54.

Hong, H. J., M. I. Hutchings, L. Hill, and M. J. Buttner. 2005. 'The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*', *Journal of Biological Chemistry*, 280: 13055-61.

Hong, H. J., M. I. Hutchings, J. M. Neu, G. D. Wright, M. S. Paget, and M. J. Buttner. 2004. 'Characterization of an inducible vancomycin resistance system in *Streptomyces coelicolor* reveals a novel gene (vanK) required for drug resistance', *Mol Microbiol*, 52: 1107-21.

Hopwood, D. 1999. 'Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico', *Microbiology*, 145: 2183-202.

Hopwood, D. A. 2007. *Streptomyces in nature and medicine: the antibiotic makers* (Oxford University Press).

Horn, C., A. Namane, P. Pescher, M. Riviere, F. Romain, G. Puzo, O. Bârzu, and G. Marchal. 1999. 'Decreased capacity of recombinant 45/47-kDa molecules (Apa) of *Mycobacterium tuberculosis* to stimulate T lymphocyte responses related to changes in their mannosylation pattern', *Journal of Biological Chemistry*, 274: 32023-30.

Houel, S., R. Abernathy, K. Renganathan, K. Meyer-Arendt, Natalie G. Ahn, and William M. Old. 2010. 'Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies', *J Proteome Res*, 9: 4152-60.

Howlett, R., N. Read, K. P. Anttonen, A. Varghese, Kershaw. C., Hancock. Y., and M. C. M. Smith. 2016. 'Intrinsic resistance to multiple antibiotics conferred through a protein-O-glycosylation pathway in *Streptomyces coelicolor.*', *Manuscript in preparation*.

Huang, T., and S. A. McLuckey. 2010. 'Gas-phase chemistry of multiply charged bioions in analytical mass spectrometry', *Annual review of analytical chemistry (Palo Alto, Calif.)*, 3: 365.

Huddleston, M. J., M. F. Bean, and S. A Carr. 1993. 'Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests', *Analytical chemistry*, 65: 877-84.

Hugonnet, J. E., N. Haddache, C. Veckerle, L. Dubost, A. Marie, N. Shikura, J. L. Mainardi, L. B. Rice, and M. Arthur. 2014. 'Peptidoglycan cross-linking in glycopeptide-resistant Actinomycetales', *Antimicrob Agents Chemother*, 58: 1749-56.

Hunter, S. W., and P. J. Brennan. 1990. 'Evidence for the presence of a phosphatidylinositol anchor on the lipoarabinomannan and lipomannan of *Mycobacterium tuberculosis*', *J Biol Chem*, 265: 9272-9.

Ichimiya, T., H. Manya, Y. Ohmae, H. Yoshida, K. Takahashi, R. Ueda, T. Endo, and S. Nishihara. 2004. 'The twisted abdomen phenotype of Drosophila POMT1 and POMT2 mutants coincides with their heterophilic protein O-mannosyltransferase activity', *Journal of Biological Chemistry*, 279: 42638-47.

Jennings, M., M. Virji, D. Evans, V. Foster, Y. Srikhanta, L. Steeghs, P. Van Der Ley, and E. Moxon. 1998. 'Identification of a novel gene involved in pilin glycosylation in *Neisseria meningitidis*', *Mol Microbiol*, 29: 975-84.

Jones, M., K. Marston, C. Woodall, D. Maskell, D. Linton, A. Karlyshev, N. Dorrell, B. Wren, and P. Barrow. 2004. 'Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract', *Infect Immun*, 72: 3769-76.

Juncker, A. S., H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen, and A. Krogh. 2003. 'Prediction of lipoprotein signal peptides in Gram-negative bacteria', *Protein Science*, 12: 1652-62.

Kang, S., I. Kim, Y. Rho, and K. J. Lee. 1995. 'Production dynamics of extracellular proteases accompanying morphological differentiation of *Streptomyces albidoflavus* SMF301', *Microbiology*, 141: 3095-103.

Kärkkäinen, J. 1970. 'Analysis of disaccharides as permethylated disaccharide alditols by gas—liquid chromatography—mass spectrometry', *Carbohydrate Research*, 14: 27-33.

Kärkkäinen, J.. 1971. 'Structural analysis of trisaccharides as permethylated trisaccharide alditols by gas-liquid chromatography-mass spectrometry', *Carbohydrate Research*, 17: 11-18.

Karlyshev, A. , P. Everest, D. Linton, S. Cawthraw, D. Newell, and B. Wren. 2004. 'The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks', *Microbiology*, 150: 1957-64.

Kawai, Y., J. Marles-Wright, R. M. Cleverley, R. Emmins, S. Ishikawa, M. Kuwano, N. Heinz, N. K. Bui, C. N. Hoyland, and N. Ogasawara. 2011. 'A widespread family of bacterial cell wall assembly proteins', *Embo j*, 30: 4931-41.

Kearse, K., and G. Hart. 1991. 'Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins', *Proceedings of the National Academy of Sciences*, 88: 1701-05.

Kieser, T., M. J. Bibb, M. J. Buttner, K. Chater, and D. Hopwood. 2000. *Practical Streptomyces Genetics* (The John Innes Foundation).

Kim, D. W., A. Hesketh, E. S. Kim, J. Song, D. Lee, I. Kim, K. Chater, and K. J. Lee. 2008. 'Complex extracellular interactions of proteases and a protease inhibitor influence multicellular development of *Streptomyces coelicolor*', *Mol Microbiol*, 70: 1180-93.

Kim, H. K., P. Brennan, D. Heaslip, M. Udey, R. Modlin, and J. T. Belisle. 2015. 'Carbohydrate-dependent binding of langerin to SodC, a cell wall glycoprotein of *Mycobacterium leprae*', *J Bacteriol*, 197: 615-25.

Kim, I., and K. J. Lee. 1995. 'Physiological roles of leupeptin and extracellular proteases in mycelium development of *Streptomyces exfoliatus* SMF13', *Microbiology*, 141: 1017-25.

Kluepfel, D., S. Vats-Mehta, F. Aumont, F. Shareck, and R. Morosoli. 1990. 'Purification and characterization of a new xylanase (xylanase B) produced by *Streptomyces lividans* 66', *Biochem J*, 267: 45-50.

Kornfeld, R., and S. Kornfeld. 1985. 'Assembly of asparagine-linked oligosaccharides', *Annual review of biochemistry*, 54: 631-64.

Kowarik, M., N. Young, S. Numao, B. Schulz, I. Hug, N. Callewaert, D. Mills, D. Watson, M. Hernandez, and J. Kelly. 2006. 'Definition of the bacterial N-glycosylation site consensus sequence', *Embo j*, 25: 1957-66.

Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. 'Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes', *J Mol Biol*, 305: 567-80.

Ku, S., B. Schulz, P. Power, and M. Jennings. 2009. 'The pilin O-glycosylation pathway of pathogenic Neisseria is a general system that glycosylates AniA, an outer membrane nitrite reductase', *Biochemical and biophysical research communications*, 378: 84-89.

Kurien, B., and R. Scofield. 2006. 'Western blotting', *Methods*, 38: 283-93.

Lairson, L., B. Henrissat, G. J. Davies, and S. G. Withers. 2008. 'Glycosyltransferases: structures, functions, and mechanisms', *Biochemistry*, 77: 521.

Larriba, G., M. Elorza, J. Villanueva, and R. Sentandreu. 1976. 'Participation of dolichol phospho-mannose in the glycosylation of yeast wall manno-proteins at the polysomal level', *FEBS letters*, 71: 316-20.

Levin, D. E. 2005. 'Cell wall integrity signaling in *Saccharomyces cerevisiae*', *Microbiology and Molecular Biology Reviews*, 69: 262-91.

Linton, D., N. Dorrell, P. Hitchen, S. Amber, A. Karlyshev, H. Morris, A. Dell, M. Valvano, M. Aebi, and B. Wren. 2005. 'Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway', *Mol Microbiol*, 55: 1695-703.

Liu, C. F., L. Tonini, W. Malaga, M. Beau, A. Stella, D. Bouyssie, M. C. Jackson, J. Nigou, G. Puzo, C. Guilhot, O. Burlet-Schiltz, and M. Riviere. 2013a. 'Bacterial protein-O-mannosylating enzyme is crucial for virulence of *Mycobacterium tuberculosis*', *Proc Natl Acad Sci U S A*, 110: 6560-5.

Liu, G., K. Chater, G. Chandra, G. Niu, and H. Tan. 2013b. 'Molecular regulation of antibiotic biosynthesis in Streptomyces', *Microbiology and Molecular Biology Reviews*, 77: 112-43.

Logan, S. 2006. 'Flagellar glycosylation–a new component of the motility repertoire?', *Microbiology*, 152: 1249-62.

Loibl, M., and S. Strahl. 2013. 'Protein O-mannosylation: what we have learned from baker's yeast', *Biochim Biophys Acta*, 1833: 2438-46.

Lommel, M., A. Schott, T. Jank, V. Hofmann, and S. Strahl. 2011. 'A conserved acidic motif is crucial for enzymatic activity of protein O-mannosyltransferases', *Journal of Biological Chemistry*, 286: 39768-75.

Lommel, M., and S. Strahl. 2009. 'Protein O-mannosylation: conserved from bacteria to humans', *Glycobiology*, 19: 816-28.

Lussier, M., A. Sdicu, and H. Bussey. 1999. 'The KTR and MNN1 mannosyltransferase families of *Saccharomyces cerevisiae*', *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1426: 323-34.

Maeda, Y., and T. Kinoshita. 2008. 'Dolichol-phosphate mannose synthase: structure, function and regulation', *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1780: 861-68.

Magnet, S., S. Bellais, L. Dubost, M. Fourgeaud, J. L. Mainardi, S. Petit-Frère, A. Marie, D. Mengin-Lecreulx, M. Arthur, and L. Gutmann. 2007. 'Identification of the L, D-transpeptidases responsible for attachment of the Braun lipoprotein to *Escherichia coli* peptidoglycan', *J Bacteriol*, 189: 3927-31.

Mahne, M., A. Tauch, A. Puhler, and J. Kalinowski. 2006. 'The *Corynebacterium glutamicum* gene pmt encoding a glycosyltransferase related to eukaryotic protein-O-mannosyltransferases is essential for glycosylation of the resuscitation promoting factor (Rpf2) and other secreted proteins', *FEMS Microbiol Lett*, 259: 226-33.

Mainardi, J. L., V. Morel, M. Fourgeaud, J. Cremniter, D. Blanot, R. Legrand, C. Fréhel, M. Arthur, J. van Heijenoort, and L. Gutmann. 2002. 'Balance between two transpeptidation mechanisms determines the expression of β-lactam resistance in *Enterococcus faecium*', *Journal of Biological Chemistry*, 277: 35801-07.

Marchler-Bauer, A., and S. Bryant. 2004. 'CD-Search: protein domain annotations on the fly', *Nucleic acids research*, 32: W327-W31.

Marchler-Bauer, A., M. Derbyshire, N. Gonzales, S. Lu, F. Chitsaz, L. Geer, R. Geer, J. He, M. Gwadz, and D. Hurwitz. 2014. 'CDD: NCBI's conserved domain database', *Nucleic acids research*: gku1221.

Martín-Blanco, E., and A. García-Bellido. 1996. 'Mutations in the rotated abdomen locus affect muscle development and reveal an intrinsic asymmetry in Drosophila', *Proceedings of the National Academy of Sciences*, 93: 6048-52.

Mescher, M. F, and J. L Strominger. 1976. 'Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*', *Journal of Biological Chemistry*, 251: 2005-14.

Michell, S., A. Whelan, P. Wheeler, M. Panico, R. Easton, A. Etienne, S. Haslam, A. Dell, H. Morris, and A Reason. 2003. 'The MPB83 antigen from *Mycobacterium bovis* contains O-Linked mannose and (1→ 3)-mannobiose moieties', *Journal of Biological Chemistry*, 278: 16423-32.

Morita, Y., C. Sena, R. Waller, K. Kurokawa, M. Sernee, F. Nakatani, R. Haites, H. Billman-Jacobe, M. McConville, and Y. Maeda. 2006. 'PimE is a polyprenol-phosphate-mannose-dependent mannosyltransferase that transfers the fifth mannose of phosphatidylinositol mannoside in mycobacteria', *Journal of Biological Chemistry*, 281: 25143-55.

Mouyna, I., O. Kniemeyer, T. Jank, C. Loussert, E. Mellado, V. Aimanianda, A. Beauvais, D. Wartenberg, J. Sarfati, and J. Bayry. 2010. 'Members of protein O-mannosyltransferase family in *Aspergillus fumigatus* differentially affect growth, morphogenesis and viability', *Mol Microbiol*, 76: 1205-21.

Muntoni, F., and T. Voit. 2004. 'The congenital muscular dystrophies in 2004: a century of exciting progress', *Neuromuscular Disorders*, 14: 635-49.

Nakatsukasa, K., S. Okada, K. Umebayashi, R. Fukuda, S. Nishikawa, and T. Endo. 2004. 'Roles of O-mannosylation of aberrant proteins in reduction of the load for endoplasmic reticulum chaperones in yeast', *Journal of Biological Chemistry*, 279: 49762-72.

Nieselt, K., F. Battke, A. Herbig, P. Bruheim, A. Wentzel, O. M. Jakobsen, H. Sletta, M. T. Alam, M. E. Merlo, J. Moore, W. A. Omara, E. R. Morrissey, M. A. Juarez-Hermosillo, A. Rodriguez-Garcia, M. Nentwich, L. Thomas, M. Iqbal, R. Legaie, W. H. Gaze, G. L. Challis, R. C. Jansen, L. Dijkhuizen, D. A. Rand, D. L. Wild, M. Bonin, J. Reuther, W. Wohlleben, M. C. Smith, N. J. Burroughs, J. F. Martin, D. A. Hodgson, E. Takano, R. Breitling, T. E. Ellingsen, and E. M. Wellington. 2010. 'The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*', *BMC Genomics*, 11: 10.

Nothaft, H., and C. Szymanski. 2010. 'Protein glycosylation in bacteria: sweeter than ever', *Nature Reviews Microbiology*, 8: 765-78.

Ogawara, H. 2015. 'Penicillin-binding proteins in Actinobacteria', *The Journal of antibiotics*, 68: 223-45.

Olden, K., J. Parent, and S. White. 1982. 'Carbohydrate moieties of glycoproteins a re-evaluation of their function', *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes*, 650: 209-32.

Olsen, J., B. Macek, O. Lange, A. Makarov, S. Horning, and M. Mann. 2007. 'Higher-energy C-trap dissociation for peptide modification analysis', *Nature methods*, 4: 709-12.

Ong, E., D. Kilburn, R. Miller, and R. Warren. 1994. '*Streptomyces lividans* glycosylates the linker region of a beta-1, 4-glycanase from *Cellulomonas fimi*', *J Bacteriol*, 176: 999-1008.

Ongay, S., A. Boichenko, N. Govorukhina, and R. Bischoff. 2012. 'Glycopeptide enrichment and separation for protein glycosylation analysis', *Journal of separation science*, 35: 2341-72.

Orlean, P. 1990. 'Dolichol phosphate mannose synthase is required in vivo for glycosyl phosphatidylinositol membrane anchoring, O-mannosylation, and N-glycosylation of protein in *Saccharomyces cerevisiae*', *Molecular and cellular biology*, 10: 5796-805.

Orlean, P., C. Albright, and P. Robbins. 1988. 'Cloning and sequencing of the yeast gene for dolichol phosphate mannose synthase, an essential protein', *Journal of Biological Chemistry*, 263: 17499-507.

Paget, M. S., L. Chamberlin, A. Atrih, S. Foster, and M. J. Buttner. 1999. 'Evidence that the extracytoplasmic function sigma factor çE is required for normal cell wall structure in *Streptomyces coelicolor* A3 (2)', *J Bacteriol*, 181: 204-11.

Petersen, T., S. Brunak, G. von Heijne, and H. Nielsen. 2011. 'SignalP 4.0: discriminating signal peptides from transmembrane regions', *Nat Meth*, 8: 785-86.

Piddington, D., F. Fang, T. Laessig, A. Cooper, I. Orme, and N. Buchmeier. 2001. 'Cu, Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst', *Infect Immun*, 69: 4980-87.

Power, P., L. Roddam, M. Dieckelmann, Y. Srikhanta, Y. Tan, A. Berrington, and M. Jennings. 2000. 'Genetic characterization of pilin glycosylation in *Neisseria meningitidis*', *Microbiology*, 146: 967-79.

Power, P., K. Seib, and M. Jennings. 2006. 'Pilin glycosylation in *Neisseria meningitidis* occurs by a similar pathway to wzy-dependent O-antigen biosynthesis in *Escherichia coli*', *Biochemical and biophysical research communications*, 347: 904-08.

Pratt, R. F. 2008. 'Substrate specificity of bacterial DD-peptidases (penicillin-binding proteins)', *Cellular and Molecular Life Sciences*, 65: 2138-55.

Pridham, T., and D. Gottlieb. 1948. 'The utilization of carbon compounds by some Actinomycetales as an aid for species determination', *J Bacteriol*, 56: 107.

Prill, S., B. Klinkert, C. Timpel, C. Gale, K. Schröppel, and J. Ernst. 2005. 'PMT family of *Candida albicans*: five protein mannosyltransferase isoforms affect growth, morphogenesis and antifungal resistance', *Mol Microbiol*, 55: 546-60.

Rabilloud, T. 2016. 'A single step protein assay that is both detergent and reducer compatible: The cydex blue assay', *Electrophoresis*, 37: 2595-601.

Rademaker, G. J., S. A. Pergantis, L. Blok-Tip, J. I. Langridge, A. Kleen, and J. E. Thomas-Oates. 1998. 'Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity', *Analytical biochemistry*, 257: 149-60.

Ragas, A., L. Roussel, G. Puzo, and M. Rivière. 2007. 'The *Mycobacterium tuberculosis* cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary C-type lectin surfactant protein A', *Journal of Biological Chemistry*, 282: 5133-42.

Reynolds, P. 1989. 'Structure, biochemistry and mechanism of action of glycopeptide antibiotics', *European Journal of Clinical Microbiology and Infectious Diseases*, 8: 943-50.

Roepstorff, P, and J Fohlman. 1984. 'Letter to the editors', *Biological Mass Spectrometry*, 11: 601-01.

Romain, F., C. Horn, P. Pescher, A. Namane, M. Riviere, G. Puzo, O. Barzu, and G. Marchal. 1999. 'Deglycosylation of the 45/47-kilodalton antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit in vivo or in vitro cellular immune responses', *Infect Immun*, 67: 5567-72.

Romano, A., and W. Nickerson. 1958. 'Utilization of amino acids as carbon sources by *Streptomyces fradiae*', *J Bacteriol*, 75: 161.

Salton, M., and J. Ghuysen. 1960. 'Acetylhexosamine compounds enzymically released from *Micrococcus lysodeikticus* cell walls: III. The structure of di-and tetra-saccharides released from cell walls by lysozyme and Streptomyces F1 enzyme', *Biochim Biophys Acta*, 45: 355-63.

Sandbrook, J., E. F. Fritsch, and T. Maniatis. 1989. "Molecular cloning: a laboratory manual." In.: Cold Spring Harbor Lab. Press, Plainview, NY.

Sanders, A., L. Wright, and M. Pavelka Jr. 2014. 'Genetic characterization of mycobacterial L, D-transpeptidases', *Microbiology*, 160: 1795-806.

Sartain, M. J., and J. T. Belisle. 2009. 'N-Terminal clustering of the O-glycosylation sites in the *Mycobacterium tuberculosis* lipoprotein SodC', *Glycobiology*, 19: 38-51.

Sartain, M. J., R. Slayden, K. Singh, S. Laal, and J. T. Belisle. 2006. 'Disease state differentiation and identification of tuberculosis biomarkers via native antigen array profiling', *Molecular & Cellular Proteomics*, 5: 2102-13.

Sauvage, E., F. Kerff, M. Terrak, J. Ayala, and P. Charlier. 2008. 'The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis', *FEMS microbiology reviews*, 32: 234-58.

Savitski, M. M., S. Lemeer, M. Boesche, M. Lang, T. Mathieson, M. Bantscheff, and B. Kuster. 2011. 'Confident Phosphorylation Site Localization Using the Mascot Delta Score', *Mol Cell Proteomics*, 10.

Schirm, M., M. Kalmokoff, A. Aubry, P. Thibault, M. Sandoz, and S. Logan. 2004. 'Flagellin from *Listeria monocytogenes* is glycosylated with β-O-linked N-acetylglucosamine', *J Bacteriol*, 186: 6721-27.

Schirm, M., E. Soo, A. Aubry, J. Austin, P. Thibault, and S. Logan. 2003. 'Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*', *Mol Microbiol*, 48: 1579-92.

Schoenhofen, I., E. Vinogradov, D. Whitfield, J. Brisson, and S. Logan. 2009. 'The CMP-legionaminic acid pathway in Campylobacter: biosynthesis involving novel GDP-linked precursors', *Glycobiology*, 19: 715-25.

Schoonmaker, M., W. R. Bishai, and G. Lamichhane. 2014. 'Nonclassical transpeptidases of *Mycobacterium tuberculosis* alter cell size, morphology, the cytosolic matrix, protein localization, virulence, and resistance to β-lactams', *J Bacteriol*, 196: 1394-402.

Seipke, R., M. Kaltenpoth, and M. I. Hutchings. 2012. 'Streptomyces as symbionts: an emerging and widespread theme?', *FEMS microbiology reviews*, 36: 862-76.

Sentandreu, R., and D.H. Northcote. 1968. 'The structure of a glycopeptide isolated from the yeast cell wall', *Biochem J*, 109: 419-32.

Sinclair, R. B., and M. J. Bibb. 1988. 'The repressor gene (c) of the *Streptomyces* temperate phage φc31: Nucleotide sequence, analysis and functional cloning', *Molecular and General Genetics MGG*, 213: 269-77.

Smedley, J., E. Jewell, J. Roguskie, J. Horzempa, A. Syboldt, D. Stolz, and P. Castric. 2005. 'Influence of pilin glycosylation on *Pseudomonas aeruginosa* 1244 pilus function', *Infect Immun*, 73: 7922-31.

Smith, G. T., M. J. Sweredoski, and S. Hess. 2014. 'O-linked glycosylation sites profiling in *Mycobacterium tuberculosis* culture filtrate proteins', *J Proteomics*, 97: 296-306.

Spiro, R. 2002. 'Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds', *Glycobiology*, 12: 43R-56R.

Spiro, R. G. 1973. 'Glycoproteins', *Advances in protein chemistry*, 27: 349-467.

Stimson, E., M. Virji, K. Makepeace, A. Dell, H. Morris, G. Payne, J. Saunders, M. Jennings, S. Barker, and M. Panico. 1995. 'Meningococcal pilin: a glycoprotein substituted with digalactosyl 2, 4-diacetamido-2, 4, 6-trideoxyhexose', *Mol Microbiol*, 17: 1201-14.

Stone, K. J., and J. L. Strominger. 1971. 'Mechanism of action of bacitracin: complexation with metal ion and C55-isoprenyl pyrophosphate', *Proceedings of the National Academy of Sciences*, 68: 3223-27.

Strahl-Bolsinger, S., M. Gentzsch, and W. Tanner. 1999. 'Protein O-mannosylation', *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1426: 297-307.

Strahl-Bolsinger, S., T. Immervoll, R. Deutzmann, and W. Tanner. 1993. 'PMT1, the gene for a key enzyme of protein O-glycosylation in *Saccharomyces cerevisiae*', *Proceedings of the National Academy of Sciences*, 90: 8164-68.

Strahl-Bolsinger, S., and A. Scheinost. 1999. 'Transmembrane topology of Pmt1p, a member of an evolutionarily conserved family of protein O-mannosyltransferases', *Journal of Biological Chemistry*, 274: 9068-75.

Strahl-Bolsinger, S., and W. Tanner. 1991. 'Protein O-glycosylation in *Saccharomyces cerevisiae* Purification and characterization of the dolichyl-phosphate-D-mannose-protein O-D-mannosyltransferase', *European Journal of Biochemistry*, 196: 185-90.

Sutcliffe, I. C., and D. J. Harrington. 2004. 'Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components', *FEMS microbiology reviews*, 28: 645-59.

Swain, S., T. Tseng, T. Thornton, M. Gopalraj, and N. Olszewski. 2002. 'SPINDLY is a nuclear-localized repressor of gibberellin signal transduction expressed throughout the plant', *Plant physiology*, 129: 605-15.

Szymanski, C., R. Yao, C. Ewing, T. Trust, and P. Guerry. 1999. 'Evidence for a system of general protein glycosylation in *Campylobacter jejuni*', *Mol Microbiol*, 32: 1022-30.

Taylor, A., O. Holst, and J. Thomas-Oates. 2006. 'Mass spectrometric profiling of O-linked glycans released directly from glycoproteins in gels using in-gel reductive β-elimination', *Proteomics*, 6: 2936-46.

Thibault, P., S. Logan, J. Kelly, J. Brisson, C. Ewing, and P. Guerry. 2001. 'Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin', *Journal of Biological Chemistry*, 276: 34862-70.

Thomas, L., D. A. Hodgson, A. Wentzel, K. Nieselt, T. E. Ellingsen, J. Moore, E. R. Morrissey, R. Legaie, W. Wohlleben, A. Rodriguez-Garcia, J. F. Martin, N. J. Burroughs, E. M. Wellington, and M. C. Smith. 2012. 'Metabolic switches and adaptations deduced from the proteomes of *Streptomyces coelicolor* wild type and phoP mutant grown in batch culture', *Mol Cell Proteomics*, 11: M111.013797.

Thompson, B., D. Widdick, M. Hicks, G. Chandra, I. C. Sutcliffe, T. Palmer, and M. I. Hutchings. 2010. 'Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium *Streptomyces coelicolor*', *Mol Microbiol*, 77: 943-57.

Tipper, D., and J. L Strominger. 1965. 'Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine', *Proceedings of the National Academy of Sciences*, 54: 1133-41.

Tissot, B., S. North, A. Ceroni, P. Pang, M. Panico, F. Rosati, A. Capone, S. Haslam, A. Dell, and H. Morris. 2009. 'Glycoproteomics: past, present and future', *FEBS letters*, 583: 1728-35.

Tomita, S., N. Inoue, Y. Maeda, K. Ohishi, J. Takeda, and T. Kinoshita. 1998. 'A homologue of *Saccharomyces cerevisiae* Dpm1p is not sufficient for synthesis of dolichol-phosphate-mannose in mammalian cells', *Journal of Biological Chemistry*, 273: 9249-54.

Torres, C., and G. Hart. 1984. 'Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc', *Journal of Biological Chemistry*, 259: 3308-17.

Twine, S., C. Paul, E. Vinogradov, D. McNally, J. Brisson, J. Mullen, D. McMullin, H. Jarrell, J. Austin, and J. Kelly. 2008. 'Flagellar glycosylation in *Clostridium botulinum*', *Febs Journal*, 275: 4428-44.

Typas, A., M. Banzhaf, C. Gross, and W. Vollmer. 2012. 'From the regulation of peptidoglycan synthesis to bacterial growth and morphology', *Nature Reviews Microbiology*, 10: 123-36.

Van Sorge, N., N. Bleumink, S. Van Vliet, E. Saeland, W. Van Der Pol, Y. Van Kooyk, and J. Van Putten. 2009. 'N-glycosylated proteins and distinct lipooligosaccharide glycoforms of *Campylobacter jejuni* target the human C-type lectin receptor MGL', *Cellular microbiology*, 11: 1768-81.

VanderVen, B. C., J. D. Harder, D. C. Crick, and J. T. Belisle. 2005. 'Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways', *Science*, 309: 941-3.

Varghese, A. S. 2008. 'Investigation of *Streptomyces coelicolor* A3 (2) glycosylation mutants', University of Aberdeen.

Varki, A., R. D. Cummings, J. Esko, H. Freeze, P. Stanley, C. Bertozzi, G. Hart, and M. Etzler. 2009. 'Essentials of Glycobiology. 2nd edition', *Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press*.

Vik, Å., F. Aas, J. Anonsen, S. Bilsborough, A. Schneider, W. Egge-Jacobsen, and M. Koomey. 2009. 'Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*', *Proceedings of the National Academy of Sciences*, 106: 4447-52.

Vollmer, W., D. Blanot, and M. De Pedro. 2008. 'Peptidoglycan structure and architecture', *FEMS microbiology reviews*, 32: 149-67.

Wacker, M., D. Linton, P. Hitchen, M. Nita-Lazar, S. Haslam, S. North, M. Panico, H. Morris, A. Dell, B. Wren, and M. Aebi. 2002. 'N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*', *Science*, 298: 1790-93.

Wada, Y., P. Azadi, C. Costello, A. Dell, R. Dwek, H. Geyer, R. Geyer, K. Kakehi, N. Karlsson, and K. Kato. 2007. 'Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study', *Glycobiology*, 17: 411-22.

Waxman, D. J., and J. L. Strominger. 1983. 'Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics1', *Annual review of biochemistry*, 52: 825-69.

Wehmeier, S. unpublished.

Wehmeier, S., A. S. Varghese, S. S. Gurcha, B. Tissot, M. Panico, P. Hitchen, H. R. Morris, G. S. Besra, A. Dell, and M. C. Smith. 2009. 'Glycosylation of the phosphate binding protein, PstS, in *Streptomyces coelicolor* by a pathway that resembles protein O-mannosylation in eukaryotes', *Mol Microbiol*, 71: 421-33.

Wells, J., and S. A. McLuckey. 2005. 'Collision-induced dissociation (CID) of peptides and proteins', *Methods in enzymology*, 402: 148-85.

Wells, L., K. Vosseller, and G. Hart. 2001. 'Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc', *Science*, 291: 2376-78.

Widdick, D., K. Dilks, G. Chandra, A. Bottrill, M. J. Naldrett, M. Pohlschröder, and T. Palmer. 2006. 'The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*', *Proceedings of the National Academy of Sciences*, 103: 17927-32.

Wietzerbin, J., B. Das, J. Petit, E. Lederer, M. Leyh-Bouille, and J. Ghuysen. 1974. 'Occurrence of D-alanyl-(D)-meso-diaminopimelic acid and meso-diaminopimelyl-meso-diaminopimelic acid interpeptide linkages in the peptidoglycan of Mycobacteria', *Biochemistry*, 13: 3471-76.

Willer, T., M. Brandl, M. Sipiczki, and S. Strahl. 2005. 'Protein O-mannosylation is crucial for cell wall integrity, septation and viability in fission yeast', *Mol Microbiol*, 57: 156-70.

Willer, T., B. Prados, J. Falcón-Pérez, I. Renner-Müller, G. Przemeck, M. Lommel, A. Coloma, M. Valero, M. de Angelis, and W. Tanner. 2004. 'Targeted disruption of the Walker–Warburg syndrome gene Pomt1 in mouse results in embryonic lethality', *Proc Natl Acad Sci U S A*, 101: 14126-31.

Wuhrer, M., M. I. Catalina, A. M. Deelder, and C. H. Hokke. 2007. 'Glycoproteomics based on tandem mass spectrometry of glycopeptides', *J Chromatogr B Analyt Technol Biomed Life Sci*, 849: 115-28.

Yankovskaya, V., R. Horsefield, S. Törnroth, C. Luna-Chavez, H. Miyoshi, C. Léger, B. Byrne, G. Cecchini, and S. Iwata. 2003. 'Architecture of succinate dehydrogenase and reactive oxygen species generation', *Science*, 299: 700-04.

Zebian, N., A. Merkx-Jacques, P. Pittock, S. Houle, C. Dozois, G. Lajoie, and C. Creuzenet. 2016. 'Comprehensive analysis of flagellin glycosylation in *Campylobacter jejuni* NCTC 11168 reveals incorporation of legionaminic acid and its importance for host colonization', *Glycobiology*, 26: 386-97.

Zimmerman, J., C. Specht, B. Cazares, and P. Robbins. 1996. 'The isolation of a Dol-P-Man synthase from *Ustilago maydis* that functions in *Saccharomyces cerevisiae*', *Yeast*, 12: 765-71.

Zimmermann, R. A., A. Muto, P. Fellner, C. Ehresmann, and C. Branlant. 1972. 'Location of ribosomal protein binding sites on 16S ribosomal RNA', *Proc Natl Acad Sci U S A*, 69: 1282-6.