INVESTIGATION INTO THE STRUCTURE AND FUNCTION OF THE GLYCOSYLATED BACTERIOCIN GccF AND THE GLYCOSYLTRANSFERASE GccA FROM *LACTOBACILLUS PLANTARUM* KW30

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Biological Sciences

at Massey University, Palmerston North

New Zealand

Evelyn Marianne Poulson

2011

ABSTRACT

Bacteriocins are typically antimicrobial proteins or peptides produced by Gram-positive and G–negative bacteria, that are capable of inhibiting the growth of other bacteria. Glycocin F (GccF) is a 43 amino acid bacteriocin produced by *Lactobacillus plantarum* KW30 which is post-translationally modified by two N-acetylglucosamine residues (GlcNAc). One of these residues is linked to a serine side-chain (*O*-linked), while the other is linked through the thiol sulphur at the C-terminal cysteine (*S*-linked). Within the gene cluster encoding GccF are a set of genes thought to be required for the maturation and secretion of GccF. The GccF gene cluster consists of six genes encoding a family 2 glycosyltransferase (GTase) thought to be responsible for the addition of either one or both of these GlcNAc groups, an ABC transporter involved in the secretion of the bacteriocin across the cellular membrane, two thioredoxin-like genes which may be responsible for the disulfide bonding pattern of GccF, a gene of unknown function, and GccF itself.

Within *L. plantarum* KW30 no other proteins modified by a GlcNAc residue were identified in the present study, making GccF the only known GlcNAcylated protein produced by this organism. Methods were developed to pull-down the proteins involved in the maturation and secretion of GccF, and to find its binding target(s) in strains suceptible to its activity. Although proteins were found to bind tightly to GccF during pull-down experiments, those that bound were mostly involved in glycolysis/gluconeogenesis which does not fit into the hypothesised mechanism of action for GccF. Fluorescent microscopy experiments on wild-type GccF and GccF that contained only the *O*-linked or *S*-linked GlcNAc residue found that localisation of the modified GlcNAcylated GccF on susceptible strains was different to what is seen for wild-type in that they appeared randomly along the cells, whereas wild-type GccF appeared to localise at the point of cell division and at the tips of the cells. These microscopy results show that the post-translational modifications appear to play a role in targeting of GccF to susceptible cells. Assays to detect and test the activity of the GTase found that it may be located within the cytosol of *L. plantarum* KW30 instead of the membrane which is where it was proposed to be due to the presence of a predicted transmembrane spanning region identified during bioinformatic analysis.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr Gill Norris and Dr Mark Patchett for all their support, guidance and encouragement throughout the duration of my thesis. Without their enthusiasm, positive attutidue, and willingness to answer my queries I would have been very lost.

I would like to thank everyone in the X-lab especially Trevor Loo and Natalie Burr, for their help, advice and good humour which made working so much fun.

I would also like to thank all staff in IMBS for their help and support during my time here.

Finally I would like to thank all of my family and friends for their continued support thoughout the duration of my studies. Special thanks to my mum and Rob for putting up with me when my emotions took the better of me, when I was stressing over finishing and when things just weren't working.

TABLE OF CONTENTS

LIST OF FIGURES

PART 1: INTRODUCTION

PART 2: GENERAL METHODS

PART 3: FUNCTIONAL CHARACTERIZATION OF THE *L. PLANTARUM* KW30 BACTERIOCIN, GCCF.

PART 4: STRUCTURAL AND FUNCTIONAL CHARACTERISATION OF THE *L. PLANTARUM* KW30 GLYCOSYLTRANSFERASE, GCCA.

LIST OF TABLES

ABBREVIATIONS

Å	Ångström (10^{-10} m)
aa	Amino acid
ABC transporter	ATP-binding cassette transporter
ACN	Acetonitrile
Amp	Ampicillin
Amu	Atomic mass unit
APS	Ammonium persulfate
ATCC	American type Culture Collection
ATP	Adenosine-5'-triphosphate
Bis-Tris	Bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
°C	Degrees Celcius
C11	The 11 C-terminal amino acids of GccF, unmodified.
C15	The 15 C-terminal amino acids of GccF, unmodified.
CAZy	Carbohydrate-Active enzyme database
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
Cm	Chloramphenicol
CV	Column Volume
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribose nucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
Efc	*Enterococcus faecalis*
et al.	*et alteri* (and others)
EtBr	Ethidium bromide
EtOH	Ethanol
FITC	Fluorescein isothiocyante
g	Gram; standard gravity ($9.81 m/s^2$)
GccF	Glycocin F
$GccF^{deFFL}$	Glycocin F without the *O*-linked *N*-acetylglucosamine residue
$GccF^{-HC}$	Glycocin F without the *S*-linked *N*-acetylglucosamine residue and the last two C-terminal amino acids
GlcNAc	*N*-acetylglucosamine
GlcNAcylated	Modified by an *N*-acetylglucosamine
GT2	Family 2 glycosyltransferase
GTase	Glycosyltransferase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HFBA	Heptafluorobutyric Acid
HPLC	High Pressure Liquid Chromatography

hrs	hours
IEF	Isoelectric Focusing
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl-β-D-Thiogalactopyranoside
Kan	Kanamycin
kb	kilobases
kDa	kilo daltons
L	Litre
LAB	Lactic Acid Bacteria
LB	Luria Bertani broth
M	Molar (mol/L)
m	Metre
min	Minute
mol	6.023×10^{23} molecules (Avogandro's constant)
MRS	de Man, Rogosa and Sharpe broth
MW	Molecular mass
m/z	Mass-to-charge ratio
NCBI	National Centre for Biotechnology Information
NH_2	Amine
Sulfo-NHS biotin	*N*-hydroxysulfosuccinimide biotin
NMR	Nuclear magnetic resonance
OD	Optical Density
o/n	Overnight
Pa	Pascal (= bar 10^{-5} = 145.04×10^{-6} psi)
PCR	Polymerase Chain Reaction
pH	Negative decadal logarithm of the proton concentration
PMSF	Phenylmethylsulfphonyl fluoride
ppm	parts per million
PTM	Post-translational modification
RNA	ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
RT	Room temperature
RP-HPLC	Reverse-Phase High Pressure Liquid Chromatography
RR	Response regulator
sec	Secretory system
s/sec	Second
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SOB	Super optimal broth
SOC	SOB with catabolite repression, indicative of the presence of glucose
sp.	Species

subsp.	Subspecies
T	Time
TAE	Tris-Acetate-EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TE	Tris-EDTA
TET	Tetracycline
TFA	Trifluoroacetic Acid
TMD	Transmembrane domain
TOF/TOF	Tandem time-of-flight
Tris	Tris(hydroxymethyl)aminomethane
trx	Thioredoxin
UDP	Uridine diphosphate
UDP-GlcNAc	Uridine diphosphate *N*-acetylglucosamine
USA	United States of America
UV	Ultraviolet light
V	Volts
v/v	Volume per volume
w/v	Weight per volume
w/w	Weight per weight
WGA	Wheatgerm agglutinin
WT	Wild-type
Yfr	*Yersinia frederiksenii*

PART 1:

BACTERIOCINS: AN INTRODUCTION

Introduction

Lactic acid bacteria (LAB) are gram-positive, acid-tolerant, microaerophilic rods, or cocci that contain a low-GC content, are generally non-sporulating, and can tolerate a lower pH range. They are defined for their ability to ferment hexose sugars to primarily produce lactic acid using either homo- or heterofermentation pathways. This ability to produce lactic acid *via* fermentation is advantageous in that it allows LAB to outcompete other organisms which have limited survival or do not grow in low pH environments. LAB occupy a wide range of environments including certain foods, decaying plant or animal matter, oral cavities, intestines, and the urogenital tract of both humans and animals. Along with the ability to produce lactic acid, many LAB are producers of small antimicrobial peptides or proteins termed bacteriocins. The first bacteriocin was discovered in 1925 and was called a colicine as it was found to kill *Escherichia coli* (Nes and Tagg, 1996). Since this discovery, bacteriocins have been the subject of intense study as their ability to be bacteriocidal against specific species is highly desired in both the food and medical industries. Nisin, a bacteriocin produced by *Lactococcus lactis* is the most widely studied bacteriocin as it has GRAS (generally recognised as safe) status, and is commonly used in a wide array of foods to control spoilage caused by a variety of microbes, including pathogens (Delves-Broughton *et al*, 1996). It was first used as a food preservative in 1951, and since its use, susceptible organisms have not been reported to have developed immunity to its activity, a trend seen with many other bacteriocins. The ability to maintain activity against susceptible species over a prolonged period of time without the development of immunity has sparked many studies into the antimicrobial activity of bacteriocins and in particular, into the mechanism of their production (Breukink and de Kruijff, 2006).

Bacteriocins

Bacteriocins are antimicrobial proteins or peptides which are produced by bacteria to kill or inhibit the growth of other bacterial organisms. They are thought to be produced by a majority of both bacteria and archaea (30-99%), and are biochemically and functionally diverse (Cotter *et al*, 2005). Bacteriocins can either act in a narrow spectrum where inhibition of growth is confined to organisms closely related to the producer strain, or a broad spectrum where the inhibitory

growth action is seen over a number of different species (Todorov, 2009; Cleveland *et al*, 2001). A bacteriocin-encoding gene is typically located in a gene cluster that can be either plasmid or chromosomally encoded, and contains all the genes required for its maturation and transport: the structural pre-bacteriocin gene, processing and transport genes (usually an ABC transporter), immunity genes, and a two-component signalling system (generally consisting of a histidine protein kinase and a response regulator) (Cleveland *et al*, 2001; Todorov, 2009). In some cases a gene encoding an accessory protein has been identified in gram-negative bacteria. This protein is required for bacteriocin translocation by the ABC transporter, and is thought to span the periplasm, facilitating the export through both the inner and outer membranes (Todorov, 2009).

The bacteriocin gene encodes a pre-bacteriocin precursor peptide containing an N-terminal leader sequence followed by the mature bacteriocin sequence. In some class II bacteriocins this N-terminal leader sequence is identifiable by a double-glycine sequence prior to the cleavage point between the N-terminal leader and the mature bacteriocin (Jack *et al*, 1995; Todorov, 2009). These leader sequences differ from typical signal secretion peptides that are involved in the direction of polypeptides into *sec*-dependent secretion pathways in both chemical struture and function. It is believed they are present as a preventative measure against the activity of the mature bacteriocin within its producer organism and also as a recognition signal for transport out of the cell, usually by an ABC transporter. It is also believed that the largely polar leader peptide may confer immunity to the producer by increasing the solubility of the prebacteriocin in water, causing its retention in the aqueous phase of the cell rather than promoting partition into the membrane. These leader peptides may also interact with mature bacteriocin peptides reducing their affinity for membranes. However, studies carried out on the pediocin AcH precursor produced by *Pediococcus acidilactici* LB42-923 have found that it is nearly as active as the mature species (~80%), indicating that the leader peptide has little effect upon the function of mature pediocin AcH (Ray *et al*, 1999). Therefore in some cases the leader sequence may not confer immunity but may only be involved in secretion.

Protection against the effects of the precursor and mature bacteriocin within the producer strain are often mediated by an immunity protein. The immunity gene is commonly located on the same operon next to the structural bacteriocin gene to ensure that immunity against the bacteriocin is present during its production. Immunity to bacteriocins can be either due to the effect of a single gene, such as *nis*I for nisin and *spa*I for subtilin which both encode single immunity proteins, or multiple genes, as in the case for some lantibiotics where immunity to their bacteriocins is the result of several proteins. Generally an immunity protein is between 50-150 amino acid residues long and is loosely associated with the producer organism's membrane. Though several immunity proteins have been identified for a number of different species, the mechanisms involved are still not well understood. It has been suggested that immunity proteins may interact with other protein receptors located on the cytoplasmic side of the producer's cell membrane thus protecting the host from its own bacteriocin activity.

Most bacteriocins are transported across the membrane of their producer strain by a dedicated ABC transporter system although several bacteriocins have been found to be secreted by the Sec-dependent system. Bacteriocin ABC transporters are composed of three domains contained within the same polypeptide chain. These are the cytoplasmic N-terminal domain (which may be proteolytic), a hydrophobic integral membrane domain, and a cytoplasmic C-terminal ATP-binding domain. Two polypeptide chains appear to be required for the function of the bacteriocin ABC transporter. For all bacteriocins that contain a double glycine leader sequence, their associated ABC transporters contain an N-terminal domain of approximately 150 amino acid residues that has proteoloytic activity specific for cleavage of the double glycine leader. Thus as the pre-bacteriocin is secreted, the leader sequence is cleaved to produce the mature bacteriocin. ABC transporters that secrete bacteriocins without a double glycine leader usually have a dedicated protease responsible for cleavage of the leader sequence.

Bacteriocin Classification

There has been much debate over the classification of bacteriocins. In the early nineties Klaenhammer proposed four major classes of lactic acid bacteria bacteriocins:

(i) Class I - lantibiotics

(ii) Class II – small, heat-stable proteins

(iii) Class III - large, heat-labile proteins

(iv) Class IV – complex bacteriocins

Class I bacteriocins are the lantibiotics, and consist of small peptides (19-38 amino acids long, <5kDa) which are heat stable. These peptides can be divided into two types, Type A which are cationic, elongated and pore forming, and Type B which are thought to be globular and more compact.

Class I bacteriocins such as Nisin contain serine, threonine, alanine and cysteine residues that have been unusually post translationally modified to form lanthionine and β-methyl lanthionine rings, and dehydrated amino acid residues (fig 1.1). The gene clusters for these lantibiotic bacteriocins contain genes which encode enzymes responsible for dehydration and cyclization, necessary for the formation of these rings. (Patton and van der Donk, 2005)

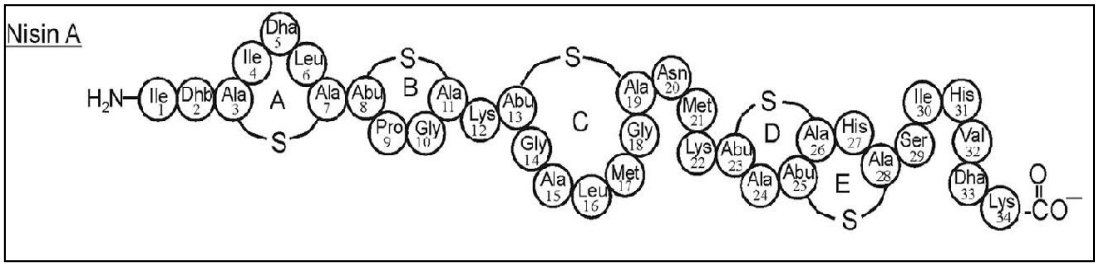


Figure 1.1 Schematic of Nisin A showing five lanthionine rings A-E (modified from Wilson-Standford *et al*, 2009).

Class II bacteriocins are small heat-stable often cationic bacteriocins that do not contain post translationally modified residues. This class can be further subdivided into three groups. Class IIa are pediocin-like anti-listerial bacteriocins, Class IIb are

two peptide bacteriocins where either one or both peptides are required for either bactericidal or bacteriostatic effects, and Class IIc are thiol-activated peptides.

Class III includes large (>30kDa) heat-labile bacteriocins, and has been the cause of some controversy, as it has been suggested that these bacteriocins are really enzymes.

Class IV includes the complex bacteriocins which contain lipid and/or carbohydrate modifications. However, this class has recently come under scrutiny and has been removed from the classification table, as little evidence has been discovered to support it (Cotter *et al*, 2005).

Characterization of bacteriocins into class types is under contention, with several new schemes for classification having been recently suggested. Table 1.1 shows a number of different classification systems which have been suggested by various groups.

Table 1.1 Comparisons of proposed classification systems for bacteriocins. (adapted from Klaenhammer, 1993; Cotter at al, 2005; Todorov, 2009).

Class	Klaenhammer (1993)	Cotter *et al* (2005)	Todorov (2009)
I	Lantibiotics	Lantibiotics	Lantibiotics
IIa	Pediocin-like, anti-listeria bacteriocins	Pediocin-like bacteriocins	Listeria-active bacteriocins
b	Two-component peptides	Two-peptide bacteriocins	Two-peptide complexes
c	Thiol-activated peptides	Cyclic peptides	The sec-dependent bacteriocins
d	-	Non-pediocin single linear peptides	Unclassified small heat-stable non-lanthionine bacteriocins
III	Large, heat-labile proteins	Bacteriolysins	Large heat-labile bacteriocins
IV	Protein complexes	-	-

Bacteriocin Mode of Activity

Different bacteriocins have varied modes of activity against target cells. Some have the ability to inhibit cell wall synthesis while others act by permeabilizing the target cell membrane or by inhibiting RNase or DNase activity. The most studied lantibiotic bacteriocin nisin is produced by *Lactococcus lactis* and studies have

shown that its effect is concentration dependent and is affected by physiological conditions such as pH, temperature, ionic strength as well as the growth phase of target cells (Rollema *et al,* 1995). As bacteriocins do not necessarily act using the same mechanism against target species, many studies have been carried out to examine the effect of different bacteriocins on a wide range of bacterial species. It has been observed that the formation of pores within target organisms is mediated by the bacteriocin 'docking' to specific molecules on the target cell membrane. The bacteriocins nisin and mersacidin clearly demonstrate this pore formation mechanism by docking with lipid II, a precursor of petidoglycan. Although both of these bacteriocins use lipid II, in the case of mersacidin the docking to this molecule allows it to inhibit peptidoglycan synthesis, whereas in the case of nisin, docking to lipid II is the first step in both pore formation and inhibition of petidoglycan synthesis (Breukink *et al*, 1999; Todorov, 2009). As a result of pore formation by nisin, biosynthesis of DNA, RNA, protein and polysaccharide is also inhibited, probably due to a de-energised cell membrane. The consequent decrease in ATP synthesis inhibits all biosynthetic processes.

Class IV - Complex Bacteriocins - Glycocin F

The bacterium *Lactobacillus plantarum* KW30 was, until recently, the only known producer of the 46 amino acid bacteriocin Glycocin F (Fig 1.2), whose bacteriocidal and bacteriostatic activity is limited to other closely related *Lactobacillus* species (Stepper *et al*, 2011).

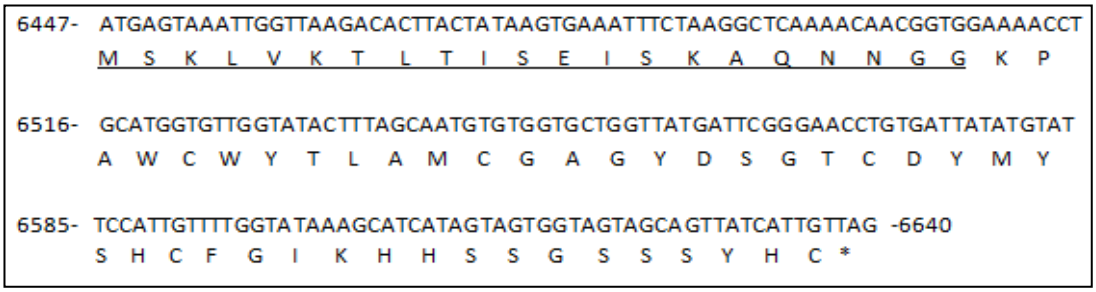


Figure 1.2 Amino acid and nucleotide sequence of Glycocin F. The leader sequence which is cleaved to produce mature glycocin F is underlined.

An almost identical bacteriocin ASM1 found in *L. plantarum* A-1 was reported in late 2010 (Hata *et al*, 2010). Investigation into the structure of glycocin F indicates that it is post-translationally modified by two N-acetylglucosamine (GlcNAc) residues, one of which is covalently linked to a serine side-chain (*O*-linked), the other to the thiol sulphur of the C-terminal cysteine (*S*-linked). The modified serine residue is part of an 8 residue turn constrained by two disulfide bonds, and is reminiscent of eukaryal *O*-glycosylation where glycosylated serines often occur in β-turns (Upreti *et al*, 2003; Stepper *et al*, 2011). The *S*-link is exceptional, as naturally occurring *S*-linked glycoproteins have only been verified in two proteins, both bacteriocins, GccF and sublancin 168. Figure 2 shows the structure of glycocin F as a cartoon and a Cα trace. The two disulfide bonds can be seen to anchor the loop structure providing a scaffold for the *O*-linked glycan. (Fig 1.3). Enzymatic removal of the *O*-glycosylated serine residue or disruption of the disulfide bonds results in a loss of antibacterial activity of the bacteriocin, indicating the importance of this carbohydrate modification in the context of correctly folded glycocin F (Stepper *et al*, 2011). Removal of the *S*-linked glycan also has an effect on the activity of GccF. Although it does not abrogate activity, it does attenuate it by ~1/3, indicating it also plays a role in the activity of GccF (personal communication, Dr Mark Patchett).

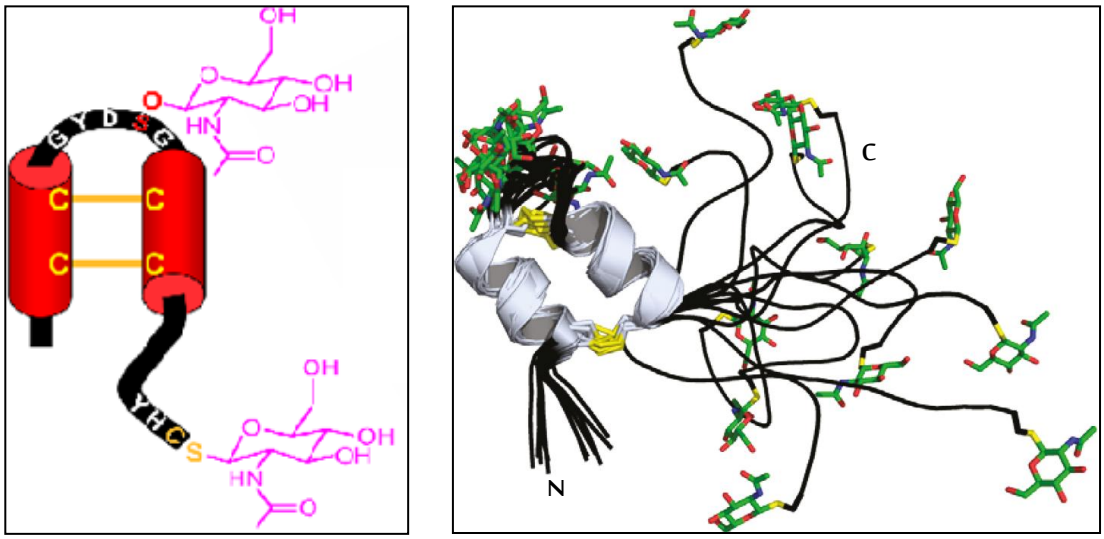


Figure 1.3 Cartoon and NMR structures of GccF showing *O*- and *S*-linked GlcNAcs. Superposition of the 12 calculated structures of GccF with the lowest energy. All sulfur atoms and the disulfide bonds between the two helices and the *S*-linked sugar at the C-terminus are shown as yellow sticks. The sugars are also shown as sticks (Venugopal *et al,* 2011).

It is hypothesised that the bacteriostatic activity caused by GccF on susceptible organisms is by its targeting of a hexosamine-phosphotransferase system transporter (likely either GlcNAc or Glucosamine). PTS systems are vital for bacteria sugar transport and in Gram-positive lactic acid bacteria comprise up to half of the sugar transport systems (Lorca *et al,* 2007). It is believed that the interaction between the PTS-transporter and GccF is intially through one or both of the GlcNAc residues on the peptide or through direct protein-protein interactions between the peptide and transporter. Interestingly, it has been found that the addition of free GlcNAc challenges the activity of GccF when incubated with susceptible strains. After attachment of GccF to the PTS-transporter, the GlcNAc residues of GccF may act as a sugar substrate and could act by blocking the channel preventing transport across the membrane. The blocking of the transporter may leave it in an open state causing leakage of cellular materials resulting in bacteriostasis and/or eventual cell death.

In order for these two residues to be GlcNAcylated there must be one of two enzymes which are capable of transferring sugar groups onto the acceptor bacteriocin. Within the glycocin F gene cluster (Fig 1.4) there is a gene encoding a family 2 glycosyltransferase (GT2), and it is almost certain that this enzyme is responsible for the addition of either one or both of the GlcNAcs onto the bacteriocin (Stepper *et al*, 2011).

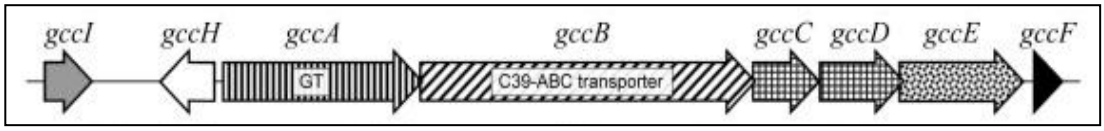


Figure 1.4 Representation of the GccF gene cluster. *gccI* – putative immunity protein; *gccH* - unknown; *gccA* – glycosyltransferase; *gccB* – C39 ABC-transporter; *gccC* – thioredoxin domain protein 1; *gccD* – thioredoxin domain protein 2; *gccE* – putative signal transduction/regulator; *gccF* – glycocin F (Stepper *et al,* 2011).

Eukaryotic Glycosylation

Protein glycosylation is found to occur in virtually all organisms. It is broadly distributed throughout the range of phylogenies, from single cell organisms to more complex cellular structures and the tissues of animals and plants.

Glycosylation is thought to be one of the most common and most diverse post-translational modifications of newly synthesised proteins (Spiro, 2002). The glycan moiety is thought to be involved in a large number of cellular processes as well as having physico-chemical roles, these include protein folding, solubility, protection from proteases, cell adhesion, cellular trafficking, and protection against bacterial and parasitic invasion, other roles include wound healing, inflammation, cancer metastasis and molecular mimicry (Haynes, 1998: Weerapana and Imperiali, 2006).

In eukaryotes *N*-linked glycosylation is one of the most common protein modifications, and is characterized by a β-glycosylamide linkage between a GlcNAc and the side chain of an asparagine residue (GlcNAcβ1-Asn). The Asn has to be part of an Asn-X-Ser/Thr sequon, where X can be any amino acid except proline and aspartic acid (Mellquist *et al*, 1998). It was once thought that only proteins that were secreted were glycosylated. Now it is known that intracellular proteins may be modified by the addition of a single GlcNAc where it acts as a signal, mimicking phosphorylation. *N*-linked glycan synthesis in mammals involves the addition of a polysaccharide (14-mer) core unit which is assembled as a dolichyl pyrophosphate precursor on both sides of the endoplasmic reticulum membrane. The nascent polypeptide is threaded into the endoplasmic reticulum and, if a sequon is encoutered by the oligosaccharyl transferase, the 14-mer glycan is transferred from the dolichyl pyrophosphate precursor to the nitrogen of the Asn sidechain. The glycan then undergoes a series of trimming and elongation events in the golgi before the glycosylated protein is transported to the membrane or is secreted out of the cell. It has been predicted that, of the multiple Asn-X-Ser/Thr sequons present within a protein, around 90% are glycosylated (Helenius and Aebi, 2001).

O-linked glycosylation occurs when a sugar is attached to the hydroxyl group of an amino acid sidechain, most commonly serine or threonine (GalNAc α1-Ser/Thr). Tyrosine, hydroxyproline and hydroxylysine can also be linked to monosaccharides (Spiro, 2002; Weerapana and Imperiali, 2006). Unlike *N*-linked glycosylation, *O*-linked glycosylation does not require a specific sequence motif in the target protein or an oligosaccharide precursor for transfer to the protein,

although it frequently occurs in proline-rich domains (Moens and Vanderleyden, 1997; Apweiler *et al*, 1999; Spiro, 2002). Phosphoglycosylation and *C*-mannosylation are two other forms of glycosylation that have only been reported in eukaryotes thus far (Spiro, 2002). Phosphoglycosylation is the formation of a bond between a phosphate group on a serine or threonine and a monosaccharide such as GalNAc or glucose. C-mannosylation involves the attachment of an α-mannose residue to the C-2 of tryptophan. The linkage results in the formation of a C-C bond and, unlike *O*- and *N*-glycosyl linkages, the functional group of an amino acid is not involved (Spiro, 2002).

Prokaryotic Glycosylation

For many years it was thought that prokaryotes did not possess the enzyme machinery to glycosylate proteins. In the mid 70's however, the first glycosylated (*O*-linked) protein was discovered in the surface layer (S-layer) of the halophile, *Halobacterium salinarium* (Mescher *et al*, 1976). Since then, similarly glycosylated proteins have been found in many other prokaryotes. These prokaryotic glycoproteins are not restricted in their localisation or functionality, as they have been observed in a variety of cellular molecules, where they are involved in pathogenesis, enzymatic activity, and cell defence (Power *et al*, 2000; Schmidt *et al*, 2003; Szymanski and Wren, 2005). Interestingly, prokaryotic glycosylation seems to mimic that in eukaryotic systems (or *vice versa*), apart from the fact that the range of monosaccharides used to make up the oligomers is much greater in prokaryotes. (Moens and Vanderleyden, 1997). Interestingly *N*-glycosylation in prokaryotes still occurs at an Asn-X-Ser/Thr sequon which often contains an Asp or Glu at the -2 positions (Nothaft and Szymanski, 2010) in contrast to eukaryotic systems where Asp residues at 'X' are not favoured. However mutational analysis of S-layer glycoproteins showed that replacement of the serine residue by valine, leucine or asparagine does not appear to prevent the occurrence of *N*-glycosylation in contrast to the more common requirements of Ser or Thr in eukaryotes (Zeitler *et al*, 1998; Schaffer and Messner, 2001). This suggests that there may be multiple *N*-glycosyltransferases, with varying specificities, present in prokaryotes, in contrast to eukaryotes. Current knowledge of *O*-glycosylation has not revealed a specific sequence motif and has shown that *O*-glycosylation often uses unusual

sugars or linkages, for example, HexNAc residues modified with methylated Asp through a phosphate linkage in *Clostridium* spp. and a rhamnose residue through which a glycosyl moiety is linked in *Pseudomonas aeruginosa* PAK and *P. aeruginosa* JJ692 (Nothaft and Szymanski, 2010).

Single Sugar *O*-GlcNAc Glycosylation in Eukaryotes

O-GlcNAc is a dynamic monosaccharide modification commonly found on serine and threonine residues in nucleocytoplasmic proteins of eukaryotes. This modification has been found in a large number of proteins such as oncogene products and tumour suppressors, transcription factors, cytoskeletal proteins and nuclear pore proteins (Wells *et al*, 2001). Although these single sugar modifications have been most widely described in vertebrates, they have been identified in other eukarya, plants, parasitic protozoa, bacteria and yeasts (Hanover, 2001). *O*-GlcNAc modifications are unusual in comparison to other forms of protein glycosylation as they do not appear to become elongated by the addition of other sugar moieties to form complex structures (Slawson and Hart, 2003). Recently *O*-GlcNAc has been thought to play a role in signal transduction as the reversible addition of this modification has been found to result in a change in cell processes (Zachara *et al*, 2004; Comer and Hart, 2000). The subsequent addition and removal of an *O*-GlcNAc modification is mediated by an *O*-GlcNAc transferase (OGT) and an *O*-GlcNAc hydrolase, both of which are regulated enzymes found within the nucleocytoplasm (Wells *et al*, 2001). Studies have reported that this single sugar *O*-GlcNAc modification may also regulate *O*-phosphorylation, as it has been observed that inhibition of kinase proteins increases the level of *O*-GlcNAc modified proteins, and the overexpression of protein kinases results in a decrease in *O*-GlcNAc modified proteins (Griffith *et al*, 1999; Slawson and Hart, 2003). Analyses of the post-translational modifications of several proteins including estrogen receptor β and c-Myc, showed that both serine and threonine could be glycosylated or phosphorylated, although not at the same time, indicating that some proteins can be regulated by both glycosylation and phosphorylation (Cheng *et al*, 2000; Kamemura *et al*, 2002). The addition or removal of *O*-GlcNAcs has also been found to play a role in protein-protein

interactions, as is the case with Sp1, or in protein activity, as shown for the tumour suppressor protein p53 (Wells *et al*, 2001).

Until recently *O*-GlcNAcation of prokaryotes was thought to only occur within oligosaccharides where the GlcNAc was linked to a sugar molecule, such as is the case with bacterial S-layer proteins, lipopolysaccharide biosynthesis and flagellin/pilin proteins. In mid-2004 the first reported β-*O*-GlcNAcylated protein, with attachment of the carbohydrate occuring directly to an amino acid residue, was discovered in flagellin proteins from *Listeria monocytogenes* (Schirm *et al*, 2004). Since then several other reports of *O*-GlcNAcylated proteins have been published including that of the bacteriocin glycocin F (Stepper *et al,* 2011).

Glycosyltransferases of Prokaryotes

Glycosyltransferases (GTases) are a large group of enzymes which transfer sugar residues from an active donor substrate to an acceptor. Approximately 1% of the open reading frames in each genome are thought to be dedicated to the synthesis of glycosidic bonds (Coutinho *et al*, 2003). These enzymes are involved in oligosaccharide and polysaccharide biosynthesis, and facilitate a wide range of functions from the storage and structure of proteins to cell signalling (Wells *et al*, 2001). GTases are primarily classified into families by the sugar that they transfer, the acceptor molecule, their specificity, and amino acid similarity (Breton *et al*, 2001). There are currently 92 families of glycosyltransferases and the numbers are still growing steadily (Carbohydrate-Active enZymes database, http://www.cazy.org/fam/acc_GT.html). These families have been further grouped into superfamilies based upon their structural similarities (Campbell *et al*, 1997; Coutinho *et al*, 2003). In eukaryotes GTases are generally located within the membrane of the endoplasmic reticulum and the golgi apparatus, and are type II transmembrane proteins, with short amino-terminal domains, a stem region, and a large globular catalytic domain. GTases have also been localised in cellular locations other than the membrane, and this occurs *via* cleavage of the stem region (Breton *et al*, 2001). There are two main catalytic mechanisms observed for the transfer of monosaccahrides by GTases (Fig. 1.5); one involves the inversion of the anomeric configuration of the donor sugar as it is attached to the acceptor (i.e. UDP-glucose or β-glucosidase), while the other involves the retention of the

anomeric configuration on attachment (i.e. UDP-glucose α-glucosidase). Both of these mechanisms refer to the C1 position of the donor sugar (Breton *et al*, 2006; Ünligil and Rini, 2000).

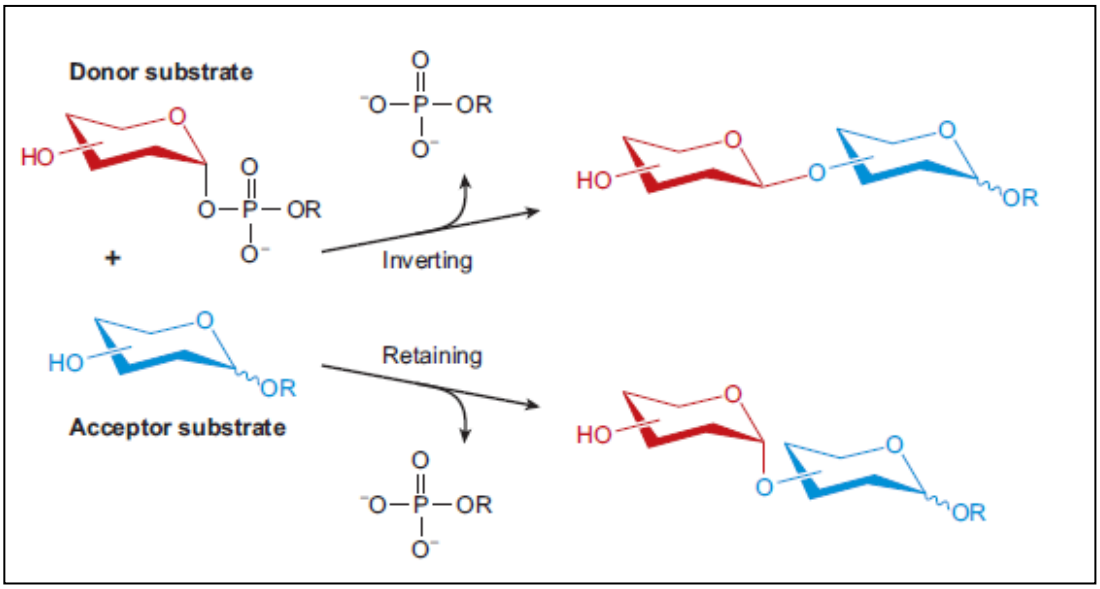


Figure 1.5 Schematic of the two main catalytic mechanisms for monosaccharide transfer by glycosyltransferases (Lairson *et al*, 2008).

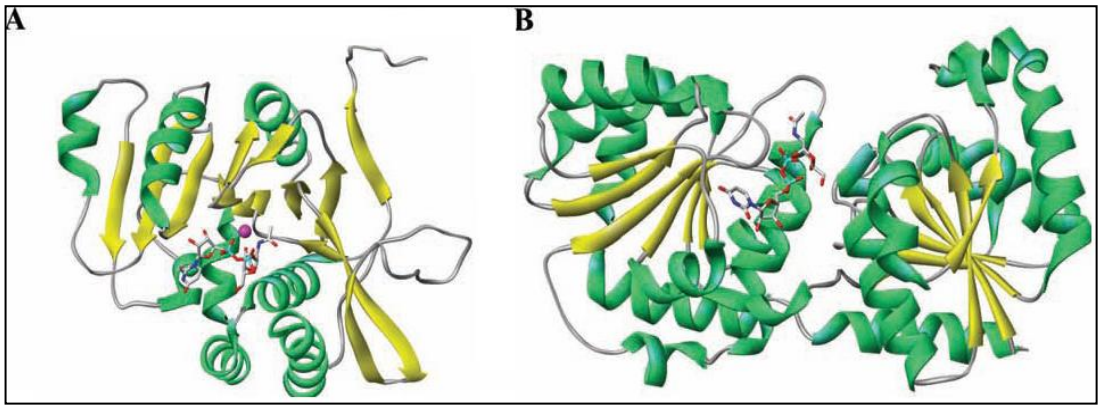


Figure 1.6 Ribbon diagram of two protein-modifying glycosyltransferases that represent the two different folds. Manganese ion indicated by purple ball, and bound nucleotide sugar represented by sticks. (A) GT-A fold, mouse α-1,4-*N*-acetylhexosamineyltransferase (EXTL2) complexed with UDP-GalNAc. (B) GT-B fold, *Escherichia coli* MurG complexed with UDP-GlcNAc (adapted from Breton *et al*, 2006).

There are two main folds described so far for GTases, 'GT-A' and 'GT-B' (Fig. 1.6). The most studied GT-A is the family 2 glycosyltransferase (GT2) SpsA from *Bacillus subtilis*. This GT2 has two highly dissimilar domains, one of which binds the acceptor and the other being involved in nucleotide binding (Charnock and Davies,

1999). GT-A folds are characterised by closely associated β/α/β domains which form a continuous sheet of at least eight β-strands, surrounded by a variable number of helices. GTases with a GT-A fold generally contain a Asp-Xaa-Asp motif, which is thought to be involved in the coordination of a divalent cation that is required for the binding of a nucleotide sugar (Coutinho *et al*, 2003; Lairson *et al,* 2008). The metal ion in this site has been shown to be essential to catalyse the formation of glycosidic bonds, and is involved in fixing the pyrophosphoryl group of the sugar donor in the active site of the enzyme (Hu and Walker, 2002; Coutinho *et al*, 2003).

The GT-B superfamily is more diverse, and includes the majority of prokaryotic enzymes which glycosylate secondary metabolites, leading to the production of products such as the antibiotics vancomycin, erythromycin, novobiocin, and daunomycin. It also includes prokaryotic enzymes which glycosylate proteins involved in primary metabolic pathways such as cell wall biosynthesis (Hu and Walker, 2002). The 'GT-B' fold was first reported in the phage T4 DNA-glucosyltransferase, and has two typical β/α/β Rossmann fold subdomains (Morera *et al*, 1999). These domains are found in all known GT-B folded GTs, and are less tightly associated than the β/α/β domain of the GT-A superfamily. The active site is located between the two Rossmann-like domains and the donor sugar is shown to interact solely with the C-terminal domain of these proteins (Coutinho *et al*, 2003).

As the result of iterative sequence searches a third glycosyltransferase fold termed `GT-C' has been predicted. The structure of this predicted superfamily is unlike that for GT-A and GT-B in that there is no β/α/β Rossmann fold subdomain. The GT-C fold is predicted to be that of a large hydrophobic integral membrane protein which is located on the plasma membrane or in the endoplasmic reticulum, which has between 8-13 transmembrane helices with the active site located within a long-loop region. Although most predicted glycosyltransferases appear to belong to one of the three superfamilies, some sequence families, such as GT76, are `orphan' families in that the glycosyltransferases in this family do not conform to any of the folds for the superfamilies (Lairson *et al* , 2008).

Few crystal structures of GTases have been solved due to their localisation to the membrane and low expression, which makes purification difficult. However it is expected that the other GTase families will adopt either the GT-A or GT-B configuration, on the basis of bioinformatic analyses of their amino acid compositions (Franco and Rigden, 2003).

The overall aims of this thesis are:
1) To discover the mode of action of GccF in regards to the producer strain *L. plantarum* KW30 and organisms sensitive to its bacteriostatic effect.
AND
2) To investigate the structure of GccA and the mechanism by which it glycosylates the bacteriocin, GccF.

PART 2:

GENERAL METHODS

2.1.1 Purified Water

Purified water was obtained from a Barnstead NanopureII water filtration system (Maryland, USA). Water filtered using this system is refered to as pure H_2O.

2.1.2 Filter sterilization equipment

Sterile syringes were obtained from Terumo Corporation (Tokyo, Japan) and sterile 0.22 µm and 0.8 µm syringe filters and filter membranes were supplied by Millipore (Massachusetts, USA).

2.1.3 Media

Media was prepared and autoclaved at 121°C and $2x10^5$ Pa for 20 minutes. Solid media was prepared by the addition of 1% (w/v) agar to the culture media prior to autoclaving. All *E. coli* strains were grown in Luria Bertani (LB) media, *Lactobacillus* strains in De Man, Rogosa and Sharpe (MRS) broth, and *Enterococcus* and *Yersinia* strains in Tryptic Soy Broth (TSB).

LB media:	25.0 g LB media
	Pure H_2O made up to 1 L
MRS media:	52.2 g MRS media
	Pure H_2O made up to 1 L
TSB media:	30 g TSB media
	Pure H_2O made up to 1 L

2.1.4 Antibiotic stock solution

Antibiotic stock solutions were prepared in pure H_2O, filter sterilized (0.22 µm syringe filter) after preparation, and then frozen in aliquots at -20°C. Antibiotics were added to autoclaved cooled (~40°C) media to the final concentrations required.

2.1.5 Glycerol stocks

Cells were streaked onto agar plates (1% agar) made up with the appropriate media and containing any necessary antibiotics. After sufficient growth, a single colony was picked using a sterile toothpick and used to inoculate 1 mL of the appropriate media which was then incubated overnight at an appropriate temperature with shaking. 0.5 mL aliquots were stored with 20% (v/v) glycerol solution in sterile screw cap tubes (NUNC ™*CryoTubes*) at -80°C. The glycerol stock strains were reactivated by scraping the frozen stock with a sterile toothpick and either streaking onto an agar plate, or by innoculating liquid media.

2.1.6 Measurement of cell optical density (OD)

Culture cell density was determined by measuring the optical density (OD) at a wavelength of 600 nm with a Smart Spec ™ Plus Spectrophotometer (BioRad Industries, Milan, Italy). Standard curves of cell number against OD_{600} have been determined previously and samples with an OD_{600} greater than 0.8 were diluted in sterile culture media. Sterile culture media was used as a blank.

2.1.7 Protein Concentration Determination

(a) Bradford Assay

A 2 mg/mL stock solution of BSA was diluted into standards: 1.6, 1.2, 1.0, 0.8, 0.6, 0.4, 0.2, 0.1 mg/mL. 5 µL of each standard and dilutions of the unknown sample were added to separate wells in a 96-well plate. 100 µL of 1x Bradford reagent was added to each well and mixed thoroughly. After 5 minutes the absorbance was read at 595 nm using a Smart Spec ™ Plus Spectrophotometer and a standard curve was used to determine the protein concentrations of samples.

Bradford reagent 5x:	100 mg Coomassie brilliant blue G-250
	50 mL Ethanol (95%)
	100 mL Phosphoric acid
	Pure H_2O up to 200 mL

(b) Absorbance at 280 nm

Protein concentration was determined using the NanoDrop ND-1000 Spectrophotometer (ThermoScientific) using ND-1000 software V3.1.0 or the Smart Spec™ Plus Spectrophotometer (Bio Rad Laboratories) where the optical density of the protein sample was measured at 260 and 280 nm and converted to mg/mL.

(c) EZQ® Protein Quantitation Kit

An EZQ® Protein Quantitation Kit (Molecular Probes™, USA) was used to determine protein concentration for samples to be subjected to 2D gel electrophoresis. A 10 mg/mL solution of Ovalbumin was diluted into standards: 5.0, 2.0, 1.0, 0.5, 0.2, 0.1, 0.05, 0.02 mg/mL, and unknown samples were diluted 5 and 20 fold. Assay paper was inserted into the microplate and 1 µL of standards and sample were spotted into separate wells on the paper in duplicate. The samples were left to dry completely and the assay paper removed from the microplate. 20 mL of methanol was placed into a tray along with the assay paper and left to incubate for 5 minutes with agitation at room temperature. The assay paper was then removed and left to air dry. 20 mL of EZQ® protein quantitation reagent was poured into a tray along with the assay paper. The paper was left to incubate for 30 minutes with gentle agitation at room temperature. The liquid was poured off and replaced with rinse buffer for 1-2 minutes. This was repeated twice for a total of three rinses. Rinse buffer was removed and the assay paper was placed in water for analysis using a FLA-5000 (Fujifilm, USA) scanner with a 473 nm laser and 580 nm long pass emission filter. Images were visualised using Image reader software V2.1 and concentration of samples was determined using Image Gauge software V4.0.

Rinse buffer:	10% Methanol
	7% Acetic acid
	H_2O

2.1.8 Agarose gel electrophoresis

DNA fragments were separated on the basis of size by agarose gel electrophoresis. Depending upon the size of wells present in the gel, 10-20 µL of DNA solution was mixed with 6x loading dye and loaded onto a 1.2% agarose Tris-acetate-EDTA gel. The gels were run at 80 V until the dye front had migrated to the other end of the gel. The gels were then stained with ethidium bromide and the DNA bands were visualised by ultraviolet (UV) transillumination (BioRad Gel Doc, BioRad Laboratories, Milan, Italy). A standard 1kb plus DNA ladder (Invitrogen, USA) was included in all gels and used to estimate the size of the bands.

6x loading dye:	2% bromophenol blue
	50% glycerol
	H_2O

Tris-acetate-EDTA:	40 mM Tris-HCl
	20 mM Acetic acid
	2 mM EDTA
	pH 8.0

2.1.9 Competent cell preparation

A single colony was seeded into 50 mL LB culture medium and grown overnight at 37°C with shaking. 1 litre of LB culture medium was inoculated with the 50 mL overnight culture in a 5 litre flask and incubated with shaking at 37°C until the OD_{600} reached 0.5-0.6. The culture was then transferred into two chilled, sterile 500 mL centrifuge bottles and was incubated on ice for 30 minutes. The cell pellet was harvested by centrifugation at 4,000 x g for 15 minutes at 4°C the supernatant discarded and the cell pellet placed on ice then resuspended in 500 mL of cold sterile water. The cells were centrifuged at 2,000 x g for 15 minutes at 4°C, the water decanted and the cell pellet placed on ice. The process was repeated using 250 mL of cold sterile water. After the final wash, the cell pellet was resuspended in 20 mL of 10% glycerol and was transferred into two chilled, sterile 50 mL centrifuge tubes and then

centrifuged at 4,000 x g for 15 minutes at 4°C. The liquid was decanted and the cell pellet was placed back on ice, then resuspended in 1 mL cold sterile 10% glycerol, and pooled in one of the 50 mL centrifuge tubes. 55 µL aliquots of the cell suspension was placed in 1.5 mL Eppendorf tubes and frozen by immersion in liquid nitrogen for ~2 minutes then stored at -80°C.

2.1.10 DNA concentration determination

DNA concentration was determined using the NanoDrop ND-1000 Spectrophotometer (Thermo Scientific). The absorbance of 1 µL of DNA sample was measured at 260 and 280 nm and converted to mg/mL using ND-1000 software V3.1.0.

2.1.11 Protoplast formation

Protoplasts were prepared following the protocol described in Morelli *et al* (1987). A 10% inoculum of a mid-log phase culture was added to 40mL of 30°C MRS broth, incubated at 30°C for 2 hours, then chilled on ice and harvested by centrifugation at 4,400 x g for 10 minutes. Cells were then washed with sterile saline solution before being resuspended in 4 mL of protoplast buffer (PB) containing 20 µL/mL of mutanolysin (~320 units). After incubation at 37°C for 40 minutes, the protoplasts were recovered by centrifugation at 3,500 x g for 10 minutes, washed three times in 1 mL PB and resuspended in 400 µL of the same buffer. The presence of protoplasts were confirmed by microscopy at 400 x magnification (Olympus CH40 Light Microscope, USA). Resuspended protoplasts were stored at -80°C.

Protoplast buffer (PB): 10 mM Tris-HCl pH 8
20 mM $CaCl_2$
0.5 M Sucrose

Protoplast lysis

1 mL of protoplast suspension was thawed on ice then centrifuged at 4, 000 x g for 1 minute. The supernatant was discarded and the protoplasts were resuspended in 876 µL H_2O and 124 µL of a stock solution of protease inhibitor (1 tablet Complete Mini, EDTA-Free; Roche dissolved in 1.4 mL water). Lysis was visualised by light microscopy and samples stored at -80°C.

Protoplast fractionation

Lysed protoplasts were placed in Eppendorf tubes and centrifuged at 3000 x g for 5 minutes and the supernatant collected. The supernatant was then centrifuged at 60,000 x g for 60 minutes. The supernatant was decanted and kept as the cytosolic fraction, the pellet was washed 3x in H_2O, resuspended in solubilisation solution and sonicated for 10 minutes to form the membrane fraction.

Solubilisation solution:	20 mM Bis-Tris (pH 7)
	1% Triton X-100
	2x protease inhibitor Complete Mini-EDTA free (Roche applied science, Germany)
	0.5 mM TCEP

2.1.12 Cell growth and collection

The growth of cells in liquid medium was carried out in flasks, with a volume of flask to liquid ratio of 5:1. Cultures were typically inoculated with 0.5 – 2% (v/v) of an overnight culture and incubated at 25-37°C with or without rotation. Harvesting of cell culture with a volume up to 2 mL was carried out by centrifugation in 1.5 mL Eppendorf tubes in a MiniSpin plus centrifuge (Eppendorf, Germany) at 13,000 x g at 4°C for 5-10 minutes. Cell cultures with a volume between 10 - 300 mL were harvested by centrifugation in a Sorvall RT7 centrifuge (Kendro Laboratory Products, Germany) for 15 minutes at 2,700 x g and 4°C. Harvested cells were then washed 3 times in 1x PBS to remove media contaminants.

Lysis of *E. coli* cells was by two passes through a French pressure cell (Wabash, Aminco Instruments Co.) at 4000 psi.

To lyse *L. plantarum* whole cells two methods were used. Cells harvested from culture volumes of less than 5 mL were lysed using a ribolyser (Hybaid Ltd, UK) where cell pellet was resuspended in 800 µL PBS and transferred to screw top microcentrifuge tubes. 200 µL of glass beads (Alltech Associates, USA) were added and the cells were ribolysed three times at maximum speed for 20 s, with 1 minute incubations on ice. For larger volumes cells were lysed by two passes through a French pressure cell (Wabash, Aminco Instruments Co.) at 4000 PSI.

10x Phosphate buffered saline (PBS):	137 mM NaCl
	27 mM KCl
	43mM KH_2PO_4
	14 mM Na_2HPO_4 $7H_2O$
	pH 7.4
	Autoclave to sterilise

Cell lysates were fractionated in the same manner as protoplast lysates (2.1.11).

2.1.13 SDS-Polyacrylamide gel electrophoresis

Proteins were separated on the basis of mass by SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE). Standard SDS-PAGE used the discontinuous buffer method originally described by Laemmli (1970). The gels (7.3 cm x 7 cm x 0.75 cm) were cast and run using the Mini-Protean® 3 system (BioRad Industries, Milan, Italy). Electrophoresis was performed at room temperature at 200 V for 45 minutes or until the dye front reached the bottom of the gel.

For the separation of small proteins, 16.5% tricine SDS-PAGE was used based on the method developed by Schägger (2006). For these gels electrophoresis was performed at room temperature. Initially 30 V was applied across the electrodes and after 30 minutes the voltage was increased to 120 V for 60 minutes or until the dye front reached the bottom of the gel.

2.1.14 Detection of proteins after SDS-PAGE or western blotting

(a) Coomassie Staining

After separating the proteins by SDS-PAGE (2.1.13) the gel was placed in a 'gel box', covered with Coomassie blue staining solution, and incubated at room temperature with gentle agitation for 30 minutes. After incubation, the stain solution was decanted and background staining removed by incubation in destaining solution for 1-3 hours depending on the intensity of the background staining.

Coomassie blue staining solution: 0.1% Coomassie blue R-250
40% Methanol
10% Acetic acid

Destaining solution: 40% Methanol
10% Acetic acid

(b) Silver Staining

After separating the proteins by SDS-PAGE (2.1.13) they were fixed by incubation for 60 minutes in four changes of 50% methanol. After this time, the gel was washed in water, transfered to a clean container and incubated for 15 minutes with agitation in solution C. The solution was decanted and the gel washed well in several changes of H_2O for 5 minutes. The gel was once more transferred into a clean gel container and incubated in a fresh preparation of solution D for 5-10 minutes, until bands started to appear. Once the desired amount of development was reached the solution was decanted and 50 mL 50% methanol containing 2-3 drops of acetic acid was added to prevent further development.

Solution A: 0.8 g Silver nitrate
4 mL pure H_2O

Solution B: 0.36 g NaOH
1.4 mL conc. NH_4OH
21 mL pure H_2O

Solution C: Add solution A dropwise with vigorous stirring to solution B. If a precipitate occurs add concentrated ammonia dropwise until solution clears. Add to 73.6 mL H_2O and use immediately.

Solution D: 2.5 mL 1% Citric acid
0.25 mL 38% Formaldehyde
Make up to 500 mL with H_2O

(c) Colloidal Coomassie Blue Staining

After separating the proteins on SDS-PAGE (2.1.13) the gel was placed in a 'gel box' and covered with 50 mL of colloidal stain solution, and incubated at room temperature with gentle agitation overnight. After incubation, the stain solution was decanted and background staining removed by incubation in wash solution (25% methanol). To store the gel the wash solution was decanted and 20% ammonium sulphate solution added.

Solution A: 2.5 g Coomassie blue G-250
Make up to 50 mL with H_2O

Solution B: 60 g Ammonium sulphate
60 g Phosphoric acid (85%)
Make up to 500 mL with H_2O

Colloidal stain: 1.2 mL Solution A
40 mL Solution B
Mix for 10 minutes then add 10 mL methanol
Use immediately

2.1.15 In-gel tryptic digest

SDS-PAGE gels were run as described (2.1.13). Bands of interest were excised with a clean scalpel, placed on clean glass plates then sliced into 1 mm slices and transferred into Eppendorf tubes. 100 µL of destaining buffer was added to cover the gel pieces and the Eppendorf tubes were incubated in a thermomixer at 50°C for 2 hours. The destaining solution was then discarded, 100 µL of water was added and the samples were vortexed for 5 s. After removal of the water 100 µL of acetonitrile was added and the sample was vortexed. The acetonitrile was removed using a pipette, and gel pieces were rehydrated with the addition of 100 µL of water. Once the water was removed 100 µL of acetonitrile was again added, vortexed and removed using a pipette. 100 µL of aqueous acetonitrile (1:1) was then added, vortexed and removed as before, and the gel particles dried under vacuum (SpeedVac Concentrator, Savant, USA) for 20 minutes. The gel particles were rehydrated in 30 µL of cleavage buffer and incubated overnight at 37°C. The samples were centrifuged (10,000 x g, 15 s) and 10 µL of acetonitrile/5% trifluroacetic acid was added. Samples were then placed in an ultrasonic bath for 10 minutes before being pelleted (10,000 x g, 15 s). The supernatant was removed, and the pellet was resuspended in 10 µL of acetonitrile/5% trifluroacetic acid. Samples were then placed in an ultrasonic bath for 10 minutes and then centrifuged (10,000 x g, 15 s). This was repeated twice more, and the supernatant fractions pooled.

Destaining Solution:	50% 100 mM 4-ethylmorpholine acetate (pH 8.1) 50% 2M acetonitrile
Cleavage Buffer:	25 mM 4-ethylmorpholine acetate (pH 8.1) 0.01% 2-mercaptoethanol 5% acetonitrile 10 ng/µL trypsin (sequencing grade, Promega).

50 µL of a slurry of Poros resin 50R2 (Applied Biosystems, California, USA) in water, made 1% with TFA, was added to a gel loading pipette tip and

equilibrated in 90% acetonitrile/1% TFA. 5 CVs of H_2O/0.1% TFA was rinsed through the in-tip column. The sample was then added to the column using a 1 mL syringe capped with a yellow pipette tip, incubated for 1 minute, then eluted with 5 CVs H_2O/0.1%TFA followed by elution with 2 CVs of 70% acetonitrile 1% TFA then 2 CVs of 90% acetonitrile 1% TFA. The 70% and 90% elutions were then frozen at -80°C and lyophilised. The samples were then resuspended in 20 µL pure H_2O and analysed by mass spectrometry.

2.1.16 Bacteriocin Purification

SP-Sephadex resin (GE Healthcare)was equilibrated by incubation overnight in 50mM formate/0.1M NaCl pH 3.5 and was stored at 4°C.

8 litres of MRS culture medium were inoculated with 1% (v/v) of an overnight MRS culture of *L. plantarum* KW30 and incubated at 25°C for 4-5 days without stirring. The cells were removed by centrifugation (16,000 x g) and the supernatant incubated overnight at room temperature with stirring with 50 mL (settled volume) of pre-equilibrated SP-Sephadex resin using an overhead stirrer. The resin was then packed into a column (Econo-Column 6 x 20 cm, Biorad Industries, USA) and washed with 10 CVs of 50 mM sodium formate buffer pH 4.5. This was followed by 10 CVs of 50 mM MOPS-NaOH pH 7.2 then 10 CVs of 25 mM ammonium bicarbonate pH 8.0 and lastly with 5 x 1.5 CVs 25 mM ammonium bicarbonate/50% acetonitrile. The eluate from each wash was collected and tested for GccF activity using a biological assay (2.1.17). Fractions showing GccF activity were freeze dried, resuspended in a total volume of 20 mL of pure H_2O/1% acetic acid, and subjected to reverse phase HPLC (Ultimate 3000, Dionex, Germany) using a Jupiter 5 µM C_4 300 Å column (250 x 10 mm, Phenomenex) using the gradient shown in Figure 2.1: A: H_2O, 0.1% TFA, and B: Acetonitrile, 0.8 % TFA. The gradient started at 10 % B and increased to 25% B over 5 minutes at a constant flow rate of 1 mL/min. The %B was then increased to 40% over 15 minutes before being further increased to 90% over 5 minutes, then held for 5 minutes until it was returned to the starting conditions. Elution was monitored using a photodiode array detector (Dionex) at 214 nm and 280 nm

and Chromeleon software V 6.80. GccF peaks were collected, pooled and freeze dried. Freeze dried GccF was then resuspended in 1 mL pure H_2O.

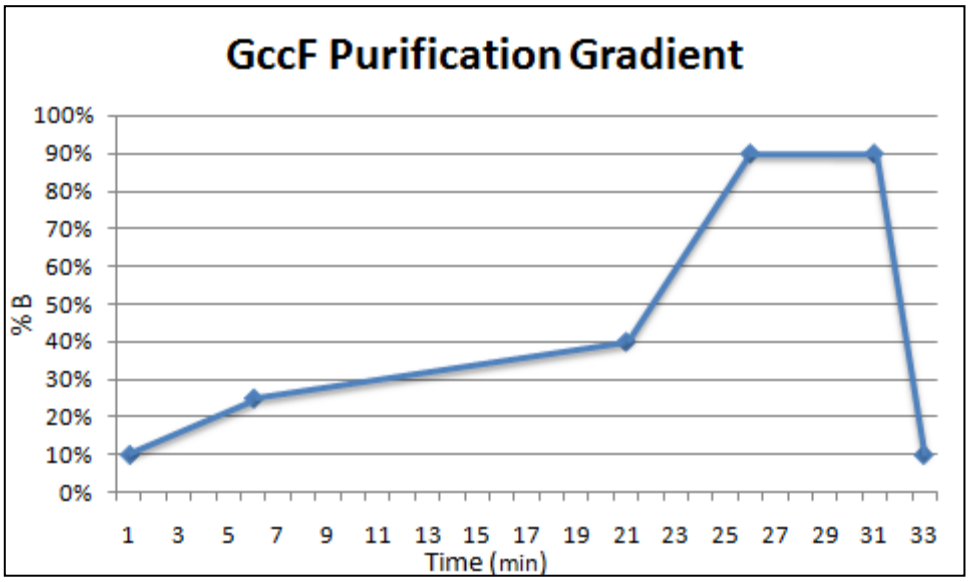


Figure 2.1 RP-HPLC gradient for GccF purification.

2.1.17 Bacteriocin bioassay

To prepare the indicator plates for the biological assay the indicator strain *L. plantarum* ATCC 14197, which is susceptible to the activity of GccF, was grown overnight in MRS media. 1 mL of culture was added to ~40 mL of 1% MRS agar at 40°C and mixed gently by inversion. This mixture was poured into petri dishes, using ~15 mL per plate, and allowed to set at room temperature. To carry out the assay, 1 - 2 µL of each sample to be tested plus 1 µL of purified GccF, as positive control, were spotted onto the indicator plate, and incubated at 25°C until a clear zone of growth inhibition was observed for the positive control.

2.1.18 Mass Spectrometry

Samples for electrospray ionisation mass spectrometry (ESI-MS) and liquid chromatography mass spectrometry (LCMS) were purified using either RP-HPLC, or a gel loading pipette tip containing Poros 50RS resin following in-gel tryptic digest (2.1.15), and stored in Maximum™ recovery tubes

(Axygen, USA). Samples were then lyophilised, and reconstituted in 50 µL of 50% ACN/HPLC grade H_2O/0.1% formic acid. Samples were sonicated for 5 minutes on high (Sonicator, Soniclean Pty. Ltd.) and then centrifuged at 16,000 x g for 5 minutes. The supernatant was transferred to a new tube and an aliquot was taken from each sample and diluted according to their A_{280} with the same solution to a desired concentration.

(a) ESI-MS

1-2 µL of sample was injected into the MS on bypass using an Agilent 1200 series LC-MS system running 50% ACN/pure H_2O/0.1% formic acid at 0.1 mL/min. The ionised species were analysed in positive ion mode by Agilent 6520 series Q-TOF running Mass Hunter (ver. B.02.01) data acquisition software and the results were analysed using Qualitative Analysis software (ver. B.03.01). Typical parameters were; Fragmentor voltage = 60 V, Skimmer voltage = 45 V, V Cap voltage = 3800 V, Gas temperature = 325°C, Drying gas = 8 L/min, Nebuliser = 35 psi.

(b) LCMS

1-4 µL of sample was loaded onto an Agilent HPLC-Chip Cube MS Interface (G4240A) using a ProtID-Chip-43 (G4240-62005, 43 mm 300 Å C_{18} chip with a 40 nL trap column) mounted on an Agilent 6520 Q-TOF LC-MS system running a linear gradient from 5%-85% solution B; 0.1% formic acid/ 5% pure H_2O/95% ACN (Solution A; 0.1% formic acid/pure H_2O) at 0.4 µL/min. The ionised species were analysed in positive ion mode using Mass Hunter (ver. B.02.01) data acquisition software and the results were analysed using Qualitive Analysis software (ver. B.03.01). Typical parameters were; Fragmentor voltage = 125V, Skimmer voltage = 60 V, V cap voltage = 1800 V, Gas temperature = 325°C, Drying gas = 8 L/min, Nebuliser = 35 psi, Collision energy used a slope of 2.5 m/z with an offset of 2 V.

PART 3:

FUNCTIONAL CHARACTERISATION OF GccF, A BACTERIOCIN PRODUCED BY *L. PLANTARUM* KW30

3.0 INTRODUCTION

Part 3 of this thesis focuses upon the functional aspects of the bacteriocin GccF first isolated from fermented corn in *Lactobacillus plantarum* KW30. It has previously been shown (Stepper *et al*, 2011) that this bacteriocin is uniquely modified with two GlcNAc residues, one *O*-linked residue through serine-18 and one *S*-linked through cysteine-43. Although the apparent maturation and modifying proteins for GccF are thought to have been identified, the mechanism by which it prevents the growth of its target microorganisms is unknown. The aim of this section of work was to investigate the potential mode of action of GccF in regards to the producer strain and organisms sensitive to its bacteriostatic effect.

3.1 OBJECTIVES AND STRATEGIES

The objectives for this section of work were:

1) Perform western blots and wheat-germ lectin pull-downs in order to ascertain whether there are any other *O*-GlcNAcylated proteins in the *L. plantarum* KW30 proteome.

2) To investigate the effect of FITC labelled modified GccF (GccF^{-HC} and GccF$^{de\mathit{O}GlcNAc}$) on an indicator strain in comparison to the effects observed with native FITC labelled GccF.

3) To identify potential receptors in the indicator strain, by immobilising GccF on CDI-agarose beads, then using the beads in pull-down experiments with fractionated indicator strains.

4) To isolate the enzymes responsible for the maturation of GccF using pull-downs as in objective 3, but this time with *L. plantarum* KW30 cell lysate.

5) To identify potential receptors in the indicator strain using biotin labelled GccF and streptavidin labelled magnetic beads.

6) Compare protein expression in untreated and GccF treated indicator strains by 2D gel electrophoresis.

3.2 METHODS

3.2.1 Growth Assay

1 litre of MRS culture medium was inoculated with 2% (v/v) of an overnight culture of *L. plantarum* KW30 and was incubated at 25°C under static conditions. 5 mL aliquots were removed at hourly intervals, and the optical density at 600nm measured, (Smart Spec™Plus Spectrophotometer). Cells from each aliquot were removed by centrifugation and stored at -80°C, and the pH of the supernatant was measured.

The bacteriocin bioassay is described in the General Methods chapter (2.1.17).

3.2.2 Protein transfer for western blot

20 µg protoplast lysate (2.1.11) and 10 µL of 0.1 mg/mL GccF was run on a mirrored tricine gel (2.1.13). 12-15% SDS-PAGE gels (2.1.13) were also trialled but resolution of lower molecular weight bands was poor. Better resolution was obtained with 16.5% tricine gels (2.1.13) as the band corresponding to the bacteriocin was clearly visible. Glycocin F was used as a positive control for protein GlcNAcation. Following electrophoresis the gel was removed from the gel apparatus and rinsed in transfer buffer to remove any contaminants. A nitrocellulose membrane was cut to the size of the gel, then immersed in transfer buffer for 5 minutes. 3 mm Whatman filter paper was cut into 4 oblongs slightly larger than the original gel size and soaked in transfer buffer along with the fibre pads belonging to the blotting apparatus (BioRad Labroratories, USA). A gel box was half filled with transfer buffer and the transfer apparatus was placed within, along with a magnetic stirring rod and a frozen insert. The gel cassette was then placed on a tray and a pre-soaked fibre pad followed by two pre-soaked Whatman papers were placed onto it. The gel was then placed on the Whatman paper followed by the nitrocellulose membrane, two more pieces of Whatman paper and a fibre pad. The gel cassette was then closed and placed in the buffer tank containing a magnetic stirrer with the membrane facing the cathode. The gel box was then filled with transfer buffer until the gel cassette was completely immerged, the magnetic stirrer turned on, and the lid put in place.

The gel was run at 100 V for 80 minutes or until adequate transfer was observed.

Transfer buffer:	25 mM Tris
	200 mM Glycine

After transfer the gel was removed and stained with coomassie blue (2.1.14) to observe any bands which had not transferred onto the membrane. The membrane was also stained with Ponceau S (b) to observe bands transferred from the gel.

3.2.3 Ponceau S Staining

After transfer of proteins onto a membrane the membrane was rinsed in H_2O before incubation in 40 mL of Ponceau S staining solution for 10 minutes with gentle agitation. After incubation the proteins can be visualised. To destain, rinse the membrane in H_2O then rapidly immerse in 0.1 M NaOH. Once the protein bands are no longer visible rinse the membrane in running H_2O for 3 minutes. After destaining the membrane can be used for protein detection.

Ponceau S Staining Solution:	0.1% Ponceau S (w/v) (Sigma Aldrich)
	5% Acetic acid (v/v)

3.2.4 Protein detection for western blot

The nitrocellulose membrane was probed with mouse anti-*O*-GlcNAc and rabbit anti-C-14 antibodies using the following protocol provided from Abcam.
The membrane was cut into two pieces and placed into a container containing 15 mL of blocking solution and incubated for 60 minutes at room temperature with agitation. Blocking is an important step in western blotting as the membranes used are chosen due to their ability to bind to proteins non-specifically. As antibodies are also proteins the membrane must be blocked to prevent non-specific binding of the antibody to the membrane causing false results. To block the membrane it is typically placed in a solution containing milk or bovine serum albumin (BSA) along with tris-buffered saline and detergent. The BSA or milk proteins then attach to the membrane in regions where the proteins transferred from the gel have not,

occluding any regions on the membrane to which the antibody can bind. During this time the primary antibodies (mouse anti-*O*-GlcNAc from Sigma Aldrich and rabbit anti-C-14) were thawed on ice then diluted (1/4000) into BSA/TBS-T. Blocking solution was decanted from the membranes which were then incubated with the antibodies for 30 minutes at room temperature with agitation (one membrane was probed with mouse anti-GlcNAc antibodies and the other membrane with rabbit anti-C-14 antibodies). After incubation the liquid was decanted and the membranes were washed in TBS-T for 5 minutes with agitation, this was repeated two times. During washing the secondary antibodies (anti-mouse and anti-rabbit HRP conjugated antibodies) were thawed on ice and then diluted (1/12000) into BSA/TBS-T. After the final wash secondary antibodies were added to the appropriate membrane and incubated at room temperature for 30 minutes with agitation. Liquid was decanted and the membranes were washed in TBS-T for 15 minutes with agitation, this was repeated two times with 5 minute washes. The membrane was finally washed in TBS for five minutes with agitation. To detect the bound antibodies 1 mL of 3,3' 5,5' -tetramethylbendizine (TMB) was incubated with the membranes for ~10 minutes or until a purple colour was observed, the TMB was then decanted and the membranes stored in water.

TBS:	20 mM Tris-HCl 150 mM NaCl pH 7.5
TBS-T:	0.05% Tween20 Made up in TBS
Blocking solution:	2% BSA Made up in TBS-T
BSA/TBS-T	0.1% BSA Made up in TBS-T

Table 3.1 Conditions trialled for western blotting optimisation.

Trials	Protein concentration for SDS-PAGE (µg)	Transfer time (~minutes)	Blocking time (minutes)	Anti-*O*-GlcNAc	Anti-C-14	1° antibody incubation (minutes)	2° antibody incubation (minutes)
1	30	45	30	1/2000	1/4000	60	60
2	10	60	30	1/2000	1/4000	60	60
3	30	60	30	1/1000	1/8000	30	60
4	10	60	30	1/1000	1/12000	30	30
5	20	60	60	1/1000	1/12000	30	30
6	20	80	60	1/1000	1/12000	30	30
7	30	80	60	1/2000	1/12000	30	30
8	20	80	60	1/4000	1/12000	30	30

Multiple trials were carried for western blotting under different conditions (Table 3.1) in order to obtain the optimum conditions for transfer and detection of proteins and protein modifications. Trial 8 seen in Table 3.1 was found to be the most suitable and was used in subsequent western blots.

3.2.5 Wheat-germ agglutinin pull-downs

Lectin-resin (Vector Laboratories, Inc., USA) was washed 3 x in 400 µL of lectin buffer and ~100 µL of WGA-agarose was mixed with 100 µL of lectin buffer. Equilibrated resin was then incubated with ~20 µg of protoplast lysate (2.1.11) for 2 hours at room temperature with rotation. After incubation the samples were centrifuged at 3,000 x g for 2 minutes and the supernatant removed using a pipette. The resin was then rinsed three times with lectin buffer, then incubated with SDS sample buffer, at 60°C for 5 minutes. Both the supernatant and WGA-agarose were analysed using tricine SDS-PAGE (2.1.13).

Lectin Buffer:	10 mM HEPES
	0.5 M NaCl
	pH 7.5

3.2.6 Fluorescein Isothiocyanate (FITC) Labelling

1 mg of FITC was dissolved in 1 mL of DMSO. 10 µL of this solution was added to 2 mg/mL of GccF in PBS pH 6.5, then incubated overnight with rotation at room temperature (20-25°C) while protected from light. To separate labelled GccF from unlabelled protein and from free FITC, samples were subjected to RP-HPLC (Ultimate 3000, Dionex, Germany) using Jupiter 5 µm C_{18} 300 Å columns (250 x 4.6 mm, Phenomenex) and the following solutions to form a linear gradient: A: H_2O, 0.1% TFA, and B: Acetonitrile, 0.8% TFA. The gradient started at 10% B and increased to 50% B over 20 minutes at a constant flow rate of 1 mL/min. The gradient was increased to 90% B and held for 3 minutes before it was returned to the starting conditions over a period of 3 minutes. Labelled bacteriocin peaks were monitored by fluoresence emission at 520 nm using a Fluorescence detector (Dionex RF2000, Germany) and Chromeleon software Ver. 6.80. Collected peaks were freeze-dried overnight then resuspended to 2mg/mL in PBS pH 6.5 and activity was tested with the bacteriocin bioassay (2.1.17). Labelled bacteriocin was stored at -80°C in tubes wrapped in foil.

3.2.7 Microscopy Assay

All FITC work was performed in the dark to prevent bleaching.

L. plantarum ATCC 8014 was grown under standard conditions (25°C in MRS media) to an OD_{600} of 0.2-0.3. At this point 1 mL of culture was taken and the cells pelleted by centrifugation at 7.7 x g for 4 minutes before being resuspended in 1 mL PBS pH 6.5. The cells were pelleted as before the supernatant discarded and the pellet resuspended in 200 µL of PBS pH 6.5. At this point, 300 nM FITC-GccF (personnal communication, Dr Shilpa Shastri) was added to the cell suspension which was then held at 30°C for 30 minutes. After incubation cells were pellet using the same conditions as before then washed in 500 µL of PBS pH 6.5. Cells were resuspended in 500 µL PBS/2.5% formaldehyde pH 6.5 and incubated overnight at 4°C. After this time cells were pelleted as before, resuspended in 500 µL of PBS pH 6.5, before being pelleted once again then resuspended in 10 µL PBS pH 6.5 and 2 µL of Vectashield (Vector Labroratories, USA). Vectashield was used as it helps to prevent rapid photo-bleaching of fluorochromes such as FITC. After incubation for 5-10 minutes at room temperature 2 µL was aliquoted onto a glass microscope slide, covered by a coverslip and sealed using nail polish around the edges of the cover slip. Cells were then analysed by fluorescent microscopy using an inverted microscope (Olympus IX71, USA) and MetaMorph meta imaging series software Ver. 7.5.2.0.

3.2.8 Production of protein/peptide – agarose affinity columns

Peptides were coupled to column following a modified method using Pierce CDI-agarose (Pierce, USA). The capacity of the beads is 2-10 mg of protein per 1 mL of resin and binding must occur at an alkali pH between 8.5-10. All steps were performed at room temperature, and for the C-11 peptide 5mM DTT was present at all times to prevent disulfide bond formation. Typically the protein for coupling (GccF or C-11) was dissolved in 50mM borate buffer pH 8.5 to a concentration of 5mg/mL. A pH of 8.5 was used to promote binding of the resin to the N-terminal primary amine on the peptide instead of the primary amine located on lysine-32. It is important to prevent binding of the resin to lysine-32 as this amino acid is located in a region thought to be essential for the activity of GccF (personnal communication, Dr Gill Norris). The resin was brought to room temperature and 1

mL aliquoted into a column (10 mL Poly-Prep®, Chromatography columns, Bio-Rad Laboratories, USA). The column was then washed with 5 mL of pure ice cold H_2O before 1 mL of protein solution was added. The reaction slurry was incubated at room temperature overnight with rotation. The next morning, excess buffer was drained from the reaction slurry and 2mL of 50mM Tris buffer (pH 9) added to the column then incubated for 5-6 hours to quench the reaction. After this time, the buffer was drained and the resin washed 3 times with 5 mL of PBS. The activity of the substituted beads was determined using the bacteriocin bioassay described in 2.1.17, except in this case a slurry of resin was added to the plate with unbound resin used as a negative control. The column was stored at 4°C for future use. Protoplasts from a chosen indicator strain were lysed (2.1.11) then subjected to ultracentrifugation at 60,000xg for 60 minutes. The supernatant was removed and collected as the cytosol fraction. The remaining pellet was resolubilised in twice the cell volume of solubilisation buffer, sonicated for 10 minutes, and collected as the membrane fraction (2.1.11). Protoplasts were fractioned into membrane and cytosol samples for use with both the GccF coupled CDI-agarose, and C-11 coupled CDI-agarose resin as it is thought that GccF must bind to proteins located in the membrane of susceptible strains due to its rapid activity, and that the machinery involved in maturation and secretion of the peptide are located within the membrane of *L. plantarum* KW30.

The concentration of each fraction was measured using the NanoDrop spectrophotometer(Thermo Scientific, 2.1.12) then adjusted to the required concentration using either water for the cytosol or solubilisation buffer for the membrane sample. Three 100 µL aliquots of resin were then placed into separate microcentrifuge tubes and incubated with 200 µL of either the membrane, cytosol or water for two hours, with rotation, at room temperature. After incubation the tubes were centrifuged at 5,000 x g for 1 minute and the supernatant removed. The beads were then washed for 5 minutes with 200µL of water with rotation before beingpelleted at 5,000 x g for 1 minute. The supernatant was removed and collected, and the washing process repeated two more times. The beads were then washed with 200µL aliqots of increasing concentrations of either NaCl (0.1M-1M) or GlcNAc(2 mM-1 M) using the same protocol as that for the water washes. Each fraction including the beads were then analysed by SDS-PAGE (2.1.13). The resin

itself was retestest for activity following this affinity procedure using the bacteriocin bioassay(2.1.17). Bands of interest were excised with a clean scalpel, placed on clean glass plates then sliced into 1mm slices and transferred into Eppendorf tubes. The bands were sent away for mass spectrometry by Dr Petr Man at the Academy of Sciences of the Czech Republic using a Bruker ultraflex II I TOF/TOF.

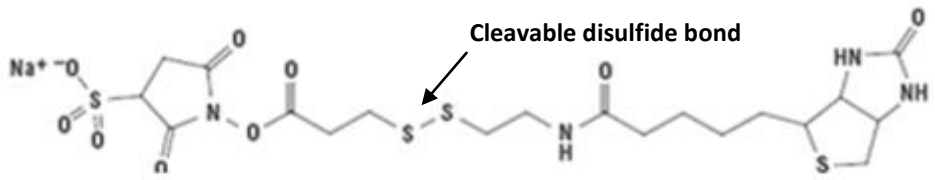
Reagents

C-11 peptide: HHSSGSSSYHC

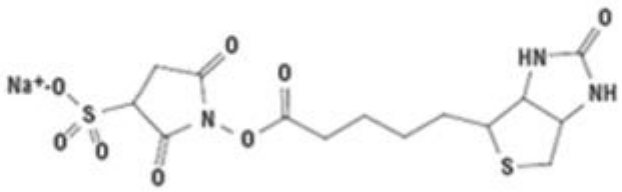
10x Phosphate buffered saline (PBS):	137mM NaCl
	27mM KCl
	43mM KH_2PO_4
	14mM Na_2HPO_4 $7H_2O$
	pH 7.4
	Autoclave to sterilise
Solubilisation buffer:	20 mM Bis-Tris (pH 7)
	1% Triton X-100
	2x protease inhibitor Complete mini-EDTA free (Roche Applied Science)
	0.5mM TCEP

3.2.9 Biotin labelling

GccF was labelled with biotin following a modified version of the protocols used by Kumar *et al* (2003) and the method for the use of EZ-Link-Sulfo-NHS-Biotin provided by Thermo Scientific (USA). Initial labelling experiments were carried out with biotin which had a long spacer arm (24.3 Å) that contained a disulfide bond making the biotin cleavable with 50 mM DTT (Fig 3.1). Upon addition of biotin to GccF for labelling the biotin and GccF precipitated out of solution and was unable to be re-solubilised at the necessary conditions for labelling to proceed. It was thought that the large size and long spacer arm of the biotin may have contributed to the insolubility. A different form of biotin with a shorter spacer arm (13.5 Å) was found to not precipitate in solution and was therefore used in subsequent labelling experiments.



Sulfo-NHS-SS-Biotin
M.W 606.69
Mass added to target 389.6 Da
Spacer arm length 24.3 Å



Sulfo-NHS-Biotin
M.W 443.43
Mass added to target 226.3 Da
Spacer arm length 13.5 Å

Figure 3.1 Types of biotin trialled for biotinylation of GccF.

GccF was dissolved in EPPS buffer (pH 9) to a concentration of 1 mg/mL in a volume of 100 µL. A 10 mM solution of biotin was made by the addition of 4.4 mg of biotin to 1 mL of H_2O. This was prepared immediately before use. 10 µL of 10 mM biotin was added to the GccF and the sample incubated with rotation at room temperature for 30 minutes. Glycine (0.1 M) was then added and the sample was left to incubate for 20 minutes to stop the labelling reaction. Acetic acid was added to a final concentration of 1% to the solution before labelled and unlabelled bacteriocin were separated using RP-HPLC. Elution was monitored by the absorbance at 214 and 280 nm, and peaks were collected, lyophilized, and stored at -80°C. Lyophilized samples were re-suspended in 20 µL of H_2O and analysed by ESI-MS (2.1.18). Bacteriocin bioassays could not be used to test activity of biotin labelled GccF as all conditions trialled were unable to separate out wild-type GccF from labeled; all fractions spotted onto a bioassay plate showed a zone of clearing similar to that of the unlabelled GccF control.

3.2.10 Biotin labelled GccF purification

The capacity of the streptavidin magnetic beads is stated as being 5-10 µg of biotinylated protein per mg of beads. 100 µL of MagPrep™ Streptavidin beads (Novagen, USA) were placed into a sterile low-binding 1.5 mL Eppendorf tube (Protein LoBind, Eppendorf, Australasia). The tube was placed on a magnetic rack and the beads were allowed to settle before the supernatant was decanted and discarded. The beads were removed from the magnetic rack washed 3 times with 100 µL of TBS pH 8.0 for five minutes, and the wash buffer was removed using the magnetic rack. After the last wash was decanted, 12 µg of the mixed species of GccF, in TBS pH 8.0, was added to the beads and incubated with rotation for 15 minutes at room temperature. A pH of 8.0 was used to promote the attachment of biotin on the N-terminal primary amine instead of the primary amine located on lysine-32. After incubation the beads were then placed on the magnetic rack and the supernatant removed after the beads had settled. The beads were then removed from the magnetic rack and washed 3 times in 100 µL of TBS pH 8.0 with each wash taking 5 minutes. The wash fractions were collected for analysis by RP-HPLC. To elute the biotinylated GccF from the streptavidin beads a slightly modified method from Holmberg, *et al.* (2005) was used in which the washed beads were resuspended in 100 µL of pure H_2O then slowly heated to 70°C, held for 1 s, then allowed to cool to room temperature. The beads were then magnetically collected, the supernatant removed and its concentration measured along with the concentrations of the wash fractions using the NanoDrop (2.1.10) to roughly determine whether all the biotinylated and wild-type GccF had been removed from the beads. Finally the beads were washed three times in 100 µL of TBS pH 8.0 and stored at 4°C for further use. The supernatant and wash fractions were then freeze dried, resuspended in a total volume of 100 µL of pure H_2O/1% acetic acid, and subjected to RP-HPLC (Ultimate 3000, Dionex, Germany) using a Jupiter 5µm C_{18} 300 Å column (250 x 4.6 mm, Phenomenex) using the following gradient: A: H_2O, 0.1% TFA, and B: Acetonitrile, 0.8 % TFA. The gradient started at 10 % B and increased to 60% B over 31 minutes at a constant flow rate of 1 mL/min. The %B was then increased to 100% over 2 minutes, and held for 2

minutes until it was returned to the starting conditions. Elution was monitored using a photodiode array detector (Dionex) at 214 nm and 280 nm and Chromeleon software V 6.80.

Tris buffered saline (TBS): 50 mM Tris
150 mM NaCl
pH 8.0

3.2.11 2D gel sample preparation

2D gel preparation was performed following a modified protocol from Wu *et al*, 2009. 30 mL of MRS was inoculated with an overnight culture of *L. plantarum* ATCC 14917 (1:50 v/v) the culture was incubated at 30°C and 1 mL aliquots were removed hourly and the OD_{600} was measured until early log phase was reached (~OD_{600} 0.3, 2.1.6). The culture was then split into two 40 mL aliquots and one aliquot was treated with GccF IC_{100} ~4nM (Stepper *et al*, 2011). The cultures were left to grow for a further 60 minutes with OD_{600} being read every 5 minutes to ensure bacteriostatic activity. The two cultures were then centrifuged at 4500 x g for 10 minutes the supernatant was removed and the pellet was washed three times in PBS. The cell pellets were then freeze-dried overnight. Once dried the pellet was ground into a fine powder in the presence of liquid nitrogen. The powdered samples were then suspended in cold acetone containing 10% TCA and were placed at -20°C overnight. Samples were then centrifuged at 13,000 x g for 20 minutes, washed three times with cold acetone and were freeze-dried overnight. The powdered samples were suspended in lysis buffer and sonicated for 5x 5 minutes intervals with 2 minute incubations on ice. After sonication the samples were centrifuged at 12,000 x g for 10 minutes, supernatant was collected then stored at -80°C. Protein concentrations were determined using either the Bradford assay or the EZQ® Protein Quantitation Kit (2.1.7). Samples containing 20 µg of total protein were subjected to SDS-PAGE (2.1.13).

Lysis buffer: 7 M Urea
2 M Thiourea
4% CHAPS (w/v)

3.2.12 Strip electrofocusing

7 cm pH 4-7 IPG strips

10 µg of sample was added to rehydration buffer in a total volume of 125 µL and pipetted into a 7 cm ceramic strip holder (GE Healthcare). IPG strips were placed gel side down in the strip holders and overlayed with 1 mL Plus One Dry strip coverfluid (Amersham Biosciences). The ceramic strip holders were then placed in an IPGphor (Ettan IPGphor II™, GE Healthcare) and the strips were electrofocused using the following program;

Voltage Mode	**Voltage (V)**	**Time (h:min)**
1 Step and Hold	300	1:00
2 Gradient	1000	1:00
3 Gradient	5000	2:00
4 Step and Hold	5000	0:30
Total		4:30

After electrofocusing the IPG strips were then incubated in 5 mL equilibration buffer/DTT for 15 minutes, the solution was decanted and 5 mL equilibration buffer/iodoacetamide was added. The strips were equilibrated for a further 15 minutes and the solution decanted. The strips were then run immediately on SDS-PAGE or stored at -80°C for later use.

18 cm pH 4-7 IPG strips

100 µg of sample was added to rehydration buffer in a total volume of 340 µL and pipetted into a 18 cm ceramic strip holder (GE Healthcare). IPG strips were placed gel side down in the strip holders and overlayed with 3 mL Plus One Dry strip coverfluid (Amersham Biosciences). The ceramic strip holders were then placed in an IPGphor (Ettan IPGphor II™, GE Healthcare) and the strips were electrofocused using the following program;

Voltage Mode	Voltage (V)	Time (h:min)
1 Step and Hold	500	2:00
2 Gradient	1000	8:00
3 Gradient	8000	3:00
4 Step and Hold	8000	2:30
Total		15:30

After electrofocusing the IPG strips were then incubated in 12 mL equilibration buffer/DTT for 15 minutes, the solution was decanted and 12 mL equilibration buffer/iodoacetamide was added. The strips were equilibrated for a further 15 minutes and the solution decanted. The strips were then run immediately on SDS-PAGE or stored at -80°C for later use.

24 cm pH 3-10NL IPG strips

100 µg of sample was added to rehydration buffer in a total volume of 450 µL and pipetted into a 24 cm reswelling tray (GE Healthcare). IPG strips were placed gel side down in the reswellign tray and overlayed with 3 mL Plus One Dry strip coverfluid (Amersham Biosciences). An Ettan IPGphor Manifold (GE Healthcare) was placed onto the IPGphor (Ettan IPGphor II™, GE Healthcare) and 108 mL of Plus One Dry strip coverfluid was added evenly to the 12 manifold channels on the manifold. The reswelled IPG strips were placed under the cover fluid, gel side up, in the manifold channels. The required number of precut paper wicks (GE Healthcare) were counted out and 150 µL of distilled water was added to each wick. The wicks were placed at either end of the IPG strips so that they overlaped the end of the gel, the electrode assembly was seated on top of the wicks and locked into the closed position. The IPG strips were electrofocused using the following program;

Voltage Mode	Voltage (V)	Time (h:min)
1 Step and Hold	500	2:00
2 Gradient	1000	7:00
3 Gradient	10000	3:00
4 Step and Hold	10000	3:30
Total		15:30

After electrofocusing the IPG strips were then incubated in 15 mL equilibration buffer/DTT for 15 minutes, the solution was decanted and 15 mL equilibration buffer/iodoacetamide was added. The strips were equilibrated for a further 15 minutes and the solution decanted. The strips were then run immediately on SDS-PAGE or stored at -80°C for later use.

Rehydration buffer: 7 M Urea
2 M Thiourea
2% (w/v) CHAPS
0.5% (v/v) IPG buffer pH 4-7 or 3-10
0.002% (w/v) Bromophenol blue
12 µL/mL of Destreak (GE Healthcare)

Equilibration buffer: 6M Urea
75 mM Tris-HCl pH 8.8
20.3% (w/v) Glycerol
2% (w/v) SDS
0.0002% (w/v) Bromophenol blue
100 mg per 10 mL DTT or 250mg per 10 mL Iodoacetamide

3.2.13 2D gel Electrophoresis

12.5% SDS-PAGE gels (2.1.13) were cast without a stacking layer using either the Mini-Protean® 3 system for the 7 cm IPG strips (BioRad Industries, Milan, Italy) or the Ettan DALT*six* system for the 18 and 24 cm IPG strips (GE Healthcare) following manufacturers' guidelines. The strips were inserted onto the SDS-PAGE gels which were overlayed with 1% agarose/0.002% bromophenol blue and run at 25°C until the dye front was ~1 cm from the bottom of the gel. Gel running conditions were as follows;

IPG strip length (cm)	**mA/gel or W/ gel**	**h:min**
7	10mA 20mA	0.15 1:30
18	2W 17W	0:45 4:00
24	2W 17W	0:45 4:00

The proteins were visualized with coomassie silver stain (2.1.14) and gels were scanned on a Typhoon FLA 9000 (GE Healthcare) and analysed using DeCyder 2D software Ver. 7.0 using the settings for coomassie blue stain.

3.3 CELL GROWTH AND GCCF PRODUCTION

L. plantarum KW30 is thought to produce GccF in early stationary phase as reported by Kelly *et al* (1996). Measuring the growth curve of *L. plantarum* KW30 is an important step in determining when the production of GccF occurs, as knowledge of the time GccF production is highest in the growth curve is necessary to maximise yields and following experiments would also benefit from an awareness of GccF production. Measuring the pH along with the bacterial growth allows for a quick method to check subsequent cultures in order to roughly estimate the growth phase of *L. plantarum* KW30.

Bacterial cell growth occurs in four phases (Fig 3.2): there is the initial lag phase in which bacteria adapt to their growth conditions; second is the exponential or logarithmic phase where the number of new bacteria appearing per unit time is proportional to the present population. If the growth of the cell is not limited for example by restricted nutrients, then doubling will continue at a constant rate; the third phase is the stationary phase in which the growth rate slows due to nutrient depletion and the accumulation of toxic materials. This phase is generally seen as a leveling out of the growth on a curve; the final phase is the death phase where the bacteria run out of nutrients and die.

Cell growth can be determined by measuring the OD_{600} of a cell culture at specific intervals and plotting them onto a graph. Initially the OD should increase proportionally as the number of cells present in the culture increases. Eventually as the stationary phase is reached the OD_{600} should level. One problem in using OD as a measure of cell growth is that it does not distinguish between live and dead cells, this means that once the 'death phase' is reached a decrease in cell number will not be observed using $OD_{600.}$ Another limitation of this method is that as the OD increases, the sample will eventually need to be diluted to be measured, introducing error into the results.

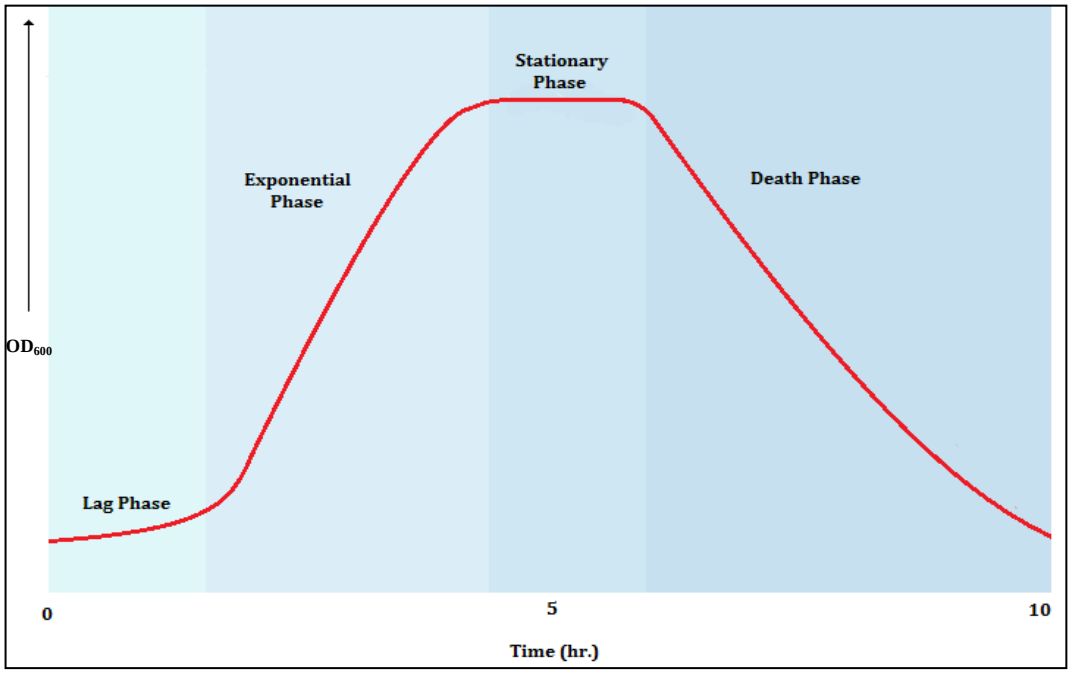


Figure 3.2 A typical bacterial growth curve showing the four growth phases.

The bacteriocin bioassay plate is the most common method used to determine and detect bacteriocin activity against other microorganisms. It involves the use of a known organism susceptible to the activity of the bacteriocin (an indicator species/strain), or an organism that you wish to test for bacteriocin susceptibility, along with a solution containing the bacteriocin. In this technique, the indicator strain is incorporated into the agar plate. The bacteriocin is applied either as 1-2 µL drops placed on the agar surface, into cuts placed into the agar, or from filter-paper discs saturated with the bacterioin which are then placed on the surface of the medium. Plates are then incubated at an appropriate temperature and a clearing around the point of bacteriocin application indicates activity. When activity assays are linked to the growth curve, the time of maximum GccF production can be determined.

3.3.1 RESULTS AND DISCUSSION

L. PLANTARUM KW30 GROWTH AND GCCF PRODUCTION

In order to determine the time period over which GccF is produced, the growth of *L. plantarum* KW30 was monitored as described in Section 2.2.1. Results are shown in Figure 3.3.

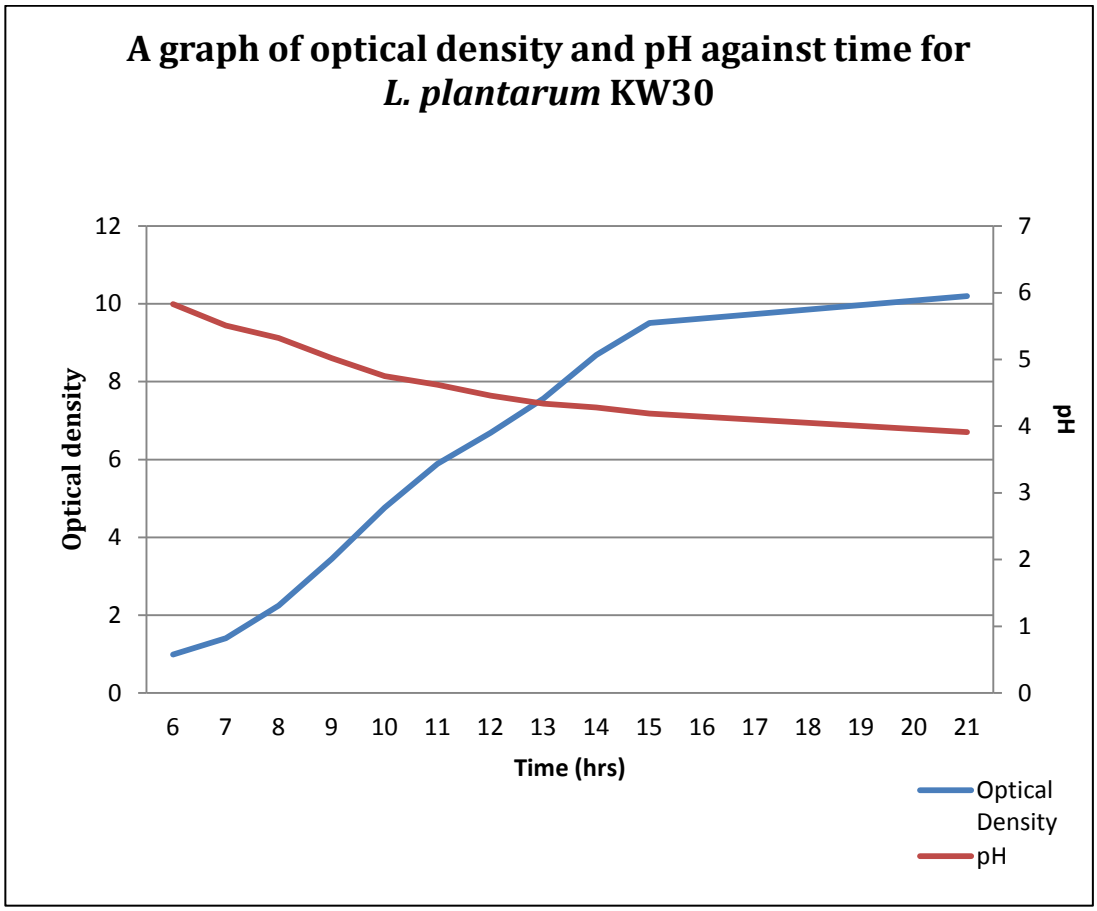


Figure 3.3 Monitoring the OD_{600nm} and pH of a *L. plantarum* KW30 culture.

The increase in optical density is accompanied by a corresponding decrease in pH which is typical for *Lactobacillus* species. *L. plantarum* KW30 apparently enters the exponential phase at approximately seven hours of incubation, entering stationary phase after fifteen hours. The pH of the culture is initially ~6, and decreases throughout growth until it reaches 4.6 when the cells are in stationary phase decreasing to around 3.8 as stationary phase progresses.

Samples of culture from growth trials (Fig 3.3) were collected and assayed for bacteriocin activity (Fig. 3.4) using a plate bioassay. *L. plantarum* ATCC 14917 was the organism used for the bioassay as it has, thus far, been found to be the most sensitive to the bacteriostatic effect of GccF (Stepper *et al*, 2011).

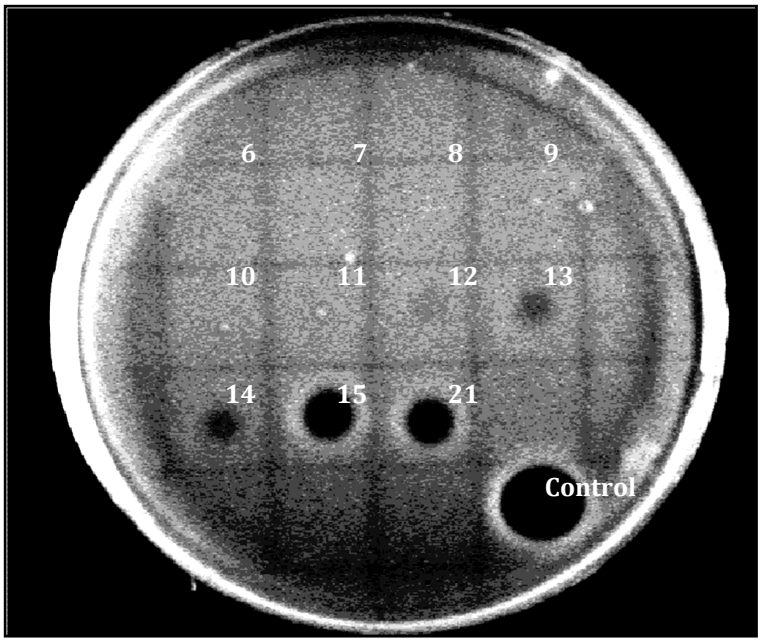


Figure 3.4 Bacteriocin bioassay using indicator organism *L. plantarum* ATCC 14917. Culture samples collected from the growth assay at times indicated (hrs) were spotted in 2 µL aliquots.

Figure 3.4 shows the results of the bioassay for GccF using *L. plantarum* 14917 as the indicator strain. The earliest a zone of clearing is observed is at 12 hours which corresponds to ~OD_{600} 6.8. When this is compared to the growth curve in Figure 3.3, it corresponds to mid-late log phase and a pH of ~4.5. As the growth continues, an increase in the size of clearing is observed. At 15 hours of growth, the biggest clearing is observed. However, due to time contraints the points between 15 and 21 hours were not collected, resulting in uncertainty about the time of maximum clearing. Direct comparison between 15 and 21 hours of growth however show that bacteriocin activity seems to have decreased after 21 hours which may indicate that over time GccF is degraded or re-absorbed. Production of GccF in late exponential phase has been reported for several bacteriocin producing microorganisms where the bacteriocin is produced as a primary metabolite (de Vuyst *et al*, 1996). The late stage of maximum production may be in response to

increased cell density and the need to protect its environmental niche as the population increases and nutrients are limited. Alternatively it may be constitutively expressed. In most microorganisms bacteriocin production occurs in response to stress such as competiton for its environmental niche from other bacterial species. (De Vuyst *et al*, 1996; Nilsen *et al*, 1998). Therefore GccF production may occur earlier if the culture is stressed but is produced at late exponential phase under laboratory conditions. As GccF is produced at a high cell density it may be that its production is regulated by a quorum sensing system where a certain cell density/threshold is required for gene expression as is the case with many bacteriocins (Diep *et al*, 2009).

3.4 DETECTION OF GlcNAcylated PROTEINS IN *L. PLANTARUM* KW30

This section investigates the possibility of other GlcNAc modified proteins being present within the proteome and secretome of *L. plantarum* KW30. Is GccF unique, with the enzymes required for synthesis and glycosylation of the protein being confined to the products of the *gcc* cluster, or are the enzymes able to process other proteins in the KW30 proteome?

Western blotting is an analytical technique which is used to detect specific proteins or modifications present within a sample that has been separated by SDS-PAGE. It is a sensitive and specific technique due to the use of antibodies to detect specific antigens. Using this technique it is possible to identify a low abundance protein, or a specific post-translational modification, in a complex mixture from a wide range of sources such as whole tissue, cell cultures, or the environment. Low abundance samples can be spotted directly onto a membrane or separated by electrophoresis before being transferred to the membrane. The two most commonly used membranes are nitrocellulose and polyvinylidenedifluoride (PVDF). To transfer the proteins from a gel to a membrane, two methods are commonly used. Passive transfer involves placing a membrane on the gel, and surrounding the sandwich in damp filter paper and applying pressure, to encourage transfer of the proteins onto the membrane. Active transfer uses an electric current to move proteins from the gel onto the membrane.

Antibodies specific to the target are used to probe the blot. A solution, usually a blocking solution, containing the antibodies is incubated with the membrane and the antibodies should bind to any proteins or post-translational modifications which they are specific to. Secondary antibodies raised against the primary antibodies are used for detection as these are generally linked to biotin or to a reporter enzyme most commonly alkaline phosphatase or horseradish peroxidase. The secondary antibody will bind to the primary antibody allowing it to be detected either by visualisation of a colour change, detection of fluorescence, or radiography (Burnette, 1980; Kurien and Scofield, 2006).

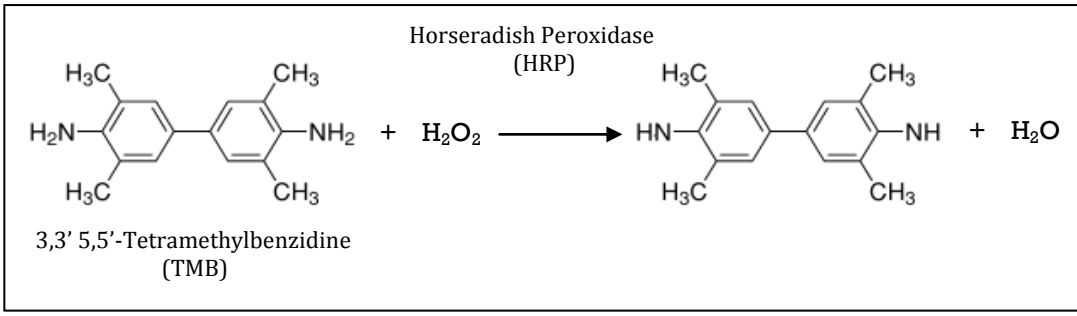


Figure 3.5 Schematic of the reaction that occurs between TMB and the HRP enzyme.

3,3' 5,5'-tetramethylbenzidine (TMB) is a commercially available substrate and when used with the enzyme horseradish peroxidase (HRP) produces a colour change from colourless to blue/purple. The colour change occurs when TMB is oxidised as a result of the oxygen radicals which are produced when hydrogen peroxide, present in the TMB solution, is hydrolysed by HRP (Fig 3.5). There are a wide variety of HRP conjugated antibodies available for purchase which have been raised against numerous animal species. Mouse-anti-*O*-GlcNAc antibodies are also commercially available and recognise *O*-GlcNAc in a β-*O*-glycosidic linkage to both serine and threonine. In this section of work the mouse-anti-*O*-GlcNAc antibodies have been used in western blotting to detect proteins modified by *O*-GlcNAc with anti-mouse-HRP used as a secondary antibody.

Lectins are sugar-binding proteins which play an important role in binding soluble carbohydrates or carbohydrate moieties that are contained within a glycoprotein or glycolipid. They are involved in important biological functions such as cell adhesion, glycoprotein synthesis, protein targeting, and immunity (Chrispeels and Raikhel, 1991; Rini and Lobsanov, 1999). Purified lectins are used in a number of applications as both medical and biochemical tools.

Lectins such as concanavalin A have been widely used in affinity chromatography methods to purify glycoproteins (Saleemuddin and Husain, 1991). In affinity chromatography an immobile phase generally made up of a gel matrix, often agarose, is exposed to cell lysate. The target molecule in the cell lysate will have a defined property which may be exploited for its purification using affinity chromatography methods. The gel matrix in the immobile phase has ligands

attached which will utilize this property capturing the target molecule within the immobile phase whereas all other molecules will not be captured. The solid medium can then be removed and washed and the target molecule eluted by changes in pH, high salt concentration, or addition of a competitive binding molecule (Chaiken, 1989).

Wheat germ lectin/agglutinin (WGA) is a 36,000 molecular weight protein which consists of two identical subunits isolated from *Triticum vulgaris*. It binds to two sugars, *N*- acetylglucosamine (GlcNAc) and *N*-acetylneuraminic acid (sialic acid), and is capable of binding to a single GlcNAc, as well as oligosaccharides containing terminal GlcNAcs. Bacterial cell wall peptidoglycans, chitin, cartilage glycosaminoglycans and glycolipids will also be recognised by wheat germ agglutinin. In this section of work agarose bound, succinylated WGA (which does not bind sialic acid residues) has been used in affinity chromatography in an attempt to capture proteins containing exposed GlcNAc residues within *L. plantarum* KW30 protoplasts.

For this work protoplasts were used, these are bacterial cells which have had their cell walls removed. For *L. plantarum* KW30 mutanolysin was used to remove the cell wall and any peptidoglycans because they contain GlcNAc which will interfere with WGA affinity chromatogrpahy as well as western blotting using anti-*O*-GlcNAc antibodies. As protoplasts are susceptible to osmotic stress they are easy to lyse, an advantage as *L. plantarum* KW30 cells become harder to lyse after longer periods of growth, and an easy and gentle method of lysis is necessary for some experiments (Salazar and Asenjo, 2007).

3.4.1 RESULTS AND DISCUSSION

L. PLANTARUM KW30 WESTERN BLOTTING

As GccF contains both an *S*-linked and an *O*-linked GlcNAc modification it was important to determine the specificity of the mouse anti-*O*-GlcNAc antibody. ^{deOGlcNAc}GccF, GccF^{-HC}, and WT-GccF were used where; ^{deOGlcNAc}GccF is GccF without the *O*-linked GlcNAc residue; GccF^{-HC} is GccF without the last two (histidine and cysteine) residues on the C- terminus and the *S*-linked GlcNAc; and WT-GccF is wild-type GccF purified from *L. plantarum* KW30.

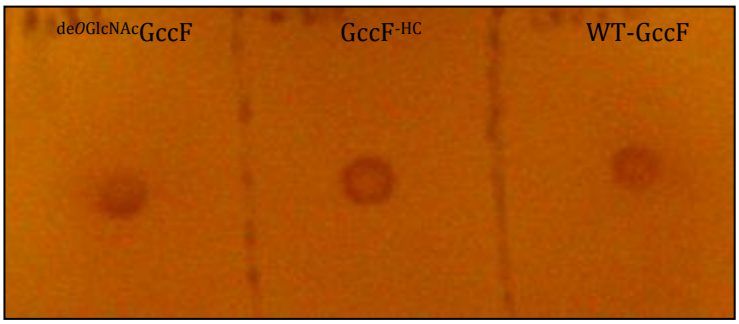


Figure 3.6 Western blot probed with mouse anti-*O*-GlcNAc antibodies to determine the specificity of the mouse anti-*O*-GlcNAc antibodies for; ^{deOGlcNAc}GccF; $^{-HC}$GccF; and wildtype GccF.

Equivalent molar ratios of ^{deOGlcNAc}GccF, $^{-HC}$GccF, and wildtype GccF were spotted onto a nitrocellulose membrane which was probed with the mouse anti-*O*-GlcNAc antibody (Sigma Aldrich). Figure 3.6 shows the result of the dot blot and as shown the antibody detects all three variants of GccF. As wild-type GccF contains two GlcNAc modifications it would be predicted that if the antibody did detect the *S*-linked GlcNAc that this would only be weakly detected which was the case. However, that the anti-*O*-GlcNAc antibody detected both the ^{deOGlcNAc}GccF and GccF^{-HC} shows that it is not specific for *O*-linked GlcNAc residues.

GlcNAc modifications are used as signalling molecules in eukaryotic organisms (Hanover *et al*, 2010). As GccF contains two GlcNAc modifications it may be that its production is mediated by a GlcNAc signalling system within *L. plantarum* KW30. *L. plantarum* KW30 protoplast lysate was separated on a 16.5% tricine gel, transferred to a nitrocellulose membrane and probed with the anti-*O*-GlcNAc

antibody from Figure 3.6 and an anti-C14 antibody raised against the last 14 amino acids from the C-terminal end of GccF.

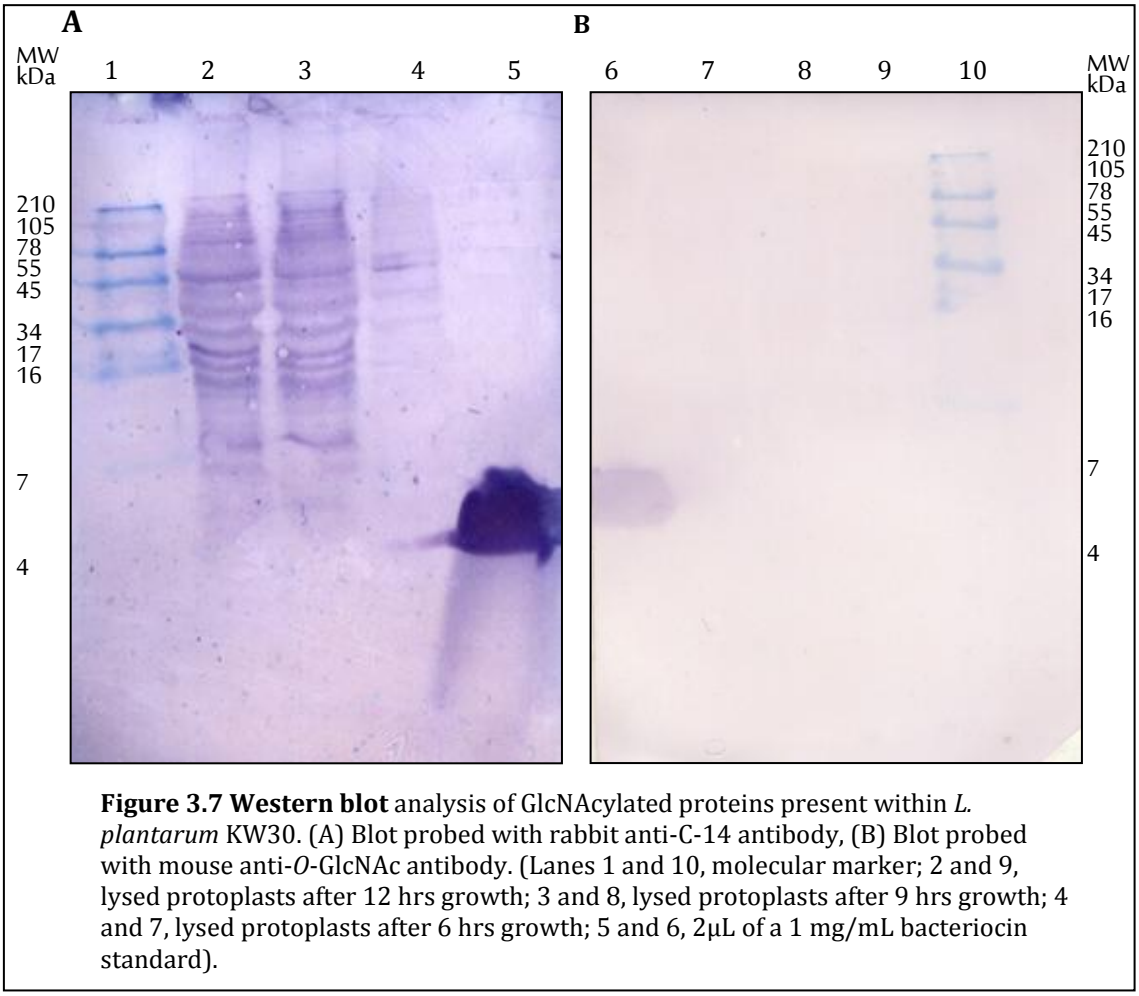


Figure 3.7 Western blot analysis of GlcNAcylated proteins present within *L. plantarum* KW30. (A) Blot probed with rabbit anti-C-14 antibody, (B) Blot probed with mouse anti-*O*-GlcNAc antibody. (Lanes 1 and 10, molecular marker; 2 and 9, lysed protoplasts after 12 hrs growth; 3 and 8, lysed protoplasts after 9 hrs growth; 4 and 7, lysed protoplasts after 6 hrs growth; 5 and 6, 2µL of a 1 mg/mL bacteriocin standard).

Figure 3.7 shows the results of a western blot of the *L. plantarum* KW30 protoplast lysate probed with both the anti-C-14 and anti-*O*-GlcNAc antibodies. GccF was used as a control in these experiments as it is known to contain GlcNAc modifications. Lane 1 and 10 shows the pre-stained molecular size marker which was transferred to the membrane to allow estimation of band size. As can be observed, the higher molecular weight bands have transferred well while the lower molecular weight bands are hard to see with the bands at 7 kDa and 4 kDa barely visible. This indicates that the lower molecular weight bands from *L. plantarum* KW30 protoplast lysate may have passed through the membrane as these generally transfer more readily than those of a higher molecular weight. Lanes 2-4 contain cell lysate collected at different growth times. As can be seen, the anti-C-14

antibody detects many proteins present in these fractions demonstrating it is rather non-specific. Lane 5 containing GccF has a large dark band present at ~7 kDa where it traditionally runs, indicating the anti-C-14 antibody has bound. Lanes 7-9 contain protoplast lysate from *L. plantarum* KW30 collected after different growth times, the same as lanes 2-4, however no bands are visible in any of these lanes. Lane 6 contains the GccF control and there is a large band visible in this lane which indicates that the anti-*O*-GlcNAc antibody is viable and that the absence of bands in lanes 7-9 is not due to a problem within the western blot itself and is instead due to the absence of GlcNAcylated proteins in these lanes. The absence of a band corresponding to the molecular weight of GccF in lanes 2-4 and 7-9 is expected as GccF is thought to be exported from *L. plantarum* KW30 once mature. Therefore the majority of GccF would be found in the cell growth medium and is likely present in concentrations too low within the cells for detection using the antibodies for western blotting. Western blots were also performed on protoplast lysates from cells grown for different times (T = 2, 4, 6, 8, 10, 12 hrs) to determine whether GlcNAcylated proteins were produced at certain stages of growth. The results from these blots also showed no bands for the membrane when probed with the anti-*O*-GlcNAc antibody. Both the gels and membrane were stained, in Coomassie blue and ponceau S respectively, after each transfer to ensure that the proteins had transferred onto the membrane. As shown in Figure 3.7 the anti-*O*-GlcNAc antibody detects both *S*- and *O*- linked GlcNAc modifications, from the results seen in the western blots there appear to be no other *S*- and *O*-linked GlcNAcylated proteins present within *L. plantarum* KW30.

L. PLANTARUM KW30 WHEATGERM AGGLUTININ PULL-DOWNS

Wheat-germ lectin which recognises *N*-linked GlcNAc residues was used in an attempt to detect any *N*-GlcNAcylated proteins present within *L. plantarum* KW30.

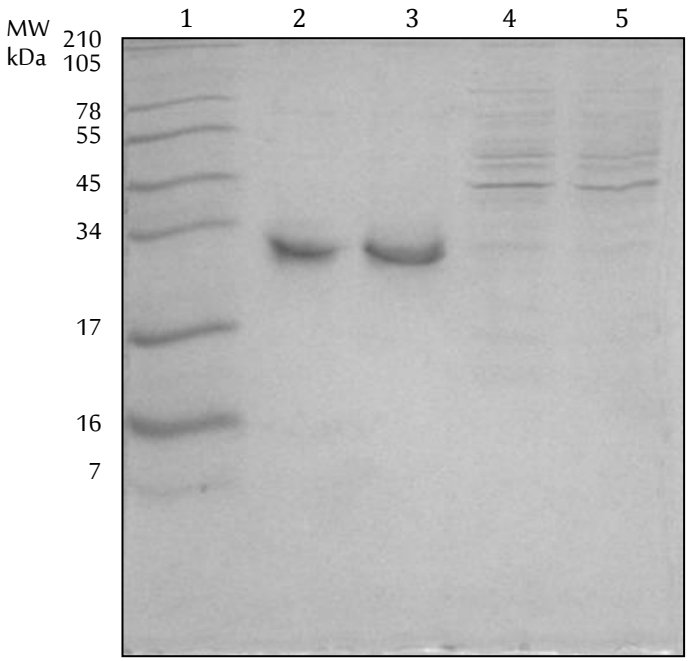


Figure 3.8 WGA agarose pull-down for *L. plantarum* KW30 protoplast lysate. (Lane 1, molecular marker; 2, WGA agarose; 3, WGA agarose after incubation with protoplast lysate; 4, first wash with lectin buffer; 5, second wash with lectin buffer).

Figure 3.8 shows the result of a WGA agarose pull-down for *L. plantarum* KW30 protoplast lysate. Lane 2 is a control for the lectin agarose resin where no protein was incubated with the agarose, there is a single dark band present at approximately 32 kDa, due to the absence of protoplast lysate in this lane the band observed must be from the wheat-germ agarose. Wheat-germ lectin consists of identical subunits with a mass of 36kDA, it may be that the band observed is lectin itself, and that the change in its mass is due to some portion of it remaining bound to the agarose resin, or that it does not run at its true mass. Lane 3 is the lectin resin incubated with protoplast lysate after washing and once again a dark band is present at approximately 32 kDa, as this band is also present in lane 2 which is a negative control, it cannot be attributed to proteins binding to the WGA agarose. Lanes 4 and 5 are wash steps to elute non-specifically bound proteins from the resin, as can be seen there are many bands present in these lanes which is expected.

3.5 INVESTIGATIONS INTO THE MECHANISM OF ACTION OF GccF

In this section FITC labeled GccF, GccF^{deOGlcNAc}, and GccF^{-HC} were used to try and dissect the role of the GlcNAc residues in the antibacterial activity of GccF.

Fluorescence microscopy is a powerful technique which uses an optical microscope to observe fluorescence and phosphorescence. Fluorescence microscopy has many applications and can be used to observe target molecules within a complex sample. Typically a sample is illuminated with light of specific wavelength which excites a specific fluorophore usually covalently attached to a bio-molecule. The fluorescence is emitted at a higher wavelength than that which was used for illumination and this can be detected and quantified. Labelling samples with fluorescent stains is a common technique used to visualize many different bio-molecules in a variety of media. There are many fluorescent molecules available which bind to specific biological molecules such as DNA, RNA, hormones and proteins. Some of these molecules, called fluorophores, can be chemically linked to a molecule and used to locate it within a sample. Samples labelled with a flurophore must be kept in the dark to prevent photo-bleaching which results in a loss of fluorescence yield over a short time period.

Fluorescein isothiocyanate (FITC) is a fluorophore which is used in a wide-range of applications. It is functionalized with an isothiocyanate group, –N=C=S, (Fig 3.9) which is reactive towards nucleophiles such as the primary amine and sulphydryl groups on proteins.

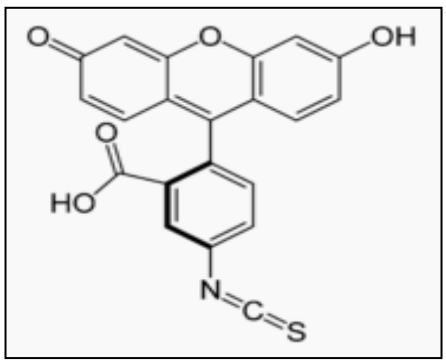


Figure 3.9 Fluorescein isothiocyante

3.5.1 RESULTS AND DISCUSSION

GCCF MICROSCOPY

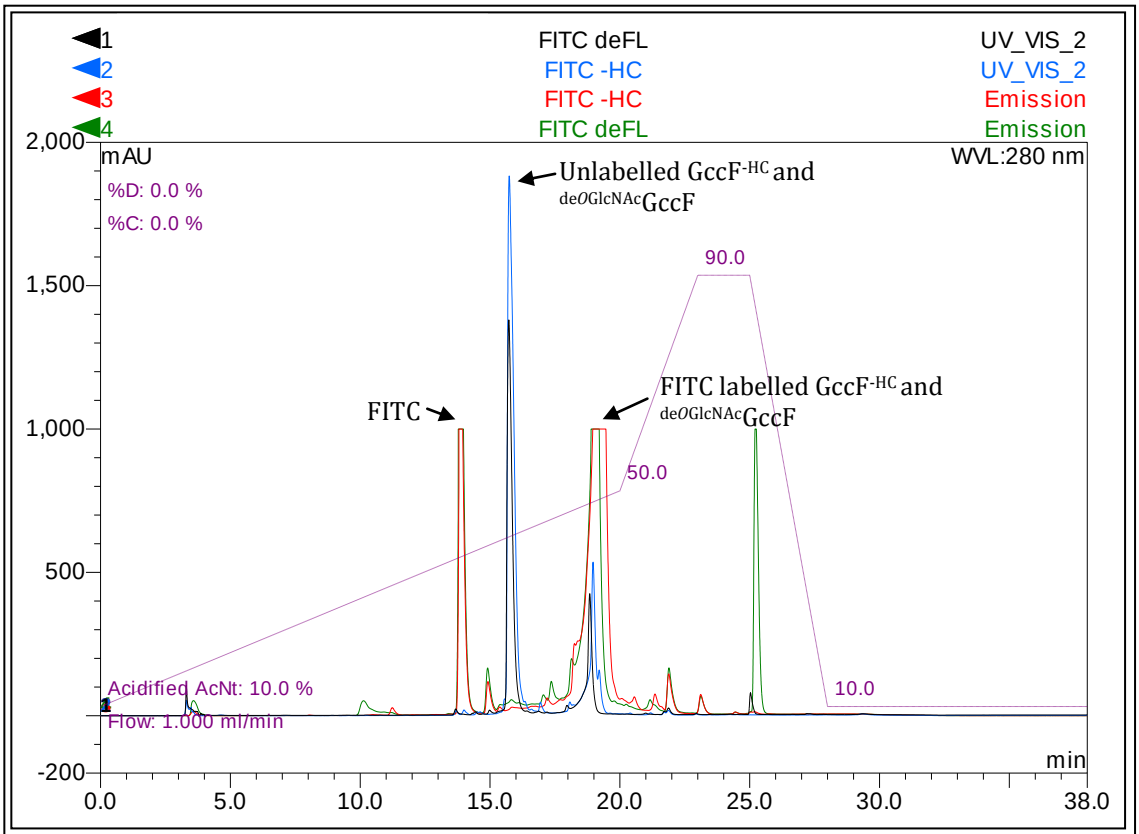


Figure 3.10 RP-HPLC purification of FITC labelled GccF^{-HC} and ^{deOGlcNAc}GccF using a C_{18} column and monitored by absorbance at 280 (UV_Vis_2) and 520 nm (Emission).

To separate out the unlabelled and FITC labelled GccF^{-HC} and GccF^{deOGlcNac} as well as free FITC after labelling, RP-HPLC was used. Emission was monitored at 280 nm to detect protein, and 520 nm to detect the fluorescence from FITC (Fig 3.10). As seen in the figure FITC labelled GccF^{-HC} and ^{deOGlcNAc}GccF elute off the column at later times than unlabelled GccF^{-HC} and ^{deOGlcNAc}GccF allowing for easy separation. The fractions collected were tested for bacteriostatic activity using a bacteriocin bioassay and no zone of clearing was present for FITC-^{deOGlcNAc}GccF and only a barely visible zone of clearing was observed for FITC-GccF^{-HC} when compared to wild-type GccF. The results of the bioassay are consistent with those seen for unlabelled ^{deOGlcNAc}GccF and GccF^{-HC} as reported in the paper by Stepper *et al*, (2011) where ^{deOGlcNAc}GccF was inactive and GccF^{-HC} was 65-fold less active than wild-type GccF.

After purification the effects of FITC labelled ^{deOGlcNAc}GccF, $^{-HC}$GccF, and wildtype GccF on *L. plantarum* ATCC 14917 or 8014 cells was examined and is shown in Figure 3.11.

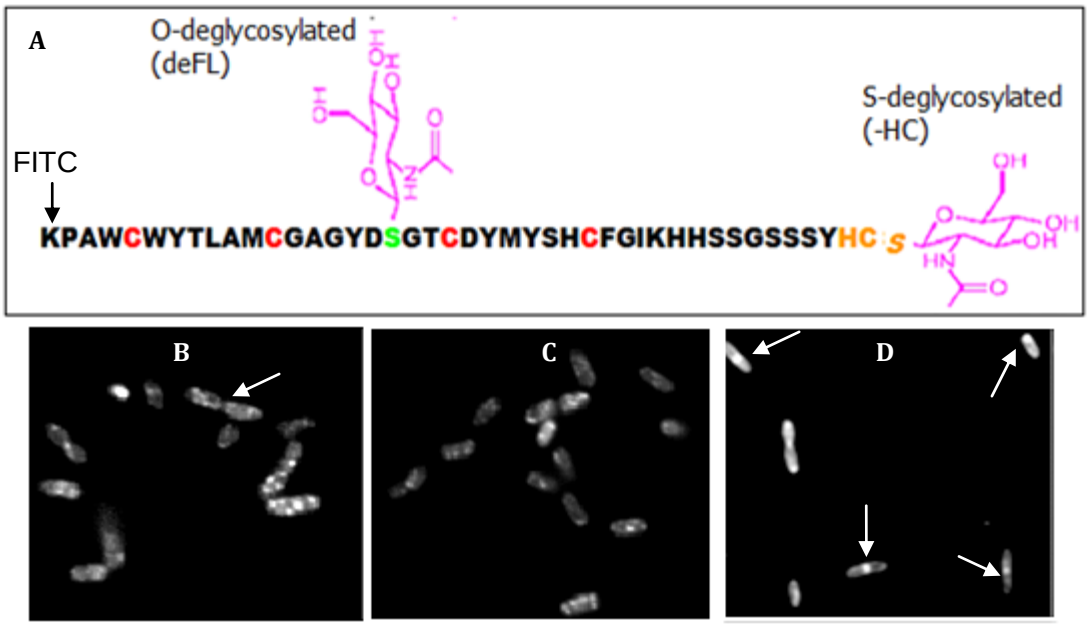


Figure 3.11 FITC Labelling (A) Schematic of GccF showing GlcNAc residues and location of FITC label. (B) FITC labelled ^{deOGlcNAc}GccF incubated with *L. plantarum* ATCC 14917; (C) FITC labelled^{-HC}GccF incubated with *L. plantarum* ATCC 14917; (D) FITC labelled wildtype GccF incubated with *L. plantarum* ATCC 8014 (courtesy of Dr Shilpa Shastri).

Figure 3.11 (A) indicates the predicted position of the FITC label present on GccF, as labelling of the N-terminal primary amine was regulated by pH. Image (B) shows the most susceptible indicator strain *L. plantarum* ATCC 14917 after incubation with 300 nm FITC-^{deOGlcNAc}GccF. The image shows spots of intense fluorescence where FITC^{deOGlcNAc}GccF has localized on the cells. It is also possible to see the point of cell division as indicated by the arrow. Image (C) shows *L. plantarum* 14917 after incubation with 300 nm GccF^{-HC} and the image shows a similar pattern to that seen in (B) where there are spots of intense fluorescence where $^{-HC}$GccF has localized on or within the cell. (D) Shows the less susceptible indicator strain *L. plantarum* ATCC 8014 incubated with wild-type GccF (Wild-type GccF work was performed by Dr Shilpa Shastri). The image shows that WT-GccF appears uniformly along the cell except at the point of cell division or at the ends of the cells, as indicated by the white arrows, where a concentrated area of

fluorescence is seen. These results imply that a protein or proteins involved in cell division may be targeted by GccF, resulting in bacteriostasis. It is possible that GccF prevents the cells from dividing causing a suspended pause in cell growth until GccF is removed or proteolysed. GccF without the *O*-GlcNAc or the two C-terminal residues and the *S*-linked glycan do not localise to this position, but rather localise along the length of the cells. It has previously been shown that *O*-deglycosylated GccF is completely inactive and that GccF^{-HC} is also less active than the full length GccF, by a factor of 65 (Stepper *et al*, 2011). The results from labelling studies therefore confirm that both the *O*-linked and *S*-linked glycans play a role in bacteriostasis, and indicate that their role may be to disrupt cell division, although the exact mechanism for this is still unknown.

3.6 FUNCTIONAL CHARACTERISATION OF GccF

Chromatography is a technique used for separation which involves dissolving a mixture in a 'mobile phase' usually some form of solid support and then passing it through a 'stationary phase' which separates the analyte which is to be measured from all other molecules within the mixture. Chromatography can be used to purifiy proteins or to analyse them.

Affinity chromatography is a method used to capture and purify proteins from complex mixtures. Usually, a molecule of interest or a molecule which binds the molecule of interest is immobilized on a solid support such as agarose beads, magnetic beads, silica resin, or synthetic resins such as nylon. These beads can be used in a number of different ways; in affinity chromatography, column chromatography, batch chromatography, or bed adsorption chromatography.

Figure 3.12 is a schematic representation of a basic batch affinity chromatography where a 'mixture' is added to a batch of resin in order to bind the target molecule. Once sufficient time has passed for binding, the remaining mixture is removed. The resin is then washed with loading buffer to remove non specifically bound proteins after which an elution buffer is added to remove the target protein from the resin. The elution buffer may be of low pH, contain a competitor, or high concentrations of salt. Batch chromatography differs from column and bed adsorption chromatography in that liquid is added to the resin in batches, usually in a container, and is removed by centrifugation or decanting.

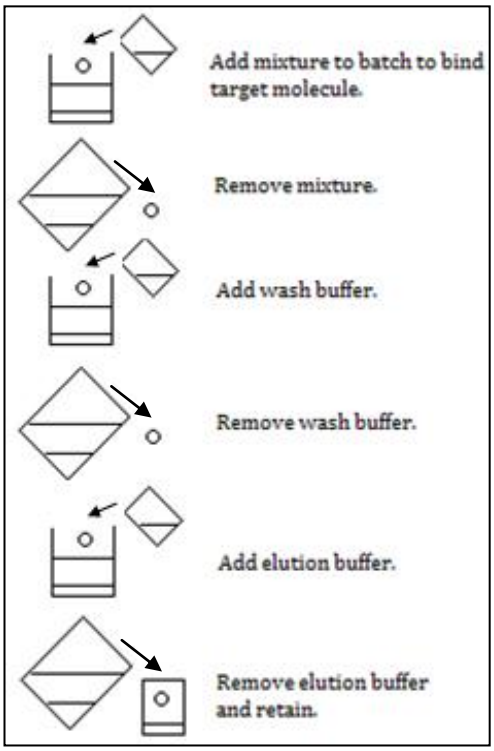


Figure 3.12 Diagram of batch affinity chromatography.

1,1'-carbonyl diimidazole (CDI) agarose contains reactive imidazole carbamates which react with primary amine-groups in aqueous solution to form carbamate linkages. (Fig 3.13). These resins are ideal for the immobilization of organic molecules, some proteins, and small peptides (Bethell and Ayers, 1987).

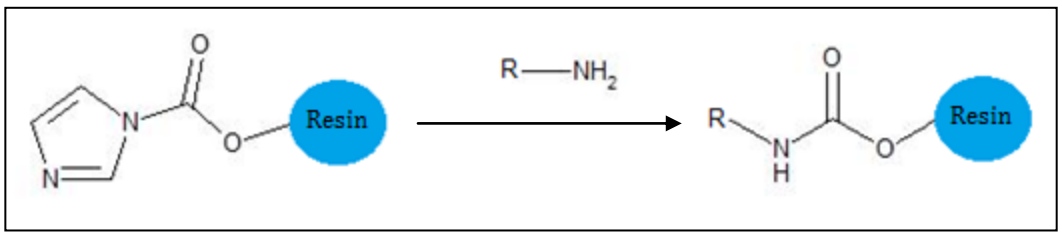


Figure 3.13 Schematic showing the reaction of primary amines with CDI-agarose support.

In order for the CDI-agarose coupling reaction to proceed, the protein of interest must be added at the required concentration and at an alkaline pH. The buffers used for the reaction cannot contain primary amines otherwise non-specific binding to buffering molecules will occur. After the coupling interaction has occurred buffer containing a primary amine such as Tris-HCl is added to the resin in order to block any sites on the resin which remain unbound. After the molecule

of interest has bound to the resin it is thoroughly washed to remove any unwanted chemicals before it can be used for affinity capture or purification.

Biotin, a B-complex vitamin, is a small water soluble molecule that plays a role in many cellular processes. As it is a small molecule (244.31 g mol^{-1}) it is often linked to proteins in a process called biotinylation for use in assays, purification and affinity experiments. It is commonly used due to its ability to bind both streptavidin and avidin with high affinity (K_d of ~10^{-14}mol/ L) and its small size helps to prevent it from interfering with the action of the target molecule when bound. EZ-link Sulfo-NHS-biotin (Fig 3.14) is a commercially available short-chain, water soluble biotinylation reagent that binds to antibodies, proteins and other molecules. It is water soluble and reacts with primary amines at pH 7-9, which enables biotinylation to proceed under physiological conditions. EZ-link Sulfo-NHS-biotin specifically binds to streptavidin and avidin in the same way as native biotin.

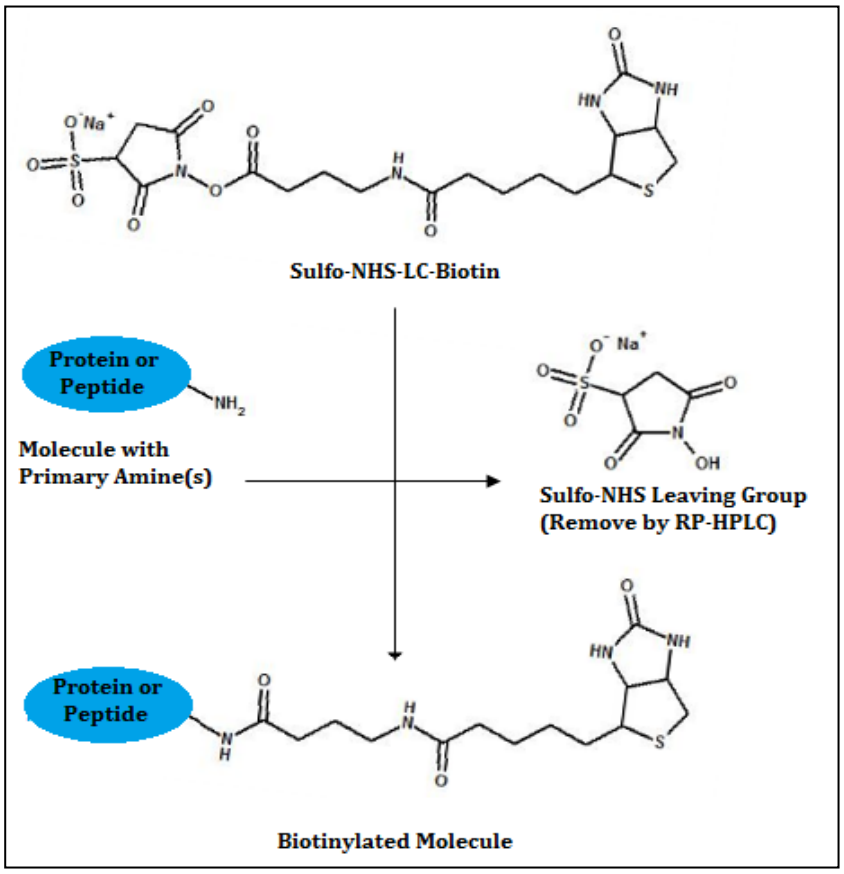


Figure 3.14 Schematic of the biotin labelling reaction (biotinylation).

Because of the strong interaction between streptavidin and biotin, streptavidin is commonly immobilised on beads where it is used as an affinity matrix for biotin containing molecules. Due to the strong nature of the binding, harsh conditions are often required to break the streptavidin-biotin interaction resulting in the denaturation of the target protein. However, experiments have shown that incubation in water at a temperature of 70°C for 5 s can reversibly break the interaction allowing reuse of both biotin and streptavidin (Holmberg *et al*, 2005).

MagPrep™ Streptavidin beads are encapsulated super-paramagnetic-polystyrol particles coated with streptavidin. They have a high binding capacity for biotin and low non-specific binding. The magnetic properties of the beads allow them to be collected using a magnetic rack or magnet, on which they will rapidly coalesce, making them suitable for affinity capture experiments.

Protoplasts were used in this section of work as proteins hypothesised to bind to GccF are located in the membrane and may be part of a protein complex. Protoplasts were chosen over whole cells as they are easy to fractionate and the method for lysis is gentle as opposed to other methods of lysis. Traditional lysis methods are harsh and may shear proteins which is undesirable for this section of work, whereas osmotic stress which is used to lyse protoplasts may allow for proteins to remain in complex which if pulled-down would provide a lot of insight into the mechanism of action of GccF.

In this section of work GccF and the C-11 N-terminal fragment of GccF were immobilised with CDI agarose and biotin for use in affinity chromatography in order to try to identify GccF binding partners/targets using mass spectrometry.

3.6.1 RESULTS AND DISCUSSION

GccF PULL-DOWNS

L. plantarum ATCC 14971

CDI-agarose pull-downs were used to try to capture proteins present in a strain susceptible to GccF that might bind to it, and thus allow identification of potential receptors. It has already been reported that free GlcNAc has a protective effect against the bacteriostatic activity of GccF (Stepper *et al*, 2011). Although the mechanism of protection is not known, it is possible that GlcNAc may act competitively to prevent GccF binding to its target in the cell. Free GlcNAc was therefore used to elute proteins from the GccF linked CDI-agarose in increasing concentrations, this was followed by high salt washes (Section 3.2.8). Wild-type GccF was bound to CDI-agarose probably *via* the N-terminal primary amine as the reaction was carried out at pH 8.5. Binding was confirmed by the bacteriocin bioassay (Fig 3.15 clearing around sample 3) because if GccF had bound to the resin at lysine-32, little or no activity would be seen on the plates in comparison to the control (personal communication, Dr Gill Norris).

Bioassay plates using indicator strain *L. plantarum* ATCC 14917 were run after the CDI-agarose pull-downs with both the membrane and cytosol fractions of washed protoplasts and the resins were spotted onto the plates to detect GccF activity (Fig 3.15).

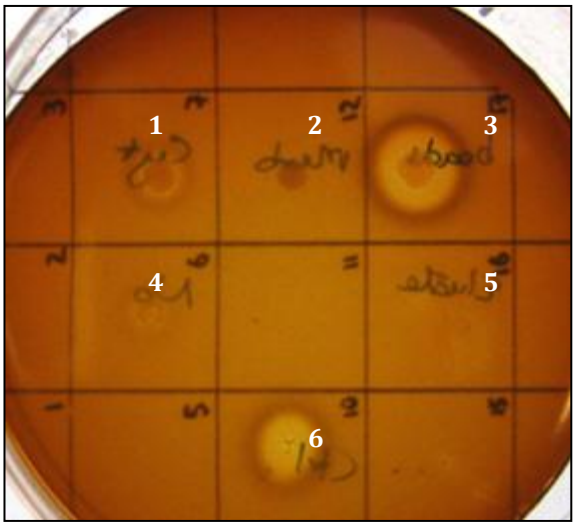


Figure 3.15 Bacteriocin bioassay using indicator organism *L. plantarum* ATCC 14917. (1, resin after cytosol pull-down; 2, resin after membrane pull-down; 3, GccF bound CDI-agarose beads before pull-downs; 4, CDI-agarose beads; 5, bead storage solution; 6, GccF control).

GccF was used as a positive control for the bioassay plate and shows a zone of clearing indicating that the assay worked properly. Sample 3 on the indicator plate shows a large halo of clearing for the GccF beads pre-incubation with protoplast lysate, confirming that the immobilised GccF is still active. Sample 1 on the plate is 2 μL of beads after incubation and elution of the protoplast cytosol fraction. The small zone of clearing visible for this fraction may indicate that some protein in the fraction is binding to GccF, decreasing but not completely eliminating its activity. As the characteristics of cell growth inhibition strongly suggest that GccF targets some protein in the cell membrane it was not expected that potential GccF targets would reside in the cytosol. It is possible that there was some contamination by membrane proteins or other cell membrane components, although this was not likely. Sample 2 is 2 μL of beads after exposure to the protoplast membrane fraction. There is no visible zone of clearing, indicating that proteins in the sample are binding to GccF in such a way as to completely stop activity, these may be possible receptors of GccF. An alternative explanation is that GccF has been lost from the beads.

Figure 3.16 shows the SDS-PAGE analysis of the elutions collected using the membrane fraction. Lane 3 Figure A shows a sample of the membrane fraction where there are many bands present. Lane 4 is the flow through from the beads indicating that most of the proteins are not specifically binding to the immobilised GccF. Lanes 5 and 6 are the initial washes with water showing that non-specifically bound proteins continue to be eluted. Lanes 8-15 also show no bands indicating that no proteins have been removed by either moderate concentrations of GlcNAc or high concentrations of NaCl. Gel B is an analysis of washes using higher GlcNAc concentrations, the maximum being 1 M. No protein bands were visible in these lanes indicating that if any proteins have bound to the immobilised GccF, they are bound very tightly. To test this hypothesis the beads themselves were incubated at 90°C with sample buffer for 4 minutes and the whole content of the tube loaded on the gel in lane 9. The gel shows several bands which are enriched compared to lanes 5 and 6 of gel A indicating that these bands may be specifically bound to the GccF CDI-agarose resin in a way that attenuates the activity of GccF (bands are seen in lane 10 but these are due to sample overflowing in lane 9). 16.5% tricine

gels (2.1.13) were also run with the samples obtained from the CDI-agarose pull-downs to detect lower molecular weight proteins.

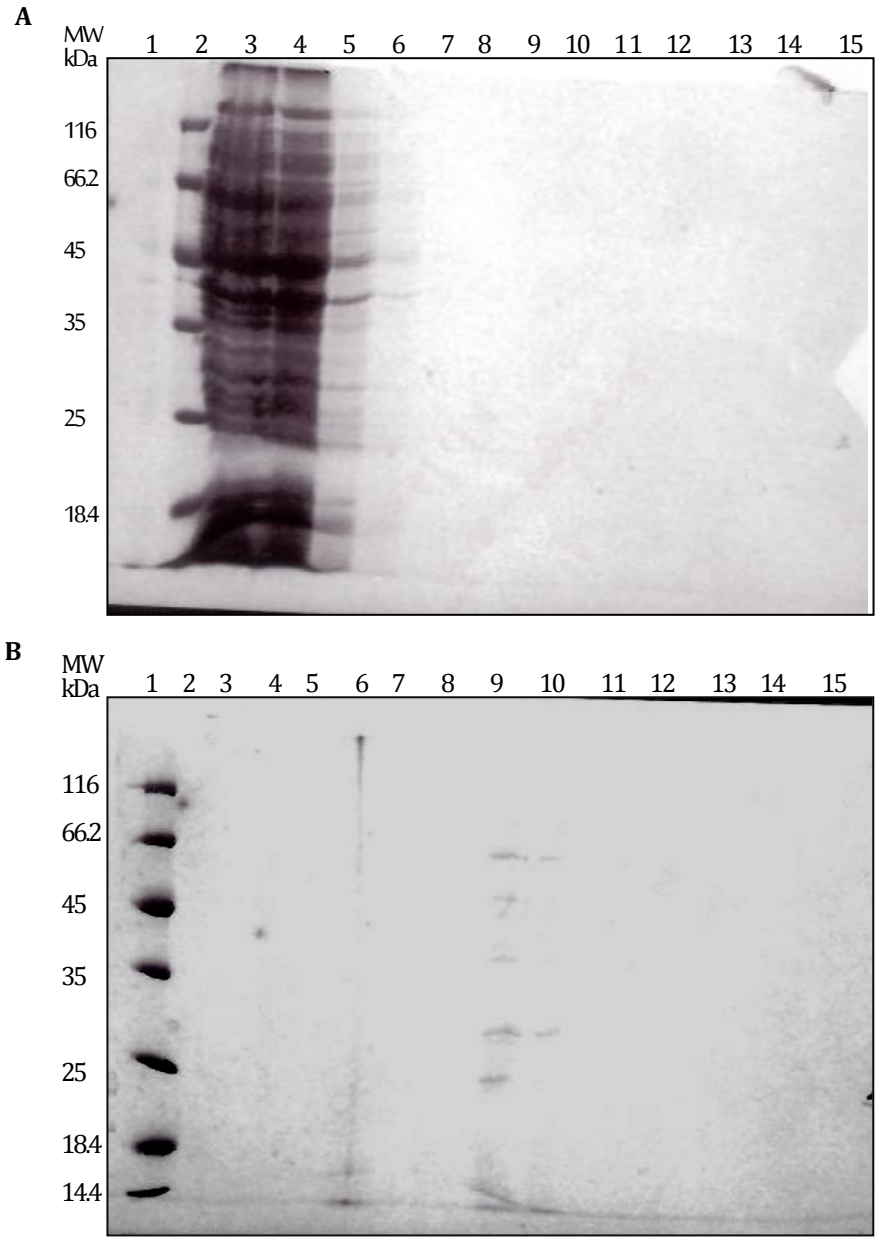


Figure 3.16 12.5% SDS-PAGE showing the results of GccF CDI-agarose pull-downs using protoplast membrane from *L. plantarum* ATCC 14917. ((**A**) Lanes 1, no sample; 2, molecular weight marker; 3, protoplast membrane fraction; 4, flow through; 5, H_2O wash; 6, H_2O wash; 7, H_2O wash; 8, 0.1 M NaCl elution; 9, 0.2 M NaCl elution; 10, 5 mM GlcNAc elution; 11, 10 mM GlcNAc elution; 12, 30 mM GlcNAc elution; 13, 60 mM GlcNAc elution; 14, 80 mM GlcNAc elution; 15, 100mM GlcNAc elution; (**B**) 1, molecular marker; 2, 200 mM GlcNAc elution; 3, 500 mM GlcNAc elution; 4, 1 M GlcNAc elution; 5, 1 M NaCl elution; 6, H_2O elution; 7 and 8, no sample; 9, CDI-agarose beads heated to 90°C).

The protoplast cytosol fraction of *L. plantarum* ATCC 14917 was also analysed in the same way. This time however, SDS-PAGE gel analysis of the washes showed no protein.

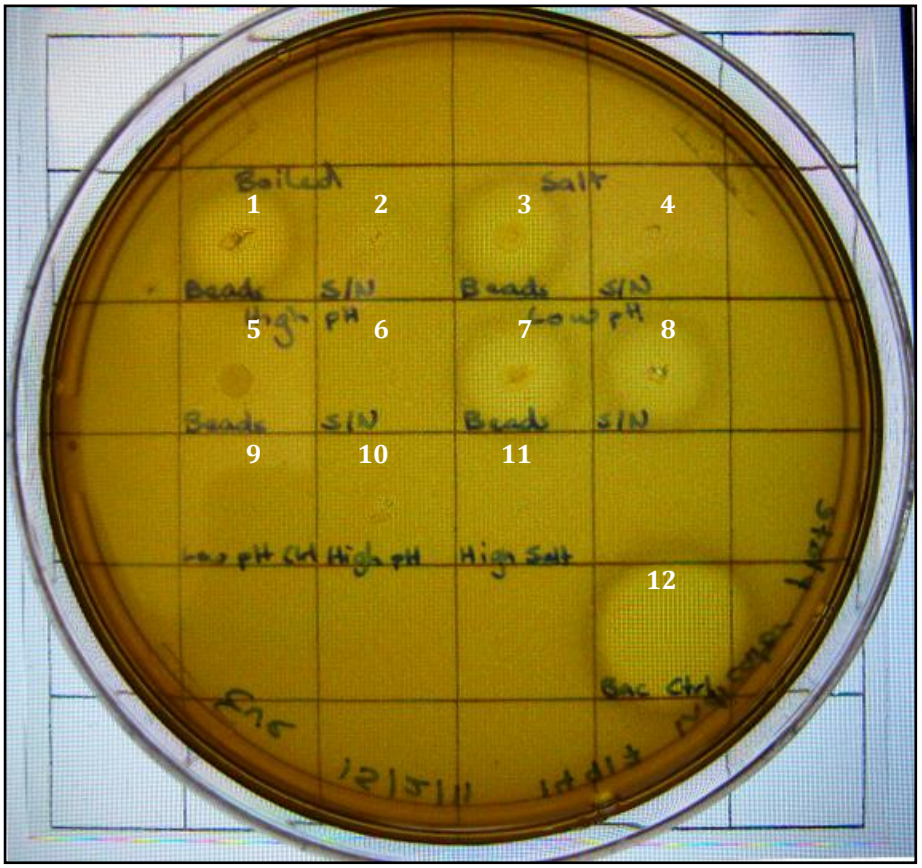


Figure 3.17 Bacteriocin bioassay using the indicator organism *L. plantarum* ATCC 14917. CDI-agarose beads after protoplast membrane pull-down seen in Fig. 3.16. Resin was added to solutions for 4 minutes at either pH 3, pH 10, 1M salt or it was boiled in H_2O, samples were then centrifuged for 1 minute at 4000 x g and the supernatant was removed. The beads and the supernatant were then analysed. (1, beads after boiling; 2, supernatant after boiling; 3, beads after 1M NaCl incubation; 4, supernatant after 1 M NaCl incubation; 5, beads after incubation at pH 10; 6, supernatant after incubation at pH 10; 7, beads after incubation at pH 3; 8, supernatant after incubation at pH 3; 9, pH 3 control; 10, pH 10 control; 11, 1 M salt control; 12, GccF control).

To determine whether the activity for the GccF CDI-agarose resin could be regained after incubation with the protoplast membrane fraction and subsequent elution of bound protein the resin from Figures 3.15 and 3.16 was treated with salt, high pH, low pH, and high temperature, methods commonly used to disrupt any protein-ligand binding. Both the supernatant and the beads were spotted on an indicator plate to detect GccF activity (Fig 3.17). The indicator plate shows that

boiling, high salt, and low pH treatment of the membrane incubated GccF CDI-agarose beads causes increased activity, indicated by the zone of clearing, against the same indicator strain. The supernatant for these treatments show no zone of clearing which demonstrates that all of these treatments do not break the primary amine bond between GccF and the beads. For treatment with high pH there is no zone of clearing seen for either the beads or the supernatant, and there is also no clearing seen for the salt, high and low pH controls. The GccF positive control shows a zone of clearing as expected. As boiling the GccF CDI-agarose beads for 4 minutes caused activity to be regained on the bioassay it is likely that the bands seen in lane 9 Figure 3.16B, where the beads were heated to 90°C to remove bound proteins, may be those molecules targeted by GccF, and thus molecules involved in bacteriostasis. It is interesting to note that as GccF is tightly bound to the resin the zone of clearing should be localized to the area directly associated with the resin. However, this is not seen and instead there is a large halo surpassing the area upon which the beads have been spotted similar to the free bacteriocin (Fig 3.15 and 3.17). This extended zone of clearing suggests that the bacteriostatic effect of GccF may not be only due to direct contact of GccF with individual cells, but that the cells affected by GccF may also signal or release a molecule which prevents the growth of or kills nearby cells.

C-11 PULL-DOWNS

L. plantarum KW30

CDI-agarose pull-downs were also performed to detect proteins present within *L. plantarum* KW30 which may bind to or modify GccF. The C-11 peptide of GccF consisting of the last C-terminal 11 amino acids of GccF minus the *S*-linked GlcNAc were bound through the N-terminal primary amine to CDI-agarose. Increasing concentrations of NaCl were used to elute any proteins potentially bound to C-11.

Figure 3.18 shows the result of a C-11 CDI-agarose pull-down for a *L. plantarum* KW30 protoplast membrane fraction. Lane 3 (A) is the protoplast membrane fraction which contains many bands whereas lane 4 is the initial flow through. The difference in intensity of the bands between lanes 3 and 4 indicates that some proteins may have been retained on the resin. In lanes 10 and 11 some faint bands are visible which are enriched compared to those seen in the initial lysate. The number of bands increase in lanes 12-14 and 3 (B), many of which are again enriched in comparison to the original fraction. Lane 7B contains the C-11 bound CDI-agarose resin after incubation and elution indicating that some proteins have bound relatively tightly to the beads. That the bands seen in these lanes are enriched suggests that they are binding to C11 and therefore may potentially interact with GccF.

The protoplast cytosol fraction of *L. plantarum* KW30 was also subjected to C-11 agarose affinity chromatography. However no bands were observed in any of the elution regimes (Not shown).

As it is possible that one of the complexed proteins was the GTase responsible for glycosylating the C-terminal cysteine, western methodology was used to detect the presence of GlcNAc on the resin using the mouse anti-*O*-GlcNAc antibodies. If a GlcNAc residue had been added to the cysteine on the C-11 peptide, a colour change should occur, that is visible to the naked eye.

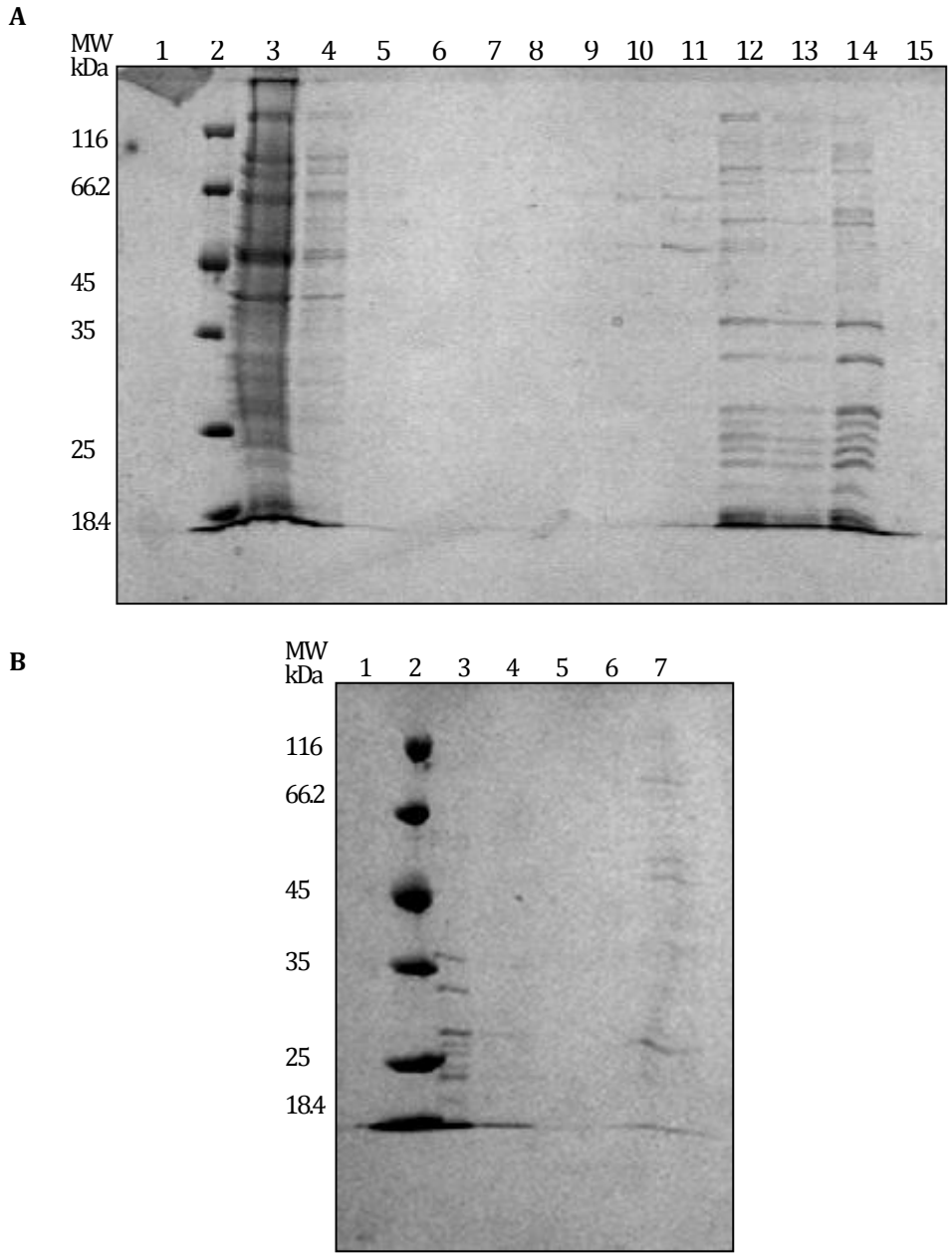


Figure 3.18 12.5% SDS-PAGE showing the results of C-11 peptide CDI-agarose pull-downs using the protoplast membrane fraction from *L. plantarum* KW30. ((**A**) Lanes 1, no sample; 2, molecular weight marker; 3, protoplast membrane fraction; 4, flow through; 5, H_2O wash; 6, H_2O wash; 7, H_2O wash; 8, 10 mM NaCl elution; 9, 20 mM NaCl elution; 10, 50 mM NaCl elution; 11, 80 mM NaCl elution; 12, 100 mM NaCl elution; 13, 200 mM NaCl elution; 14, 500 mM NaCl elution; 15, no sample; (**B**) 1, no sample; 2, molecular marker; 3, 1 M NaCl; 4, 2 M NaCl; 5 and 6, no sample; 7, CDI-agarose beads).

Figure 3.19 Modified Western blot of C-11 bound CDI-agarose beads after incubation with *L. plantarum* KW30 protoplast membrane. GccF bound CDI-agarose beads were used as a positive control.

Figure 3.19 shows the result of a modified western blot performed on the C-11 peptide bound CDI-agarose resin with GccF bound CDI-agarose used as a positive control. For the GccF bound resin a purple colour is observed indicating that either an *O*- or *S*- linked GlcNAc is present, this result is expected as GccF contains one of each modification. For the C-11 bound resin there is no colour change indicating no *O*- or *S*- linked GlcNAc modifications are present. A lack of colour change for the C-11 beads indicates that the proteins which appear to have bound to the beads in the pull-down shown in Figure 3.18 are not modifying the C-11 peptide by addition of a GlcNAc.

MASS SPECTROMETRY

To identify the proteins that were enriched from the GccF and C-11 CDI-agarose pull-downs the bands were excised from an acrylamide gel and sent away for mass spectrophotometry by Dr Petr Man at the Academy of Sciences of the Czech Republic using a Bruker ultraflex II I TOF/TOF. Figures 3.20 and 3.21 are gels which indicate the bands that were excised.

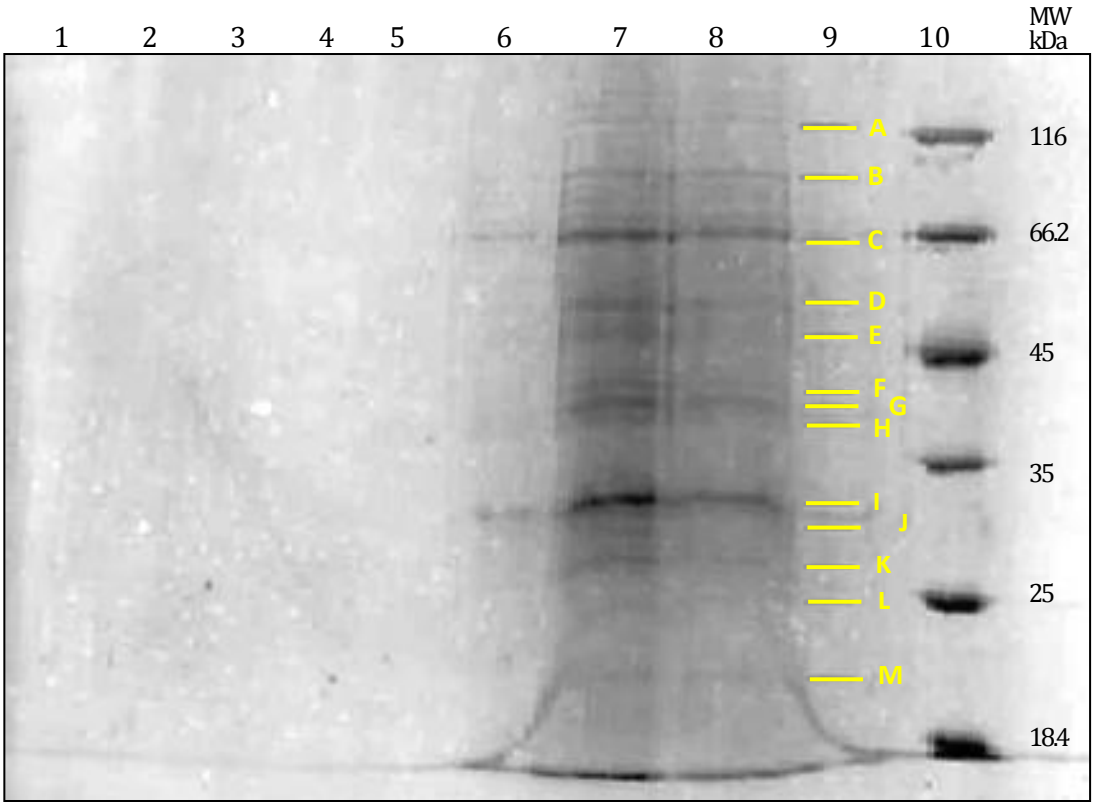


Figure 3.20 12.5% SDS-PAGE showing bands excised for mass spectrophotometry analysis for GccF bound CDI-agarose pull-downs against a *L. plantarum* ATCC 14917 protoplast membrane fraction. (Lanes 1-4, no sample; 6-9, CDI-agarose beads from Figure 3.16 after GccF pull-down with protoplast membrane; 10, molecular marker).

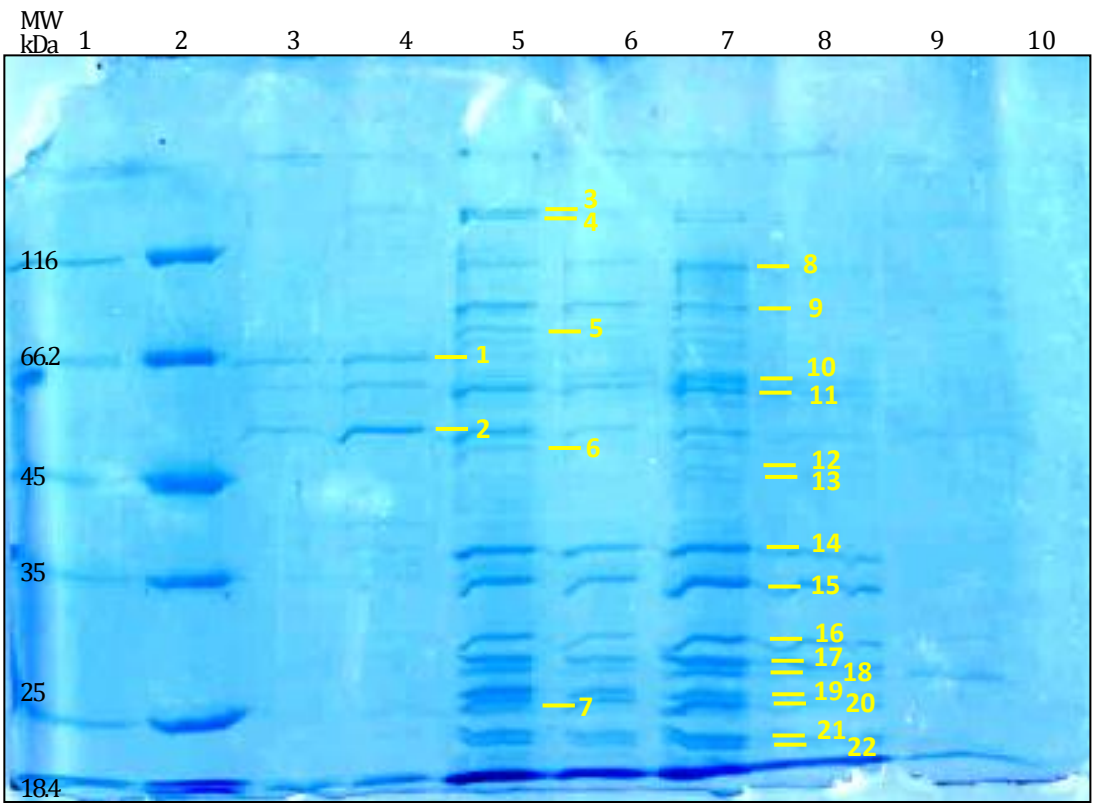


Figure 3.21 12.5% SDS-PAGE showing bands excised for mass spectrophotometry analysis for C-11 bound CDI-agarose pull-downs against a *L. plantarum* KW30 protoplast membrane fraction. (Lanes 1 and 2, molecular marker; 3, 50 mM NaCl elution; 4, 80 mM NaCl elution; 5, 100 mM NaCl elution; 6, 200 mM NaCl elution; 7, 500 mM elution; 8, 1M elution; 9, CDI-agarose beads from Figure 2.18 after C11 pull-down with protoplast membrane; 10, no sample).

Table 3.2 Results from analysis of bands from Figure 3.20 by mass spectrometry for GccF bound CDI-agarose pull-downs.

Band	Accession Number	Protein Name	Species	MW (Da)	E-value	Matched	Coverage (%)
A	114049526	PyrAb1 protein	L. plantarum ATCC 14917	116033	3.7E-02	35	34
	28378755	Pyruvate caboxylase	L. plantarum ATCC 14917	127561	2.6E+00	32	29
B	30068609	Phosphoketolase	L. plantarum ATCC 14917	88676	3.2E-05	36	43
C	28378548	Pyruvate kinase	L. plantarum WCFS1	62941	8.0E-06	17	42
D	28379083	NADH peroxidase	L. plantarum ATCC 14917	48283	1.3E-04	21	55
E	239758738	Tuf	L. plantarum ATCC 14917	30638	8.1E-04	17	50
F	28377642	Glyceraldehyde 3-phosphate dehydrogenase	L. plantarum ATCC 14917	36644	2.0E-04	17	59
	28378548	Pyruvate kinase	L. plantarum WCFS1	62941	1.7E-01	23	45
G	28378688	D-lactate dehydrogenase	L. plantarum ATCC 14917	37158	8.5E-02	16	41
H	28378686	30S ribosomal protein S2	L. plantarum ATCC 14917	30207	4.9E-01	21	61
	28378688	D-lactate dehydrogenase	L. plantarum ATCC 14917	37158	2.9E-01	17	42
	28378716	1-phosphofructokinase	L. plantarum ATCC 14917	32543	4.7E-01	12	51
	28378548	Pyruvate kinase	L. plantarum WCFS1	62941	1.8E+01	15	34
I	28377250	Fructose-bisphosphate adolase	L. plantarum ATCC 14917	31016	6.5E-04	12	49
J	28377250	Fructose-bisphosphate adolase	L. plantarum ATCC 14917	31016	4.1E+01	10	20
K	No peptide	No peptide	No peptide				
L	28377844	50S ribosomal protein L5	L. plantarum ATCC 14917	20226	1.5E+02	8	53

	28378710	Adenine phosphoribosyltransferase	L. plantarum WCFS1		8.5E+00	9	44
M	28377849 308181933	30S ribosomal protein S5 Small heat shock protein	L. plantarum ATCC 14917 L. plantarum subsp. plantarum	17250 16616	2.9E-01 2.0E+00	11 7	71 54

Table 3.2 shows the results from mass spectrophotometry of the bands from Figure 3.20 where a minimum of two peptides were matched to the predicted sequence. All of the bands matched except three, were found in *L. plantarum* ATCC 14917 which is expected as it is the strain of *L. plantarum* used for the pull-down assay. The three bands which were not within *L. plantarum* ATCC 14917 were instead located within *L. plantarum* WCFS1 which is closely related. The majority of the bands matched are proteins involved in glycolysis or gluconeogenesis with several ribosomal proteins and a Tuf (elongation factor). These protein band predictions seem unlikely to be actual binding targets for the mature GccF peptide as they appear to be proteins involved in DNA synthesis within cell cytosol. Although predicted binding targets of the GccF peptide within *L. plantarum* indicator strains are membrane proteins associated with other sugar transport, or proteins in cell division/membrane formation, processes in which none of the proteins identified are involved nothing is known about the mode of action of GccF. Thus while these proteins may bind to Gccf, they may do so non specifically, or they may be involved in the mechanism of inhibition in a way that is not yet understood. GccF has been found in previous experiments (personal communication, Dr Mark Patchett) to be 'sticky' in that it tends to bind to other proteins, therefore the bands pulled-down from *L. plantarum* ATCC 14917 may be due to non-specific interactions with GccF as opposed to specific binding.

Table 3.3 Results from analysis of bands from Figure 3.21 by mass spectrometry for C-11 bound CDI-agarose pull-downs.

Band	Accession Number	Protein Name	Species	MW (Da)	E-value	Matched	Coverage (%)
1	28377591	Chaperonin GroEL	L. plantarum WCFS1	57402	6.3E-14	34	68
2	28377654	Phosphopyruvate hydratase	L. plantarum WCFS1	48057	1.6E-08	21	53
3	300767813	DNA-directed RNA polymerase subunit beta	L. plantarum subsp. plantarum	135036	8.0E-06	52	40
4	28377826	DNA-directed RNA polymerase subunit beta	L. plantarum WCFS1	135746	3.2E-06	60	47
5	254556435	Threonine tRNA ligase	L. plantarum JDM1	73830	8.0E-11	31	50
6	28378285	Glutamate - ammonia ligase	L. plantarum WCFS1	51111	2.2E+00	21	49
	199598009	Hypothetical protein LRH_01388	L. rhamnosus HN001	7416	6.4E+02	5	
	28478605	RNA polymerase sigma factor RpoD	L. plantarum WCFS1	41418	8.0E+02	14	37
7	28378297	Elongation factor P	L. plantarum WCFS1	20353	2.0E+00	9	47
	28377598	Ribosomal protein S30EA	L. plantarum WCFS1	21712	2.0E+01	9	50
8	254555829	Ribonuclease R	L. plantarum JDM1	91624	8.0E-10	43	43
	300768184	Translation initiation factor IF2	L. plantarum ATCC 14917	89238	4.2E-02	28	36
	28378066	Leucyl-tRNA synthetase	L. plantarum WCFS1	92570	1.1E+00	24	27
9	28378761	GTP-binding protein TypA	L. plantarum WCFS1	68540	2.0E+09	23	43
10	28377485	Glytamyl-tRNA synthetase	L. plantarum WCFS1	56883	5.8E-02	20	52
	28377407	ATP-dependent RNA helicase	L. plantarum WCFS1	59003	1.5E-03	20	46
	28377434	Lysyl-tRNA synthetase	L. plantarum WCFS1	57593	2.8E+01	16	32
11	28378535	30S ribosomal protein S1	L. plantarum WCFS1	47131	8.0E-07	23	63

12	28379051	Hypothetical protein Ip_2504	L. plantarum WCFS1	43731	5.0E-13	25	68
	38478435	Hypothetical protein Ip_1756	L. plantarum WCFS1	43199	7.1E+00	12	39
13	28378435	Hypothetical protein Ip_1756	L. plantarum WCFS1	43199	8.0E-07	18	58
14	28378686	30S ribosomal protein S2	L. plantarum WCFS1	30207	1.6E-03	22	61
15	28377835	50S ribosomal protein L2	L. plantarum WCFS1	30173	1.0E-05	19	64
	28378610	GTP-binding protein Era	L. plantarum WCFS1	34263	1.7E+01	11	39
16	28377596	50S ribosomal protein L1	L. plantarum WCFS1	24755	2.5E-02	13	60
17	254556017	50S ribosomal protein L3	L. plantarum JDM1	22683	2.0E-08	17	59
18	28377838	30S ribosomal protein S3	L. plantarum WCFS1	24193	2.5E-02	15	55
19	28378911	30S ribosomal protein S4	L. plantarum WCFS1	22949	4.0E-08	21	70
20	28377833	50S ribosomal protein L4	L. plantarum WCFS1	22623	3.6E+00	8	32
	28377598	Ribosomal protein S30EA	L. plantarum WCFS1	21712	2.3E+01	6	43
	259503514	MerR family transcriptional regular	L. antri DSM 16041	8918	4.5E+04	3	
21	28377847	50S ribosomal protein L6	L. plantarum WCFS1	19228	4.4E-03	11	55
	28377844	50S ribosomal protein L5	L. plantarum WCFS1	20226	5.1E-01	14	65
22	28377847	50S ribosomal protein L6	L. plantarum WCFS1	19338	9.6E-01	10	51
	28377844	50S ribosomal protein L5	L. plantarum WCFS1	20226	8.0E-06	16	72

Table 3.3 shows the results from mass spectrophotometry of the bands from the C11 pull-down where a minimum of two peptides were matched to the predicted sequence (Fig 3.21). All the bands excised were matched to proteins within *Lactobacillus plantarum spp* but not to *L. plantarum* KW30 as its sequence is not present in any public database. None of the proteins excised match any of the proteins within the *gccF* cluster and are therefore not those involved in maturation or secretion of the peptide. Sequence alignments were performed for the different proteins but no consensus sequence was found indicating that the proteins are not binding C-11 through a common binding sequence. As C-11 is a small peptide with no structure it may be that non-specific binding of proteins to the peptide is occurring resulting in the large number of bands being pulled down.

BIOTIN LABELLING

GccF biotin pull-downs with streptavidin bound magnetic beads were also trialled as a method to identify binding partners for GccF in specific indicator strains. Obtaining singly labelled GccF biotin proved however to be extremely difficult and although many trials to obtain singly labelled active GccF were carried out, it could never be achieved under the conditions required for labelling.

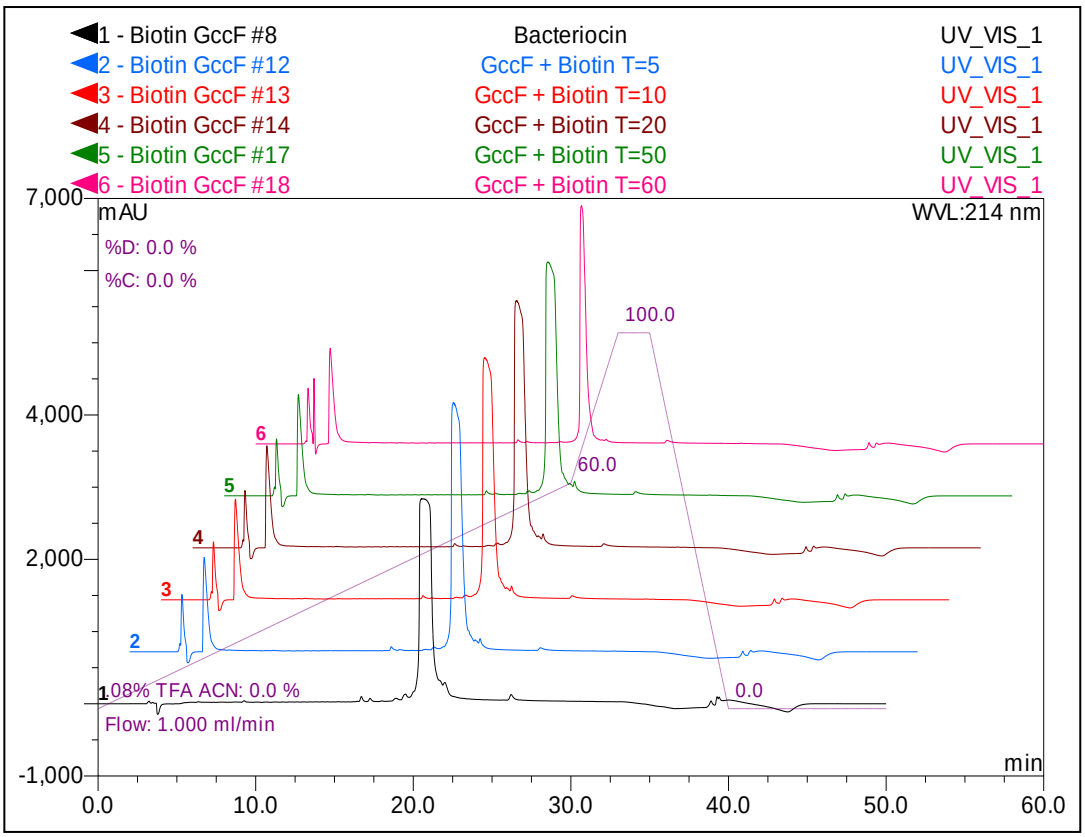


Figure 3.22 RP-HPLC results for a time trial addition of Biotin to GccF. Incubation times of 5, 10, 20, 50 and 60 minutes were investigated. 1 µL of 10 mM biotin was added to 1 mg/mL of GccF and RP-HPLC was performed with a C_{18} column at a temperature of 20°C. (Sample 1, GccF Control; 2, 5 minute GccF + biotin incubation; 3, 10 minute GccF + biotin incubation; 4, 20 minute GccF + biotin incubation; 5, 50 minute GccF + biotin incubation; 6, 60 minute GccF + biotin incubation. Samples 2-6 have been shifted along the time and absorbance axis for comparison).

Figure 3.22 shows the results of a time trial performed for the labelling of GccF. Biotin was added to GccF at approximately 20 x the molar ratio of GccF and aliquots were removed for analysis by RP-HPLC. GccF not incubated with biotin was used as a control to detect any peak shift from the addition of biotin. In the GccF control there is a peak seen at approximately 21 minutes which is GccF as

confirmed by ESI-MS. After incubation with biotin the GccF peak is clearly seen at all timepoints. There were however, no distinct peaks that were not GccF, suggesting that no covalent link was formed between biotin and GccF. MS analysis of these peaks confirmed that biotin was not linked to GccF.

Further trials were attempted using an increased molar ratio of biotin (100x). After RP-HPLC it was found that there was still a single peak present however, the peak appeared to have a shoulder on the right hand side when compared to the GccF only peak. As the addition of biotin would be expected to increase retention on the column, it was hoped that this peak contained biotin labelled GccF. The separation was repeated at a higher column temperature of 45°C which improved the resolution of these two analytes, and although a double peak could clearly be seen, it was not completely resolved. Each peak was analysed by mass spectrometry, which showed the presence of a small percentage of biotin labelled GccF. The inefficiency of the labelling could be due to an insufficient concentration of biotin in the reaction, or it could mean that biotin labelled GccF does not form gaseous ions as easily as the unlabelled peptide.

In order to determine whether a lack of biotin was limiting the production of biotinylated GccF, trials were performed using 200x and 400x the molar ratio of biotin to GccF. The reaction mixes were incubated for 10 or 30 minutes after which time, biotinylated GccF was separated from the native peptide by RP-HPLC.

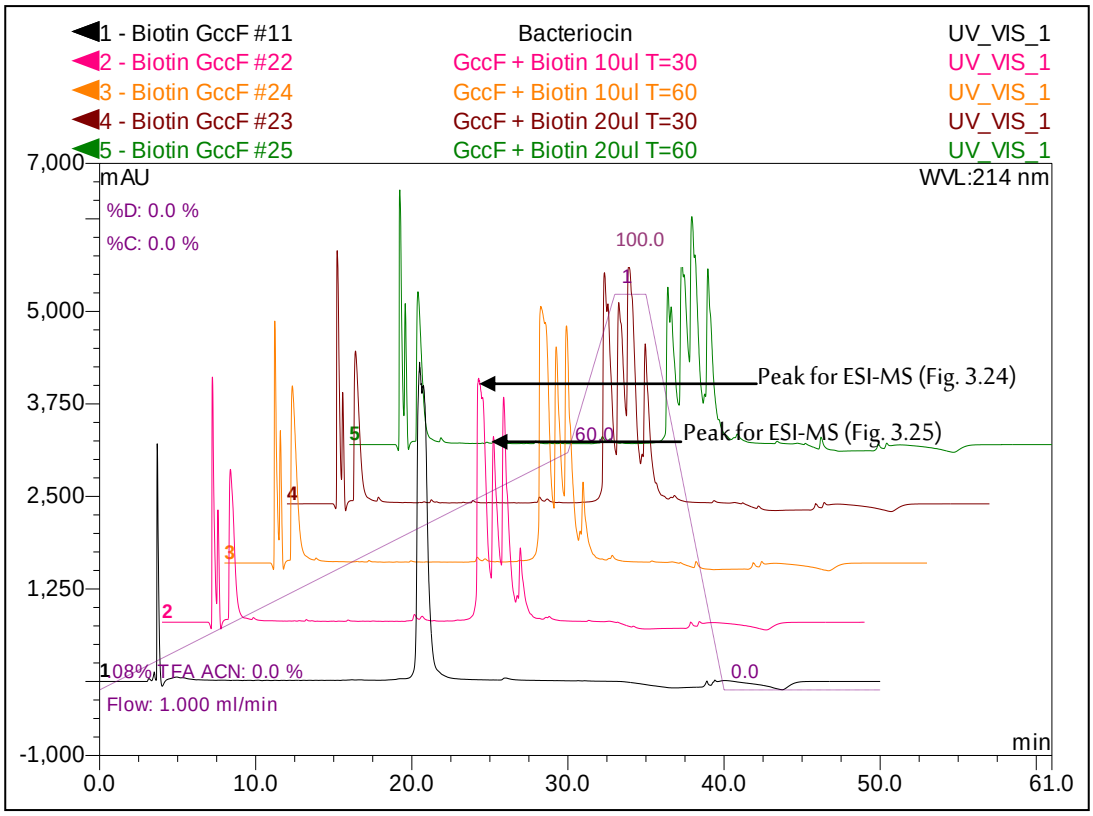


Figure 3.23 RP-HPLC of a biotin concentration trial at two different incubation time points. A C_{18} column was used and samples were run at a higher column temperature (45°C). 10 or 20µL of 10mM Biotin was incubated with 1mg/mL GccF for 30 or 60 minutes. (Sample 1, GccF control; 2, GccF + 200x molar ratio of biotin incubated for 30 minutes; 3, GccF + 200x molar ratio of biotin incubated for 60 minutes; 4, GccF + 400x molar ratio of biotin incubated for 30 minutes; 5, GccF + 400x molar ratio of biotin incubated for 60 minutes. Samples 2-5 have been shifted along the time and absorbance axis for easier comparison).

Figure 3.23 shows that once again the GccF control has a single peak at 21 minutes which corresponds to GccF. It is also seen that with an increase in the molar ratio of biotin a further 3 peaks at later time points are present, which may represent biotin labelled GccF. For all of the samples there is still a peak visible indicating unlabelled GccF, however it is considerably reduced for the reaction of 400x (20 µL) of biotin incubated with GccF for 60 minutes.

To determine whether any of the peaks from this preparation were biotin labelled, each peak was analysed by mass spectrometry (Figure 3.24 and 3.25). The peaks labelled in Figures 3.24 and 3.25 represent the most abundant isoform of the 5+ charge state of the peptides not the monoisotopic mass (example shown in the inset box for Figure 3.24). GccF alone is represented by the ion at 1040.8054 amu

$(M+5H)^{5+}$, which is the 5+ ion of GccF (Mr = 5199 amu, Fig 2.16). The sample is from the peak collected at 21 minutes and is indicated by the arrow in Figure 3.24. A peak is also present with a mass of 1086.8181 amu $(M+5H)^{5+}$ (monoisotopic mass 1086.0202 amu) which corresponds to the mass of GccF with the addition of a single biotin molecule (Mr = 5425.3 Da, the mass of biotin being Mr = 226.3 Da), there is also GccF with two biotin molecules (Mr = 5651.6 amu) added indicated by the mass 1131.38076 amu $(M+5H)^{5+}$ (monoisotopic mass 1131.328 amu). Figure 3.25 shows the results of mass spectrometry for the peak collected at 21.8 minutes indicated by the arrow in Figure 3.25. When compared with Figure 3.24 there is a decrease in the abundance of the 5+ ion of unlabelled GccF in comparison to that for the biotin labelled GccF. There is also an increase in the abundance of doubly biotinylated GccF and the appearance of a new peak which corresponds to triply biotinylated GccF. The results from mass spectrometry indicate that the peaks separated by RP-HPLC (Fig 3.23) contain mixed species of GccF and that better peak separation is required to obtain singly biotinylated GccF. The presence of doubly and triply labelled biotin suggests that biotin is labelling GccF on not only the N-terminal primary amine but also the amines on K1 and K32. Labelling on Lysine-32 is undesirable as it is located in a region of GccF which has been shown to be involved in its activity therefore labelling of this residue is likely to reduce GccF activity possibly by occluding regions of GccF involved in protein binding.

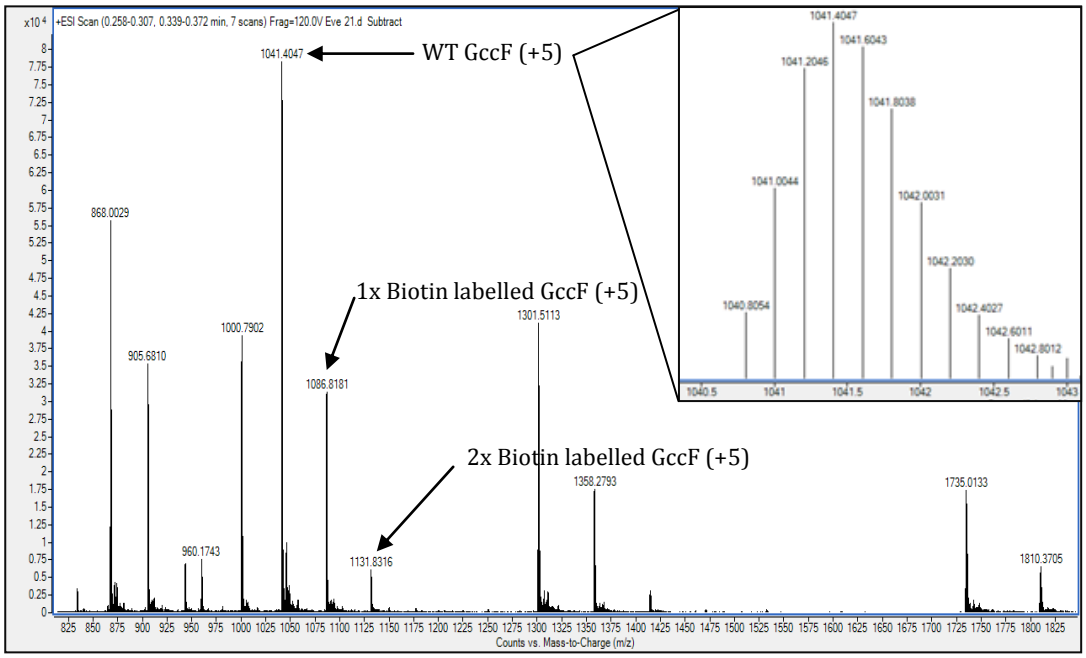


Figure 3.24 ESI-MS (positive ionization mode) results from MS for the peak with a retention time of 21 minutes (Figure 3.23). Inset box represents the isotopic cluster of the WT GccF (+5) peak showing the monoisotopc mass as 1040.80 amu $(M+5H)^{5+}$.

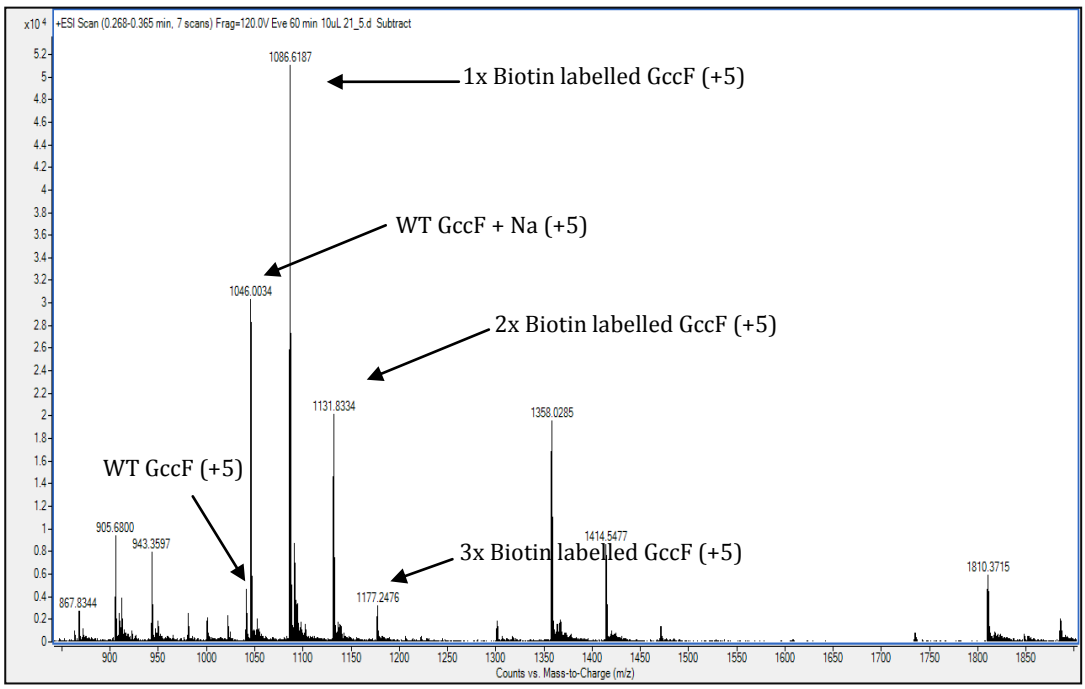


Figure 3.25 ESI-MS (positive ionization mode) results from MS for the peak with a retention time of 21.8 minutes (Figure 3.23).

In an attempt to increase peak separation in order to obtain singly labelled GccF trials were performed using the ion pairing agent heptafluorobutyric acid (HFBA) instead of trifluoroacetic acid (TFA) which was used in the previous trials.

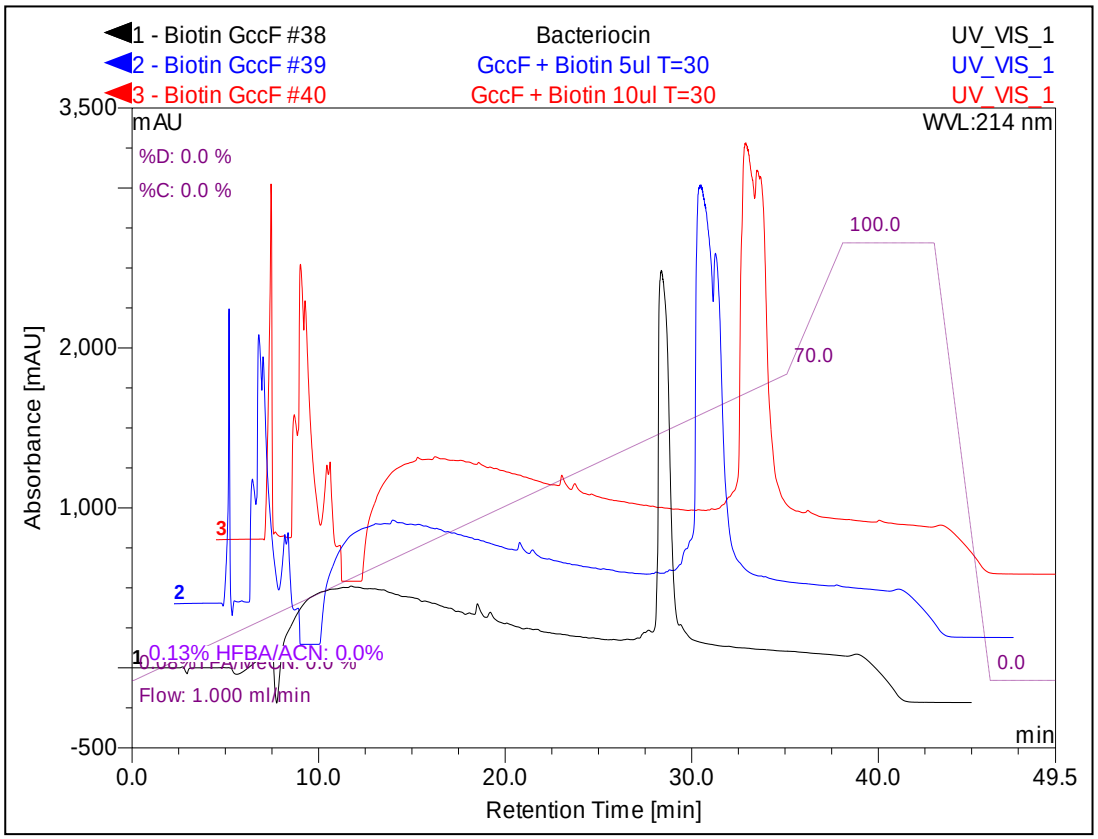


Figure 3.26 RP-HPLC results for biotin concentration trial. 5 or 10µL of 10mM Biotin was incubated with 1mg/mL GccF for 30 minutes. HFBA was used as the ion pairing agent along with the C14 column at a temperature of 25°C.

Figure 3.26 shows that the ion pairing agent HFBA increases the retention of GccF to 28.5 minutes compared to 21 minutes with TFA. The separation of labelled from unlabelled species, is however worse indicating that TFA is a better ion pairing agent. The experiment was repeated at a higher column temperature (45°C) and once again, worse peak separation was achieved (not shown). Mass spectrometric analysis of the peaks showed once again that they contained mixed species of labelled and unlabelled GccF.

The flow rate was also reduced from 1 mL/min to 0.8 mL/min in an attempt to increase the resolution of labelled and unlabelled GccF. It was found however that the slower flow rate did not improve peak separation. The mobile phase was also

changed from acetonitrile to the more polar methanol in an attempt to increase separation. At both 25°C and 45°C the peak separation was worse with MeOH. The last parameter to be changed was the solid phase. Instead of C_{18}, C_{14} and C_{10} matricies were trialled, but neither offered any improvement over the C_{18} matrix. The peak collected from RP-HPLC for mass spectometry in Figure 3.23 at 21.8 minutes was incubated with streptavidin bound magnetic beads in order to separate out unlabelled and biotinylated GccF. The concentrations of the wash and elution fractions were determined and it was found that the total concentration after the washing and elution steps was significantly lower than that initially added to the beads indicating that some of the GccF has remained bound. RP-HPLC analysis of the wash and elution fractions of GccF found that some of the biotin labelled GccF came off during the washing steps and that some unlabelled GccF remained after washing and came through in the elution fractions.

3.7 2D GEL ELECTROPHORESIS

In this section of work 2D gel electrophoresis was used to separate and compare proteins from a GccF indicator strain which has been treated with GccF with the same indicator strain not treated with GccF.

Two-dimensional gel electrophoresis is used to separate proteins on the basis of two properties. The first dimension is isoelectric focusing where proteins are separated on the basis of their isoelectric point (pI). The second dimension separates on the basis of molecular mass (Mr). Because two molecules are unlikely to be similar with respect to pI and Mr, proteins in a complex mixture are likely to be more effectively separated.

For the first dimension, proteins that have been suitably denatured are applied to an IPG strip with a set pH gradient and an electric potential is applied across the strip. Proteins will travel along the strip until they reach their isoelectric point (no net charge). Once they have reached this point they will no longer travel along the gel due to a lack of charge, but will focus into tight bands.

The second dimension is SDS-PAGE. The focused strips are layered onto a separating gel (which may be a gradient gel), sealed with agarose and the gel then subjected to an electric potential.

3.7.1 RESULTS AND DISCUSSION

L. PLANTARUM ATCC 14917 2D GEL ELECTROPHORESIS

To determine the effect of GccF on protein expression within an indicator strain 2D gel electrophoresis was used to separate and compare proteins from *L. plantarum* ATCC 14917 treated with GccF with *L. plantarum* ATCC 14917 not treated with GccF. Initially IPG strips (7cm) with a pH gradient of 4-7 were used for 2D gels, the gels showed no observable difference between treated and untreated cells. The protein spots observed were well defined, with minimal smearing and the majority of the spots were located towards the pH 7 end of the gel. A clear cut-off point at pH 7 is seen where proteins with a more alkaline pI could not migrate further indicating a larger pH range may be required for full resolution of the protein spots. However, the amount of protein spots seen on the gel was low compared to what is expected for whole cell lysate, this may be due to an issue that arose during resolubilisation where the majority of the freeze-dried cell lysate would not resuspend in the lysis buffer.

Larger IPG strips (24 cm) with a pH gradient of 3-10 were run to increase both the pH range and spread of the protein spots on the 2D gels. Figure 3.27 shows the results of 2D gels using IPG strips with a pH gradient between 3-10. For both gels (A and B) the majority of spots are seen near the alkaline end of the gel and some smearing is observed. When the two gels were analysed using DeCyder, a 2D gel analysis program, (2D software Ver. 7.0 for PC) no significant difference was found between the two gels therefore there was no difference found betweeen the untreated and GccF treated cells. There is a difference in the spread of spots between the two gels (which is common) and this is taken into consideration when analysed on DeCyder. Once again there is a low number of spots observed on the gels compared to what is expected for whole cell lysate.

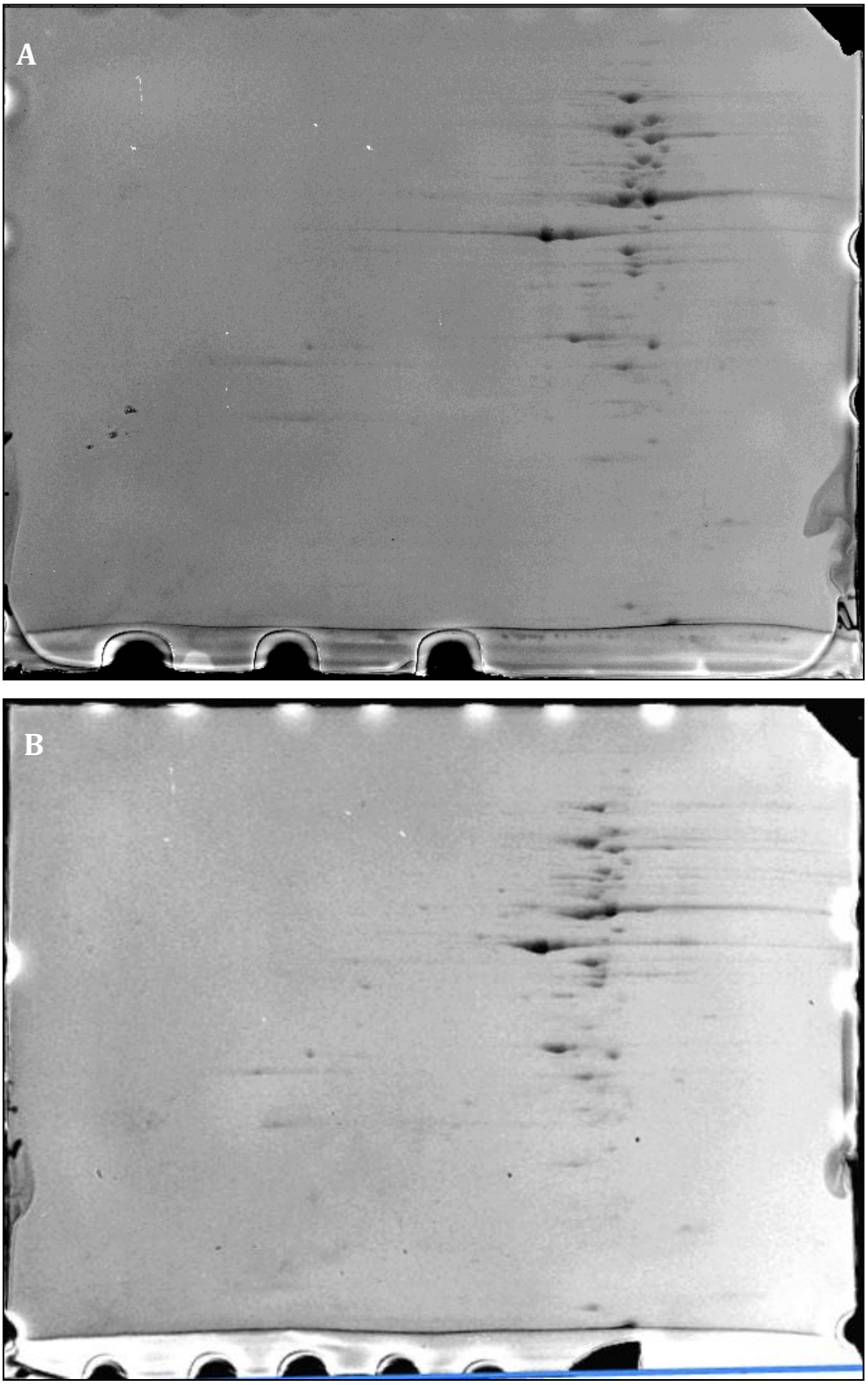


Figure 3.27 2D 24cm 12.5% SDS-PAGE gel. ((A) *L.plantarum* ATCC 14917 untreated cell lysate run on a pH 3-10 IPG strip. (B) *L. plantarum* ATCC 14917 cells lysed after incubation with 4 nM GccF run on a pH 3-10 IPG strip.)

3.8 CONCLUSIONS AND FUTURE DIRECTIONS

CONCLUSIONS

Part 3 of this thesis investigated the function of the bacteriocin GccF produced by *L. plantarum* KW30. Western blotting and lectin pull-downs detected that no other GlcNAcylated proteins are present within *L. plantarum* KW30 and from this it can be deduced that the production of GccF is unlikely to be mediated by a GlcNAc signalling system within the organism.

FITC labeled GccF^{-HC} and ^{deOGlcNAc}GccF were labelled on the primary amine found on the N-terminus. When incubated with an indicator strain it was seen that these forms of GccF localize on the indicator strain differently compared to wild-type GccF. Because wild-type GccF is found to localize to the zone of division (septum) between the cells, it is hypothesised that proteins involved in the separation of cells may be the target of GccF. As removal of either the *O*- or *S*-linked GlcNAc seems to abolish this localization of GccF it is possible that the glycans themselves are involved in targeting of GccF to a receptor on susceptible strains.

In order to understand the mechanism by which GccF inhibits cell growth, methods were developed to try and identify the receptors or targets of GccF. Similar methods were used to try and identify the KW30 proteins involved in the maturation of GccF. To do this both GccF and the unmodified C-terminal peptide of GccF were bound to CDI-agarose resin *via* their N-terminal primary amines and used in pull-down experiments against different fractions of the indicator strains *L. plantarum* ATCC 14917, and the cell lysate of *L. plantarum* KW30 respectively. Experiments investigating potential targets for GccF in *L. plantarum* 14917 showed that the majority of proteins that were bound were involved in glycolysis and gluconeogenesis. While the exact mechanism of GccF activity remains unknown, it is difficult to see how such proteins would be affected by interacting with GccF. None of the identified proteins are consistent with the currently hypothesized mechanism of action for GccF. A possible explanation is that the membrane fraction was contaminated with cytosolic proteins, or that the currently hypothesised mechanism of action is incorrect. For the C-11 peptide pull-downs of the *L. plantarum* KW30 lysate the majority of proteins that bound were identified

as being involved in DNA synthesis. As *L. plantarum* KW30 is the bacteriocin producer strain it was predicted that proteins which would bind to this relatively small peptide would be those involved in its maturation and secretion, none of the proteins identified showed any homology to these sequences.

2D gel electrophoresis found no differences between *L. plantarum* ATCC 14917 cells treated with GccF compared with cells that were not treated. However, the number of spots observed on the gels were not represenatative of the total number of different proteins present within cell lysate therefore there may be differences in protein expression that were not detected. Consequently more work must be carried out refining the method in order to support or refute the results seen so far.

Although a considerable amount of information was obtained from the experiments carried out, the overall mechanism of action of the bacteriocin GccF was not elucidated. The importance of the *O*- and *S*-linked GlcNAc residues was highlighted in the microscopy experiments, and it did appear that the membrane fraction of the protoplasts contained proteins which bind tightly to GccF as they were extremely difficult to remove. This would imply the interaction is specific. It is thus somewhat puzzling that their identities did not make sense in terms of a possible mode of production/regulation. Further work is required to identify the binding partners and regulation system involved in the function and production of GccF.

FUTURE DIRECTIONS

Although other GlcNAcated proteins were not detected using western blots and lectin pull-downs prior to the production of GccF. Further western blots should be performed to determine whether GlcNAcated proteins may be produced by *L. plantarum* KW30 after GccF production is established (i.e after 12 hours growth) as these may be linked to the GccF gene cluster.

To determine the mechanism of production for GccF further work into the gene cluster encoding the bacteriocin may reveal the proteins responsible. Continuation and optimisation of the DNA affinity work involving the hypothesized response regulator encoded within the cluster (appendix IV) may confirm that this protein is involved in regulating the production of the bacteriocin.

An approach that would prove useful in comprehending the function of GccF would be confocal microscopy on FITC-labelled GccF to conclusively determine whether the bacteriocin enters susceptible strains or whether it binds to a target in the cell envelope. A better understanding of the location of the target of GccF will help in the design of further experiments such at pull-downs.

Further investigations should be undertaken into the CDI-agarose pull-downs to determine whether there is cytosolic protein contamination within the membrane fractions using cytsolic enzyme assays. Or CDI-agarose pull-downs using the predicted immunity protein instead of GccF or the C11 peptide could be used to pull-down the proteins involved in providing immunity to GccF.

It would be interesting to investigate why when GccF is bound to the CDI-agarose beads spotted on bioassay plates that a zone of clearing is seen that surpasses the area onto which the beads are spotted. Potentially some sort of cell signalling is placed in effect by the susceptible cells directly affected by contact with GccF which halts the growth of cells within their vicinity causing the large halo. 2D gel experiments may help to determine whether there is a cell signalling effect present within susceptible cells upon contact with GccF. Determining the nature of this may elucidate the cause of the rapid bacteriostasis seen for suseptible cells when

exposed to GccF in liquid culture.

Further work on the 2D gel experiments by improving the method for resolubilisation in order to increase the number of spots as well as running triplicates to improve the reliability of the analyses by the DeCyder software will help to determine the effect if any that GccF has on proteins in susceptible cells. If interesting results are obtained further 2D gel work using differential gel electrophoresis (DIGE) would be useful to quantify any differences seen between treated and untreated cells.

It would be interesting to determine whether GccF causes an increase in membrane permeability or acts by 'jamming' open a ion channel or transporter within the cell membrane. Experiments testing the difference in membrane conductivity of susceptible strains could be undertaken to determine whether a similar mechanism of action is employed by GccF. Experiments could also be carried out to determine whether contact with GccF causes inhibition of DNA, RNA, or macromolecule synthesis on contact with susceptible strains, by using radiolabelled amino acids or base pairs and measuring the level of their incorporation.

PART 4:

STRUCTURAL AND FUNCTIONAL CHARACTERISATION OF THE *L. PLANTARUM* KW30 GLYCOSYLTRANSFERASE, GCCA.

4.0 INTRODUCTION

Part 3 of this thesis focuses on the glycosyltransferase GccA from *L. plantarum* KW30 thought to be responsible for one or both of the GlcNAc modifications on GccF. As previous attempts to produce recombinant GccA had resulted in the production of insoluble aggregates, homologues of GccA were identified using bioinformatic analyses (personnal communication, Dr Gill Norris). Attempts to clone and produce soluble recombinant protein of these homologous glycosyltransferases (which are also predicted to have a bacteriocin present within the gene cluster) were undertaken as their biochemical and structural characterization may provide insight into the mechanism of action of GccA. The aim of this section of work was to clone and produce at least one of two target glycosyltransferases and develop an activity assay.

4.1 OBJECTIVES AND STRATEGIES

The objectives for this section of work were:

1) To identify glycosyltransferases homologous to GccA that are located within bacteriocin gene clusters.

2) Clone two of these, the GTases from *Yersinia frederiksenii* and *Enterococcus faecalis,* into suitable expression systems using standard molecular cloning techniques.

3) Express and purify recombinant tagged *Y. frederiksenii* and *E. faecalis* glycosyltransferases using affinity chromatography.

4) To develop an assay to determine glycosyltransferase activity from *L. plantarum* KW30, and thus for any recombinant glycosyltransferase produced.

4.2 METHODS

4.2.1 Genomic DNA isolation

Genomic DNA was isolated from *Yersinia frederiksenii* ATCC 33641 using the Wizard Genomic DNA Purification Kit and protocol (Promega, USA).

Briefly, 150 mL MRS culture media supplemented with 1.2% glycine was inoculated with 2% of an overnight culture and incubated at 30°C until an OD_{600} of 1.0 was reached. Cells were harvested by centrifugation of 35 mL aliquots (2,700 x g) at 4°C for 10 minutes and washed 3x in 5 mL PBS. Pellets were then frozen at -80°C for later use. When required, cell pellets were thawed on ice and resuspended in 2 mL of cold solution I. They were then divided into two equal aliquots placed in 1.5mL Eppendorfs and 400 µL of 50 mg/mL lysozyme was added. The cell suspension was then mixed by inversion every 10 minutes for 50 minutes at 37°C. Cells were collected by centrifugation (1 min at 14,000 x g) the supernatant drained, and the pellet rapidly resuspended in 800 µL of ~25°C ' lysis solution'. Cells were completely lysed by three freeze-thaw cycles, where 1 minute immersions in liquid nitrogen were followed by 2 minute immersions in an 80°C water bath. After the final cycle, samples were cooled to room temperature and protein was precipiated by the addition of 270 µL of room temperature, 'protein precipitation solution'. Samples were mixed by inversion and incubated on ice for 5 minutes, before being subjected to centrifugation for 5 minutes at 14,000 x g. The supernatant was carefully transferred to sterile Eppendorf tubes, and any remaining protein precipitated using the same method until ~800 µL of protein free solution had been obtained. Isopropanol (620 µL)was then layered on top of each sample, and mixed by inversion until DNA threads formed. Samples were given a physical shock by dropping the tubes on to a bench to precipitate more DNA, which was then collected by centrifugation for 15 minutes at 14,000 x g. After removal of the supernatant, the DNA was washed twice with 800 µL 70% ethanol, then air dried and resuspended in 50 µL of TE-buffer.

TE buffer:	10 mM Tris-HCl
	1 mM EDTA
	pH 8.5

4.2.2 Plasmid DNA isolation

150 mLs of sterile MRS culture medium was inoculated with 1% of an overnight culture of *Enterococcus faecalis* ATCC 700802 and incubated overnight at 25°C. Cells were harvested by centrifugation for 15 minutes at 2,700 x g at 4°C and the resulting cell pellet was resuspended in 200 µL of 25% sucrose containing 40mg/mL lysozyme. At this point the cell suspension was transferred to a 1.5mL Eppendorf tube, then incubated at 37°C for 15 minutes. At this point, 400 µL of alkaline SDS solution was added, and the cell suspension mixed by inversion before being centrifuged at 17,000 x g for 15 minutes at 4°C. The supernatant was then transferred into a new 1.5 mL Eppendorf tube and 650 µL of isopropanol was added, mixed by inversion, then centrifuged at 17,000 x g for 15 minutes at 4°C. The supernatant was discarded and the pellet resuspended in 320 µL of H_2O and 200 µL of 7.5 M ammonium acetate containing 0.5mg/mL ethidium bromide. Following this, 350 µL of phenol/chloroform/isopropanol (25:24:1 v/v) was added to precipitate the protein, followed by centrifugation at 17,000 x g for 5 minutes at 4°C. The upper phase was transferred to a new 1.5mL Eppendorf and 1mL of ethanol (-20°C) was added, mixed by inversion and the DNA pelleted by centrifugation at 17,000 x g for 5 minutes at 4°C. DNA was washed in 70% ethanol at room temperature, then resuspended in 40 µL of TE-RNase.

Alkaline SDS solution:	3.0% SDS
	0.2 N NaOH

TE-RNase:	10 mM Tris-HCl
	1.0 mM EDTA
	0.1 mg/mL RNase
	pH 8.5

4.2.3 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify the 1.2 kb *Y. frederiksenii* ATCC 33641 glycosyltransferase gene and the 1.3 kb *E. faecalis* ATCC 700802 glycosyltransferase gene. Conditions for a typical PCR reaction were as follows:

1 ng of either *E. faecalis* or *Y. frederiksenii* template was amplified with 10 μM of both forward and reverse primers (Appendix II) depending on the tagged or untagged state required. This was carried out in a total reaction volume of 10 μL containing 0.5 units of Pwo DNA polymerase (Roche Applied Sciences, Germany), 2 mM deoxyribo-nucleotide triphosphate (dNTP) in 1x standard Pwo polymerase buffer containing $MgSO_4$ (Roche Applied Sciences).

To amplify the genes, the following thermo-cycling program was used: initial denaturation 98°C for 20 s, followed by 30 cycles of primer annealing at 61°C for 30 s, with extension at 72°C for 30 s. A positive PCR control which consisted of a proven combination of primers and template was used, along with a negative control in which no template DNA was added. After amplification the PCR products were purified using a Roche High Pure PCR Product Purification Kit and agarose gel electrophoresis was used to confirm PCR amplification. DNA concentration was determined using the NanoDrop (2.1.10).

4.2.4 PCR Mutagenesis

In order to remove restriction sites located within the target gene PCR mutagenesis was performed using the same thermocycler program as described in 4.3.3 and primers designed for mutagenesis (Appendix II).

4.2.5 Restriction Digest

100ng of PCR product or 100ng of plasmid (Appendix III) were digested with 5 units of restriction enzyme in the recommended buffer, in a reaction volume of 50μL at 37°C for 4 hours or overnight. Restriction digest products were purified with a Roche High Pure PCR Product Purification Kit. Controls to troubleshoot the reaction contained either plasmid containing no restriction enzymes or plasmid cut with one restriction enzyme only. The products of restriction digests were checked on a 1.2% agarose-TAE gel (2.1.8)

4.2.6 DNA ligation

100 ng of restriction digest (RD)-PCR product and plasmid were ligated with 0.5 units of bacteriophage T4 DNA ligase, 1x ligation buffer in a final volume of 10 μL,

at 4°C overnight. Controls were also performed and these contained reaction mixtures with RD-plasmid only, and with RD-plasmid without T4 DNA ligase.

4.2.7 Transformation

50 µL of *E. coli* BL21(DE3), XL-S, or Origami competent cells were incubated with 2 µL of the product from the DNA ligation on ice for 15 minutes. The reaction mixture was then heat shocked at 42°C for 30 seconds, placed on ice for a further 10 minutes before being plated onto LB agarose plates containing the appropriate antibiotic, and incubated overnight at 37°C. Controls included a transformation of competent cells with uncut vector and untransformed competent cells. Controls were used to ensure that transformation was succesful and to check the transformation efficiency. Single colonies were selected and grown at 37°C for 3 hours in 300 µL LB broth containing the selective antibiotic. 100 µL of each culture was removed for PCR screening.

4.2.8 PCR Screen

100 µL of transformed cells were centrifuged at 5,000 x g for 1 minute, the supernatant removed and the pellet resuspended in 50 µL of milli-Q water. After boiling for 4 minutes, the sample was centrifuged at 11,000 x g, and the supernatant collected and used as a template for PCR. PCR amplification was performed (4.3.3) using the primers for cloning. A positive control was included which contained the initial DNA template. PCR products were checked on a 1.2% agarose-TAE gel (2.1.8).

4.2.9 Plasmid Isolation

After transformation (4.2.7) and confirmation by PCR screen (4.2.8) colonies containing the insert were selected, inoculated into 4 mL of LB containing the appropriate antibiotic, and incubated for 6 hours at 37°C with shaking. After incubation cells were pelleted by centrifugation at 8,000 x g for 3 minutes at room temperature and the supernatant was decanted. Plasmid DNA was isolated using the Roche High Pure Plasmid Isolation Kit and the plasmid concentration was determined using the NanoDrop (2.1.10).

4.2.9 DNA sequence analysis

400 ng of plasmid DNA isolated from the transformed cells (4.2.9) was incubated with

3.2 pM of forward primer and 3.2 pM of T7 terminator primers and sequenced using the dye-termination method on an ABI Prism 3730 capillary sequencer. DNA sequencing was done by the Massey University Allan Wilson Centre for Molecular Evolution and Ecology Genome Service.

4.2.10 Growth of Genetically Modified Bacteria

E. coli origami cells were transformed with the proEx or pET32a plasmid containing either the *Y. frederiksenii* or *E. faecalis* glycosyltransferase gene. Cultures were grown at temperatures ranging from 25°C to 37°C both with and without agitation.

4.2.11 Induction of Gene Expression

Transformed *E.coli* origami was grown with or without agitation at 25-37°C in 200 mL of LB broth containing a final concentration of 100 µg/mL ampicillin, 15 µg/mL kanamycin, 14 µg/mL tetracycline, and 34 µg/mL chloramphenicol. When an OD_{600nm} between 0.5-1.0 was reached, gene expression was induced with IPTG at a final concentration between 0.05-0.5 mM. Following induction, the culture was grown overnight with shaking at 25°C before being harvested. As a control, a culture was grown as above, but without the addition of IPTG.

It was found that yield and solubility of the glycosyltransferase appeared best when cells were grown with agitation at 30°C till an OD_{600nm} of 0.5 was reached and when 0.5 mM IPTG was used for induction. Therefore these conditions were used for expression.

4.2.12 Harvesting and Lysis of Transformed Cells

Bacteria were harvested by centrifugation at 13,000 x g for 20 minutes at 4°C and were then resuspended in lysis buffer with or without 10% glycerol. *E. coli* cells were lysed on ice by two passes through a French pressure cell (Wabash, Aminco Instruments Co.) at 4000 psi or using an ultrasonic liquid processor (Misonix, Inc) with a 3mm tip at 3 x 60 s bursts with 60 s pauses and at an amplitude of 60 mA.

Cell lysate was incubated at room temperature for 30 minutes with 2 mg of DNase after which time a whole cell sample was collected. Cell debris was then removed by centrifugation at 13,000 x g for 20 minutes at 4°C. The supernatant was then used for IMAC purification (4.2.13). To test solubility of the glycosyltransferase a small sample of the supernatant and cell debris were resuspended in 8 M urea, and these along with the whole cell sample were analysed by SDS-PAGE (2.1.13).

Lysis buffer: 1 mM EDTA
2x Protease Inhibitor
0.5 mM TCEP
Made up in 1x PBS

4.2.13 Immobilised Metal Affinity Chromatography

Sepharose chelating resin (Amersham Biosciences, USA) was packed into a column (10 mL Poly-Prep®, Chromatography columns, BioRad Labroratories, USA), rinsed in 3 CV's of H_2O then loaded with a 1 CV of a 0.2 M solution of nickel chloride. The column was then washed with 5 CV's of H_2O and equilibrated with 2CVs of 0.5 M NaCl (pH 7.5). Protein samples were loaded onto the column at atmospheric pressure then eluted with an increasing step-wise gradient between 0 and 500mM of imidazole solutions containing 0.5 M NaCl, 50 mM PBS (pH 7.5). Purification was carried out using a low pressure BIO-RAD Econo System at room temperature. Samples at each gradient step were analysed using SDS-PAGE (2.1.13).

4.2.14 On-column Refolding

On column refolding was carried out using the protocol provided by GE Healthcare for the ÄKTA prime plus using an IMAC column.

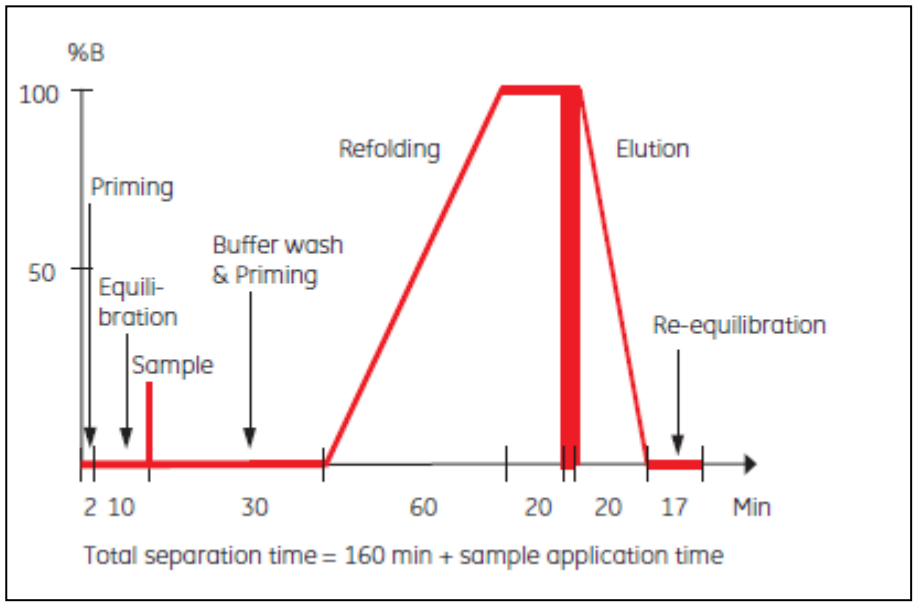


Figure 4.1 On-column refolding gradient template used (ÄKTAprime plus cue cards p. 14).

Fractions (1 mL) were collected by a fraction collector every two minutes and analysed *via* SDS-PAGE along with the resin (2.1.13).

4.2.15 Glycosyltransferase assay

GTase assays were performed following a modified method for a bacterial galactotransferase by Wakarchuk *et al*, (1998). Protoplasts or whole cells were lysed (2.1.11 or 2.1.12) and centrifuged at 3,000 x g for 5 minutes after which, 100 µL of supernatant was removed (whole protoplast/cell fraction). The remaining supernatant was subjected to ultracentrifugation at 60,000 x g for 60 minutes after which the supernatant was removed and saved as the cytosol fraction. The pellet was solubilised in 2 volumes of solubilisation buffer then sonicated for 10 minutes to produce the membrane fraction. 10 µL of the supernatant or solubilised pellet were then added to 40 µL of reaction mixture and incubated at 37^{o}C for 60 minutes. The reaction was stopped by addition of 0.5% acetic acid and samples were filtered through a 0.2 µL filter (Millipore, MA, USA) before being subjected to reverse phase HPLC (Ultimate 3000, Dionex, Germany) using a Jupiter 5µm C_{18} 300 Å column (250 x 4.6 mm, Phenomenex). Elution of the solute was achieved using the following solutions to form a linear gradient: A: H_2O, 0.1% TFA, and B: Acetonitrile, 0.8

% TFA. The gradient started at 10% B which was increased to 25% B over 25 minutes at a constant flow rate of 1 mL/min. The concentration of B was increased to 100% over 5 minutes and held for 3 minutes until it was returned to 10%. Elution was monitored using a Photodiode Array Detector (Dionex) at 214 nm and 280 nm and the instrument was controlled using Chromeleon software V 6.80. Peaks of interest were collected for further analysis by ESI-MS (2.1.18). Controls were used where peptide, substrate or 'enzyme' were excluded. The peptide substrates, C15, SubA, or SubC were also subjected to RP-HPLC without 'enzyme' or 'substrate' in order to determine their elution pattern, and eliminate their associated peaks from the reaction profile.

Solubilisation buffer:	20 mM Bis-Tris (pH 7)
	1% Triton X-100
	1 µg/µL Leupeptin
	1 µg/µL Aprotinin
	0.5mM TCEP
Reaction mixture:	25 mM Bis-Tris-HCl (pH 6.5)
	0.25% Triton X-100 (Only included for membrane fraction)
	1 mM Metal ion ($MgCl_2$, $NiCl_2$, $CaCL_2$, $CoCl_2$, $MnCl_2$, $ZnCl_2$)
	1 mM UDP-GlcNAc or GlcNAc
	1 mM Peptide (C15, SubA, or SubC)
	1 µg/µL aprotinin
	1 µg/µL leupeptin
Peptides	C-15- FGILHHSSGSSSYHC
	SubA- AGYDSGTADYL
	SubC- AGYDSGTCDYL

The C-15 peptide was chosen as it contains the C-terminal cysteine residue which is GlcNAcylated in mature GccF. The SubA and SubC peptides were chosen as they both contain the serine residue which is also GlcNAcylated in GccF. SubC contains

the residues from $GccF_{14-24}$, whereas SubA is the same as SubC except for a substitution of Cys21 for an Ala. This substitution was made in case the glycosyltransferase GlcNAcylates the cysteine as opposed to the serine residue in the peptide which this is undesirable as Cys-21 is involved in a disulfide bond in mature GccF.

4.3 BIOINFORMATICS

Due to the difficulty in cloning the family 2 putative glycosyltransferase GccA present in the GccF gene cluster in *L plantarum* KW30, bioinformatic searches were undertaken to find alternative glycosyltransferases similar to GccA residing in bacteriocin gene clusters. Identification of homologous family 2 glycosyltransferases was carried out using a protein BLAST search firstly against the *L. plantarum* KW30 GTase sequence, and secondly the *L. plantarum* WCFS1 GTase sequence. Sequences in the results list that showed a low e-value and contained a glycosyltransferase core domain were initially selected. The list was then refined by searching the regions upstream and downstream from each putative GTase for an ABC transporter as these proteins are almost always associated with bacteriocin clusters. The list was further refined by using the ORF finder from the NCBI website to detect open reading frames near the site of the putative GTase. Small open reading frames that were likely to be potential antimicrobial peptides were also investigated to identify potential bacteriocins with characteristics such as a double glycine leader sequence, along with odd numbers of cysteine residues and strings of small amino acid residues such as serine and glycines, characteristic of GccF. The purpose of this was to ensure that the glycosyltransferase genes found within the organisms genome were potentially involved in the glycosylation of a bacteriocin. The following organisms listed in Table 4.1 were selected based on the level of sequence homology with GccA, and their co-localisation with an ABC transporter and small bacteriocin-like peptides.

Organism	Accession Number	Putative Molecular Weight (Da)	% Identity	E-value
Lactobacillus plantarum WCFS1	28377316	46848	29	8e-19
Bacillus subtilis supsp. subtilis str. 168	16079204	49765	30	8e-34
Enterococcus faecalis ATCC 700802	29377941	53122	32	3e-36
Streptococcus suis 89/1591	223932213	51828	31	5e-19
Streptococcus mutans	88770853	50610	29	1e-18
Bacillus thuringiensis serovar andalousiensis BGSC 4AW1	228937261	41819	30	9e-38

Alicyclobacillus acidocaldarius sub acidocaldarius LAA1	218289680	85212	29	8e-22
Arthrospira maxima CS-328	209494006	44166	27	4e-17
Yersinia mollaretii ATCC 43961	238797019	24008	31	5e-11
Yersinia frederiksenii ATCC 33641	238789003	47572	31	1e-37

Table 4.1 Protein BLAST results against *L. plantarum* KW30.

For each organism a GTase sequence was found which contained a glycosyltransferase core domain, as well as one or several ORFs that contained small peptides with the double glycine leader sequences characteristic of known (exported) bacteriocins. In total fourteen predicted bacteriocin encoding clusters were found for the ten organisms.

CLC sequence viewer was used to carry out an initial multiple sequence alignment on the GTases from the above organisms. Each sequence was individually entered into the program and an alignment was made based upon amino acid similarity.

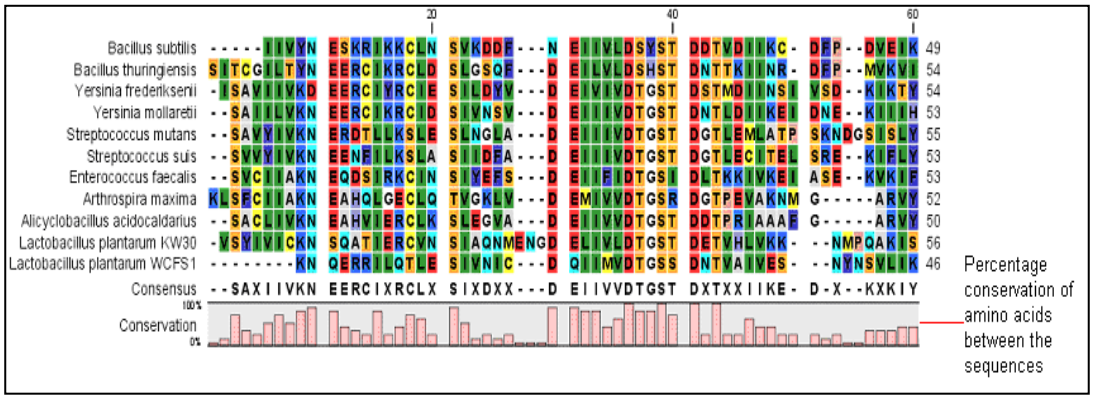


Figure 4.2 Multiple alignment excerpt from the results obtained using CLC sequence viewer.

Figure 4.2 shows a region of fairly high sequence conservation within the different GTases, which is within the predicted GTase core domain of the proteins, and may include the active or binding site of the enzymes. The results of this alignment along with the conserved domain search function on the NCBI server (Conserved domain database, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) resulted in the identification of distinct domains within these proteins as shown in Fig 4.3.

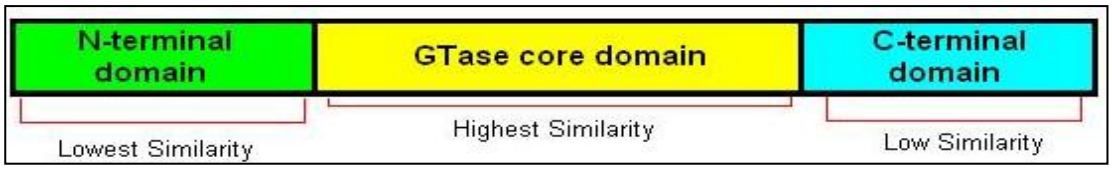


Figure 4.3 Representation of the domains and sequence similarity within the family 2 glycosyltransferase proteins.

Further multiple alignments were carried out using the glycosyltransferase core domain alone, and the N- and C-terminal domains found within each organism. This was undertaken in order to identify consensus sequences and sequence motifs that might be suitable to use in searches for GTases in other organisms. Any sequence motifs located were subjected to analysis by the program ScanProsite, which locates sequence matches in the protein database, and generates a list of results in the order in which they are found.

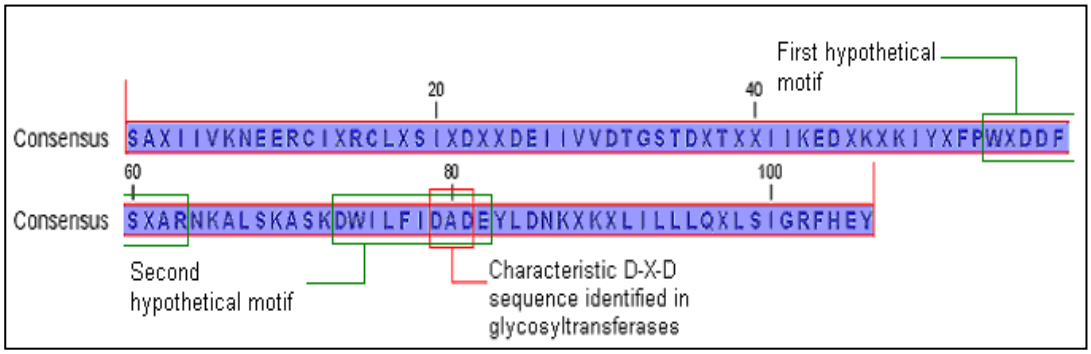


Figure 4.4 Consensus sequence obtained from CLC sequence viewer for the GTase core domain of the family 2 glycosyltransferase sequences.

Figure 4.4 highlights a D-X-D motif found in the consensus sequence of the different GTase core domains which has been described as a conserved sequon in GTases (Breton and Imbert, 1999; Ünligil and Rini, 2000; Jianghong *et al*, 2001). The CLC sequence viewer multiple alignment (not shown) indicates a high percentage of conservation in the amino acids in the N-terminal region of the GTase core domain, which decreases towards the C-terminal end of the GTase core domain. Within the N-terminal part of the GTase core domain there appear to be two conserved sequence motifs:

1 - W-$X_{(2)}$-D-[F/Y]-$X_{(3)}$-R-N

2 - [E/D]-[W/Y]-[I/L/V]-$X_{(2)}$-[L/I/V/F]-D-X-D-[E/N].

When these two sequences were run through ScanProsite (ScanProsite tool, http://prosite.expasy.org/scanprosite/), the first showed 59 matches, 28 of which were located in various GTase ORFs. The second gave 876 matches, some of which were located in sequences annotated as GTases, but most were in sequences annotated as various other proteins, some of which were hypothetical.

Figure 4.5 Consensus sequence obtained from CLC sequence viewer for the N-terminal domain of the family 2glycosyltransferase sequences.

When the N-terminal domain consensus sequence (Fig. 4.5) was used in a CLC sequence viewer, multiple alignment showed there was little conservation between the amino acid sequences of the different GTases, primarily because of the large gaps between conserved residues and the low conservation of these residues.

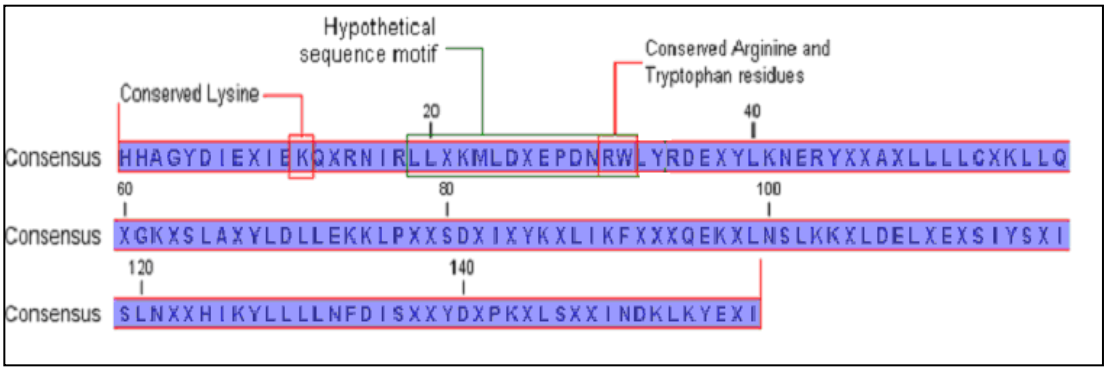


Figure 4.6 Consensus sequence obtained from CLC sequence viewer for the C-terminal domain of the family 2 glycosyltransferase sequences. A hypothetical sequence motif with highly conserved residues outlined in boxes in red and a hypothetical sequence motif outlined in a green box.

When viewed within the CLC sequence viewer there appeared to be several well conserved residues within the C-terminal domain (Fig 4.6); a lysine was conserved in all but two species, and an arginine – tryptophan couple was conserved in all but three species. There was also a long sequence motif of [L/I]-[M/L]-$X_{(3)}$-[M/I/L/Y]-$X_{(3)}$-P-X-[N/D]-R-W, which when searched on

ScanProsite matched up to 26 other proteins. Several of these proteins were hypothetical and matched the length of other GTases. Upon visual inspection, they also shared some of the potential characteristics of GccA.

As it is thought that the GTases may be localised in the membrane of the organisms, possibly forming a complex with the ABC transporter, the sequence of each GTase was assessed using three transmembrane prediction programs the first being ProtScale (ProtScale tool, http://web.expasy.org/protscale/), which predicts membrane spanning regions. It uses the amino acid hydropathicity scale of Kyte and Doolittle to assign each amino acid a value based on its hydropathicity to predict such regions. A second analysis was carried out using the Dense Alignment Surface (DAS) method (Transmembrane prediction server, Cserzö 1997), where the query sequence is plotted against a variety of non-homologous membrane proteins, and the results are analysed with a special scoring matrix. A value of or above 1.7 is the loose cut-off which signifies the location of a predicted transmembrane segment within the protein, whereas a value of or above 2.2 indicates a strict cut-off which is informative in the terms of the number of matching segments to the non-homologous membrane proteins. A third analysis was carried out using TOPCONS (Bernsel, *et al,* 2009) which includes the prediction programs SCAMP-seq, SCAMPI-msa, PRO, PRODIV and OCTOPUS which predicts transmembrane domains using a consensus from the five different topology prediction algorithms used by the programs listed above. The predictions are then put into the TOPCONS hidden Markov model (HMM), which gives a reliability score based upon the agreement of the included methods and a consensus prediction for the protein.

Table 4.2 Predicted hydropathic and transmembrane regions located within glycosyltransferases from ProtScale, DAS, and TopCons plots.

Organism	**N-terminal Domain**			**GTase Core Domain**			**C-terminal Domain**		
	ProtScale	DAS Plot	TopCons	ProtScale	DAS Plot	TopCons	ProtScale	DAS Plot	TopCons
Lactobacillus plantarum KW30	-	-	-	1	1	-	-	-	1
Lactobacillus plantarum WCFS1	-	-	-	-	-	-	2	2	-
Bacillus subtilis supsp. subtilis str. 168	-	1	-	-	1	-	1	1	1
Enterococcus faecalis ATCC 700802	-	-	-	-	-	-	-	2	1
Streptococcus suis 89/1591	-	-	-	-	-	-	1	2	-
Streptococcus mutans	1	1	-	-	1	-	-	2	-
Bacillus thuringiensis serovar andalousiensis BGSC 4AW1	-	1	-	1	1	-	1	-	1
Alicyclobacillus acidocaldarius sub acidocaldarius LAA1	-	-	-	-	-	-	2	2	2
Arthrospira maxima CS-328	-	-	-	-	-	-	1	1	-
Yersinia mollaretii ATCC 43961	-	-	-	1	-	-	-	-	-
Yersinia frederiksenii ATCC 33641	1	-	-	1	1	-	-	-	-

As the different transmembrane prediction programs use different algorithms in order to predict transmembrane regions and therefore vary considerably in their results a minimum of two out of the three programs in agreement were required to confirm the presence of a predicted transmembrane region (Table 4.2). For the putative GTase found in *B. subtilis* a transmembrane region within the C-terminal domain has been predicted by all three prediction programs. For the GTases in *E. faecalis, B. thuringiensis, S. suis,* and *A. maxima* a transmembrane spanning region has also been predicted within their C-terminal domains by two out of the three prediction programs. The GTases from both *L. plantarum* WCFS1 and *A. acidocaldarius* have two transmembrane spanning regions predicted within their C-terminal domain. A single transmembrane spanning region was predicted to occur in the GTase core domains of *L. plantarum* KW30, *B. thuringiensis*, and *Y. frederiksenii*. *S. mutans* was unusual as it was the only putative GTase to contain a predicted transmembrane spanning region within its N-terminal domain, whereas the putative GTase in *Y. mollaretii* was predicted to contain no transmembrane spanning regions. That all of the glycosyltransferases (except *Y. mollaretii*) identified are predicted to contain transmembrane spanning regions, albeit in different places in their sequences, may indicate that they are co-located in the cell membrane with an ABC transporter or within a complex where the transmembrane region is shielded.

Secondary structure predictions were then carried out on the N- and C-terminal domain sequences of the GTases using FUGUE (http://www-cryst.bioc.cam.ac.uk/fugue/) in order to determine where the secondary structural elements may be in the sequence. Results from such a prediction can be used to inform primer design for cloning. For example it is desirable that secondary structural elements are not disrupted when cloning domains. FUGUE matches the query sequence to other sequences of known structure based on amino acid composition, and then aligns and assigns the query sequence a secondary structure.

Results from FUGUE indicated that there are a number of secondary structure homologues in the C-terminal domain of the GTases listed in Table 4.2, which include, alpha-helices and tetratricopeptide repeats (TPRs). Interestingly the sequences of *Y. molaretii* and *Y. frederiksenii* did not contain any of these repeat

sequences, although they were predicted to have several alpha-helices within their C-terminal domain. In fact secondary structures predicted for each of their sequences showed no similarity at all and were not similar to those predicted for the GTases from the other organisms.

FUGUE analysis of the N-terminal regions of the different GTases predicted no secondary structure and there was no similarity between the predictions for the various GTases listed in Table 4.1.

TPRs are predicted to have a function in the recognition of intact protein substrates rather than being involved in the recognition of small peptide substrates (Lyer and Hart, 2003). This suggests that the C-terminal regions of the GTases may form a complex with the other components of the cluster, possibly the ABC transporter, allowing glycosylation of the bacteriocins before they are secreted from the cell. Clarke, *et al.* (2008) also suggested that TPRs may have a role in the active site of GTase domains. As these TPRs are located in the C-terminal regions of the putative GTases, which have also been predicted to contain hydrophobic regions, it is possible that the C-terminal domains of these enzymes may have a role in enzyme activity. The hydrophobicity of the TPR regions may mediate substrate/bacteriocin recognition, resulting in a conformational change which helps to orientate the substrate correctly. This movement would be thermodynamically favourable due to a favourable entropic gain as the hydrophobic residues in the TPR regions become shielded from the hydrophilic environment.

4.4 STRUCTURAL CHARACTERISATION OF GccA

MOLECULAR CLONING

Molecular cloning is a powerful technique which can be used as a step towards determining the structure, function, and regulation of specific genes and their products. This section reports the experiments carried out to obtain soluble and correctly folded recombinant glycosyltransferases from *Y. frederiksenii* ATCC 33641 and *E. faecalis* ATCC 700802.

The first step in molecular cloning is to isolate the target gene of interest, using polymerase chain reaction (PCR). The gene can then be cloned into an appropriate vector, which is then transformed into a heterologous host. Generally, a bacterial host expression system is used for recombinant protein production. This is due to the large number of expression systems developed for bacteria, the rapid growth rate of bacteria, and the ease of use. Bacterial expression systems may not be ideal hosts to produce recombinant proteins from some genes as is the case for certain genes from eukaryotic organisms, as the proteins produced are often post-translationally modified. Eukaryotic proteins also often require assisted protein folding which bacterial systems may not provide. For the production of eukaryotic genes or large gene products, other hosts such as yeast can be used. Problems may also arise when using a bacterial expression system to express a gene from a source which has different codon use. Such differences can be overcome by using host strains that have been adapted to translate such sequences. Selecting an appropriate host for expression is very important for the successful production of a recombinant protein.

Although there are a large number of host organisms and molecular cloning vectors available the majority of cloning experiments use the bacterium *Escherichia coli* (*E. coli*) as a heterologous host. This is because much is known about its sequence, genetic makeup, and its microbiology, and there are a number of strains that have been engineered to provide specific advantages for the production of recombinant proteins. Because of its high use there are also a large number of cloning vectors available. Small plasmid vectors are the simplest and easiest to use and typically contain: an origin of replication; promoter (often T7

promoter); multiple cloning sites; a selectable marker such as an antibiotic; and in some cases a ribosome binding site and a promoter termination sequence (Fig 4.7).

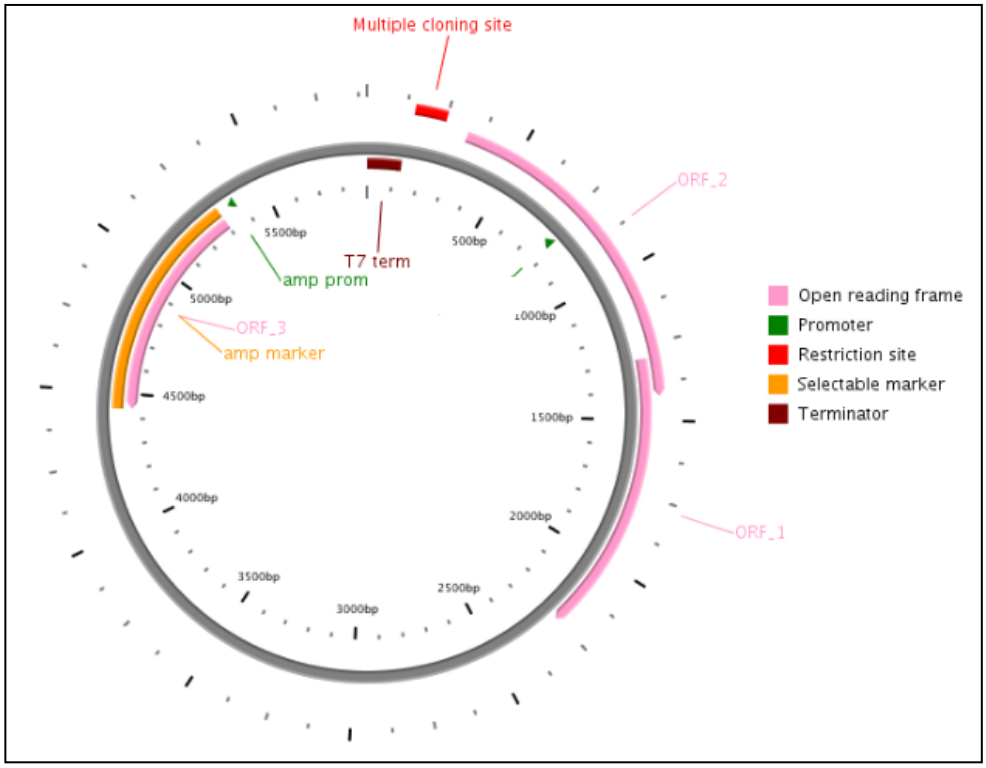


Figure 4.7 Schematic of a typical plasmid expression vector (PlasMapper ver 2.0).

To clone an open reading frame a gene is amplified using PCR in such a way that restriction enzyme recognition sites are included at both ends of the gene. It can then be inserted into a vector using the matching restriction enzyme sites within the multiple cloning site. The multiple cloning site is located within a 'cassette' which contains a ribosome binding site, a promoter, and a promoter termination sequence. Most plasmids also contain a selective marker which is used to isolate the host organisms which have been transformed with the vector. When a purification tag is desired it can be added either directly after the promoter or just before the promoter termination sequence. The addition of a tag is very convenient for purification of the recombinant protein. It can also increase the stability and solubility of the recombinant protein after translation. However, tags may need to be removed upon purification as they can interfere with correct protein folding. For this reason a cleavage site may also be included in the plasmid to allow for easy removal of the tag using a specific protease (Sambrook, 2001).

4.4.1 RESULTS AND DISCUSSION

GLYCOSYLTRANSFERASE CLONING

To investigate the proteins encoded by the predicted glycosyltransferase genes in *Y. frederiksenii* ATCC 33641 and *E. faecalis* ATCC 700802 it was necessary to clone them into suitable expression systems to produce recombinant protein. The pET32a and proEx expression vectors were trialled for expression of these genes in both tagged and untagged forms (Table 4.3).

Organism	Tag	Forward Primer Restriction Site	Reverse Primer Restriction site	Vector
Y. frederiksenii	N-terminal His-tag	NcoI	XhoI	proEx
	C-terminal His-tag	NdeI	XhoI	pET32a
	Untagged	NdeI	XhoI	proEx
E. faecalis	N-terminal His-tag	NcoI	XhoI	proEx
	C-terminal His-tag	NdeI	XhoI	pET32a
	Untagged	NdeI	XhoI	proEx

Table 4.3 The vector, restriction site and tagged state of the GTase genes from *Y. frederiksenii* and *E. faecalis* (Primer list in Appendix II).

As the *E. faecalis* GTase gene contains an NdeI restriction site, site directed mutagenesis was necessary to disrupt the internal NdeI restriction site so that C-terminal His-tagged and untagged constructs could be generated using the pET32a plasmid.

Although the pET32a vector contains a His-tag, there is a large thrombin and S-tag between the His-tag coding region and the insert region. These add a large number of amino acids at the N-terminus which is not desirable. For this reason N-terminal His-tagged constructs of the GTase genes from *Y. frederiksenii* and *E. faecalis* were cloned into the pProEx plasmid.

PCR amplifications of the 1.3 kb and 1.2 kb GTase genes from *Y. frederiksenii* and *E. faecalis* , respectively, were carried out using forward and reverse primers with flanking NdeI, NcoI, and XhoI sites depending upon the construct (Appendix II). The PCR products were then digested with the appropriate restriction enzymes along with their target vector. Figure 4.8 shows the results of a digest of the GTase genes from both *Y. frederiksenii* and *E. faecalis* along with the proEx plasmid, using the restriction endonucleases NcoI and XhoI. Figure 4.8 shows that the undigested proEx (lane 7) used as a control has a different position on the gel in relation to the single and double cut proEx plasmid (lanes 5 and 6) which indicates that the plasmid has been cut.

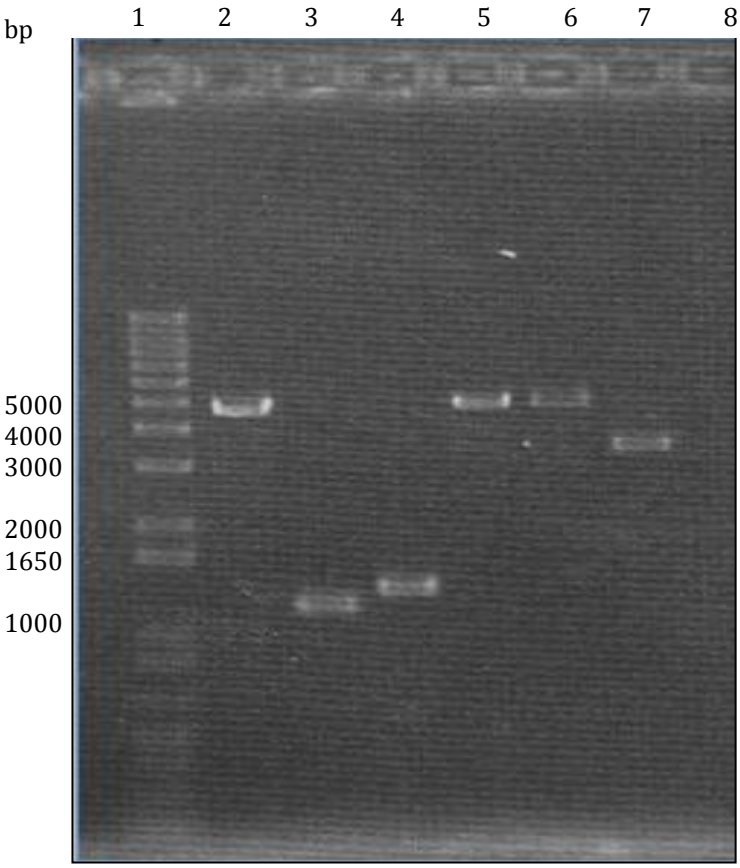


Figure 4.8 1.2 % agarose electrophoresis gel of restriction digest for glycosyltransferase genes from *Y. frederiksenii* and *E. faecalis*. (Lane 1, 1Kb plus DNA ladder; 2, proEx vector digested with NcoI and XhoI; 3, *Y. frederiksenii* GTase gene digested with NcoI and XhoI; 4, *E. faecalis* GTase gene digested with NcoI and XhoI; 5, proEx digested with NcoI; 6, proEx digested with XhoI; 7, proEx control).

The digested GTase genes and were ligated with their corresponding digested plasmids and these products of ligation were transformed into competent *E.coli*

cells which were then plated on LB media containing the appropriate antibiotics. Colonies from these plates were screened for successful transformation using PCR. Figure 4.9 shows one such PCR screen for the transformation of the *Y. frederiksenii* GTase gene. Of the many colonies picked, one in lane 7 shows a PCR product that appears to contain the GTase gene when compared to the GTase gene control in lane 8. That the insert only appears in one of the seven colonies is unusual and may suggest that there was a problem during the restriction digest such as incomplete or single cutting of some of the insert and vector.

The plasmid from the colony in lane 7 was extracted for sequencing using the forward primer for cloning and a T7 reverse primer, to ensure the insert was present and that the sequence was without error.

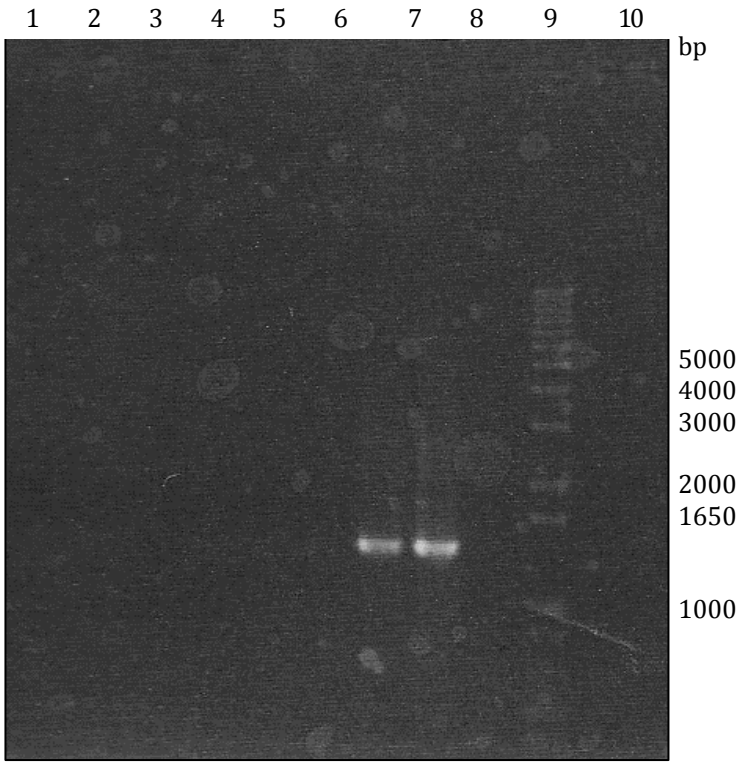


Figure 4.9 1.2% agarose electrophoresis gel of a PCR screen to check transformation. (Lanes 1-7, colonies selected from *Y. frederiksenii* GTase gene transformation plates; 8, *Y. frederiksenii* GTase control; 9, no sample; 10, 1Kb plus DNA ladder).

Sequencing results for band 7 confirmed that the band seen is the *Y. frederiksenii* glycosyltransferase and that the sequence was error free.

4.5 GENE EXPRESSION AND PROTEIN PURIFICATION

Purification involves the separation of a molecule or molecules from foreign or contaminating substances. There are many different purification techniques used in biochemistry such as affinity purification, SEC, IEX, HIC, filtration, centrifugation, and fractionation. The purification of a protein requires many considerations as to the purity needed with respect to its future use. Typically most experiments require a high level of purity and therefore several purification steps may be necessary. As proteins contain different properties these can be exploited as a means of purification.

In many cases genes are expressed to give proteins with a fusion tag which allows for purification of the protein by using the properties of the tag for affinity chromatography. Immobilised metal ion affinity chromatography (IMAC) is a type of affinity chromatography that uses a 6-histidine tag and is often used for purification as there are many available expression vectors which incorporate such a tag near the multiple cloning site. IMAC is based on the ability of a protein to form a complex with metal ions, but with recombinant proteins, a tag is expressed specifically for this purpose. Metal ions used in IMAC include; Cu^{2+}, Ni^{2+}, Zn^{2+}, Fe^{3+}, Ga^{2+}, and Zn^{3+}. Certain metal ions have interactions with different amino acid side chains, for example Ni^{2+}, which has a strong affinity for exposed histidine residues, but can also interact with cysteine or tryptophan, whereas Fe^{3+}, Ga^{2+}, and Zn^{3+} interact with phosphate groups added by post translational modification. In order to immobilize the metal ion the matrix for binding is charged with a chelating agent, such as nitrilotriacetic acid (NTA) or iminodiacetic acid (IDA), that then can be loaded with transition metal ions (Chaga, 2001).

In the case of Ni^{2+} the strength of the affinity interaction increases with the number of histidine residues. Therefore a routine way of purifying a protein using Ni^{2+} IMAC is by the addition of a 6x histidine tag. There are many commercially available resins which are able to chelate metal ions and these can be used to capture and immobilize a protein which has the ability to form a metal ion complex. Elution of the target protein is often by cleavage of the tag or by addition

of a competitive binding agent such as imidazole, whose structure resembles histidine and thus competes for position on the metal loaded resin.
Due to the innate ability of some proteins to form metal ion complexes purity of the protein of interest after IMAC may be compromised. Therefore a second purification step is often required to achieve the necessary purity. However, upon each subsequent purification step the concentration and often the integrity of the target protein is likely to decrease.

4.5.1 RESULTS AND DISCUSSION

Y. FREDERIKSENII GTase GENE EXPRESSION AND PROTEIN PURIFICATION

The three GTase gene constructs from *Y. frederiksenii* were expressed in origami cells in order to produce soluble protein for purification. Expression was trialled on a small scale (10 mL cultures) and a low level of soluble product between ~10-20%, compared to the band in the pellet was seen on SDS-PAGE (Fig. 4.10). A larger scale preparation was then carried out under the same conditions and soluble protein was once again observed at a similar percentage as seen earlier.

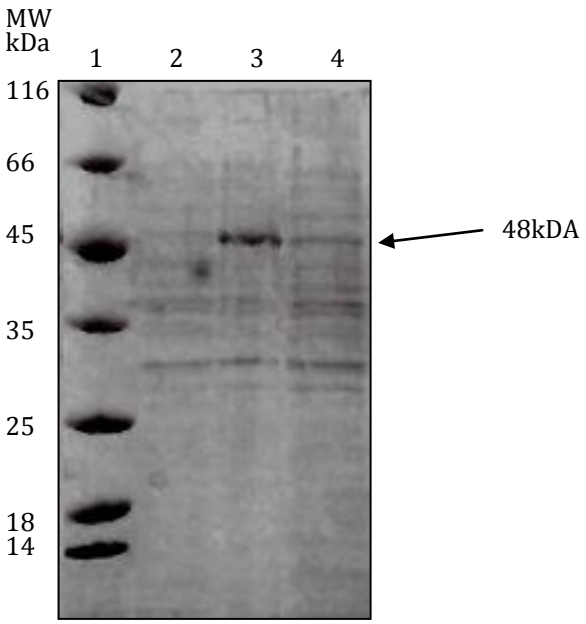


Figure 4.10 12.5% acrylamide gel showing *Y. frederiksenii* N-terminal His-tagged glycosyltransferase expression in *E. coli* grown at 30°C induced at an OD 0.5 with 0.5mM IPTG. (Lane 1, molecular marker; 2, uninduced whole cell lysate; 3, cell pellet; 4, cell supernatant).

When the supernatant was subjected to IMAC, analysis of the eluted fractions showed no band at or near 48 kDA. There are two possible reasons for this; The glycosyltransferase has been degraded (despite the addition of protease inhibitor) or that it remained bound to the resin. Both of these are indications that the recombinant protein is incorrectly folded. As such a small percentage of it appeared in the soluble fraction, this seems very likely.

As greater success was initially obtained with the glycosyltransferase produced by *E. faecalis* subsequent cloning experiments were performed on this protein.

E. FAECALIS GTase GENE EXPRESSION AND PROTEIN PURIFICATION

Trials were performed in order to determine the best conditions for *E. faecalis* glycosyltransferase gene expression.

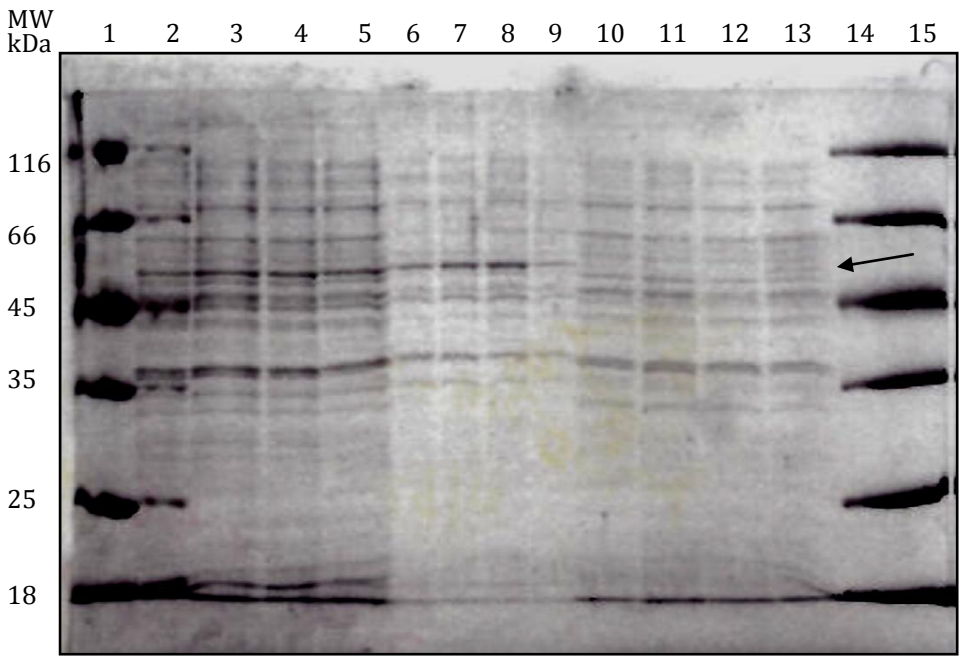


Figure 4.11 12.5% acrylamide gel showing an *E.faecalis* His-tagged glycosyltransferase expression trial in *E. coli* induced at an OD 0.5. (Lane 1 and 14, molecular marker; 2, whole cell lysate grown at 25°C induced with 0.05mM IPTG; 3, whole cell lysate grown at 25°C induced with 0.5mM IPTG; 4, whole cell lysate grown at 30°C induced with 0.05 mM IPTG; 5, whole cell lysate grown at 30°C induced with 0.5 mM IPTG; 6, cell pellet grown at 25°C induced with 0.05 mM IPTG; 7, cell pellet grown at 25°C induced with 0.5mM IPTG; 8, cell pellet grown at 30°C induced with 0.05mM IPTG; 9, cell pellet grown at 30°C induced with 0.05 mM IPTG; 10, supernatant grown at 25°C induced with 0.05 mM IPTG; 11, supernatant grown at 25°C induced with 0.5 mM IPTG; 12, supernatant grown at 25°C induced with 0.05 mM IPTG; 13, supernatant grown at 30°C induced with 0.05 mM IPTG; 13, supernatant grown at 30°C induced with 0.5 mM IPTG).

The His-tagged glycosyltransferase has a molecular weight of ~53 kDa and as can be seen in both the whole cells and pellet fractions (Fig 4.11) there is a light band present at this molecular weight. For the supernatant there is no band visible which corresponds to the expressed glycosyltransferase except for cells grown at 30°C with 0.5 mM IPTG (indicated by arrow) used for induction with agitation. These results indicate that the glycosyltransferase is largely insoluble with only a small amount being soluble in the sample grown at and induced at an OD of 0.5 with 0.5 mM IPTG at 30°C. This condition was therefore used for the growth and

induction of cells used to produce recombinant glycosyltransferase. For the uninduced control (not shown) a band at ~53 kDA is not visible in the whole cell, pellet, or supernatant fractions. This confirms that the band seen at a Mwt of ~53 kDa is most likely to be the glycosyltransferase.

As the *E. faecalis* ATCC 700802 constructs were His-tagged, affinity chromatography using Ni^{2+} bound metal chelating resin was used to enrich for the recombinant proteins which were eluted with increasing imidazole concentration as shown in Figure 4.12.

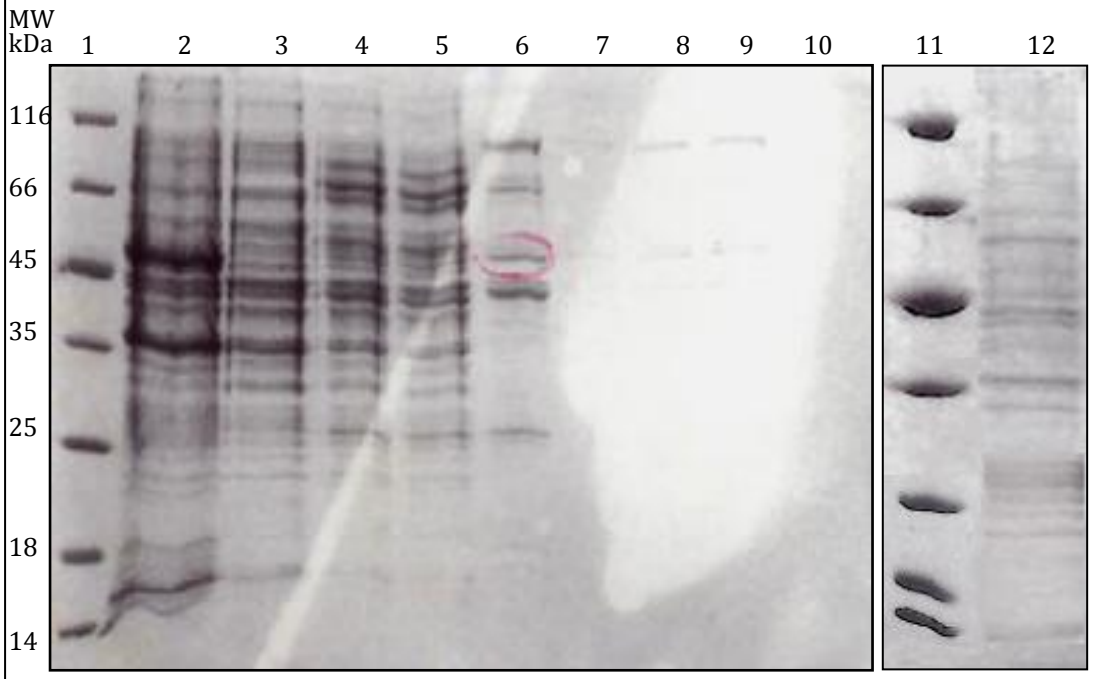


Figure 4.12 12.5% acrylamide gel showing *E. faecalis* His-tagged glycosyltransferase purification from IMAC using cells from Figure. 2.5 incubated at 30°C induced with 0.5mM IPTG. (Lane 1, molecular marker; 2, cell lysate pellet; 3, cell lysate supernatant; 4, 50mM Imidazole elution; 5, 50mM imidazole elution; 6, 250mM imidazole elution; 7, 250mM imidazole elution; 8, 500mM imidazole elution; 9, 500mM imidazole elution; 10, no sample; 11, molecular weight marker; 12, Uninduced cell lysate control).

Purification of the GTase from *E. faecalis* using IMAC showed a large dark band at ~53kDa, not present in the uninduced cell lysate control in lane 12, which corresponds to the expected weight of the GTase, indicating the glycosyltransferase may be largely insoluble. In lane 6 however there is a band eluting off at an imidazole concentration of 250 mM which corresponds to the putative GTase indicating that soluble GTase is present, and elutes off IMAC

columns with high concentrations of imidazole. Purification was attempted on a larger scale with imidazole elutions ranging from 20-500 mM and this showed that the GTase eluted off the column at all imidazole concentrations. When the solubilised pellet was analysed by SDS-PAGE it was seen that there is a heavy band corresponding to the GTase indicating that the majority of the GTase remained in the pellet. There is a second band eluting off at a ~100kDa which may correspond to dimerised GTase or may just be a protein present with *E. faecalis* which is able to bind with some specificity to Ni^{2+}. As the glycosyltransferase eluted off at all imidazole concentrations it is believed that though soluble it may be in several folded states with some being incorrectly or incompletely folded.

In an effort to further improve yield, solubility, and to prevent the apparent multiple misfolded and unfolded species of the glycosyltransferase from being produced a number of different growth conditions were trialled such as; reducing or increasing oxygen availability, and shaking or not shaking during incubation. These different growth conditions did not increase solubility or yield of the glycosyltransferase. The methods used to lyse cells were also changed and included using a needle sonicator as opposed to the French Press for lysis, including glycerol in the lysis buffer, and lysing cells in a larger volume of solution in order to decrease the protein concentration. However, these changes did not alter the amount of soluble and correctly folded protein produced.

The C-terminal His-tagged glycosyltransferase from *E. faecalis* proved to be less soluble than the N-terminal His-tagged version. It also eluted over a wide range of imidazole concentrations, in a similar manner to the N-terminal His-tagged protein.

On-column refolding was attempted for both the C- and N-terminal His-tagged glycosyltransferases. In this procedure the His-tag is used to immobilize the protein on the resin, before refolding on column to help prevent insoluble aggregates forming. The results obtained using this technique were less than satisfactory with only a small amount of the soluble glycosyltransferase being obtained.

4.6 FUNCTIONAL CHARACTERISATION OF GccA

GLYCOSYLTRANSFERASE ACTIVITY ASSAY

GccA is thought to be responsible for the addition of one or both GlcNAc residues to GccF. The development of a glycosyltransferase enzyme assay is essential to follow the purification of the enzyme and to provide evidence of putative function.

As enzymes work under specific conditions designing an enzyme assay is a difficult process as factors such as the ionic strength, temperature, substrate, and pH need to be investigated in order to determine the optimum conditions for enzyme activity. Developing an enzyme assay using crude cell lysate is also complicated as multiple activities are an issue due to the large number of other enzymes present in the solution which can make interpretation of results difficult. Other molecules present in the cell lysate may also inhibit activity of the enzyme under investigation. It is common practice to add protease inhibitors to crude solutions which introduces the risk that these may also inhibit the enzyme.

Glycosyltransferases are enzymes which are responsible for the addition of specific sugar moieties to peptide or protein molecules. As the specificities of glycosyltransferases are many, there is no standard enzyme assay that can be used to measure their activity. Peptide-glycosyltransferase assays require their particular nucleotide sugar substrate, the peptide or protein which they modify and usually a metal ion, commonly Fe^{2+}, Mn^{2+}, Cu^{2+}, and Mg^{2+}. In the case of GccA it is known that one of the substrates is GlcNAc. It is unclear, however, whether this is as GlcNAc or UDP-GlcNAc. It is also unknown if GccA requires a metal ion for activity. In the case of GccF it is not known whether GccA modifies Cysteine-43, Serine-18 or both therefore different peptide substrates are required. Assays were carried out in the presence and absence of protease inhibitor in case the addition of protease inhibitors inactivated the enzyme. Futhermore EDTA could not be used in the glycosyltransferase enzyme assay because of the requirement by most glycosyltransferases for metal ions.

4.6.1 RESULTS AND DISCUSSION

GCCF ACTIVITY ASSAY

Membrane and cytosol fractions from *L. plantarum* KW30 protoplasts and whole cells were used to try and detect GccA activity. Three small peptides modelled from both the loop region and C-terminal end of GccF were used C-15, SubA and SubC along with UDP-GlcNAc and GlcNAc. As metal ions are a necessity for glycosyltransferases different metal ions were added to the assay in case a metal ion is required for activity. If GlcNAc becomes covalently bound to a peptide, the retention time of the peptide should decrease due to increased hydrophilicity. Of the 236 assays performed only 2 showed a significant shift in retention time for the peptide peak indicating a change to its structure. The two assays both used the C-15 peptide and GlcNAc as the substrates. One contained Mg^{2+} as the metal ion while the other had no added metal ion, both produced a peak with the same retention time (18 minutes, Figure 4.13). Interestingly it was incubation with whole cell cytosol not the membrane fraction which produced this peak, which is at odds with the hypothesis that GccA is localized in the protoplast membrane.

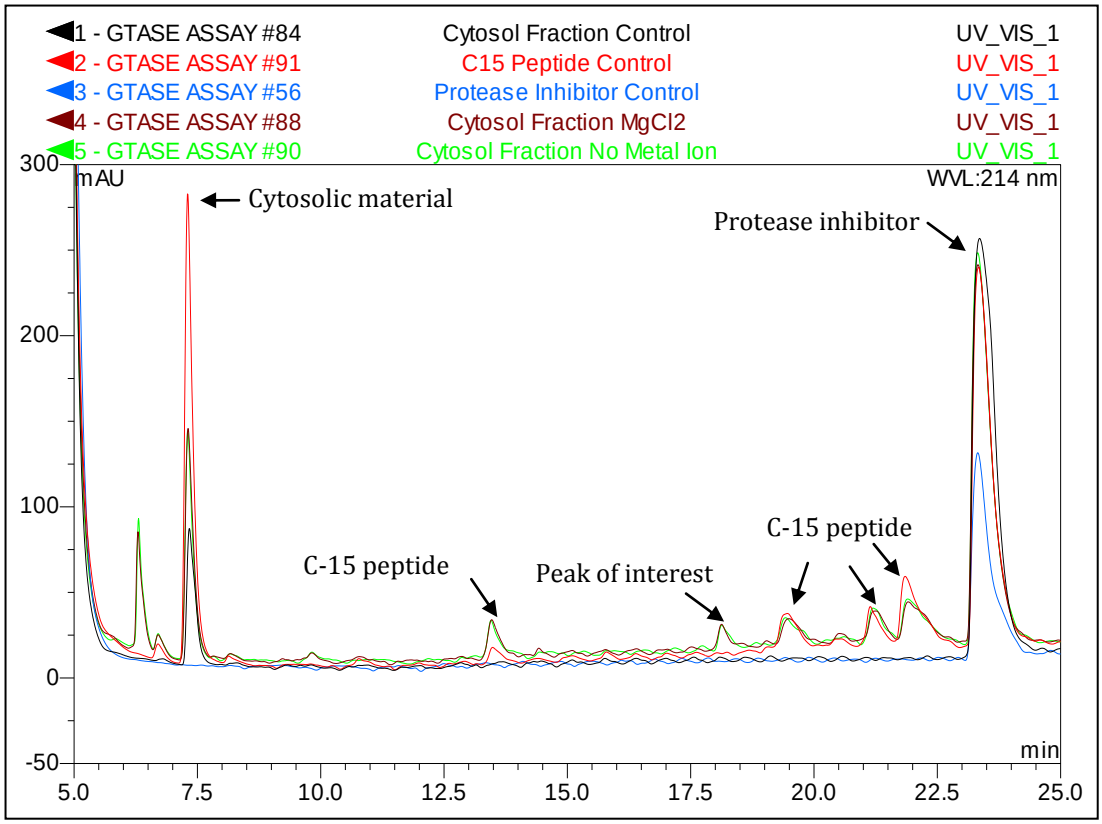


Figure 4.13 Glycosyltransferase assay for GccA using cytosol fraction of *L. plantarum* KW30 incubated with the C-15 peptide and GlcNAc. (1, cytosol fraction only; 2, C-15 peptide only; 3, protease inhibitor only; 4, Cytosol fraction incubated with C-15 peptide and $MgCl_2$; 5, Cytosol fraction incubated with C-15 peptide).

The peak of interest was collected from RP-HPLC and analysed by ESI-MS. The mass of the C-15 peptide is calculated as 1617.70 Da and its mass upon addition of a GlcNAc residue would be 1820.70 Da. Mass spectrometry showed no ion of 1820.70 Da indicating that the peak of interest is not the anticipated GlcNAcylated C-15 peptide.

4.7 CONCLUSIONS AND FUTURE DIRECTIONS

CONCLUSIONS

Part 3 of this thesis involved cloning and purifying two glycosyltransferases one from *Enterococcus faecalis* and the other from *Yersinia frederiksenii* which are located within a predicted bacteriocin gene cluster and are homologous to GccA. In order to gain some understanding of the structure of GccA attempts were made to produce recombinant glycosyltransferases upon which the structure of GccA could be modelled.

Bioinformatic analysis found several glycosyltransferases located within bacteriocin gene clusters that contained both an ABC transporter and a small bacteriocin-like peptide. Of the predicted glycosyltransferases two belonged to organisms that were obtainable and could be used for cloning experiments. Secondary structure analyses predicted that the C-terminal domains of all of the putative glycosyltransferases were likely to contain alpha helices and tetratricopeptide repeats. Another common feature of these sequences is the prediction of one of more transmembrane domains, suggesting that the GTases may be membrane anchored.

A small amount of soluble N-terminal His-tagged recombinant GTase from *Y. frederiksenii* was produced in *E. coli*. Unfortunately it could not be captured by IMAC probably because it aggregated or was cleaved due to misfolding. The glycosyltransferase from *E. faecalis* was also successfully cloned and produced in *E. coli*. The protein was also probably not correctly folded, as it appeared to be permanently captured on the resin by some sort of non specific interaction.

A discontinous HPLC-based assay designed to show the activity of the enzyme was developed using the protocols reported in the literature for the glycosyltransferase from Wakarchuk *et al,* 1998. Of the many conditions trialled, only two produced a unique peak in the HPLC. Both of these involved the C-terminal peptide and the cytosol fraction of whole cells (as opposed to protoplasts). When this peak was analysed by mass spectrometry it did not have the predicted molecular mass for the glycosylated peptide.

FUTURE DIRECTIONS

To obtain soluble recombinant glycosyltransferases from *E. faecalis* and *Y. frederiksenii* further work is required. As the glycosyltransferase from *E. faecalis* contains two predicted transmembrane domains located within the C-terminal region, truncation mutations of this domain may produce more soluble protein that folds correctly. The glycosyltransferase from *Y. frederiksenii* also contains two potential transmembrane domains, although in contrast to *E. faecalis,* one is located within the N-terminal region. N-terminal truncation mutants of this protein may also promote solubility and correct folding. An alternative method to obtain more soluble protein is to clone the two glycosyltransferases into different vectors in different expression systems or to co-express them with chaperones to assist in folding. Attaining soluble folded forms of either of these glycosyltransferases will confirm the putative function assigned to the enzyme, and allow structural studies to be undertaken. Such studies are necessary to unravel the mechanism, and to understand how these proteins are able to add a sugar to both a serine and cysteine sidechain.

The incentive behind producing a glycosyltransferase assay was to establish a method to determine the scope of GlcNAcylation of GccF by GccA; whether it adds both of the GlcNAc residues or just one. As thus far the assay has proven unsuccessful further experimental conditions could be trialled such as using different pH, temperature, incubation time, reaction time, metal ions, and different peptides. Reverse transcription quantitative PCR experiments have indicated that the transcription of the gene *gccA* is low and therefore it may have a low concentration in the cell which could be a reason for the assay not working (personnal communication, Dr Gill Norris). To counter this growing a large volume of *L. plantarum* KW30 cells and then fractioning and concentrating the cell lysate by molecular weight (to attain proteins within the molecular weight range of GccA) and using this in the assay may provide enough glycosyltransferase to GlcNAcylate the peptide for the assay.

APPENDICES

APPENDIX I DNA AND AMINO ACID SEQUENCES

L. plantarum KW30 Sequences
GccF Sequence

```
      M  S  K  L  V  K  T  L  T  I  S  E  I  S  K  A  Q  N  N  G
      ATGAGTAAATTGGTTAAGACACTTACTATAAGTGAAATTTCTAAGGCTCAAAACAACGGT
1
      TACTCATTTAACCAATTCTGTGAATGATATTCACTTTAAAGATTCCGAGTTTTGTTGCCA

      G  K  P  A  W  C  W  Y  T  L  A  M  C  G  A  G  Y  D  S  G
      GGAAAACCTGCATGGTGTTGGTATACTTTAGCAATGTGTGGTGCTGGTTATGATTCGGGA
61
      CCTTTTGGACGTACCACAACCATATGAAATCGTTACACACCACGACCAATACTAAGCCCT

      T  C  D  Y  M  Y  S  H  C  F  G  I  K  H  H  S  S  G  S  S
      ACCTGTGATATATGTATTCCATTGTTTTGGTATAAAGCATCATAGTAGTGGTAGTAGCAG
121
      TGGACACTATATACATAAGGTAACAAAACCATATTTCGTAGTATCATCACCATCATCGTC

      S  Y  H  C*
      TTATCATTGTTAG
181
      AATAGTAACAATC
```

GccA Sequence

```
      M  K  N  R  Q  N  E  I  D  S  Y  L  N  L  H  L  R  P  V  H
      ATGAAAAATAGACAAAATGAAATTGACAGTTATTTGAACCTTCATTTAAGGCCTGTCCAT
1
      TACTTTTTATCTGTTTTACTTTAACTGTCAATAAACTTGGAAGTAAATTCCGGACAGGTA

      K  S  F  D  F  G  N  L  T  N  I  D  Q  F  R  H  H  I  Y  V
      AAATCATTTGATTTTGGAAATTTAACAAACATTGATCAATTTAGACATCATATTTATGTA
61
      TTTAGTAAACTAAAACCTTTAAATTGTTTGTAACTAGTTAAATCTGTAGTATAAATACAT
```

S Y I V I C K N S Q A T I E R C V N S I

TCATATATTGTAATATGTAAAAATTCTCAAGCAACTATTGAAAGATGTGTGAATTCAATT
121
AGTATATAACATTATACATTTTTAAGAGTTCGTTGATAACTTTCTACACACTTAAGTTAA

A Q N M E N G D E L I V L D T G S T D E

GCTCAAAATATGGAAAATGGAGATGAATTGATTGTTTTAGACACTGGTTCAACTGATGAA
181
CGAGTTTTATACCTTTTACCTCTACTTAACTAACAAAATCTGTGACCAAGTTGACTACTT

T V H L V K K N M P Q A K I S V T N W K

ACTGTGCATCTGGTTAAAAAGAATATGCCACAAGCAAAAATATCAGTAACAAATTGGAAA
241
TGACACGTAGACCAATTTTTCTTATACGGTGTTCGTTTTTATAGTCATTGTTTAACCTTT

N D F S E V R N K A L K L A S K D W V F

AATGATTTCTCAGAGGTTAGAAATAAGGCACTCAAATTAGCATCTAAAGACTGGGTGTTC
301
TTACTAAAGAGTCTCCAATCTTTATTCCGTGAGTTTAATCGTAGATTTCTGACCCACAAG

Y V D S D E W L D V D D G A Q L K K I L

TATGTTGATAGTGATGAGTGGTTGGATGTTGATGATGGAGCTCAGCTAAAGAAAATTTTA
361
ATACAACTATCACTACTCACCAACCTACAACTACTACCTCGAGTCGATTTCTTTTAAAAT

F K V Q A K N F K F V I N P T F S D H S

TTTAAAGTTCAAGCAAAAAATTTTAAGTTCGTAATTAACCCAACATTTTCTGATCATTCA
421
AAATTTCAAGTTCGTTTTTTAAAATTCAAGCATTAATTGGGTTGTAAAAGACTAGTAAGT

G Q I Y Q T V G R I F P K K S S F H Y Y

GGACAGATATACCAAACAGTAGGGAGAATCTTCCCTAAAAAGTCGTCCTTTCATTATTAT
481
CCTGTCTATATGGTTTGTCATCCCTCTTAGAAGGGATTTTTCAGCAGGAAAGTAATAATA

A K I H E E V R K E D Q K L G Y D V R H

GTTAGGCAC GCAAAAATTCACGAAGAAGTAAGAAAAGAAGATCAAAAACTAGGTTACGAT
541
CGTTTTTAAGTGCTTCTTCATTCTTTTCTTCTAGTTTTTGATCCAATGCTACAATCCGTG

F A C D D I I L Y H D G Y D K E V L R D

TTCGCATGTGATGATATTATTCTTTATCATGATGGATATGATAAAGAAGTATTACGAGAT
601
AAGCGTACACTACTATAATAAGAAATAGTACTACCTATACTATTTCTTCATAATGCTCTA

K D K I K R N I R L L Q E M T C E E P Q

AAAGATAAAATAAAGAGAAACATTCGTTTATTACAAGAGATGACTTGTGAGGAGCCTCAA
661
TTTCTATTTTATTTCTCTTTGTAAGCAAATAATGTTCTCTACTGAACACTCCTCGGAGTT

N A R W P F L L A R D G F D V L P Q D K

AATGCTAGATGGCCGTTTTTACTAGCACGTGATGGATTTGATGTACTGCCTCAGGACAAA
721
TTACGATCTACCGGCAAAAATGATCGTGCACTACCTAAACTACATGACGGAGTCCTGTTT

L K Q L V K R T L D L V A S D S L Q E K

CTCAAACAGTTAGTAAAAAGGACCTTAGACTTAGTTGCTTCCGACAGTCTTCAAGAAAAA
781
GAGTTTGTCAATCATTTTTCCTGGAATCTGAATCAACGAAGGCTGTCAGAAGTTCTTTTT

Y S P F A K K L L G R I L L R E G K T T

TATTCGCCATTTGCAAAAAAATTGCTTGGCCGTATTTTATTAAGAGAAGGTAAAACCACT
841
ATAAGCGGTAAACGTTTTTTTAACGAACCGGCATAAAATAATTCTCTTCCATTTTGGTGA

Q A V L S F K D V L Q I T G G E D S D A

CAAGCAGTATTATCTTTTAAAGATGTTTTACAGATAACAGGTGGGGAAGACAGTGATGCA
901
GTTCGTCATAATAGAAAATTTCTACAAAATGTCTATTGTCCACCCCTTCTGTCACTACGT

```
      I  Y  Y  I  E  S  F  K  I  N  E  I  I  A  E  A  K  S  I  E

      TATACTATATTGAATCATTTAAAATTAATGAAATTATAGCGGAAGCTAAGTCTATAGAA
961
      TATATGATATAACTTAGTAAATTTTAATTACTTTAATATCGCCTTCGATTCAGATATCTT

      V  K  M  L  R  Y  L  N  K  H  K  G  M  I  D  V  N  S  D  I

      TAAAAATGTTGAGATATTTAAATAAACACAAAGGAATGATTGATGTTAATAGCGATATA
1021
      CATTTTTACAACTCTATAAATTTATTTGTGTTTCCTTACTAACTACAATTATCGCTATAT

      S  G  N  Y  Y  H  I  A  Q  V  I  L  E  C  D  I  I  S  A  N

      TCTGGAAACTACTATCATATAGCTCAGGTTATTTTAGAATGTGACATAATAAGTGCCAAC
1081
      AGACCTTTGATGATAGTATATCGAGTCCAATAAAATCTTACACTGTATTATTCACGGTTG

      Y  S  H  L  F  P  L  I  S  E  I  P  K  N  F  S  G  D  I  K

      TACAGCCATCTTTTCCCTTTAATATCAGAAATTCCTAAAAACTTTTCAGGCGATATAAAA
1141
      ATGTCGGTAGAAAAGGGAAATTATAGTCTTTAAGGATTTTTGAAAAGTCCGCTATATTTT

      S  S  V  K  S  A  V  K  L  Y  S  K  L  Q  G  D  S  K  N  E

      AGTAGTGTCAAATCTGCAGTGAAACTGTATTCAAAATTACAAGGTGACAGCAAAAATGAG
1201
      TCATCACAGTTTAGACGTCACTTTGACATAAGTTTTAATGTTCCACTGTCGTTTTTACTC

      N  N  *

      AATAATTAA
1261
      TTATTAATT
```

E. faecalis ATCC 700802 sequences

Predicted Bacteriocin

```
                M  L  N  K  K  L  L  E  N  G  V  V  N  A  V  T  I  D  E  L

              ATGTTAAATAAAAAATTATTAGAAAATGGTGTAGTAAATGCTGTAACAATTGATGAACTT
    1
              TACAATTTATTTTTTAATAATCTTTTACCACATCATTTACGACATTGTTAACTACTTGAA

                D  A  Q  F  G  G  M  S  K  R  D  C  N  L  M  K  A  C  C  A

              GATGCTCAATTTGGTGGAATGAGCAAACGTGATTGTAACTTGATGAAGGCGTGTTGTGCT
   61
              CTACGAGTTAAACCACCTTACTCGTTTGCACTAACATTGAACTACTTCCGCACAACACGA

                G  Q  A  V  T  Y  A  I  H  S  L  L  N  R  L  G  G  D  S  S

              GGACAAGCAGTAACATATGCTATTCATAGTCTTTTAAATCGATTAGGTGGAGACTCTAGT
  121
              CCTGTTCGTCATTGTATACGATAAGTATCAGAAAATTTAGCTAATCCACCTCTGAGATCA

                D  P  A  G  C  N  D  I  V  R  K  Y  C  K  *

              GATCCAGCTGGTTGTAATGATATTGTAAGAAAATATTGTAAATAA
  181
              CTAGGTCGACCAACATTACTATAACATTCTTTTATAACATTTATT
```

Predicted Glycosyltransferase

```
                M  Y  S  E  N  F  I  T  N  D  W  F  N  V  E  V  F  N  K  N

              ATGTATTCTGAAAATTTTATTACTAATGACTGGTTTAATGTAGAGGTATTTAATAAAAAT
    1
              TACATAAGACTTTTAAAATAATGATTACTGACCAAATTACATCTCCATAAATTATTTTTA

                K  Y  T  L  T  N  Q  E  N  K  D  V  T  E  L  W  L  Q  I  L

              AAGTATACTTTAACGAACCAAGAGAATAAAGATGTAACAGAATTATGGTTACAAATTTTA
   61
              TTCATATGAAATTGCTTGGTTCTCTTATTTCTACATTGTCTTAATACCAATGTTTAAAAT
```

K G L K F P N E L K E T V S Y S K N L K

AAAGGGCTAAAGTTCCCCAACGAATTAAAGGAAACTGTCAGTTACTCTAAAAATTTAAAA

121

TTTCCCGATTTCAAGGGGTTGCTTAATTTCCTTTGACAGTCAATGAGATTTTTAAATTTT

E L S L K T H A E V S V C I I A K N E Q

GAATTATCTTTAAAAACTCACGCAGAAGTATCTGTATGTATTATTGCTAAGAATGAACAG

181

CTTAATAGAAATTTTTGAGTGCGTCTTCATAGACATACATAATAACGATTCTTACTTGTC

D S I R K C I N S I Y E F S D E I I F I

ATTCAATAAGAAAATGTATTAATAGTATCTATGAATTTTCAGATGAAATTATATTTATT

241

CTAAGTTATTCTTTTACATAATTATCATAGATACTTAAAAGTCTACTTTAATATAAATAA

D T G S I D L T K K I V K E I A S E K V

ATACAGGATCAATTGATTTGACAAAAAAAAATAGTAAAAGAAATAGCAAGCGAAAAAGTA

301

CTATGTCCTAGTTAACTAAACTGTTTTTTTTTATCATTTTCTTTATCGTTCGCTTTTTCAT

K I F D Y T W Q D D F S D A R N Y S I Q

AAAATATTTGATTATACTTGGCAAGATGATTTTTCAGATGCGAGAAATTATTCAATACAA

361

TTTTATAAACTAATATGAACCGTTCTACTAAAAAGTCTACGCTCTTTAATAAGTTATGTT

K A S K E W I L I I D A D E Y V S S D E

AAGCAAGTAAAGAATGGATATTAATTATTGATGCAGATGAATATGTATCTTCAGATGAG

421

TTTCGTTCATTTCTTACCTATAATTAATAACTACGTCTACTTATACATAGAAGTCTACTC

L T K L R L L I D M L D R F K F K D S L

CTTACCAAATTAAGGCTCTTAATAGATATGTTAGACAGGTTTAAATTTAAAGATTCATTA

481

GAATGGTTTAATTCCGAGAATTATCTATACAATCTGTCCAAATTTAAATTTCTAAGTAAT

R V S C A I Y Q L D N V I T H G Q S R L

AGAGTTAGTTGTGCAATATATCAATTAGATAATGTTATCACACATGGCCAAAGTCGATTA

541

TCTCAATCAACACGTTATATAGTTAATCTATTACAATAGTGTGTACCGGTTTCAGCTAAT

F R N N N K I K Y Y G L I H E E L R N N

TTAGAAACAATAATAAAATTAAGTATTATGGTCTAATACATGAAGAGTTGAGGAACAAC

601

AAATCTTTGTTATTATTTTAATTCATAATACCAGATTATGTACTTCTCAACTCCTTGTTG

K G L D P I F N V E S E I T F F H D G Y

AAAGGATTAGATCCAATTTTTAACGTTGAAAGCGAGATTACTTTTTTCCATGACGGTTAC

661

TTTCCTAATCTAGGTTAAAAATTGCAACTTTCGCTCTAATGAAAAAAGGTACTGCCAATG

K E I L R K E K C E R N I R L L A K M L

AAGAAATACTTAGGAAAGAGAAGTGTGAAAGAAACATAAGGCTACTAGCTAAGATGTTA

721

TTTCTTTATGAATCCTTTCTCTTCACACTTTCTTTGTATTCCGATGATCGATTCTACAAT

E K E P D N V R W A Y L Y C R D S F S I

GAAAAAGAACCAGACAATGTTAGATGGGCATACTTGTATTGTAGAGATTCATTTTCTATA

781

CTTTTTCTTGGTCTGTTACAATCTACCCGTATGAACATAACATCTCTAAGTAAAAGATAT

N A N I D F E K I L L P F L I K N M D E

AATGCCAACATTGATTTTGAAAAAATTCTACTTCCATTTTTAATAAAGAATATGGATGAA

841

TTACGGTTGTAACTAAAACTTTTTTAAGATGAAGGTAAAAATTATTTCTTATACCTACTT

S I S Y E N I L L T N Y T H L I L F L I

AGTATATCATATGAGAATATCCTACTTACAAACTATACTCATTTAATCCTATTTCTTATT

901

CATATAGTATACTCTTATAGGATGAATGTTTGATATGAGTAAATTAGGATAAAGAATAA

T K K Y I I D G K S S L A S K C I K V L

ACTAAGAAATATATAATTGATGGGAAAAGCTCACTTGCAAGTAAATGTATAAAGGTGTTA

961

TGATTCTTTATATATTAACTACCCTTTTCGAGTGAACGTTCATTTACATATTTCCACAAT

E K M L P N S S D V T F Y K F L N K Q H

GAAAAAATGCTACCTAACTCTTCTGATGTTACTTTTTACAAATTTTTAAATAAACAGCAT

1021

CTTTTTTACGATGGATTGAGAAGACTACAATGAAAAATGTTTAAAAATTTATTTGTCGTA

S L Y E Q Q F E F L K E V I Q F R K N N

GTTTGTATGAACAACAATTTGAATTTTTAAAAGAAGTAATTCAATTTAGAAAAAATAAT

1081

TCAAACATACTTGTTGTTAAACTTAAAAATTTTCTTCATTAAGTTAAATCTTTTTTATTA

E Y D Q Y S Q I G C N L L H Y D L L I S

GAATATGACCAATATAGCCAAATAGGGTGTAATTTATTACACTATGATTTATTAATTTCA

1141

CTTATACTGGTTATATCGGTTTATCCCACATTAAATAATGTGATACTAAATAATTAAAGT

G L L F D V K S Y D Y S Y Q Y F L K L D

GGATTACTTTTTGATGTTAAGTCTTATGATTATTCATATCAATACTTTTTAAAATTAGAT

1201

CCTAATGAAAAACTACAATTCAGAATACTAATAAGTATAGTTATGAAAAATTTTAATCTA

L A N Y F S E L E I P D E Y K M L I N K

TTAGCTAACTATTTTTCTGAATTAGAGATTCCTGATGAATACAAAATGTTAATAAATAAG

1261

AATCGATTGATAAAAAGACTTAATCTCTAAGGACTACTTATGTTTTACAATTATTTATTC

Y R E N E S *

TATCGGGAGAATGAATCATGA

1321

ATAGCCCTCTTACTTAGTACT

Y. frederiksenii ATCC 33641

Predicted Bacteriocin

```
      M  R  K  P  K  D  L  L  R  E  I  K  D  E  V  I  L  A  Q  I
      ATGAGGAAACCTAAAGATTTACTGCGTGAAATTAAAGATGAAGTAATTCTAGCTCAAATA
1
      TACTCCTTTGGATTTCTAAATGACGCACTTTAATTTCTACTTCATTAAGATCGAGTTTAT

      S  G  G  S  S  D  Y  N  P  G  P  S  M  T  W  G  G  V  M  D
      TCTGGTGGGAGTAGTGATTACAATCCGGGGCCGAGTATGACTTGGGGCGGGGTAATGATG
61
      AGACCACCCTCATCACTAATGTTAGGCCCCGGCTCATACTGAACCCCGCCCCATTACTAC

      A  V  R  S  C  S  W  P  G  G  F  Y  G  D  P  N  P  S  Y  G
      CTGTTCGCTCCTGTTCATGGCCTGGTGGGTTCTATGGCGATCCAATCCAAGCTATACGGG
121
      GACAAGCGAGGACAAGTACCGGACCACCCAAGATACCGCTAGGTTAGGTTCGATATGCCC

      G  G  F  S  T  Q  R  *
      GGGTGGGTTTAGTACACAGAGATAA
181
      CCCACCCAAATCATGTGTCTCTATT
```

Predicted Glycosyltransferase

```
      M  S  D  I  N  Y  F  N  S  N  Y  L  I  R  L  K  D  F  I  L
      ATGAGTGATATTAATTATTTTAATTCAAACTATCTTATTAGGCTTAAGGATTTCATTTTA
1
      TACTCACTATAATTAATAAAATTAAGTTTGATAGAATAATCCGAATTCCTAAAGTAAAAT

      N  S  I  A  E  S  D  K  I  N  L  L  N  N  W  L  D  K  V  S
      AATTCCATTGCTGAAAGTGATAAAATAAATTTATTAAATAATTGGTTGGATAAGGTTTCA
61
      TTAAGGTAACGACTTTCACTATTTTATTTAAATAATTTATTAACCAACCTATTCCAAAGT
```

F I L E I D P R L N C F V V M D N T F H

TTTATTTTAGAAATTGACCCAAGATTGAATTGTTTTGTTGTCATGGATAATACATTTCAT

121

AAATAAAATCTTTAACTGGGTTCTAACTTAACAAAACAACAGTACCTATTATGTAAAGTA

D I Y L D K I L N D I L V L F E R I Y Q

GATATATATCTTGATAAGATACTAAATGATATTTTAGTTTTATTCGAACGAATATATCAA

181

CTATATATAGAACTATTCTATGATTTACTATAAAATCAAAATAAGCTTGCTTATATAGTT

V S L P T I S A V I I V K D E E R C I Y

GTATCCTTGCCTACTATATCTGCAGTTATTATCGTTAAAGATGAAGAGCGATGTATTTAC

241

CATAGGAACGGATGATATAGACGTCAATAATAGCAATTTCTACTTCTCGCTACATAAATG

R C I E S I L D Y V D E I V I V D T G S

CGTTGCATTGAGAGTATTTTGGATTATGTCGACGAGATTGTGATTGTAGATACAGGGTCT

301

GCAACGTAACTCTCATAAAACCTAATACAGCTGCTCTAACACTAACATCTATGTCCCAGA

T D S T M D I I N S I V S D K I K T Y S

ACTGATAGTACTATGGATATAATTAATAGTATTGTTAGCGATAAGATAAAAACCTATAGT

361

TGACTATCATGATACCTATATTAATTATCATAACAATCGCTATTCTATTTTTGGATATCA

T P W E N D F S H A R N F A K R K A K K

ACTCCTTGGGAGAATGATTTCTCTCATGCCCGCAATTTTGCAAAAAGAAAAGCAAAAAAA

421

TGAGGAACCCTCTTACTAAAGAGAGTACGGGCGTTAAAACGTTTTTCTTTTCGTTTTTTT

D W L M F I D A D E Y L D G K G D Y N E

ATTGGCTTATGTTTATTGATGCTGACGAGTATTTAGATGGAAAGGGAGATTATAATGAA

481

CTAACCGAATACAAATAACTACGACTGCTCATAAATCTACCTTTCCCTCTAATATTACTT

V K E I L L I L Q F L S I K N E M V I C

TAAAGGAAATATTATTAATTTTACAGTTTTTATCTATAAAAAATGAAATGGTTATATGC
541
CATTTCCTTTATAATAATTAAAATGTCAAAAATAGATATTTTTTACTTTACCAATATACG

P F I S N H N G Y N V T T V R R F F L N

CCTTTTATTTCTAACCACAATGGTTATAATGTCACAACAGTTAGACGATTTTTTTTGAAC
601
GGAAAATAAAGATTGGTGTTACCAATATTACAGTGTTGTCAATCTGCTAAAAAAAACTTG

N T D I N Y F G L V H E E P R I N N T K

AATACAGATATAAACTATTTTGGACTGGTTCATGAAGAGCCAAGAATTAATAACACAAAG
661
TTATGTCTATATTTGATAAAACCTGACCAAGTACTTCTCGGTTCTTAATTATTGTGTTTC

P Y Y I S V N I T F I H D G Y M H E I V

CCATACTATATTTCTGTTAATATTACTTTTATTCATGATGGTTACATGCATGAAATTGTA
721
GGTATGATATAAAGACAATTATAATGAAAATAAGTACTACCAATGTACGTACTTTAACAT

K N K R K T D R N L S L L S K M M L L E

AAAACAAAAGAAAAACAGATAGGAATTTATCTCTTTTGAGTAAAATGATGTTACTGGAG
781
TTTTTGTTTTCTTTTTGTCTATCCTTAAATAGAGAAAACTCATTTTACTACAATGACCTC

P N N L R W K Y F Y Y R D G I E V I D L

CCAAATAATTTAAGATGGAAATATTTTTACTATCGAGATGGAATTGAAGTCATAGATTTA
841
GGTTTATTAAATTCTACCTTTATAAAAATGATAGCTCTACCTTAACTTCAGTATCTAAAT

L N A E V G I K S S L I L N E Q Y D F S

TTAAATGCTGAAGTAGGAATAAAATCTTCACTTATATTAAATGAGCAATATGATTTTTCA
901
AATTTACGACTTCATCCTTATTTTAGAAGTGAATATAATTTACTCGTTATACTAAAAAGT

```
          K  S  N  I  R  E  D  E  F  T  F  A  L  L  D  L  L  A  K  N

          AAAAGCAATATTAGAGAAGATGAATTCACTTTCGCATTGTTAGATCTTCTTGCTAAGAAT
961
          TTTTCGTTATAATCTCTTCTACTTAAGTGAAAGCGTAACAATCTAGAAGAACGATTCTTA

          N  L  R  Q  S  K  F  D  D  V  D  I  I  T  D  I  M  N  C  F

          ATTTGAGACAAAGTAAGTTTGATGATGTAGATATTATAACAGACATTATGAATTGTTTT
1021
          TTAAACTCTGTTTCATTCAAACTACTACATCTATAATATTGTCTGTAATACTTAACAAAA

          L  P  E  N  S  N  S  Y  Y  Y  K  C  L  I  N  I  V  E  L  K

          CTTCCTGAAAATAGTAACTCATATTATTATAAGTGTCTAATTAACATTGTTGAATTAAAA
1081
          GAAGGACTTTTATCATTGAGTATAATAATATTCACAGATTAATTGTAACAACTTAATTTT

          S  K  Y  K  E  L  L  D  K  T  I  L  Y  R  E  R  N  I  D  P

          CAAAGTATAAAGAGTTATTAGATAAAACTATACTATATAGAGAGCGAAACATTGATCCA
1141
          AGTTTCATATTTCTCAATAATCTATTTTGATATGATATATCTCTCGCTTTGTAACTAGGT

          Q  Y  *

          CAATATTGA
1201
          GTTATAACT
```

APPENDIX II PCR PRIMERS

E. faecalis primers

Un-Tagged

Forward Primer NcoI
ACGTGCCATGGCATATTCTGAAAATTTTATTACTAATGACTG

Reverse primer XhoI with stop codon
TGTTCTCGAGTCATGATTCATTCTCCCG

C-Terminal

Forward Primer NcoI
ACGTGCCATGGCATATTCTGAAAATTTTATTACTAATGACTG

Reverse Primer XhoI
TGTTCTCGAGTGATTCATTCTCCCGATACT

Mutant Primers – to remove internal NdeI restriction site for N-terminal tagging.

Forward Primer NdeI Mutant
ATGAAAGTATATCACATGAGAATATCCT

Reverse Primer NdeI Mutant
AGGATATTCTCATGTGATATACTTTCAT

N-Terminal

Forward Primer NdeI
GGAATTCCATATGTATTCTGAAAATTTTATTACTAATGACT

Reverse primer XhoI with stop codon
TGTTCTCGAGTCATGATTCATTCTCCCG

Y. frederiksenii primers

N-Terminal

Forward Primer NcoI
ACATGCCATGGGAAGTGATATTAATTATTTTAATTCAAACTATCT

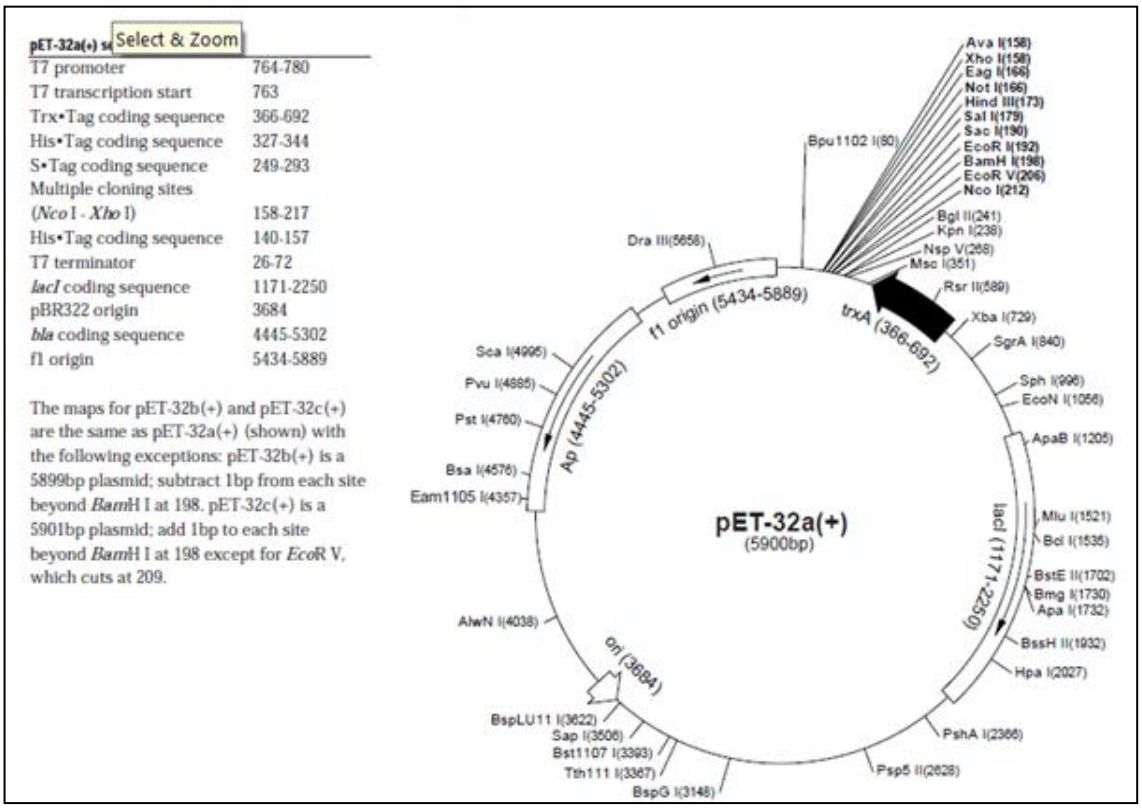
Un-tagged

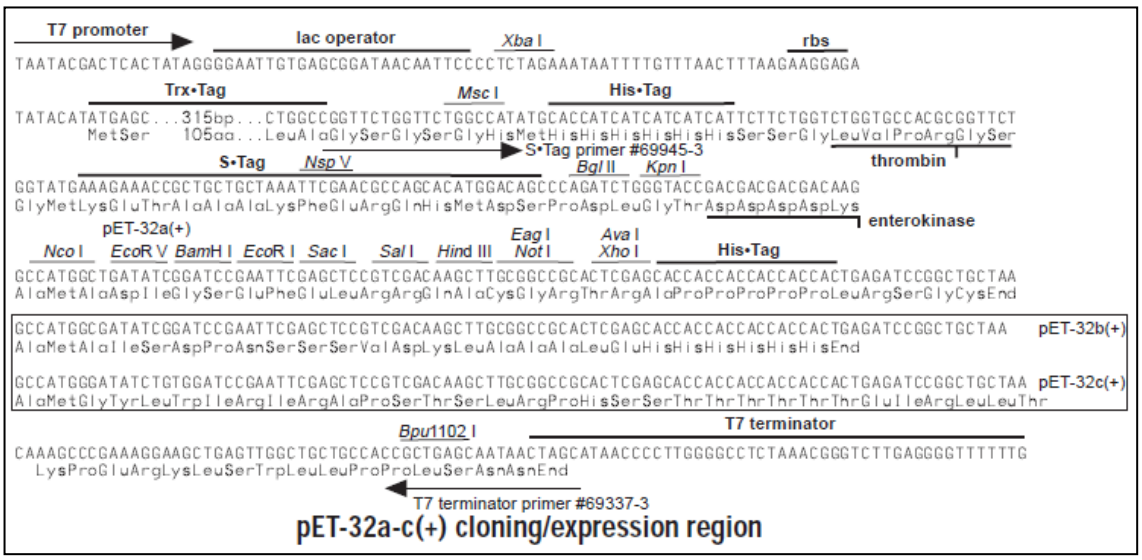
Reverse Primer XhoI with stop codon.
TGTTCTCGAGTCAATATTGTGGATCAATGTTT

C-Terminal
Reverse Primer XhoI
TGTTCTCGAGATATTGTGGATCAATGTTTCG

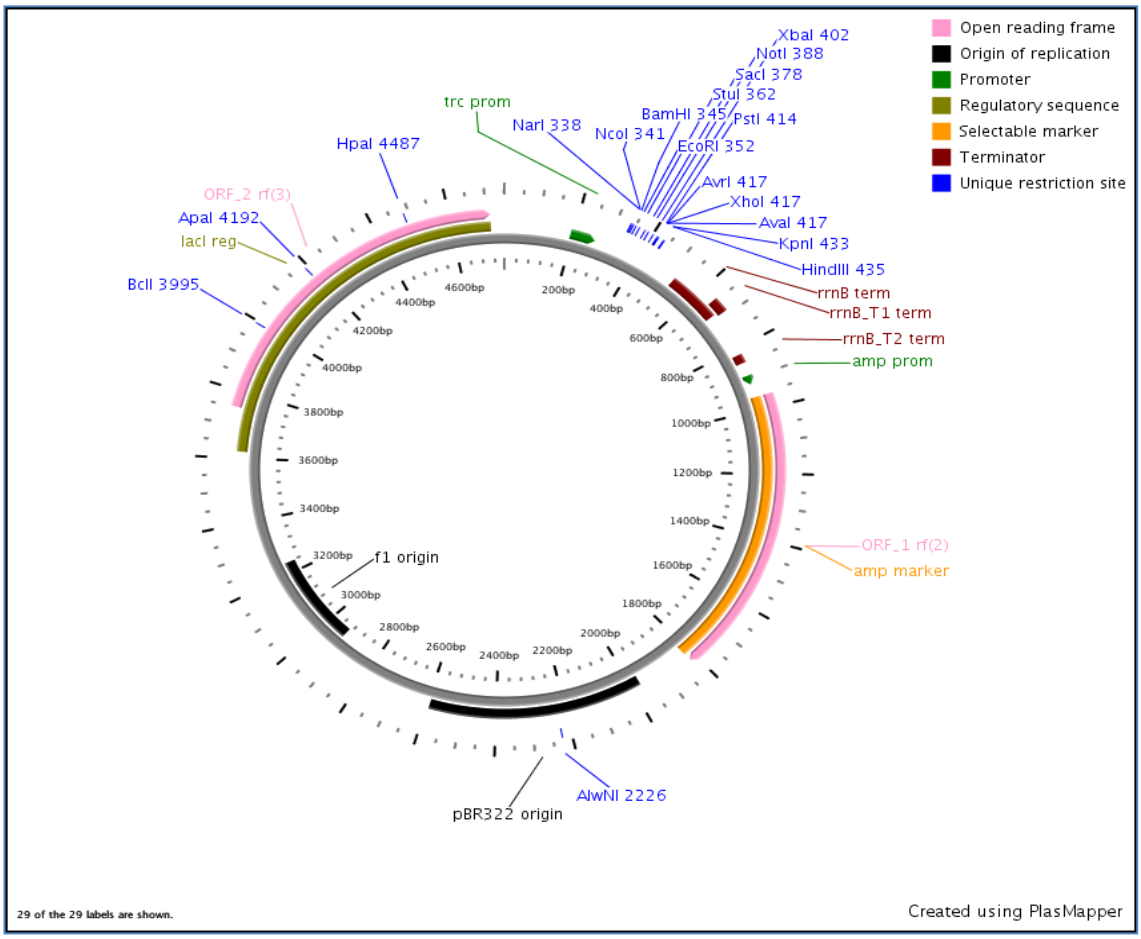
APPENDIX III CLONING AND EXPRESSION VECTORS

pET-32a-c(+) Vector (Novagen)





pProEx Htb Vector (Invitrogen)



pPROEX HTb multiple cloning site and primer binding region: 235-463

260 330

M13/pUC Reverse 23-Base Sequencing Primer

5′-AGCGGA TAACAATTTC ACACAGG -3′→

5′-AGCGGA TAACAATTTC ACACAGGAAA CAGACC ATG TCG TAC TAC CAT CAC CAT CAC CAT CAC GAT TAC GAT ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG

RBS met ser tyr tyr his his his his his his asp tyr asp ile pro thr thr glu asn leu tyr phe gln

(His)$_6$ spacer region rTEV protease cleavage site

Sty I / *Dsa* I / *Ava* I — *Ehe* I, *Nco* I, *Bam*H I, *Eco*R I, *Stu* I, *Sal* I, *Sst* I, *Spe* I — *Xma* III / *Not* I — *Nsp* V, *Xba* I, *Pst* I, *Xho* I, *Sph* I*

** GGC GCC ATG GGA TCC GGA ATT CAA AGG CCT ACG TCG ACG AGC TCA CTA GTC GCG GCC GCT TTC GAA TCT AGA GCC TGC AGT CTC GAG GCA TGC

gly ala met gly ser gly ile gln arg pro thr ser thr ser ser leu val ala ala ala phe glu ser arg ala cys ser leu glu ala cys

*Kpn*2 I

* *Sph* I has 2 recognition sites in pPROEX HT.

445

Kpn I / *Hind* III

GGT ACC AAG CTT GGC TGT TTT GGC GGA *TGA* GAGAA-3′

gly thr lys leu gly cys phe gly gly stop

APPENDIX IV DNA AFFINITY CHROMATOGRAPHY

METHOD

Biotinylated primers were used to amplify the 5'-UTR of the *L. plantarum* KW30 gene (~200 bp) that bioinformatic analyses have predicted will bind to the response regulator. To amplify the 5'-UTR the following thermo-cycling program was used: initial denaturation 98°C for 45 s, followed by 34 cycles of primer annealing at 53°C for 45 s, with extention at 72°C for 25 s. The DNA was then bound to biotinylated magnetic beads following the protocol provided by Novagen (TB262). Protoplasts were lysed and 100 µL of protein was incubated with ~100 µL DNA ligated magnetic beads at 4°C for one hour. After incubation the beads were separated from the lysates using a magnetic rack and the supernatant was removed. The beads were then incubated with 100 µL 0.1M KCL for 2 minutes and separated out using the magnetic rack, and the supernatant was collected. The beads were then incubated with 100 µL of 0.5M, 0.7M and 1.0M KCL for 2 minutes as before and the supernatant collected. The supernatant from each elution was then analysed using tricine SDS-PAGE (2.1.13)

Primers:

Forward F1 primer – AGGAACCGGTATTTTGAAAATGGTGATACT

Forward F2 primer – TTAATACCGGTGAAGACATTTTAGTTAC

Reverse (biotin) primer - TTTACCCATGGCTAATACCCTCCTTTTAATAAG

RESULTS AND DISCUSSION

In order to isolate the response regulator GccF DNA affinity purification was performed. Oligonucleotides which are predicted to bind to GccE were attached to biotinylated magnetic beads following the protocol provided by Novagen (TB262). *L. plantarum* KW30 lysed protoplast samples were incubated with the magnetic beads which were then washed with increasing concentrations of KCL to elute the proteins. Supernatants collected after each elution were analysed using tricine SDS-PAGE (2.1.13). The results are shown in Figure 1.

As the concentration of KCL is increased, the number of bands observed on the tricine gels decreased (Fig 1). Proteins bound to DNA should be eluted at ~0.5 M

KCL. At this concentration however the only bands observed were those corresponding to the control. This indicates that there was non-specific binding to the DNA bound to the beads. Experiments were repeated with more stringent washing of the beads before elution, however similar results were obtained. Inclusion of poly.(dI-dC) to the beads eliminated the non specific binding seen in the 0.5M fraction but there was still no band enrichment. These results indicate that either GccE is not binding to the oligonucleotides bound to the beads or that the interaction between GccE and the oligonucleotides is weak resulting in elution in the 0.1 M KCL wash.

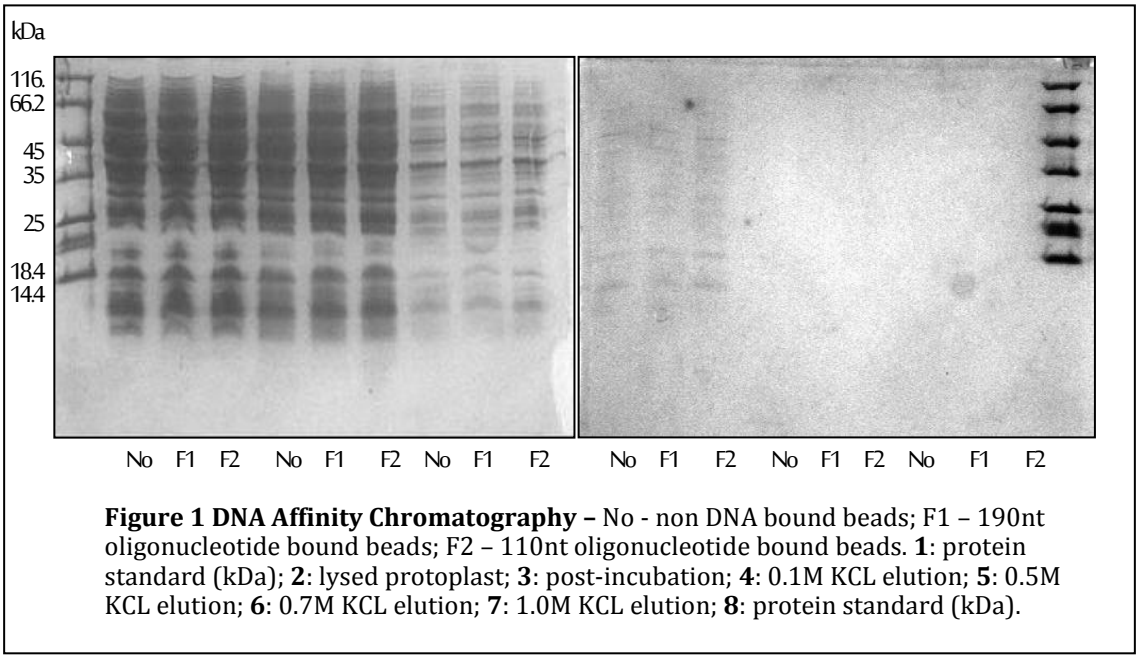


Figure 1 DNA Affinity Chromatography – No - non DNA bound beads; F1 – 190nt oligonucleotide bound beads; F2 – 110nt oligonucleotide bound beads. **1**: protein standard (kDa); **2**: lysed protoplast; **3**: post-incubation; **4**: 0.1M KCL elution; **5**: 0.5M KCL elution; **6**: 0.7M KCL elution; **7**: 1.0M KCL elution; **8**: protein standard (kDa).

APPENDIX V GLYCOSYLTRANSFERASE ASSAY TEMPLATE

Table 1 - Glycosyltransferase assay trial template

Fraction	Peptide	Metal ion	Sugar
Protoplast or whole cell (membrane or cytosol)			
A1	Protease inhibitor control (No lysate, substrate, metal ion, or peptide)		
A2	C15 control (No lysate, metal ion, or substrate)		
A3	SubA control (No lysate, metal ion, or substrate)		
A4	SubC control (No lysate, metal ion, or substrate)		
A5	Lysate control (No substrate, metal, or peptide)		
A6	Susbstrate control (No lysate, metal ion, or peptide)		
A7	No lysate control (everything but lysate, except metal ion)		
A8	C15	$MnCl_2$	UDP-GlcNAc
A9	SubA	$MnCl_2$	UDP-GlcNAc
A10	SubC	$MnCl_2$	UDP-GlcNAc
A11	C15	$MgCl_2$	UDP-GlcNAc
A12	SubA	$MgCl_2$	UDP-GlcNAc
B1	SubC	$MgCl_2$	UDP-GlcNAc
B2	C15	$ZnCl_2$	UDP-GlcNAc
B3	SubA	$ZnCl_2$	UDP-GlcNAc
B4	SubC	$ZnCl_2$	UDP-GlcNAc
B5	C15	$CoCl_2$	UDP-GlcNAc
B6	SubA	$CoCl_2$	UDP-GlcNAc
B7	SubC	$CoCl_2$	UDP-GlcNAc
B8	C15	$CaCl_2$	UDP-GlcNAc
B9	SubA	$CaCl_2$	UDP-GlcNAc
B10	SubC	$CaCl_2$	UDP-GlcNAc
B11	C15	$NiCl_2$	UDP-GlcNAc
B12	SubA	$NiCl_2$	UDP-GlcNAc
C1	SubC	$NiCl_2$	UDP-GlcNAc
C2	C15	No metal	UDP-GlcNAc
C3	SubA	No metal	UDP-GlcNAc
C4	SubC	No metal	UDP-GlcNAc
C5	C15	$MnCl_2$	GlcNAc
C6	SubA	$MnCl_2$	GlcNAc
C7	SubC	$MnCl_2$	GlcNAc
C8	C15	$MgCl_2$	GlcNAc
C9	SubA	$MgCl_2$	GlcNAc
C10	SubC	$MgCl_2$	GlcNAc
C11	C15	$ZnCl_2$	GlcNAc
C12	SubA	$ZnCl_2$	GlcNAc
D1	SubC	$ZnCl_2$	GlcNAc
D2	C15	$CoCl_2$	GlcNAc
D3	SubA	$CoCl_2$	GlcNAc
D4	SubC	$CoCl_2$	GlcNAc
D5	C15	$CaCl_2$	GlcNAc
D6	SubA	$CaCl_2$	GlcNAc
D7	SubC	$CaCl_2$	GlcNAc
D8	C15	$NiCl_2$	GlcNAc

D9	SubA	$NiCl_2$	GlcNAc
D10	SubC	$NiCl_2$	GlcNAc
D11	C15	No metal	GlcNAc
D12	SubA	No metal	GlcNAc
E1	SubC	No metal	GlcNAc

REFERENCES

Apweiler, R. et al. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWIS-PROT database. *Biochimica et Biophysica Acta* **1473:** 4-8.

Bernsel, A. et al. (2009) TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Research* **37:** 465-468.

Bethell, G.S., & Ayers, J.S. (1987) Investigation of the activation of various insoluble polysaccharides with 1,1'-carbonyldiimidazole and of the properties of the activated matricies. *Journal of Chromatography A* **219:** 361-371.

Breton, C., & Imbert, A. (1999) Structure/function studies of glycosyltransferases. *Current Opinion in Structural Biology* **9:** 563-571.

Breton, C. et al. (2001) Structural and functional features of glycosyltransferases. *Biochimie* **83:** 713-718.

Breton, C. et al. (2006) Structures and mechanisms of glycosyltransferases. *Glycobiology* **16:** 29-37.

Breukink, E. et al. (1999) Use of the cell wall precursor lipid II by a pore-forming peptide antibiotic. *Science* **286:** 2361-2364.

Breukink, E., & de Kruijff, B. (2006) Lipid II as a target for antibiotics. *Nature Reviews Drug Discovery* **5:** 321-323.

Burnette, W.N. (1980) "Western blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Analytical Biochemistry* **112:** 195-203.

Campbell, J. et al. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochemical Journal***326:** 929-942.

Chaga, G.S. (2001) Twenty-five years of immobilized metal ion affinity chromatography: past, present and future. *Journal of Biochemical and Biophysical Methods* **49:** 313-334.

Chaiken, I.M. (1989) Bioaffinity chromatography: Synergy between interactive chromatography and molecular recognition for the separation and analysis of macromolecules. *Journal of Chromatography* **488:** 145-160.

Charnock, S.J., & Davies, G.J. (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38:** 6380-6385.

Cheng, X. et al. (2000) Alternative O-glycosylation/O-phosphorylation of the murine estrogen receptor β. *Biochemistry* **39:** 11609-11620.

Chrispeels, M.J., & Raikhel, N.V. (1991) Lectins, lectin genes, and their role in plant defense.*The plant Cell***3:**1-9.
Clarke, et al. (2008) Structural insights into mechanism and specificity of *O*-GlcNAc transferase. *The EMBO Journal* **27:** 2780-2788.

Cleveland, J. et al. (2001) Bacteriocins: safe, natural antimicrobials for food preservation. *International Journal of Food Microbiology* **71:** 1-20.

Comer, F.I., & Hart, G.W. (2000) *O*-glycosylation of nuclear and cytosolic proteins. *The Journal of Biological Chemistry* **275:** 29179-29812.

Cotter, P. et al. (2005) Bacteriocins: developing innate immunity for food. *Nature Reviews Microbiology* **3:** 777-778.

Coutinho, P. et al. (2003) An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology* **328:** 307-317.

Cserzö, M. et al. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the Dense Alignment Surface method. *Protein Engineering* **10:** 673-676.

Delves-Broughton, J. et al. (1996) Applications of the bacteriocin, nisin. *Antonie van Leeuwenhoek* **69:** 193-202.

De Vuyst, L. et al. (1996) Primary metabolite kinetics of bacteriocin biosynthesis by *Lactobacillus amylovorus* and evidence for stimulation of bacteriocin production under unfavourable growth conditions. *Microbiology* **142:** 817-827.

Diep, D. et al. (2009) An overview of the mosaic bacteriocin *pln* loci from *Lactobacillus plantarum. Peptides***30:** 1562-1574.

Franco, O.L., & Rigden, D.J. (2003) Fold recognition analysis of glycosyltransferase families: further members of structural superfamilies. *Glycobiology* **13:** 707-712.

Griffith, L. et al. (1999) *O*-linked *N*-acetylglucosamine levels in cerebellar neurons respond reciprocally to perturbations of phosphorylation. *European Journal of Biochemistry* **262:** 824-831.

Hanover, J.A. (2001) Glycan-dependent signalling: O-linked N-acetylglucosamine. *The FASEB Journal* **15:** 1865-1876.

Hanover, J. et al. (2010) The hexosamine signalling pathway: O-GlcNAc cycling in feast of famine. *Biochimica et Biophysica Acta* **1800:** 80-95.

Hata, T. et al. (2010) Isolation and characterisation of plantaricin ASM1 : A new bacteriocin produced by *Lactobacillus plantarum* A-1. *International Journal of Food Microbiology* **137:** 94-99.

Haynes, J.A. (1998) Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology* **8:** 1-5.

Helenius, A., & Aebi, M. (2001) Intracellular functions of N-linked glycans. *Science* **291:** 2364-2369.

Holmberg, A. et al. (2005) The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **26:** 501-510.

Hu, Y., & Walker, S. (2002) Remarkable structural similarities between diverse glycosyltransferases. *Chemistry and Biology* **9:** 1287-1296.

Jack, R. et al. (1995) Bacteriocins of gram-positive bacteria. *Microbiological Reviews* **59:** 171-200.

Jianghong, L. et al. (2001) The DXD motif is required for GM2 synthetase activity but is not critical for nucleotide binding. *Glycobiologoy* **11:** 217-229.

Kamemura, K. et al. (2002) Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of THR-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *Journal of Biological Chemistry* **277:** 19229-19235.

Kelly, W. et al. (1996) Characterization of plantaricin KW30, a bacteriocin produced by *Lactobacillus plantarum*. *Journal of Applied Bacteriology* **81:** 657-662.

Klaenhammer, T.R. (1993) Genetics of bacteriocins produced by lactic acid bacteria. *FEMS Microbiology Reviews* **12:** 39-86.

Kumar, V. et al. (2003) Annexin II is present on renal epithelial cells and binds calcium oxalate monohydrate crystals. *Journal of the Americal Society of Nephrology***14:** 289-297.

Kurien, B.T., & Scofield, R.H. (2006) Western blotting. *Methods***38:** 283-293.

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

Lairson, L. et al. (2008) Glycosyltranferases: structures, functions and mechanisms. *Annual Reviews of Biochemistry* **77:** 512-555.

Lorca, G. et al. (2007) Transport capabilities of eleven gram-positive bacteria: Comparative genomic analyses. *Biochimica Et Biophysica Acta-Biomembranes* **1760:** 1342-1366.

Lyer S.P.N., & Hart, G.W. (2003) Roles of the tetratricopeptide repeat domain in *O*-GlcNAc transferase targeting and protein substrate specificity. *The Journal of Biological Chemistry* **27:** 24608-24616.
Mellquist, J. et al. (1998) The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation activity. *Biochemistry* **37:** 6833-6837.

Mescher, M. et al. (1976) Formation of lipid-linked sugar compounds in *Halobacterium salinarium*. Presumed intermediates in glycoprotein synthesis. *Journal of Bacteriology* **251:** 7289-7294.

Moens, S., & Vanderleyden, J. (1997) Glycoproteins in prokaryotes. *Archael Microbiology* **168:** 169-175.

Morelli, L. et al. (1987) *Lactobacillus* protoplast transformation. *Plasmid* **17:** 73-75.

Moréra, S. et al. (1999) T4 phage β-glucosyltransferase: substrate binding and proposed catalytic mechanism. *Journal of Molecular Biology* **292:** 717-730.

Nes, I.F., & Tagg, J.R. (1996) Novel lantibiotics and their pre-peptides. *Antonie van Leeuwenhoek* **69:** 89-97.

Nilsen, T. et al. (1998) An exported inducer peptide regulates bacteriocin production in *Enterococcus faecium* CTC492. *Journal of Bacteriology* **180:** 1848-1845.

Nothaft, H., & Symanski, C.M. (2010) Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology* **8:** 765-778.

Patton, G.C., & van der Donk, W.A. (2005) New developments in lantibiotic biosynthesis and mode of action. *Current Opinion in Microbiology* **8:** 543-551.

Power, et al. (2000) Genetic characterization of pilin glycosylation in *Neisseria meningitidis*. *Microbiology* **146:** 967-979.

Ray, B. et al. (1999) The pediocin AcH precursor is biologically active. *American Society for Microbiology* **65:** 2281-2286.

Rini, J.M., & Lobsanov, Y.D. (1999) New animal lectin structures. *Current Opinion in Structural Biology* **9:** 578-584.

Rollema, H. et al. (1995) Improvement of solubility and stability of the antimicrobial peptide nisin by protein engineering. *Applied and Environmental Microbiology* **61:** 2873-2878.

Salazar, O., & Asenjo, J.A. (2007) Enzymatic lysis of microbial cells. *Biotechnology Letters* **29:** 985-994.

Saleemuddin, M., & Husain, Q. (1991) Concanavalin A: A useful ligand for glycoenzyme immobilization –A review. *Enzyme and Microbial Technology* **13:**290-295.

Sambrook, J. (2001) Molecular cloning. 3rd edition. New York, Cold Spring Harbor Laboratory Press.

Schaffer, C., & Messner, P. (2001) Glycobiology of surface layer proteins. *Biochimie***83:** 591-599.

Schägger, H. (2006) Tricine-SDS-PAGE. *Nature Protocols* **1:** 16-22.

Schirm, M. et al. (2004) Flagellin from *Listeria monocytogenes* is glycosylated with β-O-linked *N*-acetylglucosamine. *Journal of Bacteriology* **186:** 6721-6727.

Schmidt, M. et al. (2003) Sweet new world: glycoproteins in bacterial pathogens. *Trends in Microbiology* **11:** 554-561.

Slawson, C., & Hart, G.W. (2003) Dynamic interplay between O-GlcNAc and O-phosphate the sweet side of protein regulation. *Current Opinion in Structural Biology* **13:** 631-636.

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

Stepper, J. et al. (2011) Cysteine *S*-glycosylation, a new post-translational modification found in glycopeptides bacteriocins. *FEBS letters* **585:** 645-650.

Szymanski, C.M., & Wren, B.W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nature Reviews Microbiology* **3:** 225-237.

Todorov, S.D. (2009) Bacteriocins from *Lactobacillus plantarum* – production, genetic organization and mode of action. *Brazilian Journal of Microbiology* **40:** 209-221.

Ünligil, U.M., & Rini, J.M. (2000) Glycosyltransferase structure and mechanism. *Current Opinion in Structural Biology* **10:** 510-517.

Upreti, R. et al. (2003) Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics* **3:** 363-379.

Venugopal, H. et al. (2011) Structural, dynamic, and chemical characterization of novel S-glycosylated bacteriocin. *Biochemistry* **50:** 2745-2755.

Wakarchuk, W. et al. (1998) Role of paired basic residues in the expression of active
recombinant galactosyltransferases from the bacterial pathogen *Neisseria meningitides. Protein Engineering* **11:** 295-302.

Weerapana, E., & Imperiali, B. (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16:** 91-101.

Wells, L. et al. (2001) Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291:** 2376-2378.

Wilson-Standford, S. et al. (2009) Oxidation of lanthionines renders the lantibiotic nisin inactive. *Applied and Environmental Microbiology* **75:** 1381-1387.

Wu, R. et al. (2009) Proteomic analysis of *Lactobacillus casei* Zhang, a new probiotic bacterium isolated from traditional home-made koumiss in inner Mongolia of China. *Molecular and Cellular Proteomics* **8:** 2321-2338.

Zachara, N. et al. (2004) Dynamic O-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *Journal of Biological Chemistry* **279:** 30133-30142.

Zeitler, R. et al. (1998) Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent N-glycosylation of the cell surface glycoprotein from *Halobacterium halobium*. *Glycobiology* **8:** 1157-1164.