Heterologous expression of the *gcc* gene cluster and subsequent characterisation of the glycocin F biosynthetic pathway

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

Master of Science

in

Biochemistry

at Massey University, Manawatu,

New Zealand

Brittany Jane Drummond

2020

In loving memory of

My grandfather, the hardest working person I have met, who passed at the beginning of this journey but inspires me to challenge myself everyday.

You cannot hope to build a better world without improving the individuals. To that end each of us must work for our own improvement, and at the same time share a general responsibility for all humanity, our particular duty being to aid those to whom we think can be most useful.

Marie Curie

Abstract

The rise of antibiotic-resistant pathogens has motivated research into new treatments to prevent the resurgence of infectious and deadly diseases. GccF is a diglycosylated, 43-amino acid, bacteriostatic peptide produced by the generally regarded as safe bacteria *Lactobacillus plantarum* KW30. It is active against many *Lactobacillus plantarum* strains as well as the pathogens *Enterococcus faecium* and vancomycin resistant *Enterococcus faecalis*, stopping the growth of susceptible species at low (1-20 nM) concentrations within 2 minutes. Understanding this novel bacteriostatic mechanism could provide a blueprint for the design of a new family of antibiotics. Reported here is the development of an easily modifiable 11.2-kbp plasmid-based heterologous expression system of the *gcc* cluster that is capable of producing active GccF in *L. plantarum* NC8 and *Lactobacillus sakei* 790. Expression of the *gccF* gene relies on the promotors found naturally within the cluster and results in the production of active GccF matching the concentration produced by the native host. Additionally, the activity of the GccF produced by this system is identical to that of the native producer with 2 nM being sufficient to inhibit the growth of *L. plantarum* ATCC 8014 by 50 %. Mutations introduced within the coding sequence of five of the cluster genes (*gccA* and *gccC-F*), confirmed their roles in the production and maturation of GccF that had previously been predicted using bioinformatic analyses.

Acknowledgements

I would like to thank my supervisor Gill Norris for encouraging an enthusiasm for science, and guiding me through this exciting project. I am grateful that you always had time to talk and checked in regularly, and even more so, with the amount of freedom and patience that you gave me during this project. I could not have asked for a better supervisor.

.....

Although not an official supervisor, the passion and knowledge of Mark Patchett was inspiring. I am incredibly lucky to have had his great minds input on this project.

.....

Thank you to Trevor Loo for advice on experimental design, teaching me how to use equipment and being patient when I kept coming to you with the same problems. I have learnt much from you, and without you my time in X-lab would have been much more daunting.

.....

Finally, being surrounded by so many kind and talented people while studying at Massey University has been an incredible experience. I am hugely thankful to have had such a fantastic lab group that supported each other in and outside of the lab, I couldn't have done this without you.

Contents

List of Figures

List of Tables

List of Abbreviations

AmBic	ammonium bicarbonate
ASM	ABC-transporter maturation and secretion
ATP	adenosine triphosphate
BCA	Bicinchoninic acid assay
BSA	bovine serum albumin
CEX	cation exchange
CFS	cell-free supernatant
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CV	column volumes
Da	dalton
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
GlcNAc	N-acetylglucosamine
GT	glycosyltransferase

IC_{50}	concentration required for 50% inhibition
IMAC	immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LAB	lactic acid bacteria
LB	Luria-Bertani medium
MeCN	acetonitrile
MOPS	3-(N-Morpholino)propanesulfonic acid
MRS	De Man, Rogosa and Sharpe
NCBI	National Centre for Biotechnology Information
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nm
ORF	open reading frame
PBS	phosphate-buffered saline
PCR	Polymerase chain reaction
PNK	polynucleotide kinase
PTS	phosphoenolpyruvate phosphotransferase system
RBS	ribosome binding site

RP-HPLC	reverse phase high pressure liquid chromatography
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SLIM	site-directed, ligase-independent mutagenesis
TCEP	tris(2-carboxyethyl)phosphine
TEMED	tetramethylethylenediamine
TFA	trifluoroacetic acid

Nucleic Acids

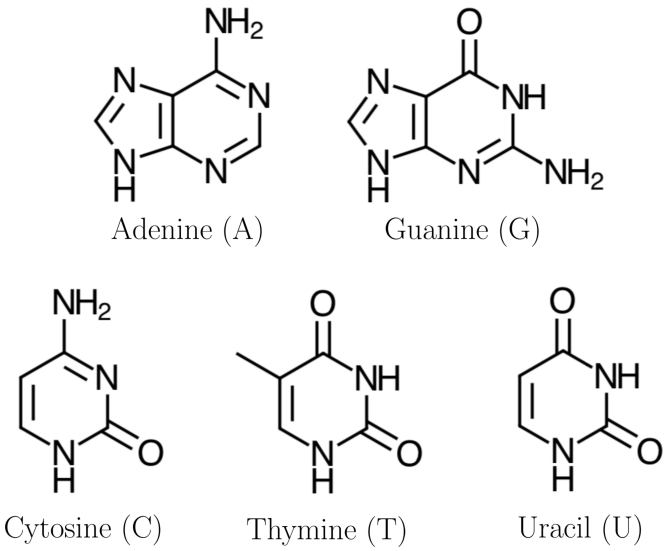


Figure 1: **Nucleic Acid Structures.**

Amino Acids

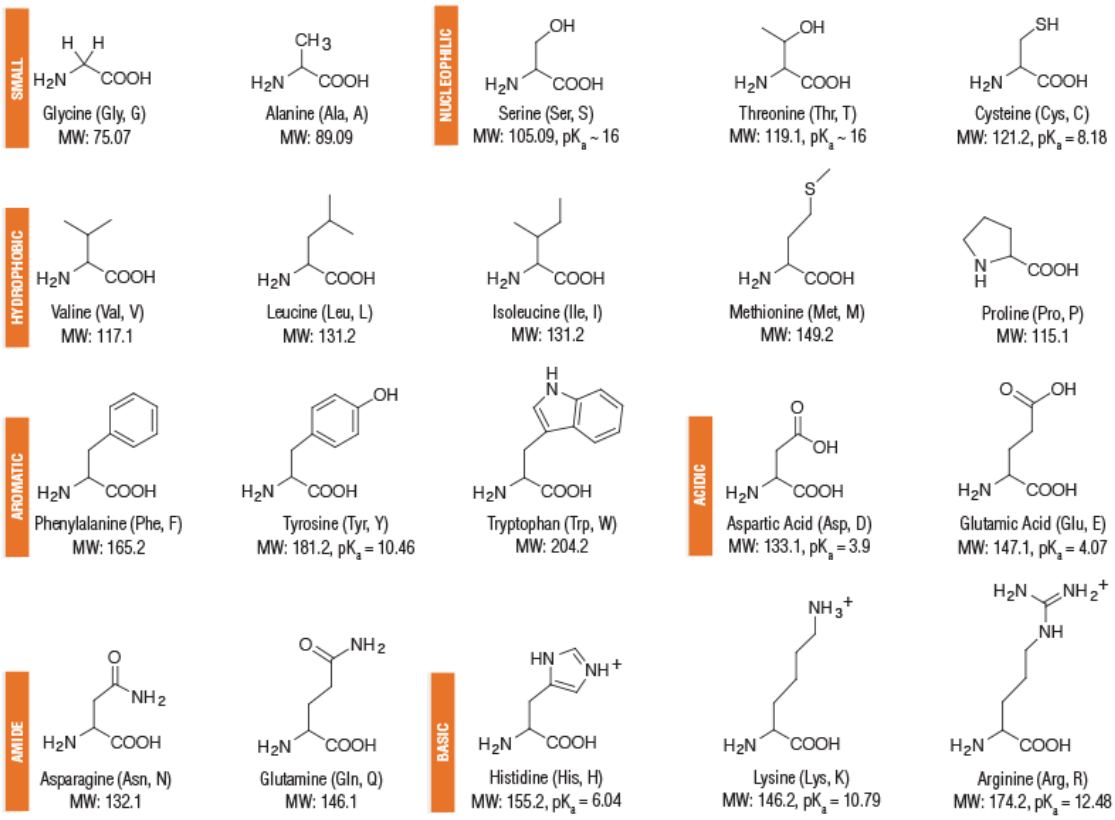


Figure 2: **Amino Acid Structures**, retrieved from www.neb.com.

Standard Genetic Code

First letter	Second letter: U		Second letter: C		Second letter: A		Second letter: G		Third letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG		UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GCA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Figure 3: **Genetic code.**

1 Introduction

1.1 Antibiotics

Until the discovery of antibiotics, bacterial infections were responsible for taking the majority of human lives. *Yersinia pestis* alone is responsible for the death of up to 200 million people. From 1347-1351 this infection killed 40% of the European population [1] earning the name "The Black Death". Although *Y. pestis* has not been eradicated it was treatable with antibiotics. However, multidrug resistant *Y. pestis* was identified in 1995 and this infection cannot yet be prevented by vaccination [2]. Terrifyingly, without treatment *Y. pestis* maintains a mortality rate of 50-90% [3]. While the potential return of plague is concerning, *Y. pestis* is not our most immediate threat as multidrug resistant Mycobacterium tuberculosis (TB) is already out of control. It is estimated that TB has killed over 1 billion people in the last 200 years [4]. In 2018 alone, ten million people fell ill and of these 1.5 million died [5]. At least 230,000 of these deaths were due to antibiotic resistant TB infections [6].

TB is not the only pathogen that has already beaten antibiotics. As of April 2019, drug-resistant pathogens are taking the lives of at least 700,000 people per year, a number that is predicted to increase rapidly. It has been estimated that by 2050 drug-resistant pathogens will cause 10 million deaths per year [6]. These attacks on public health will come in many waves with resistance already reaching concerning levels for infections caused by *Clostridioides difficile* [7], *Salmonella enterica* (typhoid) [8], *Neisseria meningitidis* (meningococcal disease) [9], *Vibrio cholerae* (cholera) [10], *Staphylococcus aureus* (staph infections) [11], *Enterococcus*, *Enterobacteriaceae*, *Pseudomonas aeruginosa*, *Acinetobacter*, *Neisseria gonorrhoeae* (gonorrhea) [12] and more.

The introduction of mainstream antibiotic use in the 20^{th} century shifted the average human life span from 47 to 79 years old by reducing the number of deaths due to infectious diseases [13]. This increase was not solely due to the successful treatment of existing bacterial infections, as the use of antibiotics was followed by advances in cancer therapy, organ transplants and treatment of chronic diseases. Unfortunately the incoming wave of antibiotic resistant pathogens has not hastened the discovery of new antibiotics; instead over the past 20 years, discovery has slowed with only three new antibiotics approved for human use [14].

Antibiotics typically target one specific process required by bacteria for survival. $\beta-$ lactams and glycopeptides inhibit cell wall synthesis by preventing peptidoglycan cross-linking facilitated by penicillin binding proteins. This results in disruption of the peptidoglycan layer and cell lysis [15, 16]. Inhibitors of protein biosynthesis target either the 30S (aminoglycosides and tetracyclines) or 50S (chloramphenicol, macrolides, lincosamides, streptogramins and oxazolidinones) subunits of bacterial ribosomes [17]. Other antibiotics inhibit deoxyribonucleic acid (DNA) gyrase preventing DNA replication (quinolones) and folic acid metabolism (sulfonamides and trimethoprim) [18]. Resistance to these antibiotics has arisen through a multitude of different mechanisms, namely changing the outer membrane permeability, up regulation or modification of efflux pumps, modification of target molecules [19], antibiotic inactivation [20], or the acquisition of an immunity protein. Although chemical synthesis has allowed optimization of existing antibiotics, rapid onset of resistance and our inability to identify new mechanisms of inhibiting their growth has hindered the discovery of new antimicrobial compounds.

1.2 Bacteriocins

In 1925 the first bacteriocin was discovered [21, 22], this predates the first antibiotic, penicillin, discovered by Alexander Fleming in 1929 [23]. Bacteriocins are ribosomally-synthesised antimicrobial peptides [24] secreted by most bacteria to protect their environmental niche. There have been many classification schemes for bacteriocins over the years [25, 26, 27]. The best studied bacteriocins are those belonging to the lactic acid bacteria, owing to their "generally regarded as safe" (GRAS) status creating an interest in their applications as food preservatives [28]. This diverse array of bacteriocins has been grouped into two main classes modified and non-modified [29, 30]. Because of the ever-growing number of new bacteriocins being discovered, the diversity of the group continues to grow, and new classification schemes continue to be developed. One recent scheme for lactic acid bacteria (LAB) proposed [31] is shown in Figure 1.1.

However, there are currently no bacteriocins that have been approved as therapies for human disease [32, 33]. This is odd considering bacteriocins have been used since 1953 as food preservatives [34]. This may be due to their large size compared to traditional small molecule antibiotics, which may impair their ability to be absorbed [35]. Additionally proteases secreted by mammalian cells could potentially destroy ribosomally-synthesised bacteriocins. For these reasons, some bacteriocins that are safe and useful for food preservation may not be practical or efficient disease therapies [36]. The manipulation of naturally produced compounds by gene modification or chemical synthesis has the ability to overcome some of these barriers and increase the effectiveness of such compounds. While the optimisation of antimicrobial activity is invaluable, of even greater importance is understanding the mechanism by which such compounds inhibit bacterial growth and the phylogenetic range of their

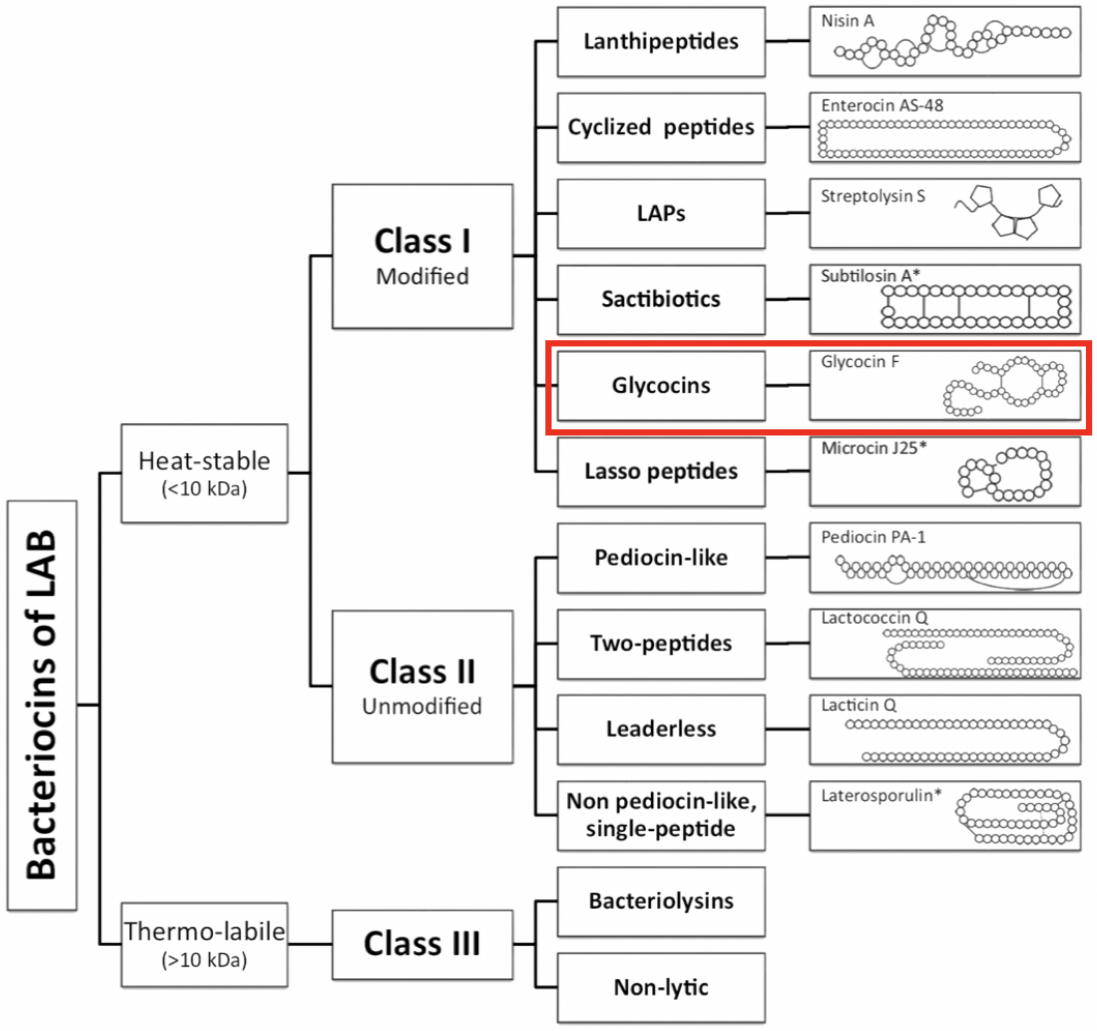


Figure 1.1: **Classifications of lactic acid bacteria bacteriocins** including (*) bacteriocins from non-lactic acid bacteria adapted from P. Alvarez-Sieiro, M. Montalban-Lopez, D. Mu, and O. P. Kuipers. Bacteriocins of lactic acid bacteria: extending the family. Appl Microbiol Biotechnol, 100(7):2939–51, 2016 [31]. With permission from Springer Berlin Heidelberg (http://creativecommons.org/licenses/by/4.0/).

activity as new drugs can then be designed for specific target organisms.

Many of the bacteriocins identified so far have similar mechanisms of action to antibiotics already in use (Figure 1.2). Nisin is a lantibiotic that targets both cell wall synthesis and the cell membrane, while colicins target DNA transcription and protein synthesis. Within a single bacteriocin class there can be multiple methods of inhibition, many of which are not well-understood [37] [24]. As almost all bacterial species secrete some type of bacteriocin [38] it is likely that there are undiscovered mechanisms that could be useful in the fight against antimicrobial resistance. Completely novel mechanisms reduce the chance of cross resistance to existing antibiotics and are therefore our best hope against multi-drug resistant pathogens.

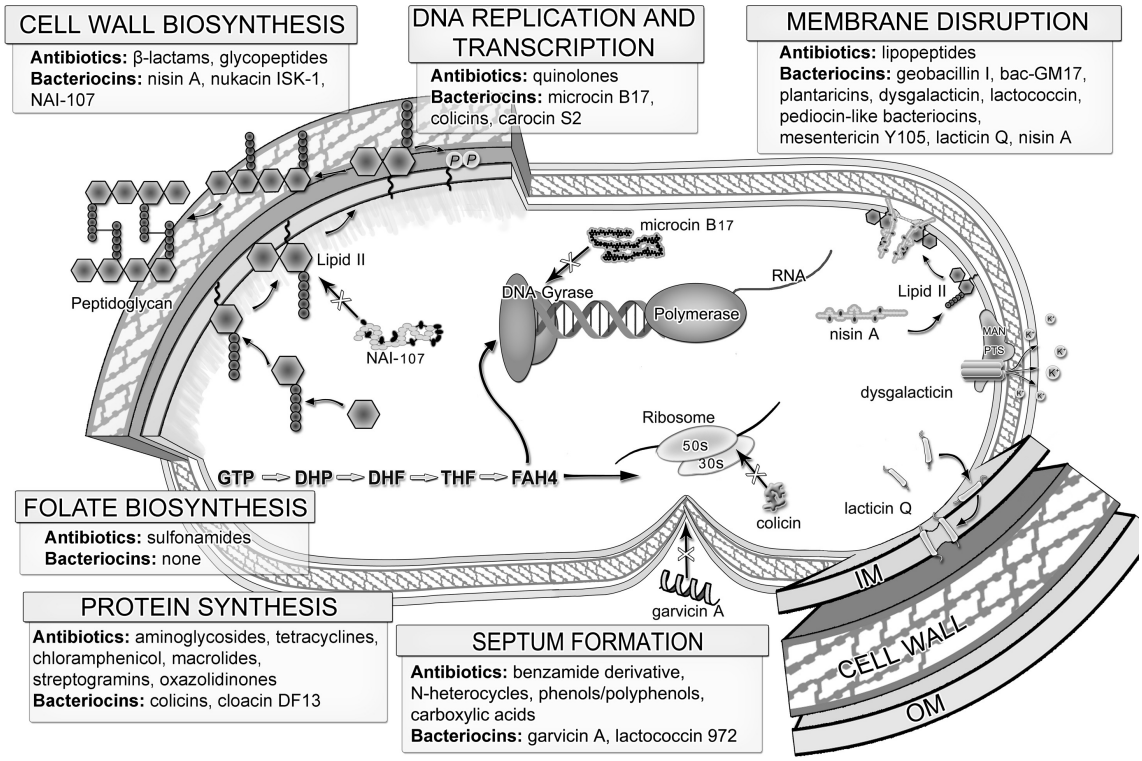


Figure 1.2: **Comparison of antibiotic and bacteriocin cellular targets.** Reprinted from V. L. Cavera, T. D. Arthur, D. Kashtanov, and M. L. Chikindas. Bacteriocins and their position in the next wave of conventional antibiotics. Int J Antimicrob Agents, 46(5):494–501, Copyright (2015) [36], with permission from Elsevier.

Although most antibiotics currently in use are broad-spectrum and bactericidal, evidence suggests that such treatments may be detrimental to the host microbiome. Additionally these treatments are responsible for an increase in secondary infections caused by an overgrowth of bacterial species usually limited by competition [39]. The use of narrow-spectrum bacteriostatic antimicrobials has several advantages that are only beginning to be understood. Firstly, they have been shown to limit significant disturbance to the microbial diversity of the host because of their narrow target range. Secondly, they invoke a weaker counter attack by target organisms who perceive there is less threat to their niche because they are not being killed, further reducing disturbance to microbial diversity [40]. Thirdly, because they do not kill pathogenic bacteria, but rather stop them growing, target organisms are less likely to develop resistance to the antimicrobial substance [41, 42]. While there are instances of organisms that have developed resistance to bacteriocins [43, 44, 45], they seem to be much fewer. For example, nisin, a bacteriocin produced by *Lactococcus lactis* has been widely used in the food industry to combat listeriosis since 1969, yet only very low levels of resistance have been detected [46]. Lastly, the possibility that any immunity developed in pathogens to combat antibiotics will be used to disable bacteriocins appears unlikely as antibiotic-resistant organisms are not usually resistant to bacteriocins [36].

It has been shown that the use of narrow-spectrum bacteriostatic antimicrobials may result in a reduced incidence of obesity [47], as well as autoimmune diseases often associated with antibiotic use such as, Crohn's disease, ulcerative colitis [48, 49] and type 1 diabetes [50]. Additionally there is a growing demand for food products containing fewer chemical preservatives. Bacteriocins can be used in place of standard preserving chemicals as they are flavourless and odourless. Once consumed proteinous antimicrobials are deactivated in the gut of mammals by host proteases

preventing them from being excreted into the environment and subsequent exposure to pathogens decreasing the development of resistance. In 2010, 63,151 tons of antibiotics were consumed by livestock alone [51], once excreted, these antimicrobials have been shown to contaminate crops and water supplies [52]. Exposure of bacteria to these low levels of antimicrobials selects for individuals with the highest tolerance to them.

1.3 Glycocins

The glycocins are a sub-class of the class I bacteriocins, that are characterised by post-translational glycosylation [53]. So far five glycocins have been experimentally identified (sublancin (SunA) [54], glycocin F (GccF) [55], ASM1 [56], enterocin 96 [57],enterocin F4-9 [58] and durancin 61A [59]) in bacterial cultures, while others have been identified though genome mining, and synthesized chemoenzymatically (thurandacin [60]), or expressed in *E. coli* using pathway refactoring systems (bacillicin CER074, geocillicin (also refereed to as pallidocin [61]), bacillicin BAG2O, and listeriocytocin [62]). The classification does not, however, specify that glycosylation is required for activity. This allows another categorisation of the glycocins into those that are glycoactive and require the glycosylation for activity and those that are not. Of the known glycocins, pallidocin [61], ASM1, enterocin F4-9, SunA and GccF, are glycoactive, while for the rest of the group, the role of the glycan modification is not clear [53]. Furthermore, GccF [63], ASM1 [64] and enterocin F4-9 [58] have been shown to be bacteriostatic, while SunA is clearly bactericidal [65]. The activity of thurandacin and durancin have not been characterised, although it appears to be bactericidal, showing that within this group of bacteriocins there is variation in the mode of action.

There has been many studies into the mechanisms of this class of bacteriocin, however their mode of action remains unknown. Visually there is little difference in the core structures of the two best studied glycocins, 43-amino acid GccF and 37-amino acid SunA (Figure 1.3). However they are thought to work in completely different ways with SunA being bactericidal and GccF bacteriostatic. Even more interestingly, GccF remains active, although activity is strongly attenuated, without its long glycosylated tail suggesting that this unstructured region may only be required to increase affinity of the glycocin to its target [66]. The difference in the mode of action of these two glycocins can be explained by their different glycosylations, with GccF harbouring two N-acetylglucosamine (GlcNAc) residues and SunA a single glucose, in combination with their different charge and loop size. Further evidence for their glycans playing a critical role in their antibacterial mechanism was the identification of the N-acetylglucosamine phosphoenolpyruvate phosphotransferase system (PTS) as the receptor for GccF [67] and another PTS being implicated in susceptibility to SunA [68]. It is therefore likely that the glycocins glycans serve to guide the glycocin to its specific PTS receptor. Once docked the glycocin potentially acts by blocking sugar uptake to that receptor. This may have complicated and poorly-understood consequences as these receptors have been implicated in a circuitry of intracellular and cell envelope processes that are thought to underpin all aspects of cellular physiology [69].

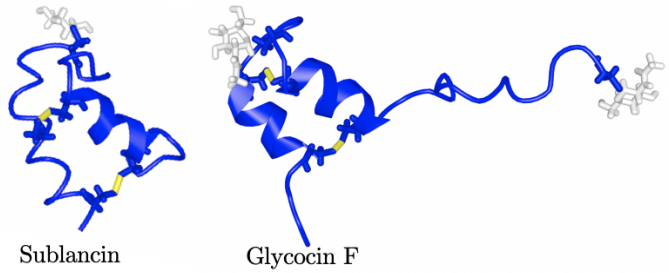


Figure 1.3: **Structures of sublancin and glycocin F.**

1.4 Glycocin F

Glycocin F (GccF) is a 43-amino acid di-glycosylated peptide produced by *Lactobacillus plantarum* KW30 [55]. Unlike many other antibacterials it does not lyse and kill cells, instead it is bacteriostatic and very efficiently prevents cell growth [55], holding target cells in stasis for up to 15 hours at room temperature preventing further infection [70]. As a consequence, the target cell contents do not leak into the surroundings, thus preventing side effects such as toxic shock syndrome [71]. The mechanism of action of GccF is as yet unknown, although some progress has been made in understanding some of the underlying mechanisms [67]. What is apparent is that it uses a completely novel method to shut down target cells at very low concentrations (*L. plantarum* 8014 growth is reduced by 50% at ∼2 nM) [55].

GccF is remarkably stable. It can withstand temperatures of 100°C for two hours and can function at pH 2 to 10 [72]. Sequence analysis showed it had a theoretical mass of 4796.92 dalton (Da) but mass spectrometry revealed the true mass to be 5199.05 Da due to the presence of two HexNAc sugars, one covalently *O*-linked to Ser18 while the other is most unusually *S*-linked to the C-terminal amino acid Cys43 [66]. GccF was the second example of *S*-linked glycosylation reported in the literature, the first being SunA one month earlier [54]. Removal of the GccF HexNAc *O*-linked to Ser18 by *N*-acetyl-α-D-glycosaminidase provided evidence that this sugar was β-linked GlcNAc [55]. This was later corroborated by nuclear magnetic resonance (NMR) studies [57]. Glycosylation was not the only post-translational modification found as Edman sequencing together with mass spectrometry identified two nested disulfide bonds Cys5-Cys28 and Cys12-Cys21 creating a hairpin structure. NMR structural analysis revealed that GccF contains two antiparallel α-helices connected by an eight-amino acid loop with little flexibility due

to the restraint enforced by the nested disulfide bonds ((C-X_6-C)$_2$ architecture)[57]. The C-terminal HexNAc was confirmed to be GlcNAc and the C-terminal amino acids were shown to be flexible [57] unlike SunA which lacks this tail but otherwise has a similar structure as seen in Figure 1.3.

GccF is active against a moderate range of gram-positive bacteria, including *Lactobacillus*, *Enterococcus* and *Streptococcus* species, some of which are pathogens [63]. Initial experiments using another *L. plantarum* as an indicator strain showed that *L. plantarum* KW30 CFS reduced the number of viable cells leading to the initial assumption that GccF was bactericidal [72]. Later research using purified GccF with this same indicator strain showed that while the number of viable cells did not decrease significantly when exposed, cell density did not increase. In other words, GccF induced rapid bacteriostasis. In order to determine the role of the post-translational modifications on the antimicrobial activity of GccF, enzymatic dissection combined with activity assays were used. These showed that when the disulfide bonds were broken or the *O*-linked GlcNAc removed, GccF was no longer active. Removal of residues 42 and 43, including the *S*-linked GlcNAc, reduced the activity 65-fold, while removal of residues 33-43 reduced it 175-fold. Interestingly both of these changes involving removal of the *S*-linked GlcNAc failed to prevent activity, while removal of the *O*-linked GlcNAc completely abrogated activity [55]. Activity assays also showed that free GlcNAc at concentrations of $\sim$ 5 mM protects target cells from bacteriostasis indicating [55] that perhaps GlcNAc competes with GccF to bind to the receptor.

Chemically-synthesised analogues have enabled further understanding of GccF's mechanism of action. Removing the tail GlcNAc and replacing the C-terminal Cys with Ser resulted in a $\sim$100-fold reduction in activity. More interestingly, swapping the GlcNAc for mannose or glucose reduced activity 1000 and 600-fold respectively

[73]. The peptide is more active with no glycosylation at Cys43 than when a mannose or glucose is attached, indicating that the C-terminal GlcNAc may be involved in localising the peptide to a particular position on the cell membrane, whereas the mannose and glucose results in a miss localisation. The role of the unusual *S*-glycosidic bond was rationalised by showing that GccF activity could be increased ∼2-fold by replacing 18Ser(GlcNAc) with 18Cys(GlcNAc) and reduced ∼5 fold by replacing 43Cys(GlcNAc) with 43Ser(GlcNAc) [74]. It is most likely that these changes reflect the susceptibility of the *O*-glycosidic link to external glycosidases [75] and explains why this unusual *S*-glycosylation is used.

Reducing the length and thereby reducing the flexibility of the tail region resulted in ∼50-fold reduction in activity. Showing that both the length of the tail and the GlcNAc residue are required for full activity. The deletion of residues 33-43 reduces activity ∼400-fold compared to the ∼100-fold reduction resulting from the removal of the *S*-linked GlcNAc alone [73], indicates that the tail length and/or sequence is also important for full function. Furthermore the structure of the GccF loop region is also of great importance as removing more than two residues from the loop reduced activity over 2000-fold [73].

The first indication that the PTS could be involved in GccF's mechanism of action came from sequencing GccF resistant mutants generated in the indicator strain *L. plantarum* 8014. These showed mutations on the ribosome binding site (RBS) of the *pts18CBA* gene, the gene coding for the GlcNAc PTS. Further sequencing of these mutants revealed other mutations in this gene. As a consequence, PTS knockouts were generated in another susceptible strain, *L. plantarum* NC8, this resulted in attenuated inhibition by GccF as expected. However, these experiments showed that knocking out this PTS did not completely prevent GccF activity [67], suggesting that there is a second target. It is possible that this secondary target is

another GlcNAc specific transporter, PTS22CBA, that is produced by *L. plantarum* NC8.

1.5 GccF synthesis

Bacteriocins are located within gene clusters that produce proteins thought to be required for their export and maturation [76]. Different species contain distinctive sets of genes that allows the creation of many structurally diverse bacteriocins each with a specific target organism or group of organisms [38, 77]. GccF is part of a seven-gene cluster (Figure 1.4) which contains a glycosyl transferase (glycosyltransferase (GT), *gccA*) thought to be required for the glycosylation of GccF, a transporter (*gccB*) required for the export of the glycocin, two thioredoxin-like proteins (*gccC* and *gccD*) thought to be required for correct disulfide bond formation of GccF, one putative transcriptional regulator (*gccE*) proposed to bind the *gccF* specific promotor, the bacteriocin itself (*gccF*) and an immunity gene (*gccH*).

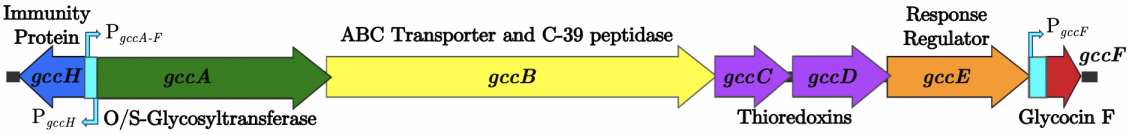


Figure 1.4: **Glycocin F gene cluster** with promotor regions coloured aqua.

1.5.1 Transcription

Previous investigations using quantitative reverse transcription polymerase chain reaction (RT-qPCR) analysis showed that *gccA-F* is transcribed as a single transcript. Further comparisons of the transcript rate of each gene showed that the immunity protein (GccH) and GccF have independent promotors, but that the rest of the cluster is constitutively transcribed. These results also showed that there was a positive correlation between *gccF* transcription and cell density [78]. It is therefore probable that increasing cell density is a signal for transcriptional activa-

tion of the cluster. In fact cell density is a common inducer of bacteriocins whose production is regulated by response regulator proteins [79, 80]. With the exception of *gccF*, each open reading frame (ORF) overlaps with at least one other ORF or promotor/RBS region making it difficult to modify genes without perturbing transcription/translation.

GccE is the most likely of the Gcc proteins to be involved in the regulation of GccF expression. Many bacteriocin gene clusters contain one or more genes required for their regulation [81, 82, 83, 84]. As GccF production increases as cell density increases [78], some sort of quorum sensing protein is likely to be involved. Bioinformatic analysis predicted GccE to be cytosolic as it does not contain a signal sequence making it unlikely to be secreted or inserted into the cell membrane [66]. The C-terminal 88 residues have sequence similarity to a LytTR DNA binding domain [79], this domain has been described as a 10-stranded elongated β–β–β fold that uses long loops rather than recognition helices to interact specifically with the major groove of DNA [85]. Further evidence for GccF regulation by a LytTR domain containing protein is found within the imperfect direct repeat (N6-AAG-N18-AAG-N6) upstream of the GccF coding region known to be recognized by LytTR domain containing proteins [86]. Additionally, other bacteriocins are known to be regulated by LytTR domain containing proteins termed response regulators [87]. Despite the 165 residue N-terminal domain having no sequence homology to any other protein [66], it is highly likely that regulation of GccF is a function of GccE but this needs to be experimentally validated.

Proteins containing LytTR domains also contain receiver domains that are required for activation of the response regulator and subsequent binding to DNA. LytTR response regulator proteins are transcriptional activators that bind to DNA creating a bend in the polymer to increase the chance of ribonucleic acid (RNA)

polymerase binding [79]. These proteins cannot bind DNA until dimerisation has occurred. As dimerisation does not occur through the LytTR domain [87], this is most likely a function of the 165 N-terminal GccE residues. It is plausible that dimerisation of GccE is induced following a conformational change brought about by the phosphorylation of an aspartic acid as seen in other LytTR containing response regulators. Phosphorylation of LytTR containing proteins is carried out by a membrane bound histidine kinase that autophosphorylates when and external stimulus is received. The histidine kinase would then transphosphorylate GccE inducing a conformational change and dimerisation facilitating binding to DNA [88]. Alternatively this unique domain may promote a new type of regulation, with glycosylation by GccA promoting DNA binding being another possibility.

1.5.2 Glycosylation

Following translation, the GccF peptide is glycosylated on Ser18 and Cys43 residues. This process is enzymatic and facilitated by a glycosyltransferase (GT) that transfers a monosaccharide from a nucleotide sugar to the peptide. The GT (GccA) produced by the *gcc* cluster is probably responsible for both the *O*-glycosylation of Ser18 and the *S*-glycosylation of Cys43, although this has not yet been experimentally verified. GccA belongs to the glycosyltransferase 2 (GT2) family in the Carbohydrate-Active enZYmes Database that is characterised by a GT-A fold and an inverting mechanism [89]. GT-A folds consist of a $\alpha/\beta/\alpha$ sandwich, a seven-stranded beta-sheet that resembles a Rossmann fold [90]. The inverting mechanism refers to the inverted configuration of the product with respect to the donor substrate. In this instance a beta glycosidic bond is formed from an alpha linked nucleotide sugar. For this process a divalent cation held within the DxD motif of the GT stabilises the negative charge formed during the departure of the nucleoside diphosphate leaving group [91]. The absolute requirement for the presence of

this divalent cation was proven by the mutation of residues within or surrounding the DxD motif which completely inhibited EntS, the enzyme responsible for the glycosylation of the glycocin enterocin 96 [92].

GccA does not appear to contain a signal peptide or transmembrane helices, therefore it is unlikely to be inserted into a membrane or to be secreted [66]. It is therefore likely that GccF is glycosylated before being transported out of the cell, although it is not known whether glycosylation is required for recognition by the transporter. *In vitro* assays showed that SunS, a homologue of GccA, is responsible for the glycosylation of Cys22 on SunA. In particular, these experiments showed that specific residues in the SunA N-terminal α-helix are involved in GT recognition [93]. Other experiments showed the leader peptide of SunA was not required for glycosylation [54] and that the peptide substrate (SunA) needed to be in the reduced form, as SunS could not glycosylate peptides containing the disulfide bonds found in the mature protein [94]. The glycocin thurandacin can also be glycosylated *in vitro* by the heterologously produced GT from its gene cluster, ThuS, another homologue of GccA. In this case, both *O*- and *S*-glycosylation of thurandacin by ThuS was confirmed [60]. These studies, however, cannot preclude the possibility, however unlikely, that other non-cluster GTs may play a role in GccF glycosylation and further investigation is required for certainty.

SunS was shown to be a homodimer *in vitro*, furthermore, the removal of the C-terminal dimerisation domain abolished all activity [94]. Although dimerisation of GccA could allow both glycosylations of the GccF peptide to be carried out simultaneously, this would not be required for the SunS substrate as SunA is mono-glycosylated. Instead, this dimerisation could allow for allosteric regulation, greater stability or self glycosylation [95]. SunS and GccA both belong to the peptide *S*-glycosyltransferase, SunS family suggesting they would have a similar structure. It

is therefore likely that GccA is also a dimer. However, there are over 100 families of glycosyltransferases [89], some of which contain both monomeric and dimeric proteins [95]. Thus further evidence is required to confirm the native conformation of GccA.

1.5.3 Export

Following glycosylation is it presumed that translocation of GccF across the membrane occurs by harnessing the energy released during adenosine triphosphate (ATP) binding and hydrolysis, facilitated by the integral membrane ATP-binding cassette (ABC) transporter. GccB was identified as an ABC transporter following bioinformatic analysis [66] and although it has not been studied, ABC transporters from other bacteriocin clusters are well-characterised. GccB belongs to the group of ABC transporters containing protease domains that are known as ABC-transporter maturation and secretion (ASM) proteins and are responsible for the unidirectional export of bacteriocins [96]. The crystal structure of PCAT, an ASM protein from *Clostridium thermocellum* showed that an alternating-access mechanism is used by these proteins to shuttle bacteriocins across the membrane. In this mechanism the absence of ATP creates an inward facing conformation where the protease domain interacts with the transmembrane domain, facilitating cleavage of the leader peptide, allowing the bacteriocin to enter the cytosolic side of the transporter. Following this, the binding and hydrolysis of ATP dissociates the protease domain from the trans-membrane domain isolating it from the transporter while concomitantly opening the channel to the outside of the membrane to release the bacteriocin [97].

Whether the activity of GccF is affected by the presence of its leader peptide has not been investigated, but other bacteriocins require its removal for full activity [98]. This indicates that GccB is not only responsible for the transport of GccF

across the membrane as it is also required for bacteriocin activation. The removal of the leader peptide is carried out by the C39 cysteine protease domain of GccB which recognises the double-glycine motif found in GccF as a cleavage site. Homology studies of other bacteriocins show that the cleavage site can consist of GA, GS or GG residues [98] and mutations within this region can result in an accumulation of prebacteriocin within the cell [99]. Interestingly SunA synthesised without its leader peptide could still be glycosylated to produce an active bacteriocin [54], showing that GccF most likely requires the presence of the leader peptide for recognition by the ABC transporter but does not require it for activity or for glycosylation. Nevertheless, as GccA is predicted to be cytosolic, it is most likely that GccF is glycosylated within the cytosol prior to the cleavage of the leader peptide on export.

1.5.4 Disulfide bond formation

It is assumed the disulfide bonds found in GccF are formed with the help of the two thioredoxin-like proteins (GccC and GccD) following its export. GccC and GccD have 66% and 47% sequence similarity respectively to TRX-superfamily thioredoxins and contain the characteristic CxxC active site motif [66]. These proteins are thought to facilitate the disulfide bond formation required for the correct folding of GccF, a prerequisite for its activity [66]. While GccC and GccD may play similar roles in GccF maturation, the relatively low sequence homology between these two proteins (25%) and other thioredoxins provokes questions about their roles in GccF production. Additionally, GccD shares 70% sequence similarity with the Pedc_BrcD family of proteins that are thought to aid the export of various class II bacteriocins and do not typically contain CxxC motifs [100]. This would imply that GccD may have a more chaperone-like role in the GccF maturation pathway. The current model has both GccC and GccD located outside the cell despite the former having a only 47% probability of containing a secretion signal in contrast

to GccD which contains a definite secretion signal [66]. This makes it possible that GccC is cytosolic, acting as an accessory protein to the ABC transporter to keep GccF reduced while it is inside the cell [66].

The SunA gene cluster contains two thioredoxin-like proteins, BdbA and BdbB that have some sequence similarity to GccC and GccD. BdbB has the most similarity to GccD at 38%, whereas GccC has 32% sequence similarity to both BdbB and BdbA. Studies showed that BdbB but not BdbA is required for the production of active SunA. However, another thioredoxin located outside of the cluster (BdbC) is able to partially restore the production of active SunA when BdbB is lost. Interestingly gene disruption of *bdbC* alone had no visible change in the production of active SunA, indicating that BdbB alone is sufficient for normal production of SunA [65]. Although disulfide bonds can form spontaneously under oxidative conditions this is usually slow and non-specific making it likely that GccC an GccD are involved in GccF production [101].

Further experiments on the Sun system showed that the absence of all four thioredoxins within *Bacillus subtilis* (BdbA-D) prevented the production of active SunA. Remarkably the heterologous expression of DsbA, a well-characterised thioredoxin from *Staphylococcus aureus* allowed partial recovery of active SunA production in the *B. subtilis* strain lacking endogenous thioredoxins. Unlike many other thioredoxin proteins that can only function as redox pairs, DsbA can function alone [102]. Typically one protein within a redox pair becomes oxidised by the transfer of electrons from a free cysteine during disulfide bond formation, the second protein reduces the first protein then transfers the electrons to the electron transport chain allowing a new cycle to begin [103]. Further investigations revealed that DsbA was only capable of folding its substrates providing free cystine was contained within the growth media of the *B. subtilis* strain lacking endogenous thioredoxins. This

suggests that DsbA was re-oxidised by the redox-active growth media rather than another protein [102]. It is likely that BdbB and BdbC can also be re-oxidised by the redox-active growth media as they too can produce active SunA without the addition of another thioredoxin protein. The addition of BdbD, a non-cluster thioredoxin that functions within a redox pair with BdbC [104], may increase the efficiency of the re-oxidation of BdbB and BdbC or enable them to function in much more reducing environments. This could be an explanation as to why the glycocin gene clusters contain two thioredoxin proteins as their functions may complement rather than overlap each other but further investigations are required to be sure.

1.5.5 Immunity

GccH was identified as the immunity protein following the observation that heterologous expression of GccH in susceptible *L. plantarum* NC8 cells provided protection from the effects of GccF [70]. A BLASTP search of this immunity protein showed the nearest match was to an uncharacterised immunity protein in a closely-related bacteriocin cluster, ASM1, and hypothetical proteins found in other *L. plantarum* species. The GlcNAc PTS18CBA is a known target of GccF and recent experiments have shown that the immunity protein may also bind the GlcNAc PTS18CBA [105], although further investigations are required to uncover the immunity mechanism. Reports of immunity proteins binding to PTS transporters have shown that the simultaneous binding of a bacteriocin and immunity protein together prevents antimicrobial activity [106]. However, in these reports the immunity protein was acting to prevent pore formation, this does not appear to be how GccF acts [63]. It is therefore probable that GccF immunity is based on a completely different mechanism.

From these studies a picture of the mechanism by which GccF and the immunity protein act has begun to emerge. It seems that GccF binds to PTS18CBA, generating a signal to modify intracellular and cell envelope processes resulting in cellular stasis. However, cells treated with higher concentrations of GccF take longer to recover from GccF treatment, adding more GccF to cells recovering from GccF stasis causes them to resume stasis [105]. This implies that the cells have not gained immunity to the glycocin, but rather the GccF bound to the PTS18CBA is degraded by the cell, possibly as part of PTS18CBA turnover. The relief of stasis can also be brought about by providing the cells with high concentrations of GlcNAc. It is possible that the turnover of the PTS18CBA could be responsible for this effect as GlcNAc at these concentrations may out compete GccF for receptor sites allowing the PTS18CBA to bind GlcNAc and transport it into the cell.

1.6 Research goals

The urgent demand to develop alternative approaches for inhibiting the growth of antibiotic-resistant bacteria motivates further analysis of bacteriocins as antimicrobial agents. An antimicrobial agent with a novel mechanism of action has the potential to act as a blueprint for a whole new family of antibiotics. The development of a system that enables the expression of the *gcc* cluster will allow changes to the sequence of GccF further revealing the residues that are important for function. This may allow manipulation of the molecule to expand the repertoire of cells challenged by GccF. Additionally such a system will help gain insight into the functionally of the remaining cluster genes with the goal of understanding the requirements for GccF production. The increased production of this bacteriocin would allow it to be trialled as a food preservative or as a treatment for disease. Preliminary research into the treatment of diseased animals using the glycocin SunA encouragingly showed it to be effective at reducing mortality when treating methicillin-resistant *Staphy-*

lococcus aureus in mice [107]. The advantages of an antimicrobial agent like GccF are many and include reduction of pressure on pathogens to develop resistance to current antibiotics, less damage to the host microbiome and a reduction of the many other side-effects suffered by those on prolonged courses of antimicrobial agents [39].

Aims

To characterise the functions of GccABCDEFH in GccF production, maturation and immunity.

This will involve the following:

1. Creating a plasmid-based heterologous expression system to produce active GccF.

2. Generating mutations or gene knockouts of *gccA*, *gccC*, *gccD*, *gccE* and *gccF* using the plasmid expression system and testing their effect on GccF biosynthesis and activity.

3. Investigation of the *Gcc* proteins for structural studies.

2 Materials and Methods

2.1 General

2.1.1 Purified water

A Barnstead^{TM} Nanopure^{TM} (Thermo Fisher Scientific; Massachusetts, USA) unit was used to obtain water largely free of bacteria, particles and dissolved solute contaminants.

2.1.2 Autoclaving

Materials that required sterilisation were autoclaved at 121°C and 2x10^5 Pa for 20 minutes.

2.1.3 Filter sterilisation

When autoclaving was not appropriate 0.2 μm ReliaPrep^{TM} Syringe Filters (Ahlstrom-Munksjö; Helsinki, Finland) or Millipore Express^{R} PLUS 0.22 μm PES membranes fitted to bottle top filters (Thermo Fisher Scientific; Massachusetts, USA) were used.

2.1.4 Sonication bath

A Fisherbrand^{TM} S-Series Heated Ultrasonic Cleaning Bath (Thermo Fisher Scientific; Massachusetts, USA) was used to dissolve solutes bound to the vessel walls, thus ensuring maximum concentrations of dissolved solutes.

2.1.5 Lyophilisation

Lyophilisation was carried out using a Freezemobile Freeze dryer (SP scientific; Pennsylvania, USA). Samples where either frozen in a -80°C freezer or with liquid nitrogen

prior to attachment. Samples remained on the freeze dryer overnight or until they appeared dry.

2.1.6 Speed vac

A Savant™ SPD131DDA SpeedVac™ Concentrator with the Savant™ Refrigerated Vapor Traps (Thermo Fisher Scientific; Massachusetts, USA) was used to dry samples smaller than 2 mL.

2.2 Microbiology

2.2.1 Bacteria

Bacterial strains and their applications in this study are listed in Table 2.1.

Table 2.1: Bacterial strains used in this study

Species	**Strain**	**Genotype**	**Application**	**Source**
Escherichia coli	EC100	*F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL (StrR) nupGDNA cloning*	DNA cloning	Lucigen; Wisconsin, USA
Escherichia coli	BL21 (DE3)	*E. coli str. B F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)*	Recombinant protein production	Novagen®; Darmstadt, Germany
Escherichia coli	BL21 (DE3) pLysS	*E. coli str. B F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS) pLysS[T7p20 orip15A](CmR)*	Recombinant protein production	Lucigen; Wisconsin, USA
Escherichia coli	C41 (DE3) pLysS	*F– ompT gal dcm hsdSB(rB- mB-)(DE3)pLysS (Cmr)*	Recombinant protein production	Lucigen; Wisconsin, USA
Escherichia coli	C43 (DE3) pLysS	*F– ompT gal dcm hsdSB(rB- mB-)(DE3)pLysS (Cmr)*	Recombinant protein production	Lucigen; Wisconsin, USA
Enterococcus faecalis	JH2-2	Wild-type	Plasmid free host	Gift from Professor Gregory Cook [108]
Lactococcus lactis	NZ9000	*pepN::nisRK*	Recombinant protein production	NIZO; Netherlands [109]
Lactobacillus plantarum	KW30	Wild-type	Wild-type GccF producer	Gift from Dr Bill Kelly [72]
Lactobacillus plantarum	NC8	Wild-type	Plasmid free host	Gift from Dr Lars Axelsson [110]
Lactobacillus plantarum	ATCC 8014	Wild-type	GccF indicator	Gift from Dr Bill Kelly [111]
Lactobacillus sakei	Lb790	Wild-type	Non-bacteriocin producing host	[112]

2.2.2 Growth media

Solid media was made by adding 1% (w/v) Agar Bacteriological (Agar No. 2) (Neogen Corporation; Michigan, United States) or 1% agarose (Bioline; London, UK) to liquid media. Both liquid and solid media were sterilized by autoclaving. *E. coli* strains were grown in 25 g/L Luria-Bertani medium (LB) (Invitrogen; Massachusetts, USA) and gram-positive strains in 55 g/L De Man, Rogosa and Sharpe (MRS) medium (Neogen; Michigan, United States).

2.2.3 Antibiotics

Table 2.2: Antibiotic stocks and working concentrations

Antibiotic	**Stock concentration**	**Solvent**	**Working concentration**
Ampicillin	100 mg/mL	H_2O	100 μg/mL
Chloramphenicol	50 mg/mL	95% EtOH	50 μg/mL
Erythromycin	100 mg/mL	95% EtOH	10 μg/mL
Kanamycin	50 mg/mL	H_2O	5 μg/mL

2.2.4 Glycerol stocks

A single colony was picked from solid plated media and cultured overnight in liquid media containing appropriate antibiotics. Cells were pelleted from 3 mL of this culture and resuspended in 1 mL of fresh media with 20% (v/v) glycerol inside a sterile 2 mL screw cap microtubes (Thermo Fisher Scientific; Massachusetts, USA) at -80°C.

2.2.5 Bacterial culture optical density measurement

The cell density of a culture grown in liquid media was estimated by optical density at 600 nm (OD_{600}). This was measured using a Smart Spec^{TM} Plus Spectrometer (BioRad; California, USA) or using Greiner 96 well plates (Merck; New Jersey, USA) on a MultiSkan GO plate reader (Thermo Fisher Scientific; Massachusetts, USA).

2.2.6 Chemically-competent *E. coli*

50 μL of an overnight bacterial culture grown in selective media was used to inoculate 50 mL of fresh selective media and aerated with shaking at 200 RPM and 37^{o}C until an OD_{600} of 0.5 was reached. The culture was then cooled on ice for 15 minutes before being centrifuged in 50 mL Nunc^{TM} tubes (Thermo Fisher Scientific; Massachusetts, USA) at maximum speed for 20 minutes in an Eppendorf^{TM} 5702 Series Centrifuge (Thermo Fisher Scientific; Massachusetts, USA) at 4^{o}C. The cell pellet was then resuspended in 18 mL of RF1 buffer (100 mM RbCl, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$ x 6 H_2O, pH 5.8 with acetic acid) and placed on ice for a further 30 minutes before pelleting again. This pellet was resuspended in 4 mL of ice cold RF2 buffer (10 mM RbCl, 10 mM 3-(N-Morpholino)propanesulfonic acid (MOPS) (free acid), 75 mM $CaCl_2$ x 6 H_2O, 15 % (v/v) glycerol, pH 5.8 with NaOH) and 100 μL aliquots were dispensed into sterile microcentrifuge tubes and stored at -80^{o}C until used.

2.2.7 Transformation of chemically-competent *E. coli*

100 μL aliquots of chemically-competent *E. coli* cells were thawed on ice. Once thawed 100-500 ng of plasmid DNA in a maximum volume of 10 μL was added, and the cells were left on ice for 30 minutes. The mixture was then heated rapidly to 42^{o}C for 90 seconds before being placed back on ice for a further 2 minutes. This

was followed by the addition of 500 μL of liquid media and incubation at 37°C with shaking for 60 minutes. Finally 100 μL of transformed cells were plated on selection media and left inverted at 37°C overnight.

2.2.8 Electrocompetent gram-positive cells

An overnight culture of gram-positive cells was diluted 100-fold into 50 mL of MRS broth and incubated at 30°C until an OD_{600} of 0.2 was reached. The cells were then placed on ice for 10 minutes before being centrifuged in 50 mL Nunc tubes at 3000 x g in an Eppendorf™ 5702 Series Centrifuge for 15 minutes at 4°C. The supernatant was discarded and the pellet was resuspended in 20 mL of 10 mM $MgCl_2$. The cells were pelleted once again and washed with 4 mL of fresh 30% (w/v) poly-ethylene glycol before the final resuspension in 200 μL of fresh 30% (w/v) poly-ethylene glycol. This made two 100 μL aliquots of electrocompetent cells that were kept on ice until transformation.

2.2.9 Electroporation of gram-positive cells

Gene Pulser® MicroPulser™ Electroporation Cuvettes with a 0.2 cm gap (BioRad; California, USA) were chilled on ice before pipetting 10 μL of plasmid (100-500 ng DNA) and 100 μL of fresh competent cells onto the inner surface of the cuvette. The Bio-Rad Gene Pulser Electroporation Porator System (BioRad; California, USA) was set to 2.0 kV, 400 Ω and 25 μF and the electroporation cuvette was pulsed before 1 mL of recovery media (0.4 M sucrose, 0.1 M $MgCl_2$ in MRS) was added to the cuvette. The cuvettes containing transformed cells were incubated at 30°C for 2 hours before 100 μL was spread onto MRS agar plates. These were inverted and allowed to grow at 30°C for 2-4 days.

2.2.10 Biological assay for bacteriocin activity

Indicator plates were prepared by embedding 100 μL of an overnight culture of GccF-susceptible *L. plantarum* ATCC 8014 per 10 mL of MRS agar. Once the plates had set, a specific volume (2-5 μL) of a known concentration of purified GccF or CFS was spotted on top, the plates were then inverted and incubated at 30°C overnight. Sample dilutions were carried out with either MRS broth or water. Alternatively, indicator tubes were prepared where 10 mL of MRS agarose containing ∼100 μL of an overnight of culture *L. plantarum* ATCC 8014 cells was carefully poured into 40 mm x 6 mm glass tubes to a depth of 15–20 mm. Once they were set, up to 50 μL of CFS was layered on top, and left to absorb into the agarose. The tubes were sealed with Para-film and incubated at 30°C overnight.

2.2.11 IC_{50}'s

An overnight culture of *L. plantarum* ATCC 8014 was diluted 20-fold in MRS media and allowed to reach an OD_{600} of 0.1 while being incubated at 30°C. 150 μL of this culture was added to each well on a 96-well plate. This was diluted to a starting OD_{600} of 0.05 by adding 150 μL of MRS media containing 2x the GccF concentration for that assay. All assays were carried out in triplicate were OD_{600} was measured every 15 minutes for 15 hours at 30°C. The growth of inhibited and non-inhibited cells was compared to produce plots of GccF inhibition over time. IC_{50} values were calculated by plotting the maximum percentage inhibition reached by each concentration of GccF compared to an untreated (no GccF) control and interpolating the concentration predicted to produce 50% growth inhibition.

2.2.12 Gram staining

Gram staining was used to help visualise cells following treatment with GccF. The reagents used were contained within the gram staining kit (BD; Berkshire, UK), this was used by heat fixing 20 μL of cell culture to a microscope slide and flooding it with crystal violet. The slide was allowed to sit for 1 minute before the slide was rinsed with water and flooded with iodine for another minute. Another water rinse was followed by de-colourising with ethanol before yet another water rinse and flooding with safrin for 45 seconds. One more water rinse of the slide was carried out before the slide was allowed to dry. Slides where first examined on a Olympus CHA microscope under the 100x oil immersion lens then imaged using the Leica SP5 DM6000B Scanning Confocal Microscope driven by LAS AF V2.7.3.9723.

2.3 DNA

2.3.1 Plasmids

Table 2.3: Plasmids used in this study

Plasmid	**Notes**	**Use**	**Source**
pET28a	T7lac promoter, adds N-terminal His tag, thrombin cleavage site, internal T7 epitope tag, C-terminal His tag; kanamycin resistance; restriction enzyme cloning	*E. coli* recombinant protein production	Thermo Fisher Scientific; Massachusetts, USA
pET28a_*gccH*	As above with *gccH* from *L. plantarum* KW30	Recombinant GccH production	This study
pSIP412	*spp*-based expression vector; *sppKR* expression driven by *ermB* read-through and cognate promoter; 256_{rep}; Em^R with $SH71_{rep}$ and P_{orfX} : : *pepN*	Gram-positive recombinant protein production	[43]
pRV610	MCS-αlacZ, RepA, Em^R, Ap^R	Gram-positive recombinant protein production	Gift from Dr Anne-Marie Le Coq [113]
pRV610*gcc* and mutants	As above with *gccABCDEFI* from *L. plantarum* KW30	Recombinant production of GccABCDEFH	This study
pProEX_HTb_*gccH*	trc promotor, Amp^R, N-His-tag, TEV protease cleavage site with *gccH* from *L. plantarum* KW30	*E. coli* recombinant expression of GccH	Gift from Mr Trevor Loo
pKS-*gcc*	MCS-αlacZ, Ap^R, f1 ori, T7 promotor and *gccABCDEFHI* from *L. plantarum* KW30	*E. coli* plasmid containing *gccABCDEFHI*	Gift from Dr Mark Patchett

2.3.2 DNA concentration

DNA quality and quantity was measured using the DeNovix DS-11 Series Spectrophotometer / Fluorometer (DeNovix Inc.;Delaware, USA) dsDNA application. A 260/280 nm ratio between 1.8 and 2.0 indicates little protein contamination and reliable DNA quantification.

2.3.3 Agarose gel electrophoresis

Following DNA manipulation such as PCR or plasmid extraction, agarose gel electrophoresis was used to visualize DNA size and purity. The percentage of agarose varied from 0.8 to 2% (w/v) in 1x Tris-Acetate-Ethylenediaminetetraacetic acid (EDTA) (TAE) buffer (50x TAE consists of 2.0 M Tris/HCl, 5 mM EDTA, 57.1 mL glacial acetic acid, pH 8.0 in 1 L pure H_2O). Agarose was dissolved in TAE by microwaving the solution, this was then allowed to cool before casting and running the gel in the Mini-Sub GT cell DNA electrophoresis system (BioRad; California, USA). The 1 kb Plus DNA Ladder (Thermo Fisher Scientific; Massachusetts, USA) and samples were mixed with 10x loading dye (0.2 g Bromophenol Blue and 6 mL of 50% glycerol in 4 mL pure H_2O) and run at 100 V for 40 minutes using BioRad power pack 300 (BioRad; California, USA). Once run the gel was stained with ethidium bromide solution (0.5 μg/mL) and then dipped into pure H_2O to destain before visualizing with UV-light using the UV-Trans-Illuminator Gel Doc XR system (BioRad; California, USA).

2.3.4 Agarose gel extraction

The GeneJET Gel Extraction Kit (Thermo Fisher Scientific; Massachusetts, USA) was used with an altered protocol for maximum DNA yield. All excess gel was removed from 0.8% agarose gel slice containing the desired DNA before binding buffer (1:1 (w/v)) was added and incubated at 60°C for 20 minutes with frequent vortexing. This solution was applied to the spin column provided and centrifuged at top speed in a Heraeus^{TM} Fresco^{TM} 17 Microcentrifuge (Thermo Fisher Scientific; Massachusetts, USA) for 1 minute. The column was then washed with 700 μL of wash buffer and centrifuged at top speed for another minute. The empty column was dried by centrifuging for 2 minutes before 20 μL of elution buffer heated to 65°C

was placed on the column for a further 2 minutes and finally eluted by a one minute centrifugation.

2.3.5 Plasmid extraction

3 mL of an overnight culture grown in selective media was pelleted for plasmid extraction using the Monarch® Plasmid Miniprep Kit (New England Biolabs; Massachusetts, United States) according to manufacturer's protocol. Gram-positive bacteria required a pre-treatment of lysozyme (100 mg/mL dissolved in resuspension buffer) at 37°C for 10 minutes followed by two washes with 5% glucose [114] prior to the standard protocol.

2.3.6 DNA sequencing

400 ng of plasmid DNA and 4 pmol of primer was diluted to a total volume of 20 μL in H_2O and sent to the Massey Genome Service. Here they used the BigDye™ Sequencing Ready Reaction mix and cycle sequencing PCR followed by the X-Terminator® system (to remove unincorporated fluorescent ddNTPs) and capillary separation with the ABI3730 DNA Analyzer.

2.3.7 Restriction endonuclease digestion

This was carried out using enzymes from either Invitrogen (Invitrogen; California, USA), New England Biolabs® (New England Biolabs; Massachusetts, USA), or Roche (Roche; Basel, Switzerland). These enzymes were used with the manufacturer's recommended buffers, temperatures, DNA concentrations and times whenever possible.

2.3.8 Ligation

Ligation was typically carried out using a 3:1 insert to vector ratio with T4 DNA Ligase (Roche; Basel, Switzerland) according to the manufacturer's instructions. Reactions were usually carried out overnight at 16°C, however the standard protocol was often modified if it was unsuccessful.

2.3.9 Polymerase chain reaction

Polymerase chain reaction (PCR) was routinely carried out as it was required for plasmid modification. For this application the Phusion® High-Fidelity DNA Polymerase (New England Biolabs; Massachusetts, United States) was used. Routine PCR reactions used for DNA identification were carried out with *Taq* DNA Polymerase (New England Biolabs; Massachusetts, United States). The primer annealing temperatures were chosen using polymerase recommendations and Integrated DNA Technologies, Inc. (IDT) oligo analyser tool (IDT; Iowa, USA) and further optimisations were carried out when required.

Table 2.4: Phusion polymerase reaction conditions

Component	**per 20 μL reaction**	**Final concentration**
H_2O	to 20 μL	
5x HF Phusion buffer	4 μL	1x
10 mM deoxyribonucleotide triphosphate (dNTP)s	0.4 μL	200 μM
10 μM forward primer	1 μL	0.5 μM
10 μM reverse primer	1 μL	0.5 μM
Template DNA	variable	< 12.5 ng/μL
Phusion polymerase	0.2 μL	1 unit per 50 μL reaction

Table 2.5: Phusion polymerase standard protocol

Step	**Temperature**	**Time**
Initial denaturation	98°C	30 seconds
30 cycles	98°C	10 seconds
(annealing)	45-72°C	30 seconds
(extension)	72°C	30 seconds per kbp
Final extension	72°C	10 minutes
Hold	16°C	∞

Table 2.6: *Taq* polymerase reaction conditions

Component	**per 25 μL reaction**	**Final concentration**
H_2O	to 25 μL	
10x standard *Taq* buffer	2.5 μL	1x
10 mM dNTP's	0.5 μL	200 μM
10 μM forward primer	1 μL	0.4 μM
10 μM reverse primer	1 μL	0.4 μM
Template DNA	variable	$<$ 40 ng/μL
Taq polymerase	0.125 μL	1.25 units per 50 μL reaction

Table 2.7: *Taq* polymerase standard protocol

Step	**Temperature**	**Time**
Initial denaturation	95°C	30 seconds
30 cycles	95°C	30 seconds
(annealing)	45-68°C	60 seconds
(extension)	68°C	1 minute per kb
Final extension	68°C	5 minutes
Hold	16°C	∞

2.3.10 Colony PCR

Colony PCR was used to identify desirable plasmids within *E. coli* following ligation reactions or point mutations introduced using the SLIM method [115]. Individual colonies were picked and boiled in H_2O for 15 minutes before a pulse spin in a

Heraeus™ Fresco™ 17 microcentrifuge. 1 μL of the supernatant was then used as template in a PCR reaction with *Taq* polymerase.

2.3.11 Inverse PCR mutagenesis

A single pair of primers was designed so that the entire plasmid would be amplified with one primer incorporating the desired mutation. As the location of the mutations did not allow the introduction of restriction sites to facilitate re-ligation, polynucleotide kinase (PNK) (Thermo Fisher Scientific; Massachusetts, USA) was instead used to phosphorylate the 5' end of the PCR product required for ligation. Poly nucleotide kinase was added to a standard ligation reaction along with *Dpn*I to remove the original construct used as the template for the PCR reaction, or used on the primers themselves prior to the PCR reaction.

2.3.12 PCR clean up

GeneJET PCR Purification Kit (Thermo Fisher Scientific; Massachusetts, USA) was used according to manufacturer's protocol with an elution volume of 20 μL.

2.3.13 SLIM

The SLIM method [116] required two complementary PCR reactions, each with one long tailed primer (F_T and R_T) including the desired mutation (insertion, deletion or substitution) and a standard primer (F_S and R_S) to amplify the entire plasmid (Figure 2.1). Once amplified, 1 μL of *Dpn*I (New England Biolabs; Massachusetts, USA) and 5 μL of D-buffer (20 mM $MgCl_2$, 20 mM Tris-HCl pH 8.0 and 5 mM dithiothreitol (DTT)) were added to the PCR reaction and kept at 37°C for 60 minutes to remove the original plasmid template before the hybridization reaction was carried out. 10 μL of each PCR reaction, 20 μL of H_2O and 10 μL of H-Buffer (750 mM NaCl, 125 mM Tris pH 9.0, and 100 mM EDTA pH 8.0) were mixed for

the hybridization reaction (99°C for 3 minutes followed by three cycles of 65°C for 5 minutes and 30°C for 40 minutes). 10 μL of this mix was then transformed into chemically-competent *E. coli*.

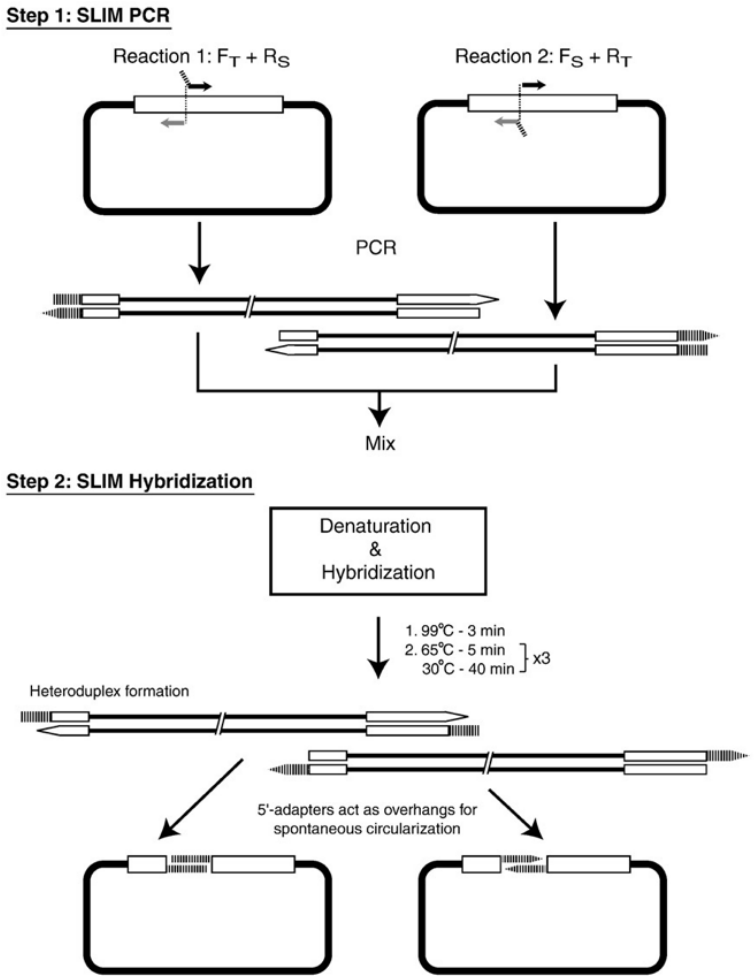


Figure 2.1: **Illustration of the SLIM method.** Reprinted from J. Chiu, D. Tillett, I. W. Dawes, and P. E. March. Site-directed, ligase-independent mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. Journal of Microbiological Methods, 73(2):195–8, Copyright (2008) [116], with permission from Elsevier.

2.4 Protein

2.4.1 Cell lysis

Phosphate-buffered saline (PBS)

PBS was commonly used as a lysis buffer for samples that were not to be purified by immobilised metal affinity chromatography (IMAC). 10 x PBS is made as follows: 8 g sodium chloride, 0.2 g potassium chloride, 1.44 g disodium hydrogen phosphate, 0.24 g potassium dihydrogen phosphate dissolved in 60 mL H_2O, adjusted to pH 7.4 with HCl or NaOH and then H_2O was added to 100 mL.

Lysis buffer for GccH purification from *E. coli* BL21 (DE3) pET28a-*gccH*

Due to the free cysteine residue in GccH, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) was included in the lysis buffer (either PBS or the IMAC equilibration buffer (method 2.4.8)) to prevent aggregation by disulfide bond formation. It is also thought that GccH may be membrane bound, so to aid its removal from the insoluble membrane fraction 1% 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) was also included in the lysis buffer (PBS only).

Sonication

For small gram-negative cultures lysis was facilitated by sonication. Up to 5 mL of cell culture was pelleted and resuspended to 300 μL using lysis buffer. The cells were kept on ice while they were sonicated with three 30 second cycles using a 1 mm micro-tip at 30% amplitude in the Virtis Virsonic 600 sonicator (Sp scientific; Pennsylvania, USA) with 2 minutes cooling on ice between each cycle.

French press

When large gram-negative cultures where lysed a French press was used. Cultures as large as 2 L where centrifuged and the pellet resuspended to a total volume of 40 mL with lysis buffer. The cells where then passed through a French pressure cell (Aminco Instruments Co; USA) cooled to 4°C at 4 kPa using a Wabash hydraulic press (Wabash; USA), twice.

Bead mill

For small scale cell lysis of gram-positive bacteria a bead mill was used. 1-2 mL of cell pellet resuspended in lysis buffer was transferred to screw top tubes pre-filled with high impact zirconium beads with a 0.1 mm diameter (Benchmark Scientific; New Jersey, USA). Once filled these tubes where cooled on ice before lysis in the Hybaid Ribolyser (Thermo Fisher Scientific; Massachusetts, USA) by three cycles of 45 seconds, 6500 rpm, with 2 minutes of cooling on ice between.

2.4.2 Protein concentration

Bradford method

This assay was the first choice for estimating the protein concentration of various solutions. Each assay was carried out in triplicate and the typical Bradford method [117] was followed using a 96-well plate and MultiSkan GO plate reader. For each well 250 μL of Bradford reagent was combined with 10 μL of a sample of unknown concentration, or 10 μL of bovine serum albumin (BSA) standard.

Bicinchoninic acid assay (BCA)

The Pierce^{TM} BCA Protein Assay Kit (Thermo Fisher Scientific; Massachusetts, USA) was used following the manufacturers protocol with a 96-well plate and MultiSkan GO plate reader. This assay was selected when greater concentration accu-

racy was required or when sodium deoxycholate was used at concentrations greater than 0.05%, as this is incompatible with the Bradford method. The BCA assay was preformed in triplicate and a protein standard curve produced with BSA was used to calculate the concentrations of unknown solutions.

Purified GccF quantification

The concentration of GccF resuspended in pure H_2O was determined with UV absorption at wavelengths of 280 and 205 nm using a Cary 300 UV-Visible spectrophotometer (Agilent Technologies Inc; California, USA). Quartz cuvettes (Hellma Analytics; Müllheim, Germany) of different pathlengths (1 mm - 1 cm) were used with molar absorptivity values of 220,030 L mol^{-1} cm^{-1} at 205 nm and 18, 700 L mol^{-1} cm^{-1} at 280 nm to calculate GccF concentration.

2.4.3 PepN assay

An overnight culture was diluted 25-fold to 50 mL in selective media and allowed to grow at 30°C until an OD_{600} of 0.3 was reached. The cells were then induced with 50 ng/mL SppIP (Auspep; Victoria, Australia) and incubated until an OD_{600} of 1.8 was reached. The cells were pelleted and resuspended in 2 mL of 0.1 M Tris-HCl pH 7.5 and disrupted using the bead mill (see subsection 2.4.1). Cell debris was removed by centrifugation at 13,000 x g and 4°C for 10 minutes. Protein concentration was then determined using the BCA method 2.4.2, and 25 μg was kept to load on a polyacrylamide gel (PepN appears at 95 kDa after induction), as this can be used quantify PepN within the cell lysate. 250 μg of cell lysate protein, 50 μL of 100 mM L-lysine p-nitroanilide (Sigma-Aldrich; Missouri, USA) and 0.1 M Tris-HCl buffer to a total volume of 500 μL was allowed to incubate at 30°C for 10 or 30 minutes. The reaction was stopped by the addition of 500 μL ethanol and any precipitate was removed by centrifugation at 10,000 x g for 10 minutes. Absorbance was then read

at 410 nm and the spontaneous hydrolysis of Lys-PNA (control without cell lysate) was subtracted to determine the hydrolysis due to PepN expression.

2.4.4 Gel electrophoresis

Gels were cast in a vertical apparatus (Mini-Protean II system(BioRad; California, USA)) where a 4% acrylamide stacking gel was laid on top of a higher percentage resolving gel. Samples were loaded with 5x sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer (250 mM Tris-HCl pH 6.8, 10% sodium dodecyl sulfate (SDS), 30% glycerol, 10 mM DTT, 0.05% bromophenol blue) and run in the appropriate tank buffer for the gel type at 200 V until the dye reached the bottom of the gel. The gels were removed from the apparatus and fixed (in 40% ethanol and 10% acetic acid) for 15 minutes before staining for a minimum of 4 hours with Colloidal Coomassie stain (stock solution: 1.2% (v/v) phosphoric acid, 10% (w/v) ammonium sulfate, 0.1% (w/v) coomassie brilliant blue G250; methanol was added to 20% to create the working solution) and destaining with H_2O. Protein size was estimated by comparison with the Precision Plus Protein ^{TM} Dual Xtra standards (BioRad; California, USA).

SDS-polyacrylamide

The components of the gels are listed in Table 2.8. Gels were run in 1x SDS tank buffer (the 5 x stock solution consists of 72 gL^{-1} glycine, 15 gL^{-1} Tris, and 5 gL^{-1} SDS).

Table 2.8: Polyacrylamide gel components

Component	**4% stacking gel**	**15% resolving gel**
H_2O	6.3 mL	3.54 mL
0.5 M Tris-HCl, pH 6.8	2.5 mL	-
1.5 M Tris-HCl, pH 8.8	-	2.5 mL
10% (w/v) SDS	100 μL	100 μL
Acrylamide/Bis-acrylamide 40% (Sigma-Aldrich; Missouri, USA)	1 mL	3.76 mL
10% ammonium persulfate	100 μL	100 μL
tetramethylethylenediamine (TEMED)	10 μL	5 μL

SDS-tricine polyacrylamide

For low molecular weight proteins tricine gels were used as this enabled better separation [118]. Table 2.9 shows the gel recipe, AB-3 consists of 1.5 g bisacrylamide and 48 g acrylamide made up to 100 mL with H_2O, the recipe for the gel buffer is listed in table 2.10 along with the tank anode and cathode buffers.

Table 2.9: Tricine gel components

Component	**4% stacking gel**	**16% resolving gel**
H_2O	8 mL	7 mL
Glycerol	-	3 mL
AB-3	1 mL	10 mL
Gel buffer	3 mL	10 mL
10% ammonium persulfate	90 μL	100 μL
TEMED	9 μL	10 μL

Table 2.10: Tricine gel and electrophoresis buffers

Component	**Anode buffer 10 x**	**Cathode buffer 10 x**	**Gel buffer 3 x**
Tris	1 M	1 M	3 M
Tricine		1 M	
HCl	0.225 M		1 M
SDS		1%	0.3%
pH	8.9	8.25	8.45

2.4.5 Tryptic digestion

In-gel

In-gel tryptic digestion was carried out to examine the identity of proteins embedded in SDS-PAGE gels by mass spectrometry. Sample manipulations were carried out in a lamina flow hood to limit keratin contamination, and LoBind microcentrifuge tubes (Eppendorf; Hamburg, Germany) were used to prevent protein loss. Target protein bands were excised and individual gel slices were chopped fine using a scalpel and destained with 300 μL washes of destain solution (50 mM ammonium bicarbonate

(AmBic), pH > 7.9, 50% MeOH at 45°C) for 1-2 hours. The pieces were then dehydrated with 300 μL of 80% acetonitrile (MeCN) for 1 minute, the liquid was removed and the sample speedvaced dry. To reduce proteins in the dry sample, 300 μL of reducing solution (10 mM DTT, 50 mM AmBic pH >7.9) was applied to the gel pieces for 1 hour at 37°C. Supernatant was removed and the pieces were washed with 100 μL of 50 mM AmBic for 5 minutes. Supernatant was removed and the pieces were dehydrated as before removing excess liquid and speedvaccing them dry. For subsequent alkylation, 30-100 μL of alkylation solution (20 mM iodoacetamide in 50 mM AmBic) was placed on the gel pieces for 30 minutes in the dark at room temperature. The supernatant was removed and the gel pieces were dehydrated with 2 washes of 100 μL of 80% MeCN for 1 minute. The pieces were then rehydrated with 100 μL of 50 mM AmBic for 5 minutes. Supernatant was removed and the pieces were dehydrated twice as before. The liquid was removed and the pieces were speedvaced until dry. For digestion the pieces were rehydrated in the minimum volume of digestion solution (20 ng/μL trypsin (Sigma-Aldrich; Missouri, USA) in 50 mM AmBic, pH > 7.9) and incubated for 10 minutes before checking that more solution was not required. The samples were incubated for a minimum of 4 hours at 37°C before they were centrifuged and sonicated in a ultrasonic bath for 2 minutes. The collected supernatant was pooled with supernatant from the following pellet washes; i) 50 μL of 0.1% formic acid in 50% MeCN for 2 minutes followed by sonication, and ii) 50 μL 0.1% formic acid in 80% MeCN for 2 minutes followed by sonication. The pooled supernatant sample was then concentrated to 30 μL using the speed vac.

In-solution

For in-solution tryptic digestion, up to 20 μL of protein-containing solution was added to an equal volume of 100 mM AmBic. If reduction and alkylation was

required then this was carried out by the addition of 4 μL of 200 mM DTT and incubation for 1 hr at 37°C followed by 4 μL of 100 mM iodoacetamide incubated for 30 minutes in the dark at room temperature. A final 4 μL of 200 mM DTT was used to neutralise the excess iodoacetamide. A 100:1 ratio of protein to trypsin (Sigma-Aldrich; Missouri, USA) was used in a overnight incubation at 37°C. Finally the sample was reduced to 40 μL using the speedvac, small particles where removed from the solution prior to mass spectrometry analysis by centrifugation at 13000 x g for 15 minutes, and the supernatant sample was acidified by the addition of formic acid.

2.4.6 GccF purification

Cell culture supernatant was separated from a 3-day old culture of *L. plantarum* NC8 pRV610*gcc* or other GccF-producing mutants by centrifugation at 5,000 x g for 30 minutes. 25 mL of SP-Sephadex C-25 (GE Healthcare; Illinois, USA), equilibrated in 50 mM sodium formate pH 4.0 was gently stirred into the supernatant overnight. This was then packed into a 5 x 20 cm glass Econo-Column® (BioRad; California, USA), washed with 1 L of equilibration buffer and then 1 L of 50 mM free acid MOPS pH 7.2 before elution with 50 mM AmBic in 70% MeCN. 20 mL fractions were collected and tested for activity using an indicator plate assay. Active fractions were concentrated by lyophilisation then resuspended in a minimum amount of buffer A (0.1% trifluoroacetic acid (TFA), 2% MeCN) before further purification by reverse phase high pressure liquid chromatography (RP-HPLC), on a Stationary phase, Phenomenex Jupiter C18, 250 x 2.00 mm, 5u, 300Å; column, using a linear gradient increasing from 100% buffer A to 100% buffer B (0.08% TFA, 98% MeCN) over 60 minutes at a flow rate of 1 mL/min). The elution was monitored by the absorbance at 280 and 214 nm, peaks were collected manually and tested for activity using an indicator plate or tube assay. Active fractions were pooled and lyophilised. Finally,

the purified peptide product was weighed and resuspended in a minimum volume of water for analysis by mass spectrometry.

2.4.7 Induction and optimisation of expression

All the *E. coli* protein expression vectors used in this study were inducible with Isopropyl β-D-1-thiogalactopyranoside (IPTG). An overnight culture was diluted 100-fold with fresh selective media and the OD_{600} was allowed increase until it reached 0.8. IPTG was then added to a final concentration of 1 mM and growth continued for the desired time. Once this time was reached the culture was placed on ice for 10 minutes before harvesting the cells by centrifugation and washing the pellet twice with PBS. Finally the cell pellet was lysed. To optimise protein production different time points for cell harvesting where screened. Additionally, for proteins that appeared to be expressed but were in a insoluble form, different induction temperatures were tested with 16°C being the lowest tried.

2.4.8 His-tagged proteins purification

Both $GccF_{HIS}$ produced by *L. plantarum* NC8 pRV610*gcc*-$GccF_{His}$ (found in cell culture supernatant), and GccH expressed in *E. coli* contained a His-tag to enable IMAC purification. A nickel IMAC column was made by charging 2 mL of iminodiacetic acid Sepharose® (Sigma-Aldrich; Missouri, USA) with nickel chloride until the flow through ran green. To remove excess nickel ions, 10 column volumes (CV) of H_2O followed by 5 CV of 50 mM sodium acetate in 0.3 M NaCl at pH 4.0 and a further 10 CV of H_2O were gravity fed through the column. To equilibrate the column, 5 CV of 500 mM NaCl, 20 mM Tris-HCl, 5 mM imidazole at pH 7.4 was used. The cells lysed in the same buffer used for equilibration, or cell culture media supernatant (with cells removed) containing 5 mM imidazole adjusted to pH 7.4 with NaOH, was fed into the column. The column was then washed with 10 CV of

the same buffer used to equilibrate the column. After washing the column, proteins were eluted using a linear gradient of 0.05-0.5 M imidazole in 500 mM NaCl, 20 mM Tris-HCl, pH 7.4. Fractions were analysed by SDS-PAGE to identify when the His-tagged protein eluted from the column. Fractions containing the desired recombinant protein where pooled, lyophilised and resuspended for further purification steps.

2.4.9 Mass spectrum analysis

2 μL samples were injected into an HPLC (Dionex UltiMate 3000 Binary RSLCnano) on bypass mode and analysed by a Q Exactive Plus mass spectrometer equipped with a high-energy collision-induced dissociation (HCD) collision cell, Orbi mass analyser and a heated HESI ion source (Thermo Fisher Scientific; Massachusetts, USA). Further information concerning the chromatography and mass spectrometry settings is provided in Appendix 6.

2.4.10 Circular dichroism

A Chiroscan spectrometer (Applied Photophysics; Surrey, UK) was used to collect far ultraviolet circular dichroism spectra. A 0.1 mm pathlength Quartz Suprasil precision cell (Hellma Analytics; Müllheim, Germany) was used with protein suspended in H_2O, measurements were taken from 185-250 nm with 0.5 seconds time per point, bandwidth 1 nm, step size 0.5 nm, and 10 repeats. The molar ellipticity of individual circular dichroism spectra were normalised by concentration and molecular weight.

3 Results and Discussion

3.1 Expression system construction

Despite numerous attempts over many years to knock out, modify or express the *gcc* genes alone or in in combination with tags, chaperones, thioredoxins, secretion signals or other Gcc proteins in *E. coli*, *Lactococcus lactis* and *Saccharomyces cerevisiae*, none of these were successful. This was either due to no production of protein, the production of insoluble protein or protein toxic to the host. It was thought that a possible reason for this could be that the individual gene products were stabilised by association with other *gcc* gene cluster products. This created the need for an alternative system to test gene function and/or investigate the structure of the gene products. A significant breakthrough was made by Drower (2015) [67] who was able to knockout a GccF receptor encoded by the gene *pts18cba* in *L. plantarum* NC8. He was subsequently able to complement the knockout cells with a copper inducible plasmid pRV613 containing the *pts18cba* gene (pRV613:*pts18cba*-FLAG). The same system was later used by Bailie (2017) [70] to produce the immunity protein, GccH in *L. plantarum* NC8. Therefore a strategy that involved cloning the entire *gcc* cluster into a plasmid for heterologous expression was adopted.

As *E. coli* are commonly used to express heterologous proteins, a plasmid that was compatible with this species was initially trialled. To see if any of the Gcc proteins were produced and soluble in *E. coli* the pKS-*gcc* plasmid was transformed into *E. coli* BL21(DE3), BL21(DE3)pLysS, C41(DE3)pLysS and C43(DE3)pLysS cells. Only *E. coli* C43(DE3)pLysS was capable of retaining a reasonable level of growth once transformed with the pKS-*gcc* plasmid. Unfortunately, activity analysis, after IPTG was used to activate the pKS promotor, revealed that this cell line was not producing active GccF. Additionally none of the proteins produced by the cluster

could be identified using SDS-PAGE. When production of the Gcc proteins in *E. coli* proved to be unsuccessful the cluster was cloned into pRV610.

3.1.1 GccF expression system

In contrast to common expression plasmids the *gcc* genes were not placed under the control of an inducible promotor in the pRV610 plasmid as the first gene in the cluster, the immunity gene (*gccH*) is transcribed in the opposite direction to all the other *gcc* transcripts. To locate the -10, -35 and transcription start points of the promotor, the algorithm BPROM [119] was used, this found a divergent promotor between *gccH* and *gccA* responsible for the constitutive expression of the cluster. A second promotor located upstream of *gccF* is most likely responsible for higher concentrations of GccF produced when the cell density of the host organism increases. This promotor is possibly activated by the constitutionally expressed LytTR domain containing protein GccE. Both these predicted promotor regions were included in the *gcc* insert cloned into pRV610 [113], a shuttle vector capable of transforming both gram-positive and gram-negative bacteria.

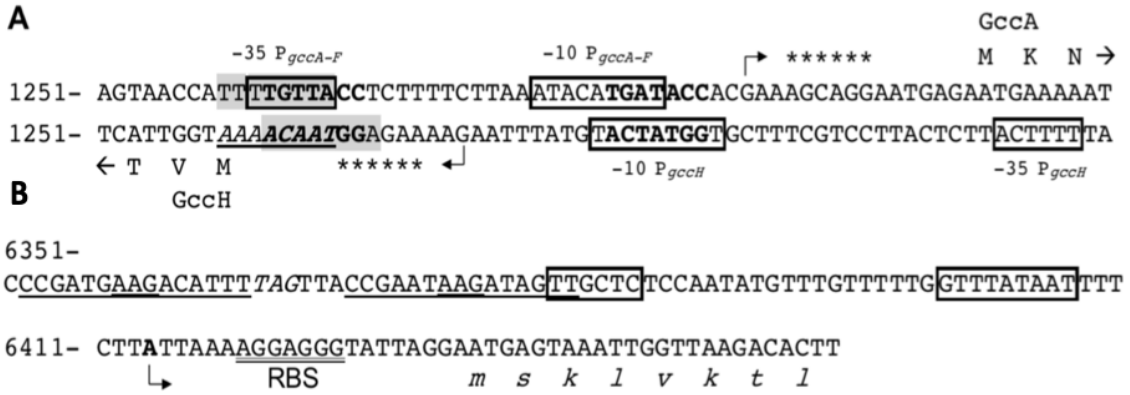


Figure 3.1: **Overview of putative *gcc* promotors.** Numbering is from Genbank accession GU552553. A) The *gccA-H* promotor. The -10 and -35 elements of the overlapping divergent promotors *PgccA-F* and *PgccH* are boxed, and bent arrows indicate predicted transcription start points. Oligonucleotide sequences highlighted in grey are predicted catabolite repressor protein (Crp) binding sites, and the underlined sequence is a predicted AraC binding site. Consensus ribosome binding sites for *gccA* and *gccH* are marked with asterisks. Bold letters denote an imperfect heptanucleotide repeat (16 nt spacing). B) The *gccF*-specific promotor *PgccF*.

The pRV610*gcc* vector was constructed by using PCR to incorporate complementary restriction sites to either end of the *gcc* cluster. The plasmid (pRV610) and the PCR product (containing the *gcc* genes) were digested individually and then ligated together to form pRV610*gcc* (Figure 3.2). This allowed the ligated plasmid to be rapidly transformed into *E. coli* EC100 cells, propagated and extracted to yield high concentrations of DNA for analysis by restriction enzyme digest (Figure 3.3) and sequencing. For protein expression, a plasmid free host *L. plantarum* NC8 [120], closely-related to the native producer was chosen to ensure that the promotors originating from *L. plantarum* KW30 were recognized by the cell. The vector pRV610*gcc* was successfully incorporated into *L. plantarum* NC8 cells that were then tested for GccF production.

Figure 3.4 indicates that the native producer and heterologous producer *L. plantarum* KW30 and *L. plantarum* NC8 pRV610*gcc* produce a similar amount of active GccF. However, from these assays alone the possibility that another endogenous *L. plantarum* NC8 bacteriocin [121] had been up-regulated by the production of the *gcc* gene products could not be ruled out. To show this activity is due to GccF, GccF was purified from *L. plantarum* NC8 pRV610*gcc* CFS using (cation exchange (CEX)) followed by RP-HPLC (Figure 3.5). During RP-HPLC a peak eluted at 37% acetonitrile, characteristic of GccF, this was collected and analysed by tricine gel electrophoresis and mass spectrometry (Figure 3.6) showing an identical mass to wild-type GccF confirming heterologous production of active GccF in *L. plantarum* NC8 pRV610*gcc*.

Following mass spectrometry, the amount of GccF produced from a 1 L culture was quantified. The purified sample was lyophilised then weighed, showing 3.75 mg of GccF was produced. However, based on absorbance at 205 nm, the yield was calculated to be 0.72 mg of protein. This large difference is likely due to residue H_2O

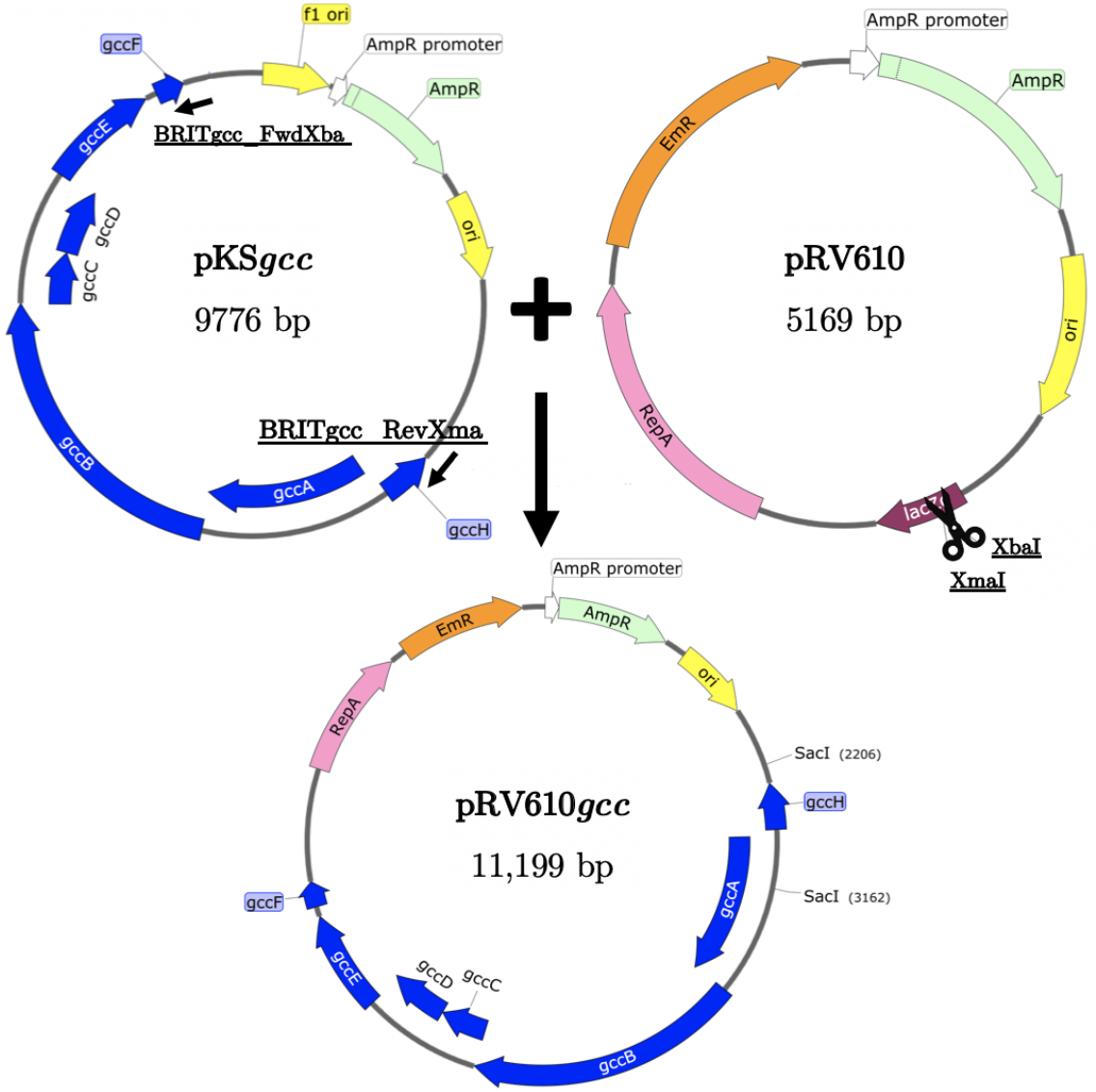


Figure 3.2: **Construction of pRV610*gcc*.** Showing the primer binding sites on pKS-*gcc* and the restriction enzyme digestion sites of pRV610 used to construct pRV610*gcc*.

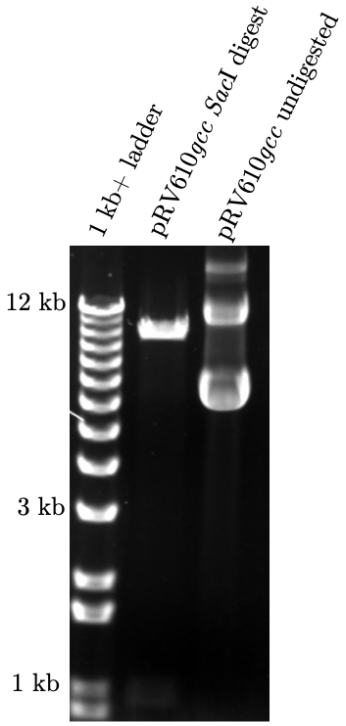


Figure 3.3: **pRV610*gcc* digested with restriction enzyme *Sac*I**, yielding bands of the expected size (10.3-kbp and 950-bp).

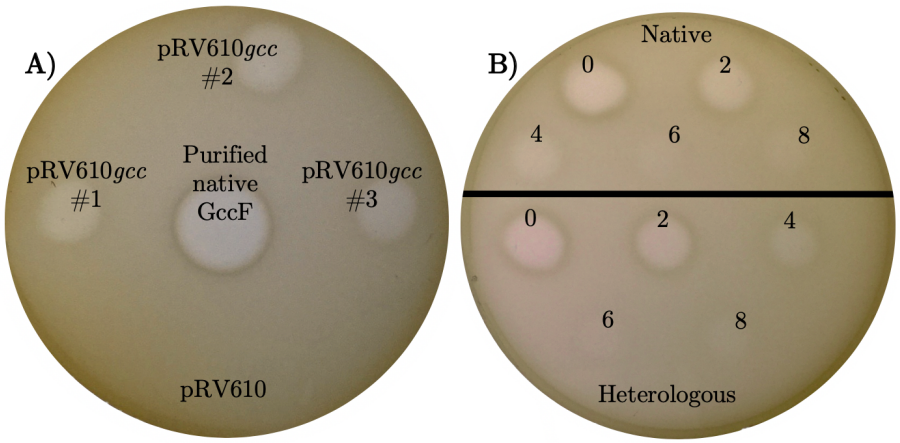


Figure 3.4: **GccF activity assay comparing heterologous and native producers.** Indicator plates with GccF purified from the native producer *L. plantarum* KW30 as a positive control. A) 4 μL of cell culture supernatant from *L. plantarum* pRV610 as a negative control, and 4 μL of cell culture supernatant three independent cultures from GccF producing *L. plantarum* NC8 pRV610*gcc* (heterologous) and *L. plantarum* KW30 (native). B) 0, 2, 4, 6, and 8-fold dilutions showing both cultures have a similar concentration of GccF in their CFS.

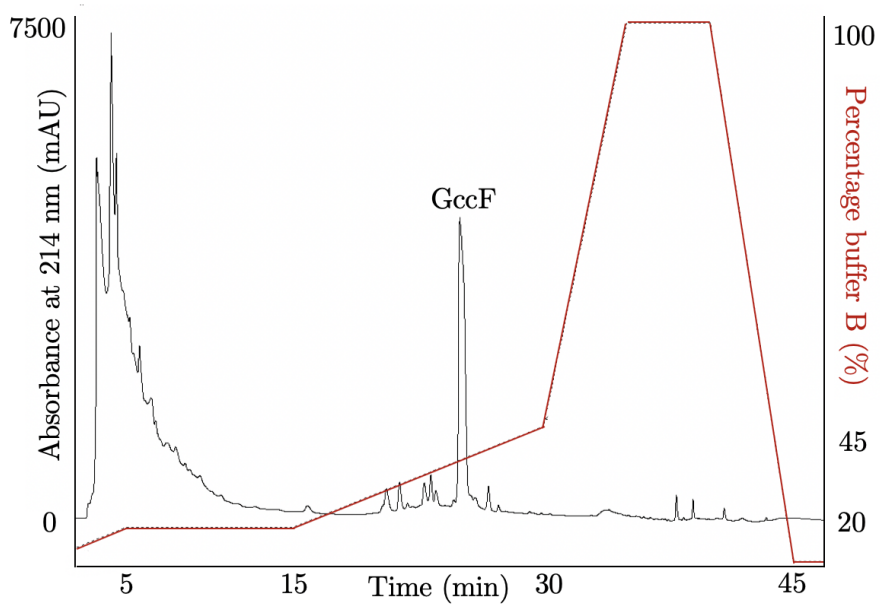


Figure 3.5: **RP-HPLC trace of GccF extracted from *L. plantarum* NC8 pRV610*gcc*.** This was first purified by CEX.

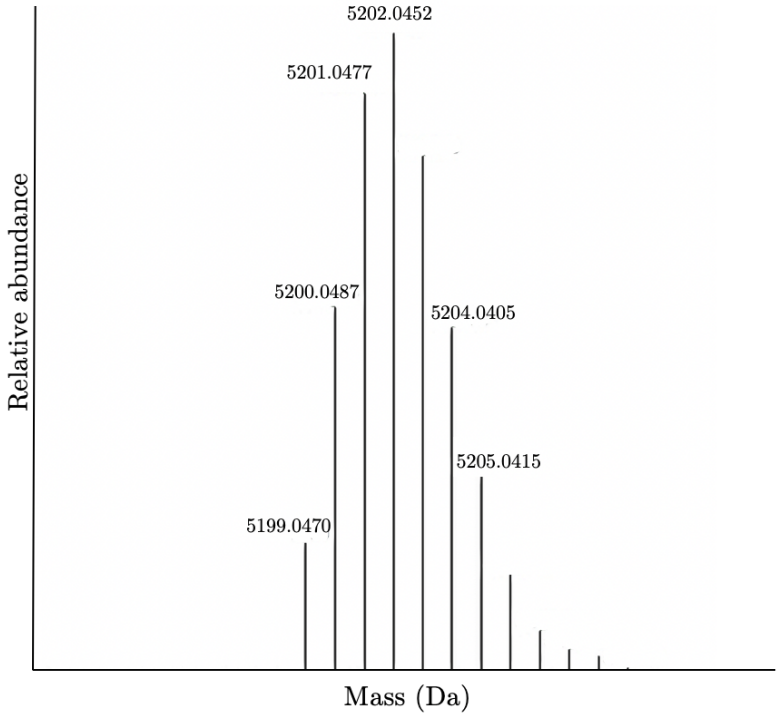


Figure 3.6: **Deconvoluted mass spectrum of GccF extracted from *L. plantarum* NC8 pRV610*gcc*.** This was purified by CEX and RP-HPLC showing a monoisotopic mass of 5199.05 Da (4796.92 (theoretical monoisotopic peptide mass) + (2 x 203.0794) (GlcNAc mass) + (2 x -2.01565) (disulfide bond mass) = 5199.048 Da (total protein mass))

and impurities such as tween, a detergent contained within the culture media containing GccF, as no other proteins appear to be contaminating the sample as shown in Figure 3.7. The gel shows semi-purified GccF with a wavy edge characteristic of GccF obtained using a preparative column (G. Norris personal communication), this can occur when detergent is present. Taken together with Figure 3.4, these results suggest that the yield of GccF from *L. plantarum* NC8 is similar to that from *L. plantarum* KW30 (0.25 - 1 mg/L) [66]. The concentration of recombinant GccF required to inhibit the growth of *L. plantarum* ATCC 8014 by 50 % (IC_{50}) was calculated to be 2.0 ± 0.3 nM compared to 2.2 ± 0.3 nM for the wild type under identical conditions (Figure 3.8). Furthermore, the antimicrobial activity of recombinant GccF purified from *L. plantarum* NC8 could also be attenuated by exposing the sensitive strain to 5 mM of GlcNAc (Figure 3.9) a property shared with the wild type peptide [55].

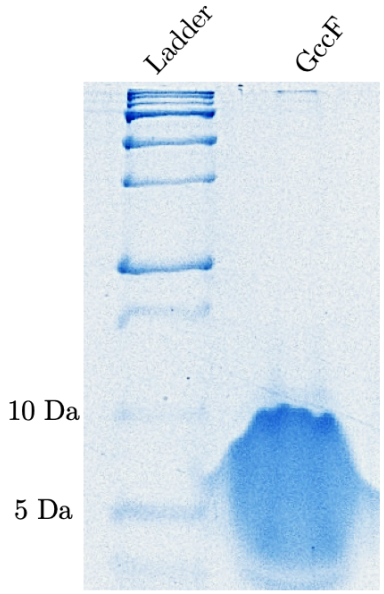


Figure 3.7: **Tricine gel with GccF purified from *L. plantarum* NC8 pRV610*gcc*** CFS by CEX and RP-HPLC. The gel used was a 16% tricine gel.

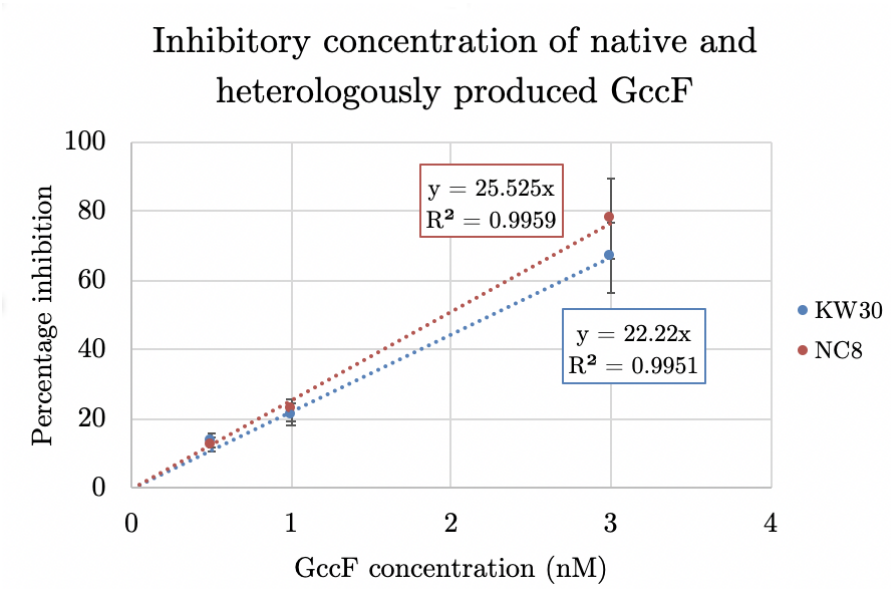


Figure 3.8: **IC_{50}'s of native and heterologously produced GccF.** Concentrations of native (*L. plantarum* KW30) and heterologously (*L. plantarum* NC8 pRV610*gcc*) produced GccF corresponding to percentage inhibition of the indicator strain *L. plantarum* ATCC 8014.

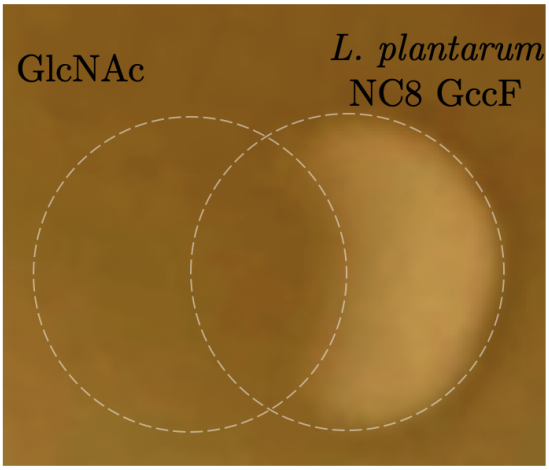


Figure 3.9: **Inhibition of GccF produced in *L. plantarum* NC8 pRV610*gcc* by GlcNAc.** Indicator plate assay 20 μL of heterologously produced GccF was spotted next to 20 μL of 5 mM GlcNAc. The region of overlap shows no inhibition of cell growth indicating GlcNAc protects against GccF purified from the heterologous system.

These results prove that the native promotors within the cluster are sufficient for the heterologous expression of GccF. The concentration of GccF produced suggests that either the GccF production is controlled by the proteins encoded by the cluster itself, or that the copy number of pRV610 is similar to that of the plasmid containing the *gcc* cluster in the native producer *L. plantarum* KW30. As the copy number of pRV610 is low it is possible that the native host has a similar number of copies as whole genome sequencing has shown that this cluster is plasmid-based (M. Patchett, personal communication). The successful production of GccF using the pRV610*gcc* vector has provided a means to verify the function of the cluster genes as well as engineering GccF itself to better understand its structure function relationship.

To help identify the requirements for GccF production, the *gcc* cluster was transformed into other closely-related species that did not produce bacteriocins that were active against the GccF indicator strain *L. plantarum* ATCC 8014. This allowed activity assays to be carried out to determine if the cells were expressing GccF without the interference of bacteriocins produced by the host strain. To begin with *Lactobacillus sakei* 790 was transformed with pRV610*gcc* and shown to produce active GccF at what appears to be the same concentration as both *L. plantarum* NC8 pRV610*gcc* and *L. plantarum* KW30 (Figure 3.10). This shows that not only *L. plantarum*, species, but possibly all *Lactobacillus* species have the ability to read the *gcc* cluster promotors originating from *L. plantarum* KW30. It also shows that the immunity protein, GccH, is sufficient for protection of the host from GccF as all transformed species were originally susceptible. To test whether the entire *Lactobacillaceae* family was capable of producing GccF from this plasmid, an attempt was made to transform pRV610*gcc* into a plasmid-less species, *Lactococcus lactis* NZ9000. Unfortunately this proved to be unsuccessful.

Figure 3.10: **GccF activity assay comparing CFS from *L. sakei* 790 and *L. plantarum* native and heterologous producers of GccF.** CFS or MRS media (20 µL) were used as negative controls.

Enterococcus faecalis JH2-2 from the order *Lactobacillales* was successfully transformed with pRV610*gcc*, but the culture supernatant did not show any activity against the standard indicator strain *L. plantarum* ATCC 8014 showing GccF was not produced by these cells. Following a plasmid extraction it was obvious that although the plasmid was being maintained by the cells, either the promotors were not being recognised or perhaps some or all of the Gcc proteins were not functional in this species. In an attempt to find possible reasons for the lack of GccF production, *Enterococcus faecalis* JH2-2 pRV610*gcc* cells were embedded in MRS agarose and used in an activity assay (Figure 3.11). Clearings showed that the cells were still susceptible to GccF indicating the immunity protein was not being produced at all, or was being produced in such small quantities that protection from GccF was lacking. The difference in clearing size is likely due to the different cell densities within the agar as these were not normalised prior to plating, instead a 100-fold dilution of each overnight culture was made. Subsequently pRV613 gccH a copper inducible

plasmid, that had been previously used to identify GccH as the immunity protein [70] was transformed into *Enterococcus faecalis* JH2-2 cells. IC50 assays showed that *E. faecalis* JH2-2 pRV613_gccH cells were immune to GccF (personal communication S. Bisset). Because GccH can be produced in these cells when under control of the inducible promotor contained within pRV613, it indicates that the problem with production begins with the promotors within the *gcc* cluster. It is possible that GccF could be produced in these cells if the promotors identified within the *gcc* cluster were exchanged for promotors compatible with *Enterococcus faecalis* JH2-2, but for reasons already discussed this would be a very laborious procedure. From these experiments it was concluded that at least some *Lactobacillus* species could host this gene cluster making it feasible that this system could be introduced into food grade *Lactobacillus* species for therapeutic use or alternatively, used within a probiotic strain of bacteria to support a healthy microbiome.

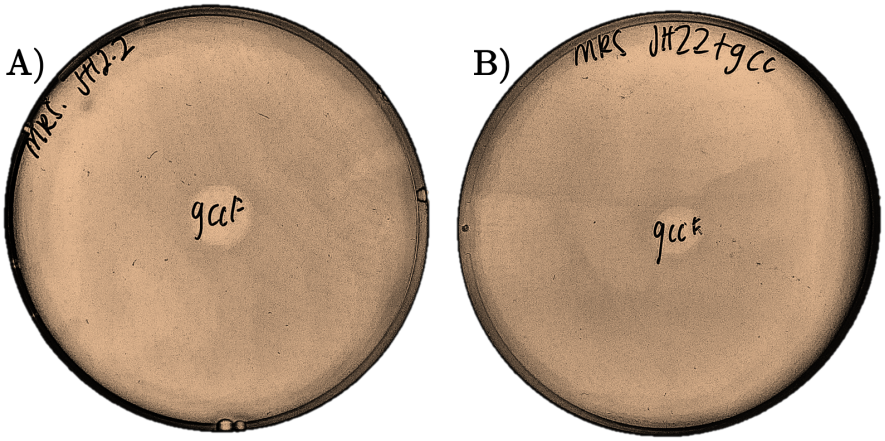


Figure 3.11: ***E. faecalis* JH2.2 susceptibility to GccF.** GccF activity assay with 5 μL of purified 10 μM GccF spotted on top creating a clearing where no cells have grown, A) *E. faecalis* JH2.2 cells. B) *E. faecalis* JH2.2 pRV610*gcc*

3.1.2 GccF expression system mutations

Following the successful construction of pRV610*gcc* mutations were introduced into cluster genes to help understand their roles in GccF production. Initially inverse PCR mutagenesis [122] was chosen as a method of mutating a plasmid without the addition of unwanted DNA that may interfere with the gene coding sequence. In order to do this, two primers were designed to amplify the entire cluster, one of which included the desired mutation. PCR was carried out using Phusion high-fidelity polymerase to create blunt ends and minimise errors that could result in other mutations affecting the production of GccF. PCR was followed by *Dpn*I treatment to remove the original plasmid, followed by PCR clean up. The PCR product was then phosphorylated using PNK and ligation was carried out overnight at room temperature. Transformation into chemically-competent *E. coli* EC100 cells the following day did not produce any transformants, necessitating troubleshooting (Table 3.1).

The only troubleshooting step that resulted in transformants produced plasmids that varied in size and lacked some of the expected restriction enzyme sites. The plasmids produced were therefore most likely a result of either insufficient blunting, or exonuclease contamination nibbling away the ends and exposing a free phosphate to facilitate ligation. There were only a few reports in the literature where blunt-ended ligation has been used with this method [122, 123, 124], and none of these were trying to ligate a plasmid of this size (11.2-kbp), suggesting that this may be a common problem. As including a restriction site to produce sticky ends was not an option without changing the coding sequence of the protein this method was abandoned in favour of a different method with much higher efficiency.

The SLIM method [116] replaced inverse PCR mutagenesis, and was successfully used to introduce mutations into the pRV610*gcc* plasmid. Phusion high-fidelity

Table 3.1: Troubleshooting carried out for inverse PCR mutagenesis

Potential problem	**Solution**	**Successful transformation**
Insufficient blunting	Increase extension time and DNTP concentration of PCR	No
*Dpn*I treatment of PCR product	*Dpn*I was not used and PCR clean-up was relied on for removal of original plasmid	No
PNK ineffective on PCR product	Primers, not the PCR product, were phosphorylated using PNK	Yes, but the plasmids were not of the correct size. Likely caused by exonuclease contamination as this was not repeatable
Transformation mix	pRV610*gcc* was added to the ligation mix	Yes (this shows the transformation mix is not inhibitory)
Ligation conditions	Overnight at 4°C and then at 37°C for 1 hr with varied concentrations of insert and vector	No

polymerase was used in both inverse PCR mutagenesis and successfully in the SLIM method, suggesting that there was no problem with amplification. Therefore, the most likely failure of the inverse PCR mutagenesis method was the inefficiency of blunt-ended ligation in a large construct. As the same ligase had been used to create pRV610*gcc* months earlier and had been used successfully prior to these experiments, it is unlikely that the problem lay with the ligase itself.

Each mutation introduced using the SLIM method required two 40 bp+ primers to be designed that included the desired mutation as well as two standard primers. Due to the AT rich sequence (most of the primers having a GC content below 30%) this was difficult and often resulted in low (48-55°C) annealing temperatures. The SLIM method was chosen because it had been used successfully to add a flag

tag to pRV613_*gccH* to create pRV613_flag-*gccH* along with other small insertions and mutations [105]. The mutant constructs created during this study were transformed into chemically-competent *E. coli* EC100 cells, propagated, extracted and sequenced. An example of the information obtained from sequencing can be seen in Figure 3.12. The identification of the correct mutation in the plasmid DNA allowed transformation of the construct into *L. plantarum* NC8 where the activity of GccF could be tested using the biological assay for bacteriocin activity. Furthermore, changes to GccF structure or production following these mutations was analysed using RP-HPLC, mass spectrometry, Circular Dichroism and GccF yield. The introduction and investigation of these mutations made it possible to carry out an investigation into the maturation pathway of GccF.

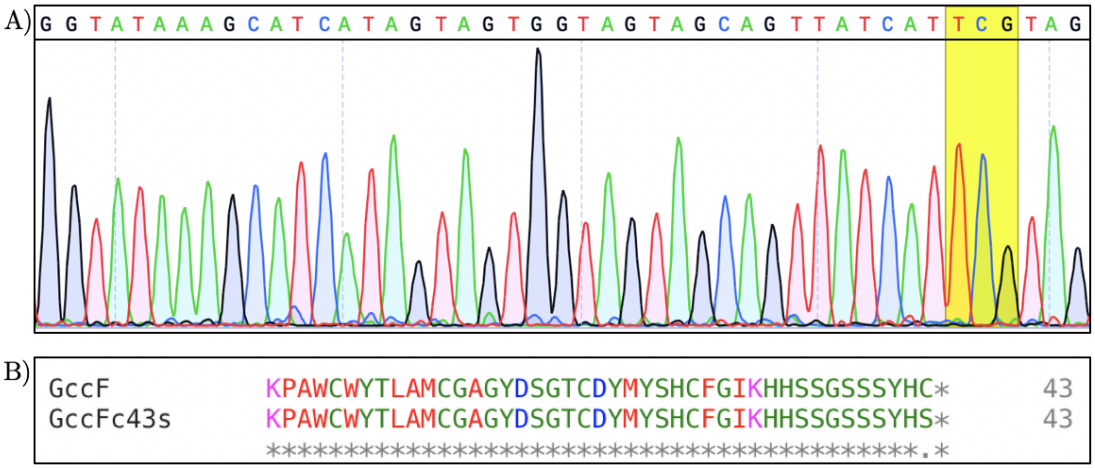


Figure 3.12: **GccF$_{C43S}$ mutation sequence.** A) Partial chromatogram of the pRV610*gcc*_GccF$_{C43S}$ sequence with the mutated codon highlighted. B) Protein sequence alignment of wild-type GccF and GccF produced by pRV610*gcc*_GccF$_{C43S}$.

In order to confirm that the transformation of bacteria with a given plasmid was successful the plasmid was extracted from the cells and subjected to restriction enzyme analysis. This was of particular importance when mutations resulted in no GccF production as the production of GccF itself was sufficient to confirm transformation. However the procedure does not come without challenges as gram-positive

bacteria are notoriously difficult to transform and to extract plasmids from in comparison with *E. coli* cells. Initially cells were pre-treated with 30 mgmL^{-1} lysozyme in 25% sucrose prior to using a standard plasmid prep kit. The extraction of plas-mid using this method resulted in a smear of DNA on a agarose gel electrophoresis (Figure 3.13A) making it difficult to identity the plasmid. In 2018 a new method for extracting plasmids from *L. plantarum* species was reported [114]. The only difference between this method and the one previously used was that the lysozyme treatment was carried out in the plasmid extraction resuspension buffer rather than in 25%sucrose and that the cells were washed twice with 5% glucose following the lysozyme treatment. Following a this new protocol and a restriction enzyme digest there was a huge decrease in genomic DNA contamination making the presence of the plasmid more obvious (Figure 3.13B). The decrease in genomic DNA contam-ination when using the new protocol is likely due to the absence of lysozyme in the solution during cell lysis. It is therefore possible that the presence of lysozyme during cell lysis decreases the effectiveness of the reagents in the plasmid extraction kits.

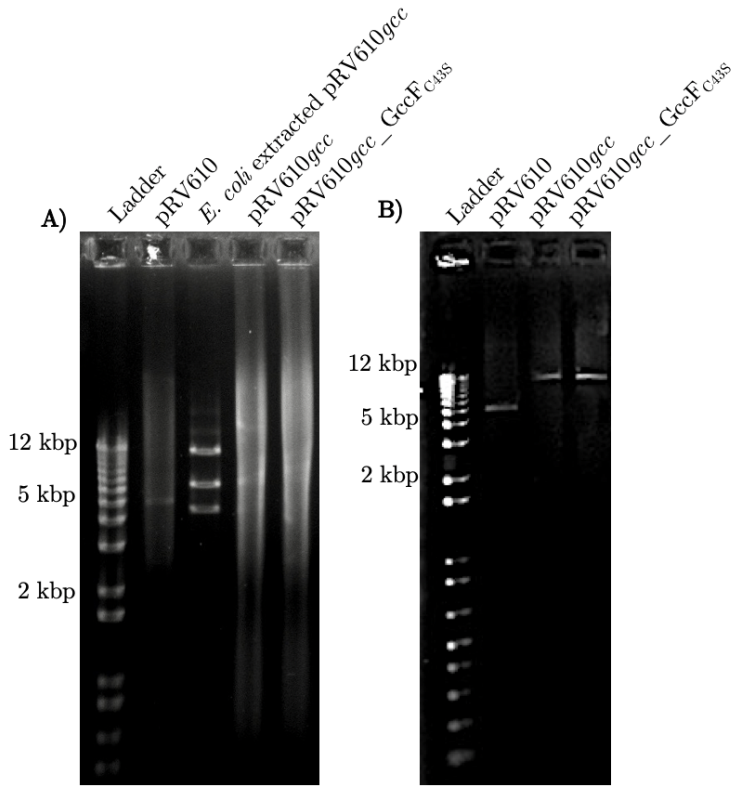


Figure 3.13: ***L. plantarum* NC8 plasmid transformation conformation.** A) Plasmid extraction using a commercial kit with sequenced *E. coli* extracted plasmid pRV610*gcc* as positive control for *Eco*RI restriction digest. B) Plasmid extraction using the new protocol [114] with *Xho*I restriction digest producing a single cut, linearising the 11,199-kbp plasmid.

3.1.3 GccH expression system

Although pRV610*gcc* was shown to be capable of producing GccF within *L. plantarum* cells, this system did not produce large quantities of the Gcc proteins as they could never be detected by SDS-PAGE. The structure of GccF is well-characterised but the structures of the remaining Gcc proteins are completely unknown. This is largely due the expression trials in both *E. coli* and *S. cerevisiae* being unsuccessful. Gaining insights into the structure of these proteins would reveal critical details concerning the maturation pathway of GccF. Interestingly, another group working on a closely-related cluster of genes required to produce the glycocin SunA have been able to produce recombinant SunS (the glycosyltransferase within the SunA cluster) protein in *E. coli* using the pET28a vector [54]. There are several differences between this plasmid and pProEx_HTB, the plasmid which the *gcc* genes had been cloned into. These are antibiotic selection, His-tag location and tag cleavage site. Additionally pET28a contains the *rop* gene to restrict the plasmid copy number. For these reasons it was decided that the *gcc* genes should be cloned into this plasmid for expression trials.

To identify the best target for crystallisation, bioinformatic analysis of each protein was carried out. XtalPred [125] was used to estimate the crystallisation propensity (where a lower score indicates a higher chance of crystallisation) and PONDR [126] to predict disorder (the more disordered a protein is, the less likely is will crystallise) and the percentage of methionine residues was considered to enable selenomethionine labelling. Table 3.2 depicts the results from this analysis and highlights that GccH, the immunity protein, is the best candidate for structural studies. Additionally, this protein is arguably the most intriguing within the cluster as it is not well-understood how it protects the producer from GccF, and it has little sequence similarity to other characterised proteins. It is also predicted to be cytosolic

and has not been the subject of multiple expression trials in various plasmids as the remaining cluster genes had been (personal communication G. Norris) making *gccH* easily the most appealing candidate.

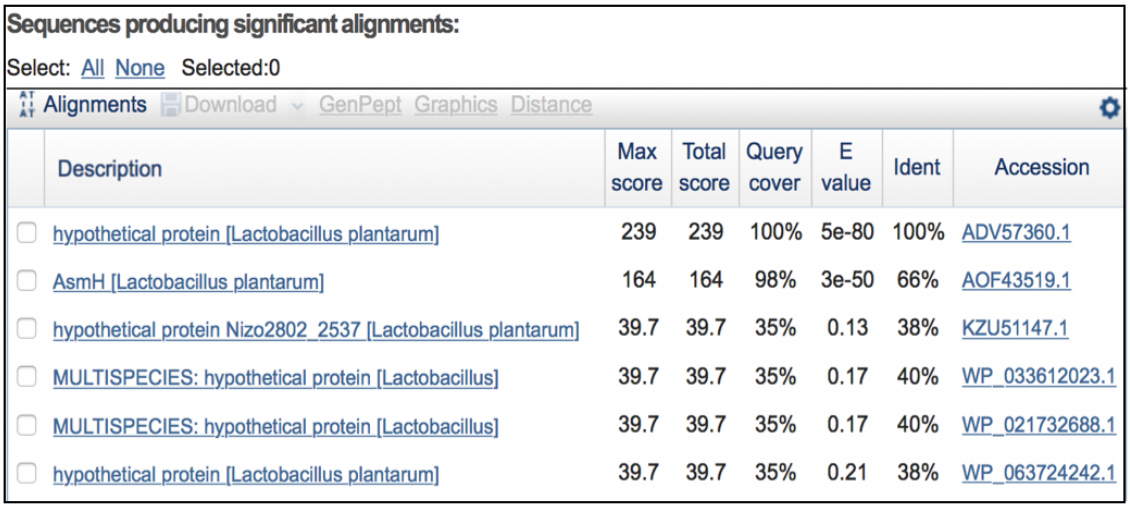
Table 3.2: Identification of best protein target for crystallisation

Protein	**Xtalpred EP class**	**Disorder (%)**	**Methionine (%)**
GccH	2	8.47	1.7
GccA	4	12.56	1.4
GccB	5	6.55	2.4
GccC	5	5.76	0.7
GccD	5	11.30	1.7
GccE	4	7.75	2.3

A BLASTP [127] search of GccH shows GccH as the number one hit at 100% sequence identity followed by ASMH with 66%. ASMH is from the gene cluster of the most closely-related bacteriocin to GccF, ASM1 [53] (Figure 3.14). The conservation of these proteins is not as high as one would expect as GccF and ASM1 have 88% sequence similarity and either immunity protein can provide protection from both bacteriocins indicating much of this sequence does not need to be conserved for its function [64]. It is worth noting that the other proteins identified with sequence similarity originate from *L. plantarum* species and are found downstream of the N-acetylglucosamine transporter that has been frequently mutated in GccF resistant cells [67] and thought to be a receptor for GccF [105].

pProEx_HTb_*gccH*

The *E. coli* BL21 (DE3) pProEx_HTB_*gccH* expression system developed by T. Loo was shown to produce His-tagged GccH using proteomic methods. Before cloning GccH into the pET28a plasmid it was worthwhile checking that the *E. coli* BL21 (DE3) pProEx_HTB_*gccH* could not be used to produce reasonable concentrations of GccH. The transformed *E. coli* BL21 (DE3) cells were grown to an OD_{600}



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance

Description	Max score	Total score	Query cover	E value	Ident	Accession
hypothetical protein [Lactobacillus plantarum]	239	239	100%	5e-80	100%	ADV57360.1
AsmH [Lactobacillus plantarum]	164	164	98%	3e-50	66%	AOF43519.1
hypothetical protein Nizo2802_2537 [Lactobacillus plantarum]	39.7	39.7	35%	0.13	38%	KZU51147.1
MULTISPECIES: hypothetical protein [Lactobacillus]	39.7	39.7	35%	0.17	40%	WP_033612023.1
MULTISPECIES: hypothetical protein [Lactobacillus]	39.7	39.7	35%	0.17	40%	WP_021732688.1
hypothetical protein [Lactobacillus plantarum]	39.7	39.7	35%	0.21	38%	WP_063724242.1

Figure 3.14: **BLASTP search of GccH.**

of 0.6 before induction with IPTG. Samples were taken at 0, 3, 6, 10 and 24 hours, lysed by sonication and separated into soluble and insoluble fractions. It was important that all buffers used contained 2 mM TCEP as GccH contains a single cysteine residue that needed to be kept reduced to be ensure there was no aggregation caused by the formation of disulfides. These fractions were analysed by SDS-PAGE (Figure 3.15), His-tagged GccH has a theoretical molecular weight of 16.3 kDa, however there was no difference in the concentration of proteins with this molecular weight after induction. A band did appear in the cell lysate at 14.2 kD (see appendix for retention factors used to estimate protein size 6.2) that increased with time after induction. It was most concentrated between 6 and 10 hours in the whole cell lysate fraction. The identity of the band was investigated using proteomic methods and shown to not include GccH.

To ensure that GccH was not being produced at low concentrations, a 1 L culture was grown in selective media to an OD_{600} of 0.6 before being induced with IPTG for a further 24 hours at 37°C. These cells were pelleted by centrifugation, washed twice in lysis buffer and lysed using the french press before the soluble

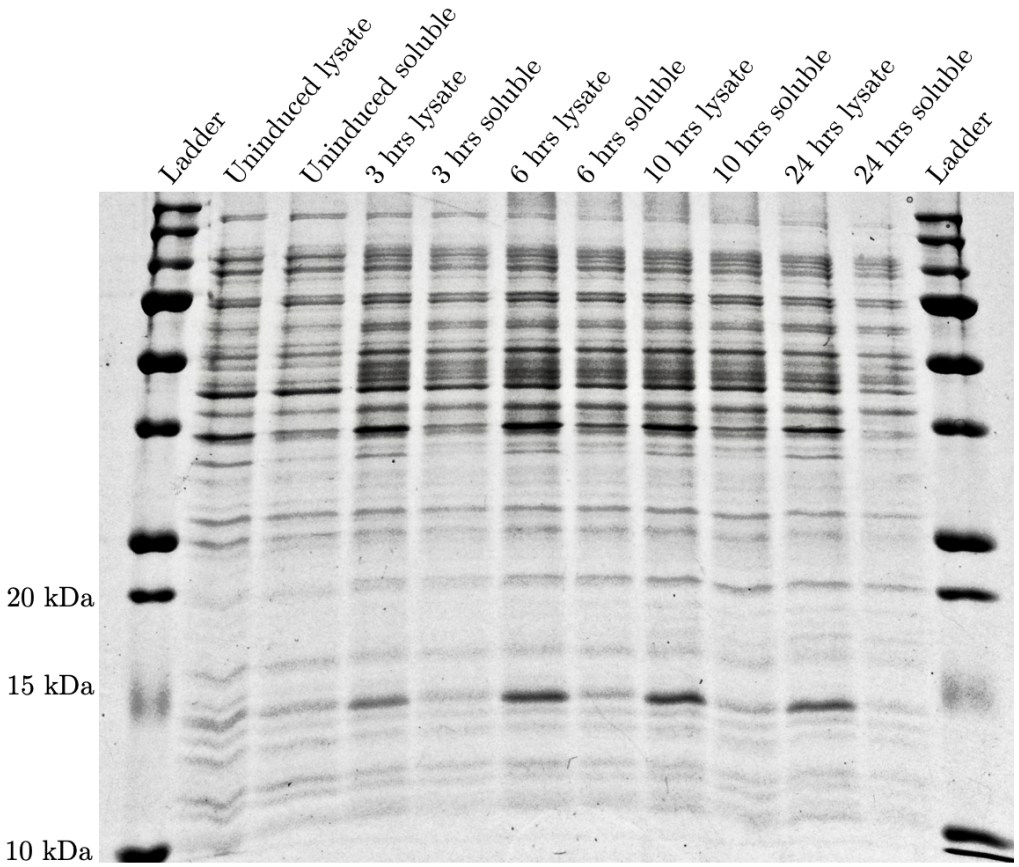


Figure 3.15: ***E. coli* BL21 (DE3) pProEx_HTb_*gccH* induction trial.** The cells were induced for different lengths of time at 20 °C before lysis. With lysate representing all cellular proteins and soluble fractions containing proteins found in the supernatant following centrifugation.

and insoluble fractions were separated. The soluble fraction was then subjected to IMAC. Elution of proteins was followed by absorbance at 280 nm, unfortunately no peaks were eluted following the initial binding and washing steps. This was unsurprising as GccH is a relatively small protein that contains a single tryptophan residue (the biggest absorber of light at 280 nm) and was not concentrated enough to identify within the soluble fraction on the polyacrylamide gel. To check if any of these fractions contained GccH a selection of imidazole washes where chosen and analysed by gel electrophoresis (Figure 3.16). The imidazole concentration increased linearly from A5 to C9 with the majority of the *E. coli* proteins eluting at low imidazole concentrations (A5-15). The absence of concentrated bands at

approximately 16 kDa following elution from the IMAC column prevented the need for further proteomic investigation.

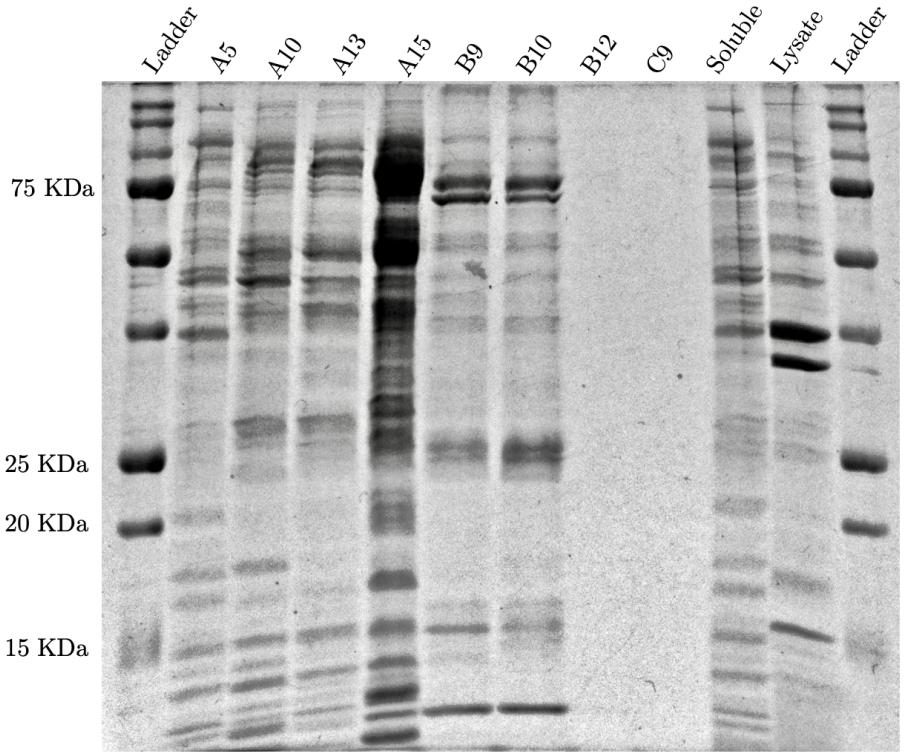


Figure 3.16: **IMAC purification of *E. coli* BL21 (DE3) pProEx_HTb_*gccH*** induced for 24 hrs at 20 °C. With lysate representing all proteins and soluble fractions obtained following centrifugation to pellet insoluble proteins. Increasing imidazole concentration from A5 to C9.

pET28a_*gccH*

At this point it was decided that moving GccH to another plasmid to try increase the concentration of soluble GccH produced by the cells was the best way forward. This was done by cutting the pET28a plasmid with restriction enzymes and amplifying *gccH* from pProEx_HTb_*gccH* using PCR to add restriction sites complementary to the vector before digestion. Once digested the small fragments were removed using a PCR clean up kit before ligation of the vector and insert and transformation into *E. coli* EC100 cells (Figure 3.17). Transformation was successful and in the first instance it was important to check the total size of the

plasmid was correct, this was achieved by digesting with *Nae*I. The fragmentation pattern matched what was expected if the plasmid contained *gccH* (1588, 3044, 850 and 103 bp) (Figure 3.18). To confirm the plasmid contained the correct sequence, it was sent for sequencing. Results from sequencing indicated that the ligation had been successful and that no mutations had arisen within the region amplified (Figure 3.19).

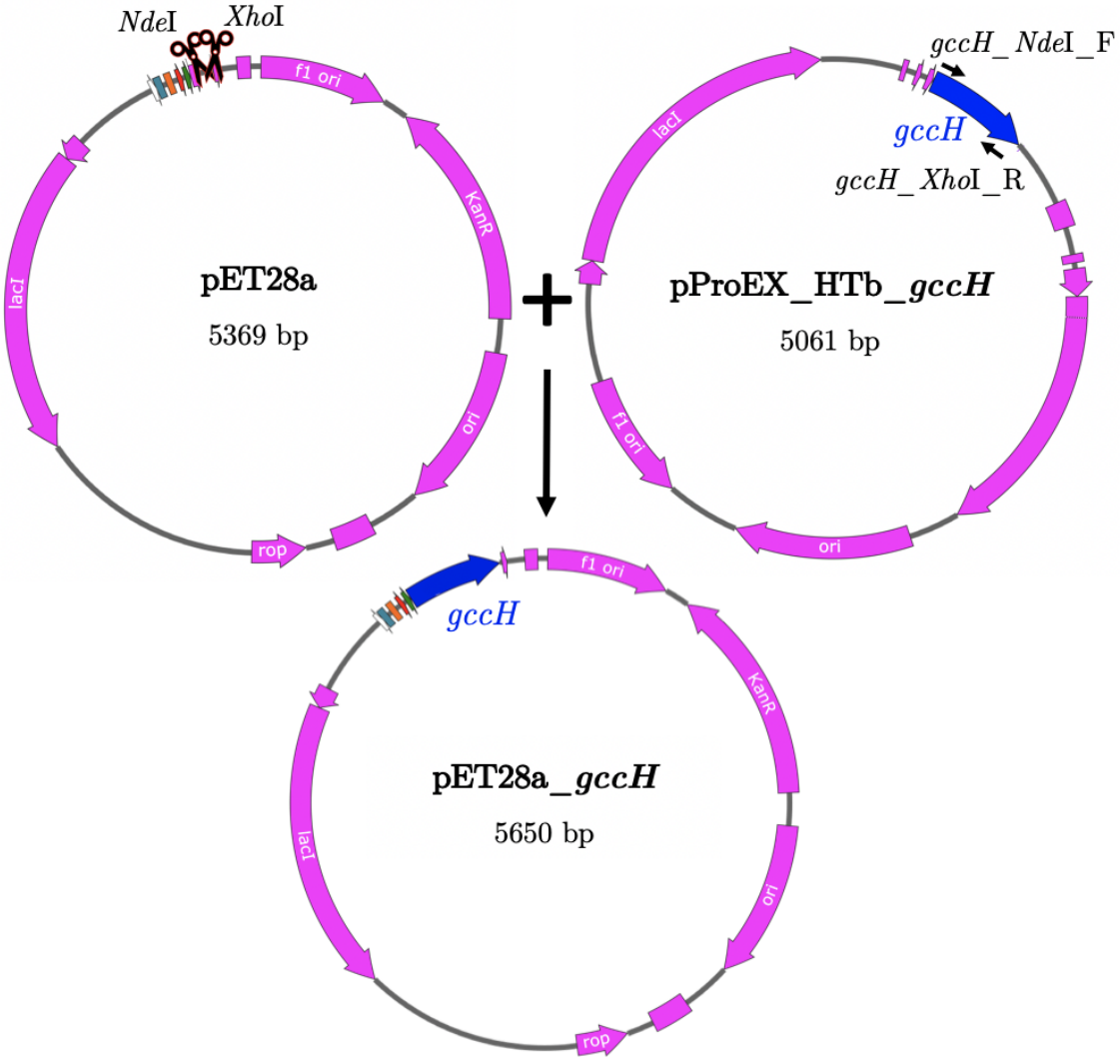


Figure 3.17: **pET28a_*gccH* construction.** This plasmid was created by adding the target restriction sites *Xho*I and *Nde*I to the *gccH* insert by high-fidelity PCR. Complementary double digests were followed by ligation of the PCR product and the vector together. pET28a_*gccH* is depicted with *gccH* coloured blue, the His-tag is coloured red, the IPTG inducible promotor white, the RBS orange and the thrombin site green.

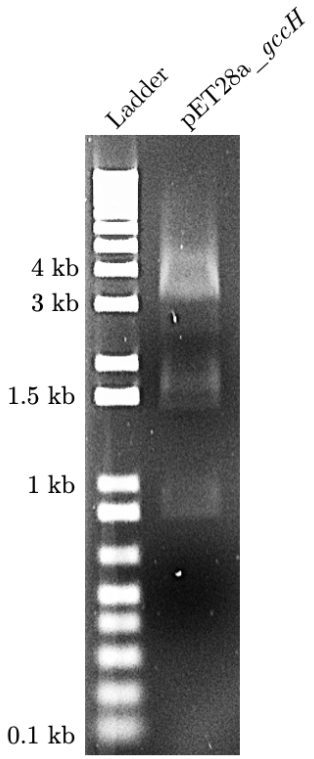


Figure 3.18: **pET28a_*gccH* *Nae*I restriction enzyme digest.**

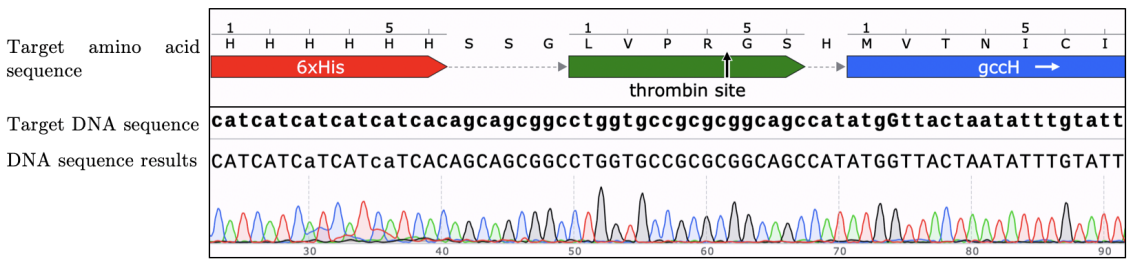


Figure 3.19: **pET28a_*gccH* sequence alignment.** 100 % match over the *gccH* insert region.

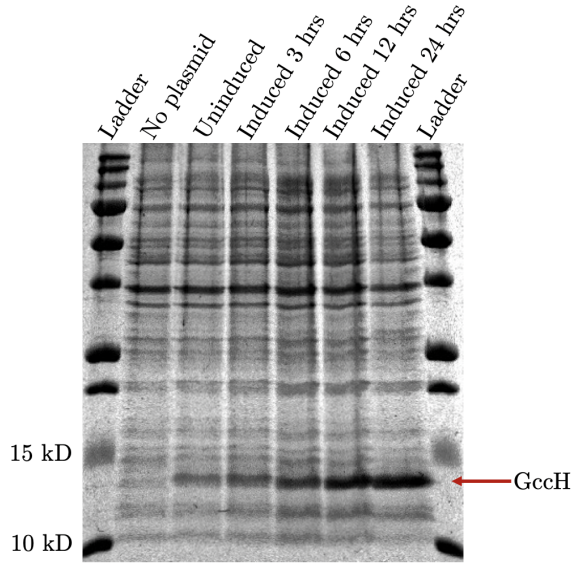


Figure 3.20: **GccH expression trial with *E. coli* BL21 (DE3) pET28a_*gccH* cells.** With uninduced, plasmid-less *E. coli* BL21 (DE3) cells as a control.

Finally the pET28a_*gccH* plasmid was transformed into the expression strain *E. coli* BL21 (DE3). These cells were grown to an OD_{600} of 0.8 at 30°C before being induced with 1 mM IPTG. Cells were harvested at 0, 3, 6, 12 and 24 hours. This time a additional control, plasmid-less *E. coli* BL21 (DE3) cells were also used, clearly showing that GccH accumulated in the cell and was most concentrated at 24 hours (Figure 3.20). This was a promising result, however the question still remained as to if this protein was soluble. To check solubility the cells induced for 24 hours were taken and separated into soluble and insoluble fractions prior to running on SDS-PAGE. Unfortunately this indicated that no soluble protein was produced (Figure 3.21).

Insoluble proteins are likely incorrectly folded and will require solubilisation prior to crystallisation, this means that any structural information gained would be less reliable as the protein may not be folded in the same way that it would be *in vivo*. As solubilisation can be a tricky and time consuming process this project was not continued. Had GccH been soluble it would have been purified for crystallisation trials.

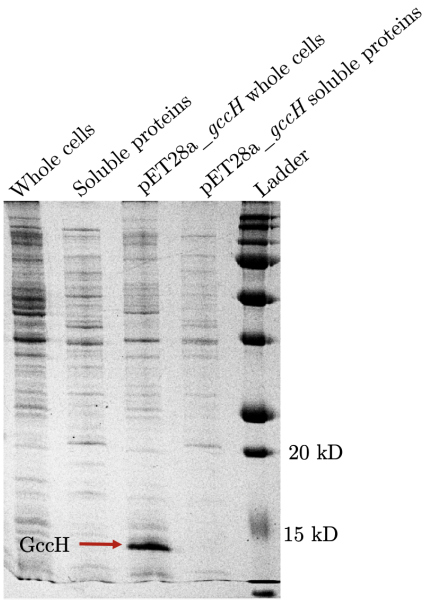


Figure 3.21: ***E. coli* BL21 (DE3) pET28a_*gccH* solubility trial.** Whole and soluble cell fractions of plasmid-less *E. coli* BL21 (DE3) cells as a control and induced *E. coli* BL21 (DE3) pET28a_*gccH* with arrow pointing to insoluble GccH.

3.2 Protein characterization

The fastest way to confirm the involvement of the *gcc* cluster genes in GccF production was to inactivate the gene products and observe the phenotypic outcome. Mutations were introduced to pRV610*gcc* using the SLIM method [116] and the resulting changes to the production of GccF were observed in transformed *L. plantarum* NC8 cells. Complete deletion of the *gcc* genes was not possible due to overlapping promotor regions at the translation start site of *gccA*, *gccF* and *gccH* and the open reading frame of *gccE*. The promotor located upstream of the *gccA* translation start site was of particular concern as this promotor drives transcription of the entire cluster. Additionally in the *gccABCDE* operon, the open reading frame of all adjacent genes overlap making it difficult to remove a single open reading frame (Figure 1.4). Instead, premature stop codons were placed as close to the start codon as possible without interfering with the sequence required for the initiation or termination of transcription and translation. The location of these stop codons resulted in severely truncated proteins, rendering them non-functional, which should produce the same phenotypic outcome as complete gene deletions. While this method is useful for determining if the Gcc proteins are not required for the production of GccF, loss of GccF activity could be caused by polar effects following the introduction of premature stop codons within the polycistronic mRNA. Therefore this method alone can only be used to determine which proteins are not required for GccF production.

3.2.1 Activity assay development

A new, more sensitive test was developed to accurately access the production of active GccF following mutations that resulted in a significant loss of activity in CFS. The standard test for activity was embedding an indicator strain within plated agar as shown in Figure 3.22. Following this purified GccF, or CFS from a GccF

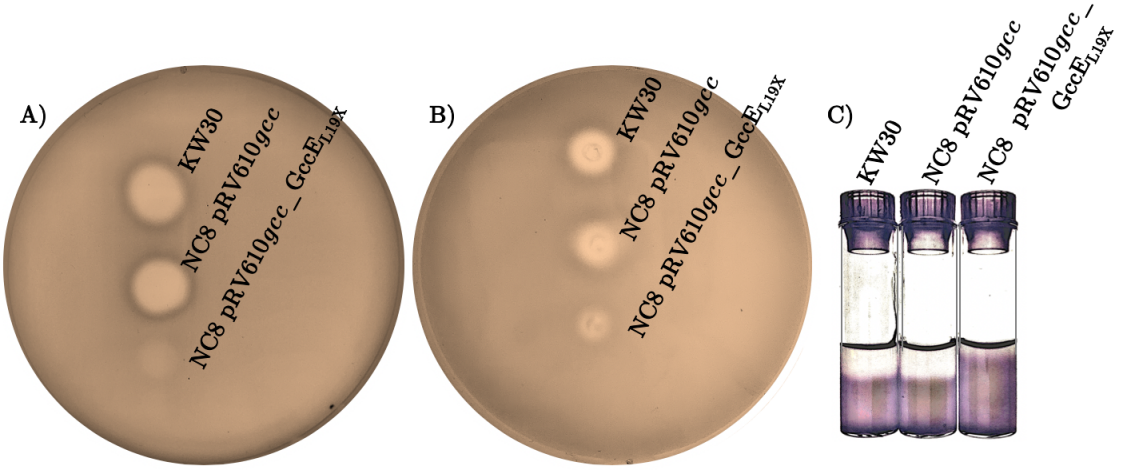


Figure 3.22: **Plate vs tube activity assays.** *L. plantarum* KW30, *L. plantarum* NC8 pRV610*gcc* and *L. plantarum* NC8 pRV610*gcc*_GccE$_{L19X}$ CFS; A) 5 μL on plate; B) 20 μL in wells; C) 40 μL in tubes.

producing strain was spotted on top. When CFS from *L. plantarum* KW30 or *L. plantarum* NC8 pRV610*gcc* were used, clearing could easily be observed with as little as 5 μL. This is not the case for mutants producing smaller volumes of active GccF such as the GccE knockout (Figure 3.22) that will be discussed later in this chapter (3.2.3). To increase the local concentration of GccF, wells with a diameter of 5 mm that can hold approximately 20 μL were made in the plates. To further increase the concentration of GccF, the first 20 μL of GccF solution to be tested was allowed to evaporate under a flame before another 20 μL was applied to the same well. This could take upwards of an hour, increasing the risk of contamination. Further complications were caused by inconsistencies in well shape, size and depth that occurred because they were hand-made, which made it difficult to accurately compare the activity of CFS from different mutants.

A new method was developed that used small tubes (diameter 6 mm) filled with the indicating agar rather than plates. Larger volumes of CFS could be placed on top of the indicating agar covering an area with a similar diameter to a standard well. This method was tested with great success; it was easy to perform and the

Figure 3.23: **GccF standards tube activity assay.** 20 μL of purified GccF diluted in MRS or *L. plantarum* NC8 pRV610*gcc* CFS was applied to each tube.

results were more consistent. Volumes up to 50 μL could be applied to the surface, thus increasing the local concentration of GccF. Unlike the plate method where the diameter of clearings was compared, the tube method enabled clearings to be measured by depth. This proved to be a more sensitive method of detection that allowed relative quantification of active GccF (Figure 3.23).

From comparing the purified GccF standards to the CFS of *L. plantarum* NC8 pRV610*gcc* the concentration of GccF within the CFS could be estimated. This estimation is based on the assumption that the activity of a single molecule of GccF is identical within CFS and when purified. However, it is likely that changes to the GccF core peptide will result in an activity change, as observed when testing the chemically-synthesised analogues [105]. Instead, this comparison could be better described as an estimation of the total activity relative to purified GccF. Figure 3.23 showed that *L. plantarum* NC8 pRV610*gcc* CFS had activity equivalent to 1.25 μM purified GccF. Additionally it demonstrates that the activity of purified

GccF could be detected at concentrations as low as 0.16 μM, indicating this assay can identify up to an 8-fold reduction in CFS activity. Furthermore, analysis of these tubes indicates that in 1 L of *L. plantarum* NC8 pRV610*gcc* CFS there should be 6.5 mg of GccF. This is much higher than the 0.7 mg of GccF purified from a 1-L culture of *L. plantarum* NC8 pRV610*gcc* and gives a yield of 11%. A yield this low is not surprising as there are multiple steps during the standard purification protocol where GccF could be lost. However, it indicates that there is potential to increase the yield from these cells without further genetic manipulation.

3.2.2 GccF

The construction of an easily-modified expression system for GccF mimicking the native producer allowed mutations to be introduced to *gccF*. Recent experiments with chemically-synthesised analogues of GccF served as a good starting point for the design of new mutations as they had already been characterised in terms of activity. Although the results of these experiments, advanced understanding of the relationship between the structure and function of GccF, the process was costly. The production of GccF mutants using an expression system, such as the one developed in this study, was not only more cost effective, it also allowed information to be gained about the process of GccF maturation.

Loop modifications

The loop region of GccF is of great interest as it contains the *O*-linked GlcNAc that is essential for activity. Figure 3.24 highlights the regions modified in the construction of $\text{GccF}_{\text{Y16F}}$, $\text{GccF}_{\text{D17N}}$, $\text{GccF}_{\text{S18C}}$ and $\text{GccF}_{\text{G16del,S18 G19insG}}$ mutants. While this image highlights the individual amino acid substitutions made, it does not take into account the possible consequences of these mutations. For example, the $\text{GccF}_{\text{S18C}}$ mutant should contain an *S*-linked GlcNAc rather than the usual *O*-linked

GlcNAc depicted here, providing this residue can still be glycosylated by GccA. It cannot be known until each GccF mutant has been purified and analysed by mass spectrometry what structural changes have occurred to the peptide. Unfortunately, due to time restraints it was not possible to isolate and purify every mutant, so assumptions were made on the basis of activity. The GccF$_{\text{G16del, S18_G19insG}}$ mutation was the only loop mutation that required the modification of more than one residue and involved the deletion of Gly15 and the insertion of another glycine residue between Ser18 and Gly19.

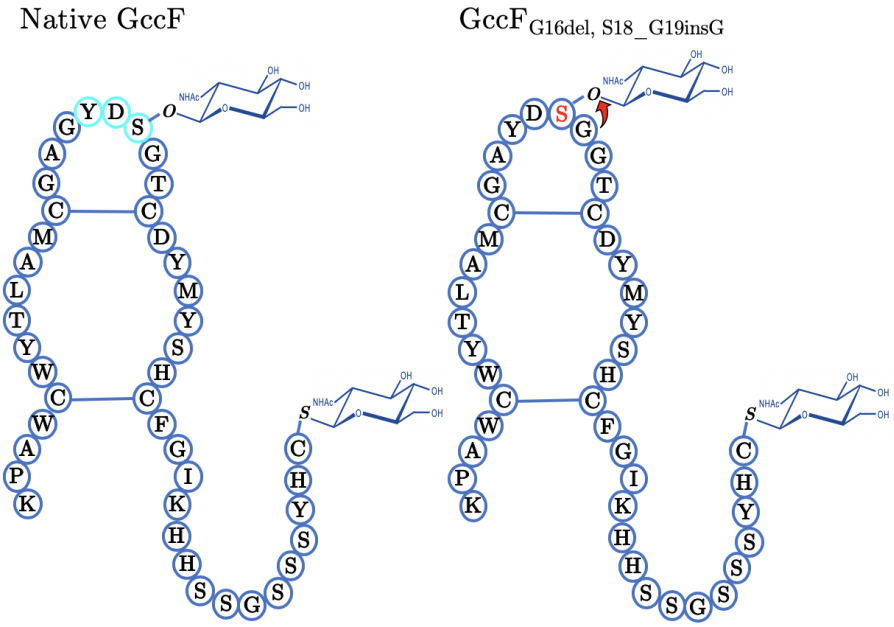


Figure 3.24: **GccF loop mutations.** Native GccF with point mutation sites coloured turquoise beside the GccF$_{\text{G16del, S18_G19insG}}$ mutation with shifted serine residue coloured red and arrow representing the change in location of the *O*-linked GlcNAc.

GccF$_{\text{G16del, S18_G19insG}}$

This mutant was created in order to see if there was any change to the activity of GccF when the relative position of the glycan was moved one residue towards the centre of the loop, retaining both the number of residues and the identity of the residues required for glycosylation within the loop region (Figure 3.24). It had

already been shown by chemically-synthesised GccF analogues that removal of one residue between Cys12 and Cys21 results in a 40-fold reduction in activity, while removal of two residues resulted in a 2000-fold reduction in activity [105]. The deletion of residues in these chemically-synthesised analogues resulted in a decrease in the overall size of the loop region. It is not unreasonable to assume that the deletion of residues within the loop region put additional strain on the disulfide bonds securing the loop while also changing the shape of the peptide exposed for recognition by the GT. While showing that the shape of the loop was important for activity, probably because of recognition or docking of GccF to the receptor, the chemically-synthesised analogues provided no information about the requirements of the GT for successful glycosylation of Ser18. The activity of the $GccF_{G16del,\ S18\_G19insG}$ mutant would also provide further information about the requirements of this glycosylation for presentation to the receptor the PTS18CBA.

Astonishingly, the $GccF_{G16del,\ S18\_G19insG}$ mutant did not show any activity when subjected to the tube activity assay (Figure 3.25). Without the mass of GccF produced from these cells, it could not be determined if there was any structural change to GccF such as the loss of the GlcNAc on Ser18. However, there are two possible explanations for this lack of activity; the first is that the serine is not glycosylated and the second is that the glycosylation proceeded as per usual but the difference in structure prevented GccF binding to its receptor. This is unlikely as the chemically-synthesised analogues retained some activity, even though the loop region would have had a larger structural difference. Another, more likely explanation, is that the activity of this mutant is outside of the detection limits of the tube assay. The tube assay can detect down to an 8-fold reduction in activity, however the most similar chemically-synthesised analogue had a 40-fold reduction in activity. These results together show that the specific location of the GlcNAc is important

and not just the size and shape of the loop. While time did not allow purification and mass spectrometric analysis, this is required to confirm that the recombinant protein is indeed glycosylated.

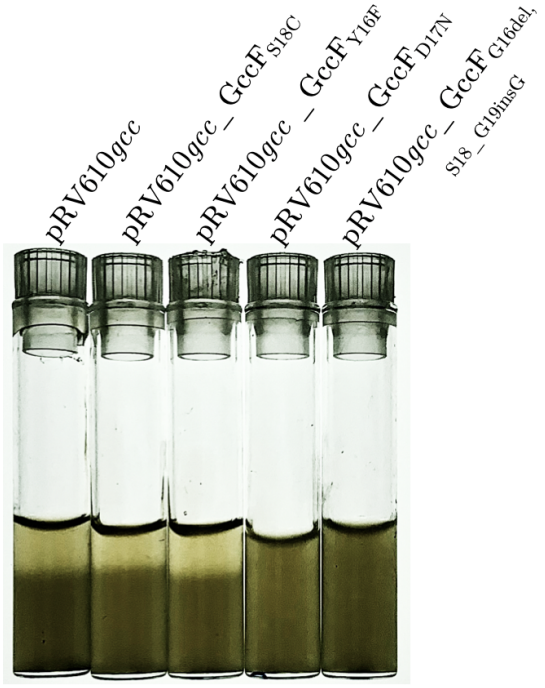


Figure 3.25: **GccF loop mutations activity assay**. 20 μL *L. plantarum* NC8 pRV610*gcc* native and mutant CFS.

GccF$_{Y16F}$, GccF$_{D17N}$ and GccF$_{S18C}$

The remaining three loop mutations were chosen to investigate the possible roles these residues play in the glycosylation of Ser18. Similar mutations to the SunA peptide did not appear to result in a significant change in activity, instead they slowed or prevented glycosylation of Cys22 [93], showing they must play some role in the glycosylation mechanism. The GccF$_{Y16F}$ and GccF$_{D17N}$ mutations were chosen to change residue functionality in the most conservative way possible. Asp17 was changed from a negatively charged residue (in the cell and once exported into the MRS growth media, pH 6.4) to a polar one. The substitution of Tyr16 for phenylalanine was chosen to maintain hydrophobicity while removing the polar hydroxyl

group. $GccF_{Y16}$ and $GccF_{D17}$ were chosen as they had been shown to be completely conserved between the putative glycosins identified through genome mining that have a $(C\text{-}X_6\text{-}C)_2$ core with an 8-residue glycosylated loop and flexible modified tail [53]. Whereas the $GccF_{S18C}$ mutation was chosen to see if the GT was capable of placing a *S*-linked glycan on this residue as it does in the tail. Interestingly, the GYDS loop sequence resembles the SYHC tail sequence where the underlined residue is glycosylated. This might be a sequence required for recognition by the glycosylase; a small residue (S or G) followed by an aromatic (Y) and thirdly a charged residue (D or H).

Of these three mutations introduced to the loop region of GccF the only one to greatly reduce or prevent activity was $GccF_{D17N}$ (Figure 3.25), indicating that the charge on the aspartic acid is required for recognition by the GT. The $GccF_{Y16F}$ mutation appeared to have no effect on GccF production (Figure 3.25) indicating that the polar side chain of the aromatic residue is not required for GT recognition or function. A follow up experiment was designed to replace this residue with glutamine, a large but non-aromatic amino acid. However the construction of this mutant was not successful due to non-specific amplification during PCR that could not be prevented by temperature optimisation.

In contrast to the chemically-synthesised analogue $GccF_{S18C}$ where the activity of GccF increased 2.5-fold upon the substitution of an *O* for an *S*-linkage [105] the CFS of *L. plantarum* NC8 pRV610*gcc*_$GccF_{S18C}$ showed a reduction in activity (Figure 3.25). This was intriguing, and to follow up a recombinant GccF was purified from a 2 L culture of *L. plantarum* NC8 pRV610*gcc*_$GccF_{S18C}$. Mass spectrometry of the $GccF_{S18C}$ produced from this culture showed the mass to be 5215.02 Da (Figure 3.26) indicating that both *S*-linked GlcNAc residues were present on the molecule. Furthermore there was no evidence of partially glycosylated peptides in

this solution indicating that this is a pure population. This reduction in activity would most likely be caused by the decreased production of GccF, probably due to reduced glycosylation efficiency. However the yield of GccF from this batch (0.5 mg/L) was within the range expected from the native producer. In order to be certain of this reduction in activity, the IC_{50} was measured and showed that the S18C mutant had at least a 10-fold reduction in activity (see appendix 6.3). This contrasts the tube activity assay, which indicates a 2-fold reduction in activity. The concentrations of purified GccF were calculated using absorption at 205 nm and although there is potential for interfering substances to create an error in the calculated concentration the results are considered more accurate than those provided by the tube activity assay. Nevertheless, it is now known that the GT can glycosylate Cys18. Further investigation is required to see if $GccF_{S18C}$ produced by this system has the same activity as the chemically-synthesised analogue.

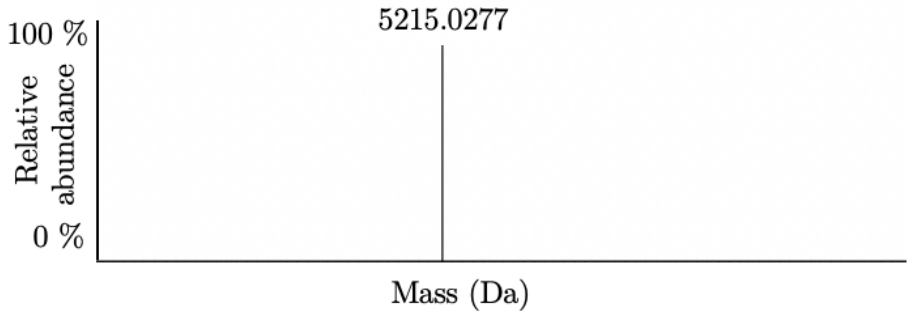


Figure 3.26: **Mass spectrum of $GccF_{S18C}$**. This was purified by CEX and RP-HPLC from *L. plantarum* NC8 pRV610*gcc*_$GccF_{S18C}$ CFS.

Tail modifications

Although the tail region is not required for GccF's activity, it does increase it dramatically and based on the results of various synthetic analogues is proposed to target GccF to its receptor on the susceptible cell membrane. Unlike the structural core of GccF, residues 32-43 are unlikely to have an effect on secondary or tertiary structure, however, increasing the length of the tail had not been tested. The tags

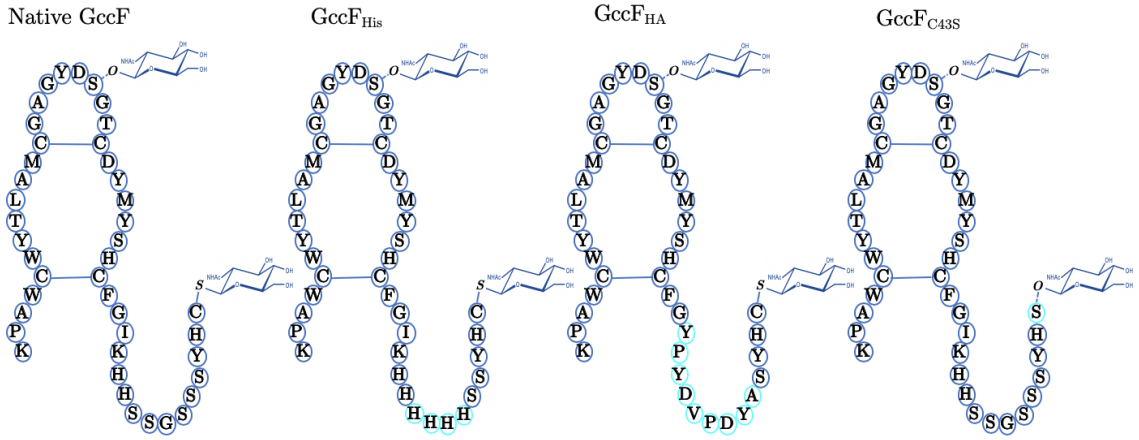


Figure 3.27: **GccF tail mutations.** Native, His-tagged, HA-tagged and C43S mutant GccF with differences coloured turquoise.

chosen were relatively small and could be used to replace existing residues so that the 43 residue length was retained, additionally the tags were placed in such a position that they were thought to minimise change in charge and residue type. His and HA tags were inserted within this region, creating the mutants GccF$_{His}$ and GccF$_{HA}$ (Figure 3.27). Providing GccF remained active following the insertion of the tags a simplified purification protocol could be developed that could potentially increase the yield of GccF but also could be used to identify GccF using the western blot technique following the construction of mutations where activity was lost in the CFS. Finally the last tail modification designed substituting Cys43 for Ser43, was made to study the functionality of GccA.

GccF$_{His}$ and GccF$_{HA}$

Interestingly the insertion of the HA tag resulted in inactive GccF according to the tube activity assay (Figure 3.28). This tag, while small, would have dramatically reduced the flexibility of the tail with the introduction of two proline and three tyrosine residues into GccF$_{HA}$. Removal of three small uncharged residues in the tail decreased the activity of the chemically-synthesised analogue approximately 30-fold in [105]. Substituting the native sequence with that of an HA tag changed the net charge of the mutated region from +3 to −2, which could have changed the affinity

of the region for the GccF target. In contrast, the addition of the His-tag did not completely abolish activity, showing that the substitution of four positive charges for four neutral charges in the tail region was not as fatal to the activity of GccF as the substitution of an inflexible negatively charged HA tag.

To see if the His-tag affected the glycosylation of GccF, a larger (1 L) culture was grown and the cell culture media split into two, allowing half the culture to be subjected to the typical purification protocol while the other half was subjected to IMAC. The pH of the CFS containing the His-tagged GccF was first adjusted to 7.4 with NaOH to neutralise the His residues and allow binding to the IMAC column. After washing (10 CV of 500 mM NaCl, 20 mM Tris-HCl, 0.05 M imidazole at pH 7.4), the protein was eluted with a linear gradient of imidazole (0.05-0.5 M). The elutent was collected in 2 ml fractions and when tested for activity using the tube activity assay, none appeared to be active. Following this, all the fractions collected after the initial wash were combined into two pools (0.05-0.25 M and 0.25-0.5 M) which were then lyophilised. After resuspension in 1 mL of H_2O both fractions showed weak activity indicating GccF was present. Interestingly, the sample collected from CEX was more active so it was used in the next purification step, RP-HPLC.

Initially the standard RP-HPLC protocol was used on $GccF_{His}$ purified by CEX, however as there was no peak at the usual place on the chromatogram, the void volume, which had been collected as one fraction was subjected to an activity assay to check if GccF had been retained on the column. As the six histidine residues would be positively charged at pH $\sim$2, the molecule would be very hydrophilic while in the standard RP-HPLC buffers. As expected, the void was active, showing that GccF was not retained on the column. To counter this effect, the pH of the mobile phase was increased to 8 using 10 mM AmBic in 2% MeCN as buffer A and 10

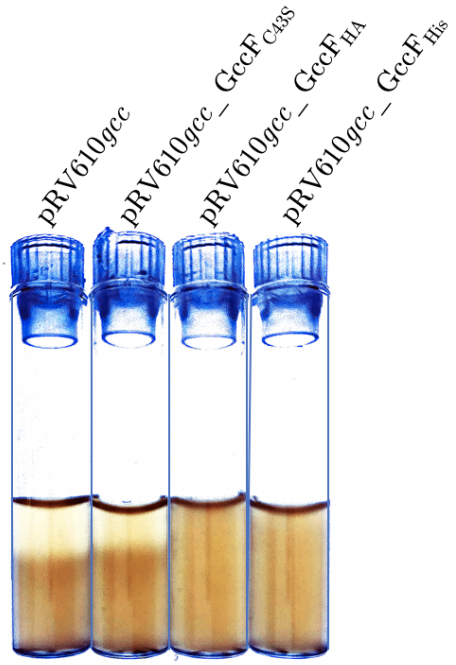


Figure 3.28: **GccF tail mutations activity assay**. 20 μL *L. plantarum* NC8 pRV610*gcc* native and mutant CFS.

mM AmBic in 70% MeCN as buffer B. At this pH the histidines should be neutral, whereas the two aspartic acids will have a negative charge and two lysines a positive charge, increasing the hydrophobicity of the peptide. A very small impure peak was eluted at a longer retention time compared to that of the wild-type peptide (Figure 3.29). The broad $GccF_{His}$ peak was collected and further analysed by mass spectrometry which showed the presence of $GccF_{His}$ along with several impurities, the mass of 5429.2 Da (Figure 3.30) shows that both glycans and the disulfides are present in $GccF_{His}$. As $GccF_{His}$ retained both glycosylations and some GccF activity, it could be used to isolate other GccF mutants. The fact that the concentration of GccF recovered was much lower than expected (less than 0.1 mg), is most likely due to losses inured in failed purification attempts. The low concentration meant that a IC_{50} for this mutant could not be measured. However, as the tube activity assay is limited to an 8-fold reduction in activity of the native producers CFS and the CFS of $GccF_{His}$ had detectable activity, it is likely that the loss is not significantly

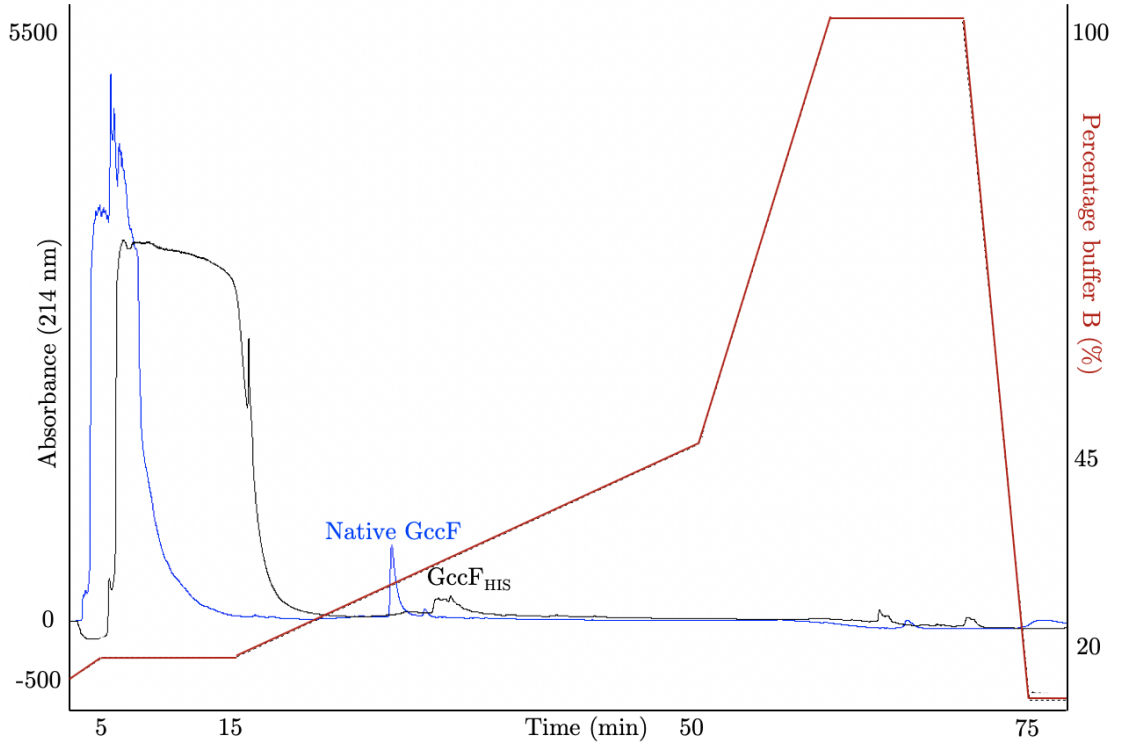


Figure 3.29: **RP-HPLC of GccF$_{His}$.** This was first purified from *L. plantarum* NC8 pRV610*gcc*_GccF$_{His}$ CFS by CEX before a modified RP-HPLC protocol was used. The GccF$_{His}$ chromatogram is overlaid with a standard GccF trace.

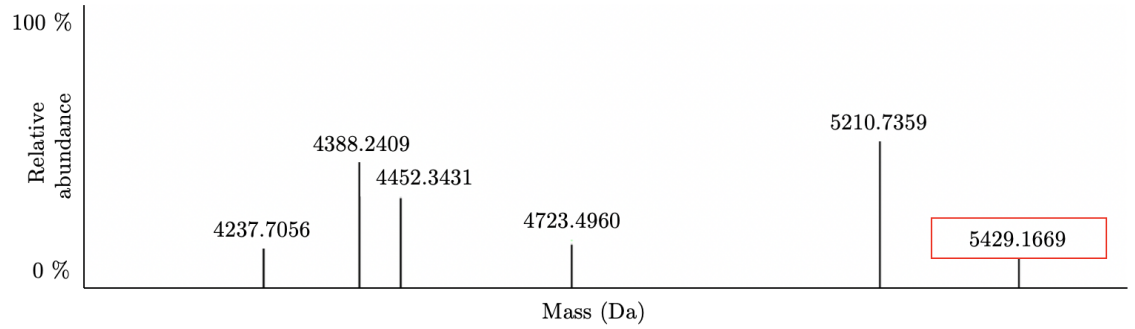


Figure 3.30: **Mass spectrum of GccF$_{His}$.** GccF$_{His}$ was purified from *L. plantarum* NC8 pRV610*gcc*_GccF$_{His}$ CFS by CEX and RP-HPLC showing a monoisotopic mass of 5429.32 Da (5027.04 (peptide mass) + (2 x 203.0794) (GlcNAc mass) + (2 x -2.01565) (disulfide bond mass) = 5429.32 Da (total protein mass)). It should be noted that this mass spectrometry was not carried out on a pure solution as indicated by the presence of other peaks.

greater than this.

GccF$_{C43S}$

Following the GccF$_{C43S}$ mutation the activity of the CFS appears to be reduced (Figure 3.28). This mutation had previously been characterised using a chemically-synthesised analogues of GccF. These analogues showed an 8-fold reduction in IC_{50} when compared to the native form, and when the sugar was lost from the serine this increased to a 70-fold reduction [105]. The CFS from *L. plantarum* NC8 pRV610*gcc*_GccF$_{C43S}$ also has activity lower than *L. plantarum* NC8 pRV610*gcc*, although this appears to be closer to a 4-fold reduction when compared to the standards (See appendix, Figure 6.1). It is clear the GccA is able to glycosylate a serine as well as a cysteine at this position. To check the activity and confirm glycosylation of GccF$_{C43S}$, it was purified using the standard protocol. Initially the chromatograph from RP-HPLC showed the presence of multiple unresolved peaks. The mass spectrometry analysis of these combined peaks confirmed that this was due to a mixed population of GccF, where some of the peptides were mono-glycosylated and others di-glycosylated (Figure 3.31).

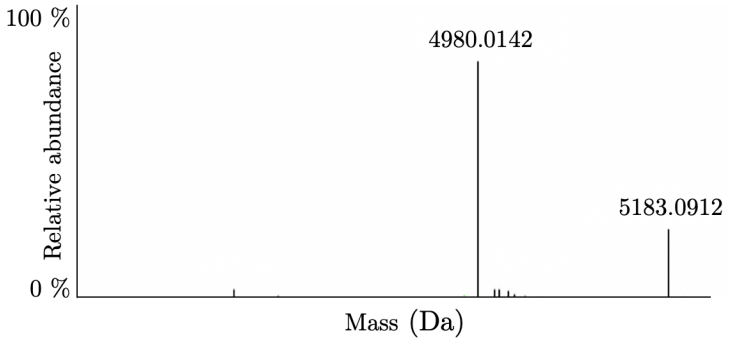


Figure 3.31: **Mass spectrum of GccF$_{C43S}$.** Recombinant GccF$_{C43S}$ was purified from *L. plantarum* NC8 pRV610*gcc*_GccF$_{C43S}$ CFS by CEX and RP-HPLC. Di-glycosylated GccF$_{C43S}$ with a mass of 5183 Da (4781.9498 (peptide mass) + (2 x 203.0794) (GlcNAc mass) + (2 x -2.01565) (disulfide bond mass) = 5184.077 Da (total protein mass)) and Mono-glycosylated GccF$_{C43S}$ with a mass of 4980 Da (5184.077 - 203.0794 (GlcNAc) = 4980.998).

To better understand the population of GccF$_{C43S}$ species, the typical elution protocol was modified by lengthening the gradient used to elute GccF (Figure 3.32).

This improved peak resolution to allow individual collection and subsequent analysis by mass spectrometry. The results showed that the first peak was di-glycosylated $GccF_{C43S}$ and the second peak mono-glycosylated $GccF_{C43S}$. The ratio of mono to di-glycosylated $GccF_{C43S}$ was calculated by integration the area of each peak from five chromatographs (Figure 3.32), to give an average ratio of 2:1.

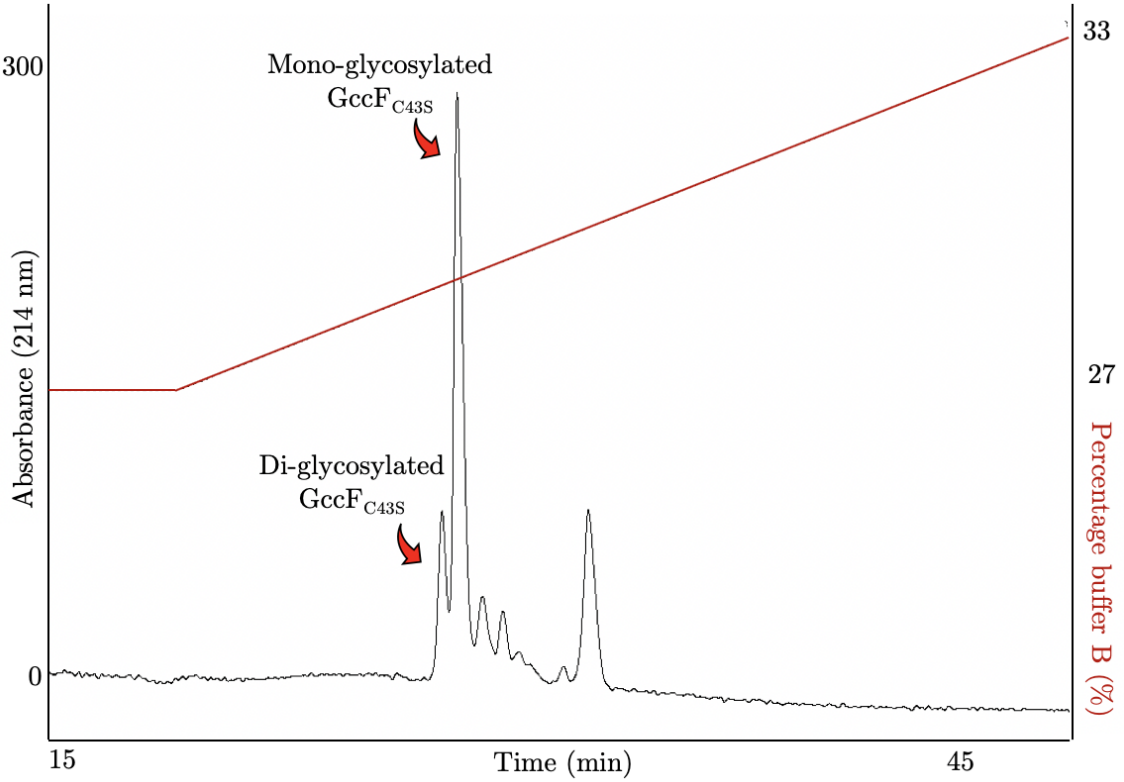


Figure 3.32: **RP-HPLC chromatogram of $GccF_{C43S}$.** $GccF_{C43S}$ was purified from *L. plantarum* NC8 pRV610*gcc*_$GccF_{C43S}$ CFS by CEX before a modified method of RP-HPLC was used to separate the peaks.

A tube activity assay carried out using the mono-glycosylated $GccF_{C43S}$ showed it to be active, strongly suggesting that the single glycan was located on the GccF loop. To confirm this, mono-glycosylated $GccF_{C43S}$ was digested with trypsin and the fragments analysed by mass spectrometry. The sample was not reduced or alkylated prior to cleavage with trypsin to maintain the disulfide bonds [55]. This allowed for a single cleavage at residue 32, resulting in the mass spectrum seen in

Figure 3.33. These results clearly showed that the GlcNAc was located on the loop and not the tail (Figure 3.34). There are two possible reasons for this observation; the first is that the less reactive serine at position 43 was not as efficiently glycosylated, resulting in twice as many mono-glycosylated compared to di-glycosylated peptides being exported. The second explanation is that GccF was glycosylated at both positions prior to export from the cell, this is likely as the *O*-glycosidic bond is more susceptible to glycosidases, which are also secreted by Lactobacillus, than the *S*-glycosidic bond [75, 128]. This de-glycosylation by glycosidases within the cell culture media has been previously proposed to be a cause of the reduction in activity exhibited by the chemically-synthesised $GccF_{C43S}$ mutant [105]. Regardless of the cause of the mono-glycosylated mutant, this experiment showed that GccA is capable of glycosylating a serine or cysteine at either location.

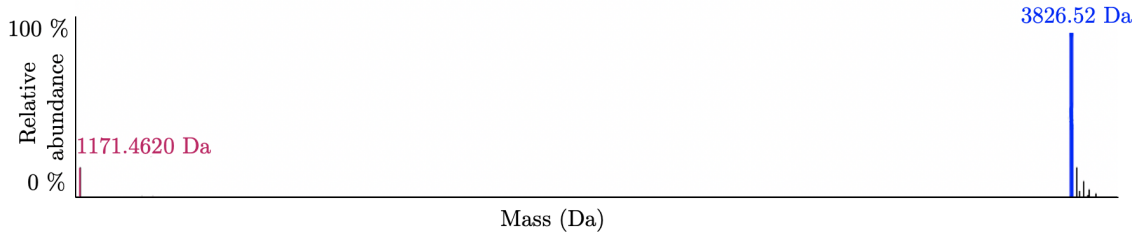


Figure 3.33: **Mass spectrum of tryptic digested mono-glycosylated $GccF_{C43S}$.** The sample was purified from *L. plantarum* NC8 pRV610*gcc* $GccF_{C43S}$ CFS by CEX and RP-HPLC, then digested with trypsin without reduction or alkylation revealing masses shown that match monoisotopic peptide masses and identify GlcNAc on loop (KPAWCWYTLAMCGAGYDS-GTCDYMYSHCFGIK 3628.4961 (peptide mass) + (2 x -2.01565) (disulfide bond mass) + (1 x 203.0794) (GlcNAc mass) = 3827.544 Da) but absent from the tail (HHSSGSSSYHS 1172.4715 Da (peptide mass)).

3.2.3 GccE

Homology studies of the LytTR DNA binding domain of GccE suggest that this protein is somehow involved in the regulation of GccF production. However the N-terminal domain has no similarity to any protein in the publicly available databases except for that of AsmE, the putative regulator found within the ASM1

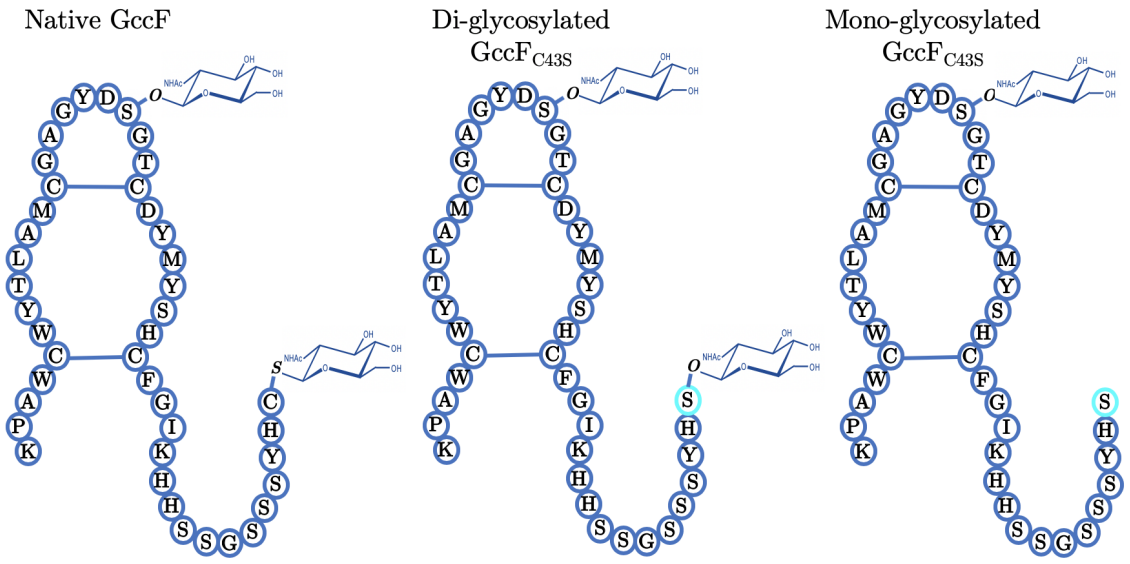


Figure 3.34: **Illustration of $GccF_{C43S}$ structures.** The native GccF structure compared to the two observed $GccF_{C43S}$ structures.

gene cluster. This suggests that the N-terminal domain of GccE has a novel structure and an unknown function. In order to investigate the function of GccE in the production of GccF by the *L. plantarum* NC8 expression system, mutations were made to the *gccE* sequence. Two different point mutations were chosen, the first of these $GccE_{L19X}$ introduced a stop codon near the start of the ORF to create a severely truncated and non-functional protein. The result of this mutation was a reduction, but not a complete loss of activity (Figure 3.35).

A second mutation $GccE_{L148X}$ was designed to retain the N-terminal region while removing the LytTR DNA-binding domain. Surprisingly, the $GccE_{L148X}$ mutation dramatically decreased the production of active GccF (Figure 3.35). As the coding sequence of *gccD* overlaps with the first 31 bp of *gccE* and the promotor for *gccF* extends into the 3' end of *gccE* it was not possible to completely delete *gccE* without extensive modifications being made to the plasmid. While these two mutations were not expected to have different consequences, this can be explained by the use of an alternative translation initiation site within GccE downstream of the L19X

Figure 3.35: **GccE mutations activity assay.** Using 20 μL *L. plantarum* NC8 pRV610*gcc* native and mutant CFS.

mutation allowing a low level of production of an almost full length and somewhat active GccE. This kind of initiation following a stop codon has been reported and can result in functional protein [129]. Translation could be re-initiated at methionine 36 retaining the entire LytTR domain and much of the uncharacterised domain. In contrast, the L148X mutation resulted in the separation of the two domains of GccE so that even if translation was re-initiated following this mutation the effect would have been much greater as observed.

To confirm that the loss in activity seen in these mutations was due to decreased production and not structural modification of GccF, mass spectrometry was used to analyse GccF purified from *L. plantarum* NC8 pRV610*gcc*_GccE$_{L148X}$. This showed the intact mass of GccF to contain both disulfide bonds and both *O* and *S*-linked GlcNAc (Figure 3.36). The results strongly suggest that GccE is a tran-

scriptional activator of the GccF promotor responsible for the production of the high levels of GccF produced in the native strain and is not otherwise involved in GccF maturation. The low levels of active GccF purified following the L148X mutation can be attributed to the transcriptional read through from the promotor located upstream of *gccA* that has been shown to be responsible for the transcription of *gccABCDEF* [78].

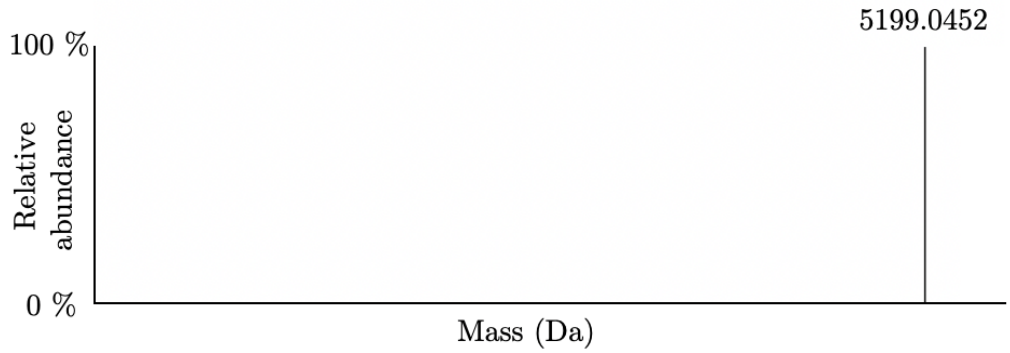
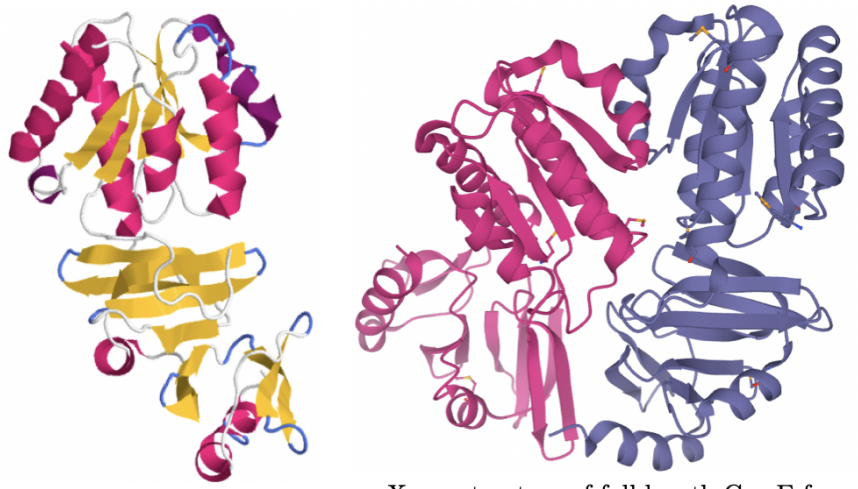


Figure 3.36: **Mass spectrum of GccF purified from *L. plantarum* NC8 pRV610*gcc*_GccE$_{\mathbf{L148X}}$** showing the same intact mass as the wild-type GccF.

In an attempt to predict GccE's function its structure was modelled using the protein modelling software phyre [130] (Figure 3.37). This modelled 94% of the residues with 90% confidence and predicted that the structure that GccE has the most similarity to is the response regulator ComE [88]. ComE interacts as a dimer where each chain binds one of the direct repeats within the DNA. It is therefore likely that GccE also functions as a dimer to bind the repeats upstream of *gccF*.

To better visualize this structure, GccE was subjected to spring on-line [131] which produced a dimeric model with 69% coverage (Figure 3.38). Within the modelled regions, there was 20% identity to the ComE dimeric structure. Interestingly phosphorylation of D58 activates ComE by inducing dimerisation [88], this residue is conserved in the protein alignments produced by spring on-line 6.4. Based on the



Predicted structure of GccE

X-ray structure of full-length ComE from Streptococcus pneumoniae

Figure 3.37: **Predicted structure of GccE obtained from phyre intensive mode [130] and X-ray structure of ComE.**

structure predicted by this software, D73 of GccE is in a similar location to D58 of ComE and is therefore the most likely target of phosphorylation. If phosphorylation at this site is sufficient to facilitate activation of GccE, then it may be possible to trap GccE in the active state by introducing the substitution D73E which would mimic the phosphorylation state as exhibited by the $ComE_{D58E}$ mutant [88]. The $GccE_{L148X}$ mutation provided evidence that GccE is an activator of *gccF* transcription, hence a phosphorylation mimic may be capable of producing larger quantities of GccF protein. If GccE is actually a repressor of transcription then simply replacing the phosphorylated residue with another which cannot be phosphorylated should be sufficient to prevent dimerisation and therefore activation of the protein.

ComE is part of a two component signalling system and relies on a membrane bound histidine kinase to respond to a stimulus, auto-phosphorylate and subsequently transphosphorylate ComE [88]. It is likely that GccE will need to be phosphorylated in order to be activated, however unlike many other two component signalling systems, some of which regulate bacteriocin production [109], there appears

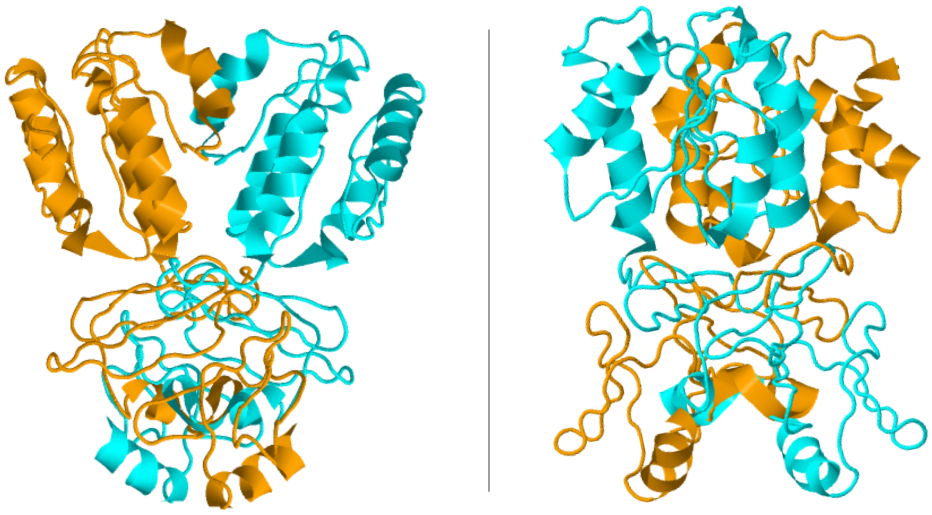


Figure 3.38: **GccE dimeric structure predicted by spring on-line** [131].

to be no histidine kinase coding sequence within the *gcc* gene cluster. Although *L. plantarum* KW30 is not yet fully sequenced, *L. plantarum* NC8 pRV610*gcc* produced the same concentration of active GccF as the wild-type system so it can be assumed gene expression is regulated in the same way. For this reason a BLASTP search with the query protein ComD, the histidine kinase responsible for activating ComE, was run on the *L. plantarum* NC8 genome. This search identified two proteins with significant sequence similarity, a putative histidine kinase (accession: CAA75398) and a two-component system histidine protein kinase (accession: EHS84485). These findings support the possibility that GccE is activated by histidine kinase dependent phosphorylation but cannot preclude the possibility that a completely different mechanism is responsible for the dimerisation and activation of GccE.

In order to confirm GccE's role as a transcriptional activator of the *gccF* promotor, a reporter gene assay was designed. The reporter gene, aminopeptidase N (PepN), was selected as it was contained within the pSIP412 plasmid that was

already transformed into *L. plantarum* NC8. A simple preliminary aminopeptidase activity assay was carried out using *L*-lysine *p*-nitroanilide as the substrate. This clearly showed that plasmid-less *L. plantarum* NC8 cells did not have any ability to hydrolyse *L*-lysine *p*-nitroanilide when compared to the transformed and induced *L. plantarum* NC8 pSIP412 cells (Figure 3.39). Following the conformation that this activity assay could be used within *L. plantarum* NC8, primers were designed to place the 2.5-kbp *pepN* gene in place of *gccF* within the pRV610*gcc* plasmid. Such a system would not only provide conformation that GccE is a regulatory protein, but could also be used to check if GccF can function as an self-inducing peptide. Other bacteriocins have been shown to induce their own expression [109] and the read through of the *gccA-F* promotor [78] may provide levels of GccF high enough for self-induction.

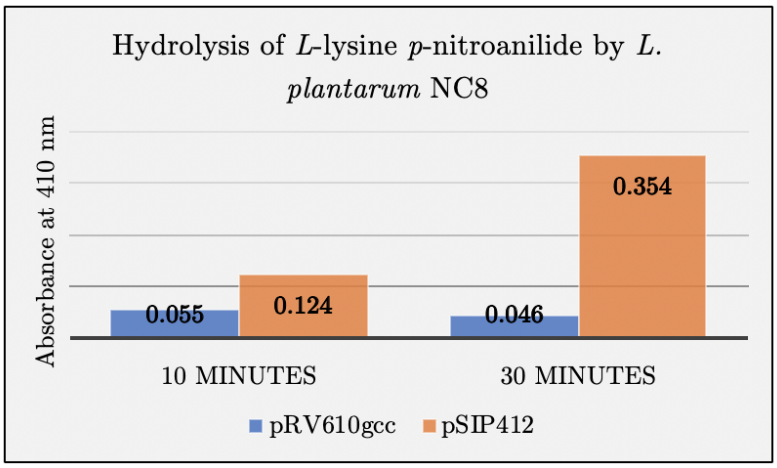


Figure 3.39: **Preliminary PepN assay.** Hydrolysis of *L*-lysine *p*-nitroanilide by cell lysate followed by change in absorbance at 410 nm revealing that there is a large increase in the hydrolysis of substrate when PepN is produced (*L. plantarum* NC8 pSIP412 cells).

Unfortunately the construction of this reporter gene vector was not successful. The process of constructing this plasmid involved amplifying both the insert (*pepN*) and the entire pRV610*gcc* plasmid without the *gccF* gene. PCR was used so that complementary restriction sites could be added to both the vector and insert, but

due to the large size of these combined sequences (13.5-kbp) there was little choice of restriction enzymes that would not cut within this plasmid. One of the restriction sites used was blunt-ended (*Msc*I) which is known to reduce ligation efficiency [132]. Ligation was attempted a total of eight times. Each attempt involved modifying at least one part of the procedure such as using new restriction enzymes, different ratios of vector to insert, different ligation concentrations, different DNA purification procedures following PCR and a different brand of ligase. Despite the many attempts, none where successful. It was decided that construction of this vector in this manner was unlikely to be successful and instead Gibson assembly [133] should be tried, but unfortunately time did not permit this.

3.2.4 GccA

The cluster GT, GccA, has a high level of identity to other glycocin GTs making prediction of the residues required for activity possible. This allowed mutations to be made to residues required for activity, producing an inactive protein rather than a truncated protein. In contrast to the stop codon mutations previously discussed that could allow an alternative start codon to be used resulting in a low level of expression of a functional protein to remain, this type of mutation is designed to prevent all enzymatic activity. Similar mutations could not be introduced to the remaining proteins because their function was uncertain, making it impossible to predict active site residues. In order deactivate GccA, two point mutations were introduced into the DxD motif, creating the mutant $\text{GccA}_{\text{D123N,D125N}}$. Figure 3.40 shows that this mutation resulted in a complete loss of active GccF production, suggesting that at least Ser18 had not been glycosylated.

To confirm the loss of activity in the $\text{GccA}_{\text{D123N,D125N}}$ mutant CFS was due to the loss of glycans on GccF, the resulting protein was purified and subjected to

Figure 3.40: **GccA mutation activity assay.** GccF activity assay with 40 μL of *L. plantarum* NC8 pRV610*gcc* native and mutant CFS.

analysis by mass spectrometry. Because it was uncertain if the GlcNAc on Cys43 was present, another already characterised mutant $GccF_{C43S}$ was used in this construct to avoid the possibility of a free cysteine causing protein aggregation. As it has already been shown that Ser43 could be glycosylated, the loss of a glycan at either position would show that GccA was responsible for that glycosylation. Mass spectrometry (Figure 3.41) confirmed the presence of both disulfide bonds and the absence of both GlcNAcs, showing that GccA is responsible for the glycosylation of Ser18 and Cys43 of GccF *in vivo*. The secondary structure of this mutant was analysed using circular dichroism, which showed the presence of secondary structure similar to what was seen for de-glycosylated GccF (Figure 3.42) confirming that the helical secondary structure was maintained despite the loss of GlcNAc [66]. The presence of unglycosylated GccF in the CFS was surprising as the model for maturation predicted the GccF peptide should be glycosylated within the cytosol of the cell

prior to export [78]. Finding unglycosylated GccF outside of the cell proves that the glycans are not required for export. Bioinformatic analysis of GccA did not detect any secretion signal or transmembrane domains suggesting it is unlikely to be exported or to be an external membrane protein, strongly suggesting it is a cytosolic protein. On this basis glycosylation is most likely to have occurred before export, although this remains to be verified.

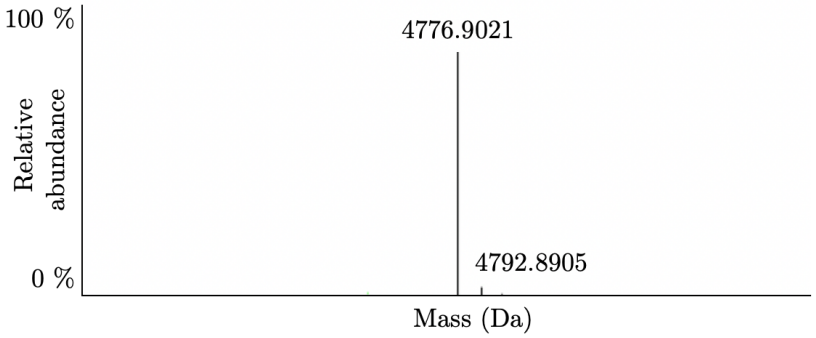


Figure 3.41: **Mass spectrum of unglycosylated $\mathbf{GccF_{C43S}}$.** The product was purified from *L. plantarum* NC8 pRV610*gcc*_$GccA_{D123N,D125N}$_$GccF_{C43S}$ CFS by CEX and RP-HPLC. The mass of 4776.9021 indicates disulfide bonds, but no GlcNAc present (4780.94 (peptide mass) + (2 x -2.01565) (disulfide bond mass) = 4776.909 Da).

Following the confirmation that GccA was the enzyme responsible for both *S* and *O*-linked glycosylations of GccF, other mutations were made to probe the structural features of GccF necessary for glycosylation to occur. From the $GccF_{Y16F}$, $GccF_{D17N}$, $GccF_{S18C}$ and $GccF_{C43S}$ mutants it was already known that this enzyme could be inhibited by the loss of the charged negatively residue D17. It was also shown that GccA could glycosylate either serine or cysteine at position 18 and 43 (allowing production of $GccF_{S18,\ C43}$, $GccF_{S18,C43S}$ and $GccF_{S18C,C43}$). To complete this set of mutations $GccF_{S18C,C43S}$ was constructed and the recombinant protein shown to be active in a tube activity assay (Figure 3.40). Unsurprisingly, $GccF_{S18C,C43S}$ had lower activity than both $GccF_{S18C}$ and $GccF_{C43S}$. It is possible that the reduced activity is due to a lower glycosylation efficiency which was also suspected following

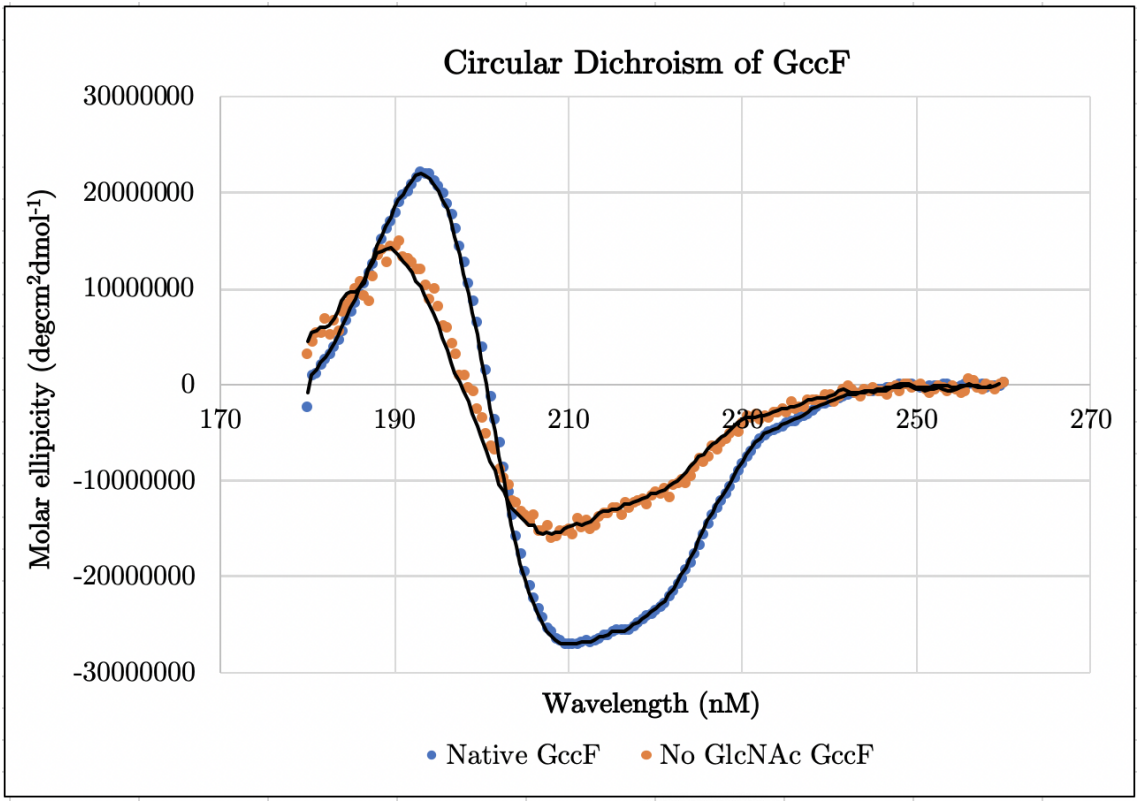


Figure 3.42: **Circular dichroism of GccF.** GccF produced by *L. plantarum* NC8 pRV610*gcc* in blue and from *L. plantarum* NC8 pRV610*gcc*$^{-}$GccA$_{D123N,D125N}$$^{-}$GccF$_{C43S}$ in orange.

the GccF$_{C43S}$ mutation, alternatively these mutations may have an effect on the IC$_{50}$ of GccF. To confirm the cause of the reduction in activity seen in the tube activity assay, (Figure 3.40) GccF$_{S18C,C43S}$ needs to be purified and the IC$_{50}$ calculated.

3.2.5 GccB

GccB is a large protein with multiple domains making the presence of internal translation initiation codons much more likely. Rather than introducing a mutation into the transporter itself, a mutation was introduced to the GccF pre-peptide sequence that was designed to prevent export of GccF. Bioinformatic analysis strongly suggests GccB is an ASM protein containing a C39 cysteine protease domain that recognises the double-glycine motif within the GccF leader peptide as a cleavage site. To test this hypothesis, the glycine residues were mutated to threonines, which

are not known to act as a cleavage signal (Figure 3.43. As expected, this mutation resulted in the complete loss of GccF activity in the CFS (Figure 3.44), indicating that the double-glycine is motif is indeed the cleavage site and without this GccF is not exported out of the cell. Presumably this mutation has trapped GccF within the cytosol of the cell as seen in similar experiments with other bacteriocins [99]. Purification of GccF from the CFS was not attempted, as the addition of the leader peptide would have altered the properties of GccF preventing it from being purified using the standard protocol. As it was unknown whether the pre-peptide would be glycosylated and retain activity, there would be no easy way to follow the purification. One solution would be to use the His-tagged GccF mutant discussed earlier in this chapter. Unfortunately purification using this tag was not successful and requires further optimisation.

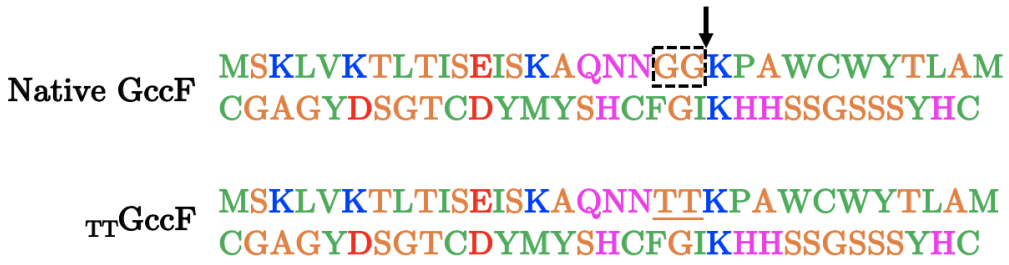


Figure 3.43: **Amino acid sequence of native pre-GccF and pre-$_{TT}$GccF**. An arrow points to the cleavage site that is recognized by the ABC transporter in the native protein. The mutated residues are underlined in the pre-$_{TT}$GccF sequence.

3.2.6 GccC and GccD

As the cytosol of any cell is highly reducing it is unlikely the disulfide bonds of GccF are formed prior to export. For this reason it is thought that GccF is exported unfolded following the cleavage of the leader peptide. Upon export the process of disulfide bond formation is thought to be facilitated by the thioredoxin-like proteins GccC and GccD which are most likely located on the exterior of the cell [78]. The CxxC motif is a requirement for the activity of most thirodoxin-

Figure 3.44: **$_{TT}$GccF activity assay.** 20 μL *L. plantarum* NC8 pRV610*gcc* native and mutant CFS.

like proteins, and although GccC and GccD have domains containing this motif the possibility that these proteins may be involved in another process could not be eliminated on the basis of sequence alone. Subsequently, premature stop codons were placed near the beginning of their sequences. The $GccC_{K3X}$ mutation count not be rescued by the use of an alternative start codon as analysis of this sequence with Open Reading Frame finder on the National Centre for Biotechnology Information (NCBI) website indicated that there were no alternative start codons in the correct frame. In contrast, the $GccD_{L4X}$ mutation could be rescued at residue 55, which is predicted to be part of a transmembrane domain. If alternative initiation was to occur then the CxxC motif required for activity would be retained, however, without a transmembrane domain it should not be able to function as an external thioredoxin.

The thioredoxin mutations were introduced into the pRV610*gcc* constructs both individually and in tandem (Figure 3.45). Unsurprisingly all of these mutants were still able to produce active GccF as it had been reported that neither of the

SunA cluster thioredoxins were required to produce active SunA, providing other thioredoxins were produced by the cell [102]. Bioinformatic analysis of *L. plantarum* genome identified the presence of six thioredoxins TrxA1, TrxA2, TrxA3, TrxB, TrxH as well as an unnamed protein with a thioredoxin-like sequence. It is therefore likely that one or more of these proteins is able to rescue the activity of GccF following the inactivation of the cluster thioredoxins. To check that the structure of GccF produced from the cluster containing these mutations remained unchanged, GccF was purified from *L. plantarum* NC8 pRV610*gcc*_GccC$_{K3X}$. While CEX was successful, RP-HPLC failed to produce a peak in the usual location (approximately 36% buffer B)(Figure 3.46). Instead, a new peak eluted much later at 98% acetonitrile, indicative of a much more hydrophobic sample. While this peak was slightly active, most of the activity was found in the fraction collected during the initial wash step. Unfortunately, these collected samples were not pure or concentrated enough for analysis by mass spectroscopy. Nevertheless, the different elution times suggest the presence of a number of different GccF structures. It is known from chemically-synthesised GccF analogues that the absence of disulfide bonds causes a decrease in activity that would be impossible to detect using the tube assay [105]. Thus, to interpret the results of this mutation fully, mass analysis is required.

While the consequences of the mutations to GccC and GccD on GccFs structure are unknown, it can be assumed that the peptide retains the loop glycosylation required for activity. Additionally, this study indicates that two genes, *gccC* and *gccD* are somewhat expendable for the production of active GccF within this expression system, as a significant level of activity is retained within the CFS following these mutations (Figure 3.45). In order to see if the activity of the CFS could be increased following exposure to cells that produce wild-type GccC and GccD, a

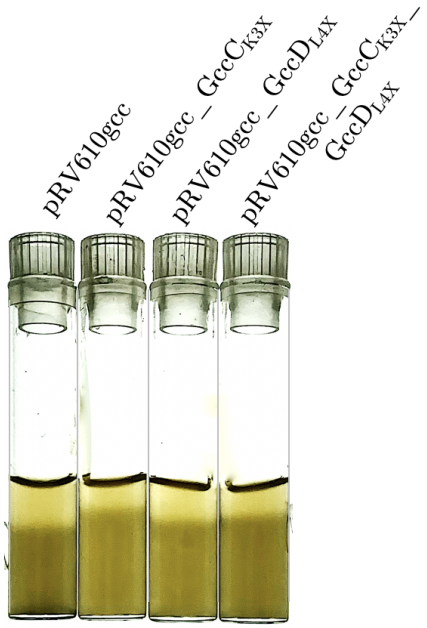


Figure 3.45: **Thioredoxin mutants GccF activity assay.** 20 μL of *L. plantarum* NC8 pRV610*gcc* native and mutant CFS was used.

complementation-like assay was set up. If GccF was incorrectly folded within the CFS following the thioredoxin mutations it may be able to be refolded by wild-type thioredoxin proteins found on the exterior of the non-GccF producing *L. plantarum* NC8 pRV610*gcc*$_{\text{TT}}$GccF cells. In this experiment CFS was harvested from *L. plantarum* NC8 pRV610*gcc* along with that from the GccC$_{\text{K3X}}$ and GccD$_{\text{L4X}}$ mutants. Each CFS was then used to wash a cell pellet from a 2 mL culture of *L. plantarum* NC8 pRV610*gcc*$_{\text{TT}}$GccF three times before resuspending the cells in 50 μL of the same CFS. Once resuspended the cells were left shaking for 24 hours before this CFS activity was compared to that of the CFS that had not been exposed to the $_{\text{TT}}$GccF cells with wild-type thioredoxins. The decrease in activity observed when active CFS was exposed to GccF immune $_{\text{TT}}$GccF cells for 24 hrs was unexpected (Figure 3.47). On reflection this is could be due to the action of secreted proteases or glycosidases. Or, as the activity appears to be decreasing at the same rate in all the samples, it is could be that the thioredoxins found on the surface of *L. plantarum* NC8

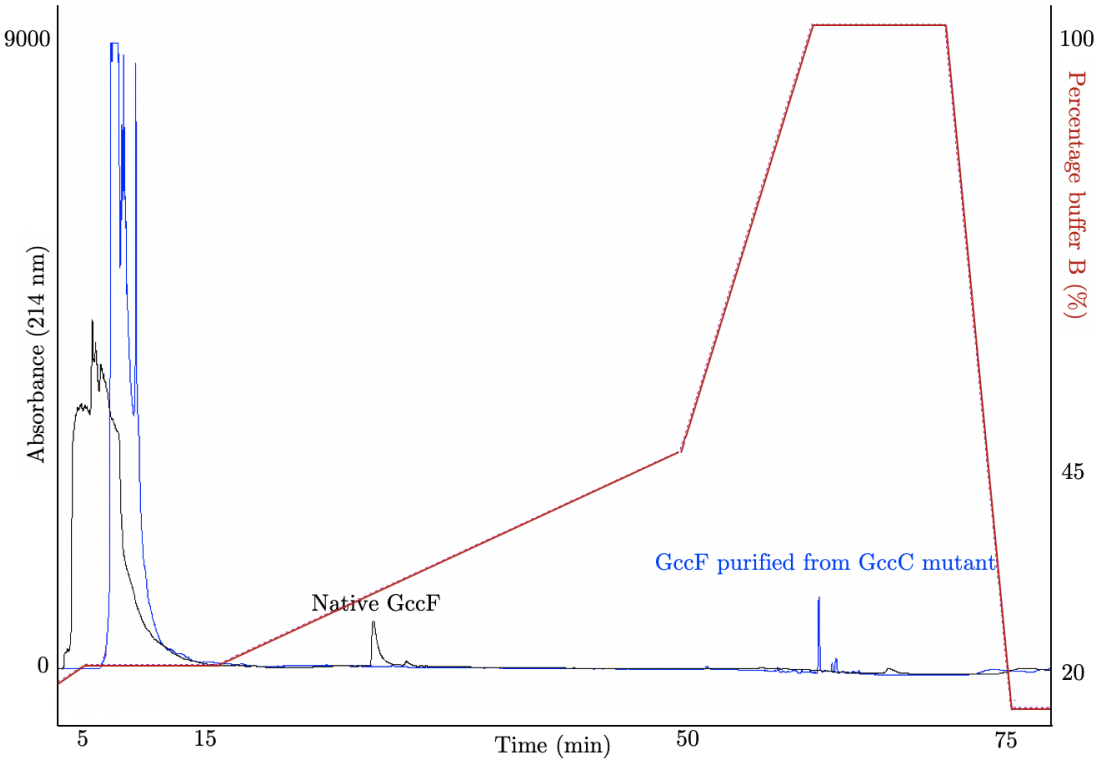


Figure 3.46: **RP-HPLC of GccF following GccC$_{K3X}$ mutation.** CEX purified GccF (blue) from *L. plantarum* NC8 pRV610*gcc*_GccC$_{K3X}$ is overlaid with a native GccF trace (black).

pRV610*gcc*$_{-TT}$GccF are unable to refold any miss-folded GccF contained within the thioredoxin mutant CFS, or alternatively the rate of refolding may be to inefficient to detect. While this was not an expected finding it is interesting as it is the first evidence that GccF may be hydrolysed or absorbed by *Lactobacillus* cells that have immunity to GccF.

3.2.7 GccH

Modifying GccH, the immunity protein within the cluster, to better understand its role was not possible in this expression system. This is because the loss of immunity to GccF in a producing stain would be catastrophic to the cells. If a mutation inactivating GccH was introduced while the cells were producing GccF it would prevent cell growth, making it impossible to tell the difference between an

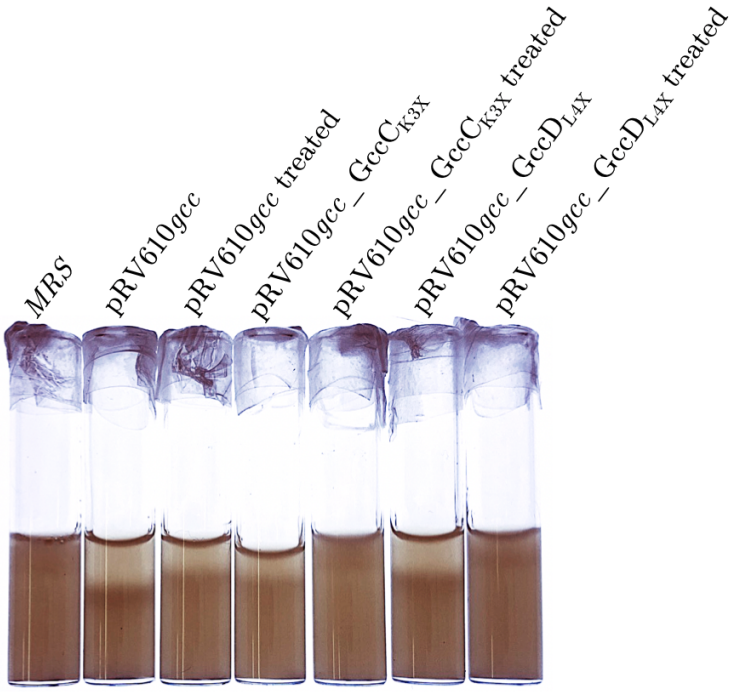


Figure 3.47: **Thioredoxin mutants complementation activity assay.** 20 μL of *L. plantarum* NC8 pRV610*gcc* native and mutant CFS was layered on to media containing the indicator strain. Treated refers to CFS that had been incubated for 24 hours with cells thought to produce native GccC and GccD but that were not sensitive to GccF although incapable of secreting GccF into the culture media.

unsuccessful transformation and an inactive immunity protein. If mutations were to be made to this protein then they would need to occur in a non-GccF producing system like *L. plantarum* NC8 pRV613*gccH* that was used to identify GccH as the immunity protein[70]. Mutations could then be introduced with no obvious phenotypic change to the cells until they were exposed to GccF allowing standard transformation and activity assays to be carried out. However as this system was not useful for analysis of the production of GccF, introducing mutations to GccH was outside of the scope of this project.

3.3 GccF mechanism studies

Following an experiment where the growth of *L. plantarum* ATCC 8014 was monitored in the presence and absence of GccF, two cultures were placed into the fridge in case further analysis was required. They remained there for several weeks and when observed, a much larger cell pellet was seen in the culture that had not been treated with GccF. This was unexpected as the cultures had a similar cell density prior to refrigeration. Although the cell density at the time the cells were placed into the refrigerator was unknown, those that had been treated with GccF had an OD_{600} of approximately 1.0 after several weeks had passed. As GccF is bacteriostatic, it was plausible that this was the OD_{600} of the cells at the time they were placed in the fridge. The accidental observation sparked an interest into the cause of the difference in cell density. Had the untreated cells had continued to grow at 4°C or had the cells treated with GccF lysed under these conditions?

To answer this question a fresh culture was grown until an OD_{600} of 1.0 was reached. It was then split into two 10 mL samples, one treated with 200 μmol of GccF, and left at 4°C. Table 3.3 shows how the OD_{600} of the culture treated with GccF remained constant while the density of the untreated cells increased at 4°C. Additional to the OD_{600} measurements, 100 μL of a $1x10^6$ dilution of cells were plated at certain time points allowing colonies to be counted the following day. This was considered to be a measure of the number of viable cells within the culture. The untreated cells had an increase in the number of colony-forming units that could be explained by an increase in cell density. However, the GccF-treated sample showed a dramatic decrease in colony-forming units over the week long experiment, this was unexpected as the cell density remained constant.

Table 3.3: GccF treatment of cells at 4°C

Time	**Untreated OD_{600}**	**Colonies untreated**	**Treated OD_{600}**	**Colonies treated**
0 hr	0.971	399	0.971	399
1 hr	1.817		1.011	
16 hr	2.062		1.011	
24 hr	2.267	TMTC	1.053	782
48 hr	2.528	TMTC	0.996	468
72 hr	2.677	TMTC	1.005	70
1 week	2.901	TMTC	0.986	6
2 weeks	3.504	TMTC	0.989	0

The number of colonies were counted following plating of 100 μL of a 1x10^6 dilution of cells.

TMTC - To many to count.

This assay was repeated, but only 1 hr and 1 week measurements were taken, showing the same decrease in colony-forming units from 420 to 2. Interestingly, Table 3.3 shows that after 24 hours the colony-forming units for culture treated with GccF had doubled even though the OD_{600} was indicative that the culture was in stasis. The decrease in GccF concentration in the media by serial dilution prior to plating appeared to effectively relieve stasis after the 24 hour incubation and doubled the number of colony-forming units within the culture. After treatment for longer periods, it appeared that this stasis could not be effectively relieved by dilution alone. Out of interest the cells were stained and observed under a microscope to see if there were any differences in the morphology of the bacteria following treatment (Figure 3.48). The images show that the untreated cells were actively replicating unlike the cells treated with GccF. Over 400 colonies were counted in both treated and untreated samples to show that 93% were in the process of dividing in the untreated sample but only 9% in the GccF-treated sample. This would suggest that

the addition of GccF to the media was able to slow or prevent cell division.

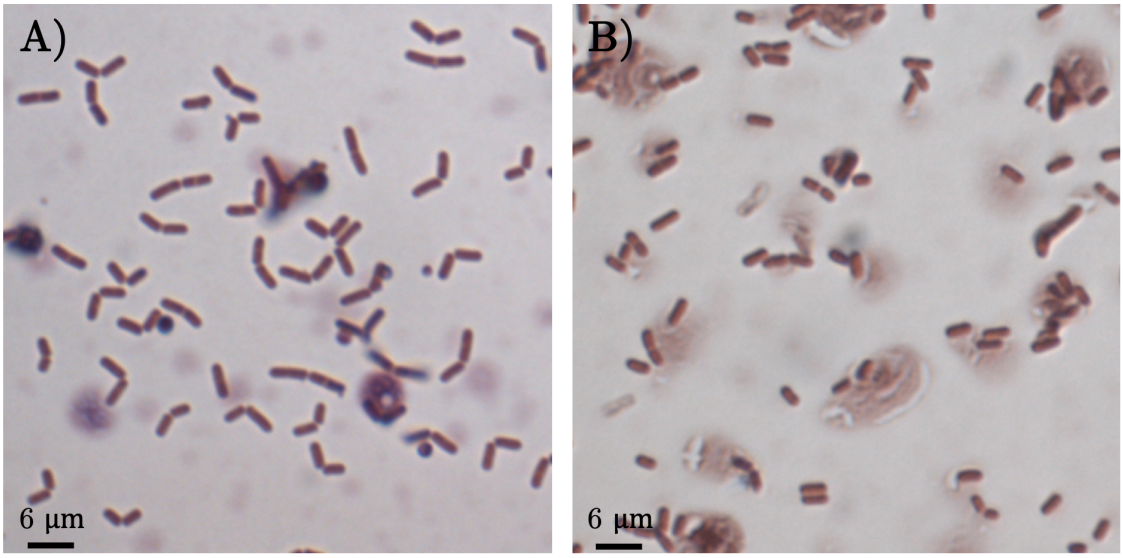


Figure 3.48: **GccF treatment of cells at 4°C.** Two week old culture of *L. plantarum* ATCC 8014 A) untreated B) treated with GccF.

To check if this result was temperature-dependent, the growth assay was to be repeated at room temperature. The same starting OD_{600} of 1.0 and the same concentration of GccF were used. After 1 hour the OD_{600} of the untreated cells had more than doubled to an OD_{600} of 2.222 while the treated cells had an OD_{600} of 1.086 indicating the addition of GccF had placed these cells in stasis. However it was obvious by 16 hrs these cells were no longer in stasis as the OD_{600} of both cultures had surpassed 3.0. This would suggest that either the cells treated with GccF had sufficient time to metabolise it or resistance had developed. Regardless this assay was no longer suitable to observe the effect of GccF. Maintaining the temperature at 4°C allowed GccF to sustain bacteriostasis much longer than at room temperature which is likely due to the slower metabolism of the cells.

As previous studies have proven that bacteriostasis induced by GccF can be relieved by the addition of GlcNAc [55] the experiment was repeated at 4°C to test if GlcNAc could relieve prolonged stasis. Following a 1 week incubation at 4°C all the

cells were brought to 30°C for 1 hour. Prior to the 30°C incubation, half of the cells that had been treated with GccF were pelleted and resuspended in media containing GlcNAc. Once plated it could be seen that following this week long incubation over 800 colonies grew if no GccF was added to the media, 52 if it was and 286 if GccF was washed off with GlcNAc prior to plating. This indicated that while the number of viable cells appeared to be decreasing following GccF treatment that this was possibly not the case as they could be at least partially recovered by exposure to GlcNAc and higher temperatures indicating that instead, these cells were in stasis. This shows that simply categorising a bacteriocin as bacteriostatic or bactericidal should not be dependent on plating alone.

4 Conclusions and Future Directions

The only glycocin gene clusters that have proteins other than the bacteriocin itself functionality annotated are the sublancin (*sun*) cluster and the thurandacin (*thu*) cluster. From these, the most functionally characterised proteins are the glycosyltransferases which were cloned and the resulting recombinant enzymes analysed *in vitro*. Before this study, the *gcc* cluster had only been bioinformatically characterised, despite many attempts to clone individual *gcc* genes and express recombinant proteins heterologously or modify them in the native host. In this study, a plasmid-based heterologous expression system was constructed that enabled gene function to be validated as well as the requirements for GccF glycosylation to be tested. Initially, expression of the *gcc* genes was attempted in *E. coli* but was not successful. When the seven cluster genes were transferred into a plasmid, pRV610*gcc*, that was compatible with *Lactobacillus* cells, namely *L. plantarum* NC8 and *L. sakei* 790, an antimicrobial compound was secreted. This was shown to be structurally and functionally identical to wild-type GccF leading to the conclusion that all the proteins required for GccF expression and maturation must be present in the cluster or be part of the cellular machinery of these three *Lactobacillus* species. To test this, pRV610*gcc* was transformed into *E. faecalis* JH2.2 cells but they did not produce active GccF, confirming that there is a difference in cellular machinery between *Lactobacillus* and *Enterococcus* that is required for the transcription of the *gcc* cluster. Following on from these experiments other *Lactobacillales* such as *Lactococcus* and *Streptococcus* could also be transformed with this plasmid to narrow down the requirements for production of the Gcc proteins.

The pRV610*gcc* vector will provide a tool to test the ability of the various Gcc proteins to synthesise other glycocins. For example, the *gcc* cluster should be able to produce ASM1 identical to the wild-type peptide, as there are only five residue

differences between ASM1 and GccF. Although ASM1 was expected to have an identical structure to GccF, once purified from its native producer *L. plantarum* A-1, ASM1 was shown to have a one-turn helix on the C-terminal side of the loop, in contrast GccF which has two α-helices as shown in Figure 4.1 [53]. To investigate the cause of the difference in three-dimensional structure of these two glycocins and the functionality of the *gcc* genes, *asm1* could be inserted in place of *gccF* on the pRV610*gcc* plasmid to see it the *gcc* genes are capable of producing ASM1 with an identical structure to GccF. If ASM1 produced using pRV610*gcc* had a different structure to the wild-type peptide then it would likely have a different interaction with bacteria and this may prove to be a useful method of changing a glycocins target species and antimicrobial potency.

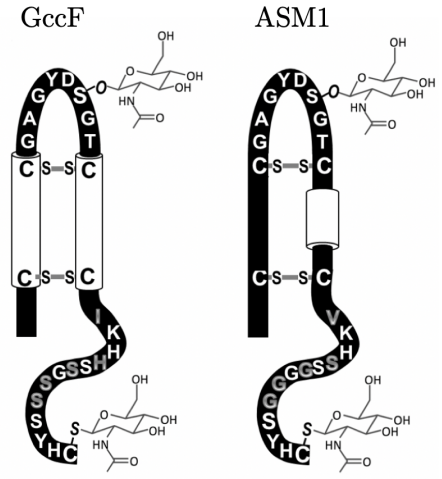


Figure 4.1: **Structural details of GccF and ASM1.** Primary amino acid sequence showing helical regions as cylinders. Reprinted from G. E. Norris and M. L. Patchett. The glycocins: in a class of their own. *Current Opinion in Structural Biology*, 40(2016):112–119, Copyright (2016) [36], with permission from Elsevier).

The mutations designed to test the functionality of GccA-E were informative, confirming previous bioinformatic predictions [66]. Mutations were designed to inactivate six of the Gcc proteins, as leaving the immunity protein (GccH) functional was necessary if the cells were to produce active GccF. Complete deletion of the *gcc* genes was not possible due to overlapping promotor regions at the translation start

site of *gccA*, *gccF* and *gccH* and the open reading frame of *gccE*. Additionally as in the *gccABCDE* operon, the open reading frame of all adjacent genes overlap it is difficult to remove a single open reading frame. Instead, premature stop codons were placed as close to the start codon as possible or within sites required for enzyme activity without interfering with the sequence required for the initiation or termination of transcription and translation. The location of these stop codons appears to have resulted in non-functional proteins which should have produced the same phenotypic outcome as complete gene deletions. While all the mutations introduced to the *gcc* genes were informative, to confirm that the results observed were not due to unintentional changes, complementation studies should be carried out by transforming the mutated cells with a second plasmid containing the wild-type protein under the control of an inducible promotor. Restoration of function following the expression of the wild-type protein would confirm that the mutations made were responsible for the change in phenotype observed.

The production of GccF appeared to be regulated by a protein contained within the Gcc cluster as *L. plantarum* NC8 pRV610*gcc*, *L. sakei* 790 pRV610*gcc* and *L. plantarum* KW30 all produced similar concentrations of active GccF. Mutations inactivating GccE reduced the concentration of GccF within the CFS but did not alter its PTMs, suggesting that GccE is unlikely to have a role in GccF maturation. It is therefore likely that GccE is a transcriptional activator as predicted by bioinformatic analysis. To confirm this, an experiment was designed to insert a reporter gene into the cluster in place of *gccF* within pRV610*gcc*. Unfortunately the construction of this vector failed, the large size of pRV610*gcc* limited restriction enzyme choice which resulted in unsuccessful blunt-ended ligation. Alternatively, Gibson assembly could be used to construct the vector thus avoiding blunt ended ligations.

Mutations that were made to the GccF peptide gave some unexpected results enhancing the current understanding of the sequence/requirements of the glycosyltransferase. The $GccF_{G16del,\ S18\_G19insG}$ mutation analysis showed that the specific location of the target serine on the loop could be important. This was in contrast to what was found for sublancin, where SunS was able to glycosylate the target cysteine at a number of different positions within the loop [93]. Of course, the loss of activity following the $GccF_{G16del,\ S18\_G19insG}$ mutation could also be due to GccF not being able to correctly dock into its target with the GlcNAc in a different position on the loop. Furthermore, if $GccF_{G16del,\ S18\_G19insG}$ is glycosylated then calculating its IC_{50} would reveal insights into the importance of the loop GlcNAc's location for GccF's interaction with its target. To resolve these two possibilities, the gene product of $GccF_{G16del,\ S18\_G19insG}$ first needs to be purified and subjected to mass spectrometry analysis to ascertain the presence of the correct post-translational modifications.

This study confirmed that the glycosyltransferase (GccA) is the sole protein responsible for the di-glycosylation of GccF, linking GlcNAc to either γ cystine or γ serine at residues 18 and 43. Furthermore, the lack of antimicrobial activity following the $GccF_{D17N}$ mutation suggests that this mutation prevents glycosylation of Ser18, although to be certain, mass analysis is required. This is in contrast to SunS where a charged residue next to Cysteine 22 markedly reduced the amount of glycosylation [93]. Sequence analysis of the GccF type glycocins characterised by a flexible modified tail and 8-residue glycosylated loop, showed that the aspartic acid upstream of serine, or cysteine 18 is conserved [53]. Therefore, it is likely that for the GTs of this type of glycocin, a charged residue positioned immediately upstream of the residue to be glycosylated is a requirement for recognition of the residue to be glycosylated by the GT. Interestingly a residue with an ionizable sidechain, histidine, is found adjacent to cysteine 43 which is also glycosylated in GccF, although this was not conserved

across all the GccF type glycocins. Furthermore, both glycosylated residues of GccF have a tyrosine residue positioned two residues N-terminal to them. Whether this is significant or not remains to be seen, as the Y16F mutation did not appear to affect the activity of GccF. To be sure that the requirement is not merely for an aromatic residue, a mutation to a nonpolar residue such as glutamine would be informative. Additionally, an experiment where Asp17 is replaced with a histidine residue to mimic the tail sequence would show if the the loop and tail have the same sequence requirements for glycosylation. Finally, finding non-glycosylated GccF in the cell media following the $GccA_{D123N,D125N}$ mutation was unexpected, creating uncertainty about the order of events during GccF maturation. The current model of GccF maturation [78] suggests that glycosylation of the peptide occurs intracellularly and could be a prerequisite for export. Clearly this is not the case as non-glycosylated GccF was purified from the cell media. As GccA is predicted to be a cytosolic protein due to a lack of recognisable secretion signal sequence, GccF should be glycosylated before it is exported from the cell [66]. However, to clarify where glycosylation takes place, GccF needs to be purified from the cells prior to secretion and analysed for glycosylation.

Although mutations were not made to GccB, the protein responsible for GccF export, its functionality was tested by introducing a mutation to the leader peptide of GccF, designed to prevent its cleavage by the C39-protease domain of GccB. The lack of activity following this mutation was a strong indication that the leader peptide of GccF must be cleaved for export by GccB as expected, as the cleavage of the SunA leader peptide was also found to be a prerequisite for activity [93]. The substitution of the double glycine cleavage site with two theronine residues also reduced the probability of GccF export by the Sec pathway from 0.3 to 0.1 (SignalP-5.0) [134]. While the Sec pathway is responsible for the export of some

bacteriocins [135], it was not considered as a likely exporter of GccF due to the lack of Sec signal sequence within the GccF pre-peptide. It remains unknown if the GccF pre-peptide is glycosylated or if the oxidized pre-peptide is active. Purifying GccF from the cell lysate under both reducing and non-reducing conditions and subsequent characterisation by mass spectrometry would clarify the order of events in the GccF maturation pathway. The His-tagged construct could be used to isolate $_{TT}$GccF from the cells to overcome the difficulty of following the purification of a peptide with no activity. It would also be of interest to purify enough $GccF_{His}$ to calculate its IC_{50} and to determine the effect the addition of charged residues to the tail of GccF has on activity. Additionally, purifying GccF using a tag may also allow more efficient isolation of GccF from the cell culture supernatant.

It appears that neither cluster thioredoxin protein (GccC and GccD) is essential for the production of active GccF under laboratory conditions as was reported for the production of SunA [65], although their presence does increase the activity of the cell culture supernatant. As purification of GccF following the $GccC_{K3X}$ mutation was not successful this needs to be repeated in order to confirm that no structural change occurred as a result of these mutations. GccF produced from cells following the $GccE_{K3X}$ mutation did not elute in the normal location during RP-HPLC, suggesting that perhaps the disulfide bonds had not been correctly formed. If this is the case, then maintaining GccF in a reduced state would be essential to prevent the formation of intermolecular disulfide bonds for an efficient purification.

The expression of soluble Gcc proteins was only successful when using the gram-positive expression system. Unfortunately, mg quantities of these proteins were not produced in this system, making it unappealing for protein structural studies. However, when an acceptable quantity of GccH was produced within *E. coli* it was not soluble. It is likely that some of the Gcc proteins are part of a

complex and/or are associated with the cellular membrane in order to be correctly folded and soluble, which may not be possible in *E. coli* due to the differences in membrane structure. For this reason, an alternative system for mass production of the Gcc proteins in *Lactobacillus* is the most promising option. Very few plasmids are available for high levels of expression in these systems and further difficulties arise from purifying proteins associated with the gram-positive membrane. Alternatively, the pRV610*gcc* expression system could be used with Gcc proteins bearing an affinity tag to aid purification. The purified protein could then be analysed using mass spectrometry to identify any post-translational modifications of the proteins which may be important for their function.

Following up from an observation that indicated GccF may cause cell lysis in cultures kept at 4°C a small study was conducted. This study clearly showed that GccF was not lytic under these conditions as cell density remained constant. However, this experiment raised further questions when an apparent decrease in cell viability as measured by colony counting was seen following the GccF treatment. Bacteriostasis was found to be the source of this reduced viability as incubation with GlcNAc restored viability, showing the cells were not killed. These findings only serve to add to the mystery of exactly how GccF functions, a question that remains to be answered. However, the reduced viability of GccF-treated bacteria seen in the plating assay could have resulted in a miss-classification of GccF as a bactericidal agent showing the importance of live/dead type cell assays for determining if a antimicrobial is a bactericidal or bacteriostatic agent.

To conclude, the heterologous expression of GccABCDEFH has progressed the understanding of GccF biosynthesis and mutation studies have showed that GccC, GccD and GccE are expendable for the production of active GccF. The remaining genes are required for production of active GccF and could be placed under the

control of inducible promotors in an attempt to increase GccF concentration within the cell culture media if large concentrations of GccF were to be produced for trials as a disease treatment or preservative. Finally, the ability to manipulate GccF's structure has provided a valuable tool for future work focused on understanding GccF's mechanism of action, or modifying its target range/potency.

5 Bibliography

[1] R. D. Perry and J. D. Fetherston. *Yersinia pestis*–etiologic agent of plague. *Clinical Microbiology Reviews*, 10(1):35–66, 1997.

[2] T. Welch, F. Fricke, P. McDermott, D. White, M. Rosso, D. Rasko, M. Mammel, M. Eppinger, M. Rosovitz, D. Wagner, L. Rahalison, J. LeClerc, J. Hinshaw, L. Lindler, T. Cebula, E. Carniel, and J. Ravel. Multiple antimicrobial resistance in plague: An emerging public health risk. *Public Library of Science One*, 2(3):1–6, 2007.

[3] D. Lippi and A. Conti. Plague, policy, saints and terrorists: A historical survey. *Journal of Infection*, 44(4):226 – 228, 2002.

[4] S. Kaufmann, J. Weiner, and von Reyn. Novel approaches to tuberculosis vaccine development. *International Journal of Infectious Diseases*, 56:263 – 267, 2017. Special Issue: Commemorating World Tuberculosis Day 2017.

[5] World Health Organization. *Global tuberculosis report 2013*. World Health Organization, 2013.

[6] Interagency Coordination Group on Antimicrobial Resistance. No time to wait: securing the future from drug-resistant infections, 2019. *https://www.who.int/antimicrobial-resistance/interagency-coordination-group/IACG_final_report_EN.pdf*

[7] J. Marsh, M. Pacey, C. Ezeonwuka, S. Ohm, D. Snyder, V. Cooper, L. Harrison, Y. Doi, and M. Mustapha. Clostridioides difficile: a potential source of NpmA in the clinical environment. *Journal of Antimicrobial Chemotherapy*, 74(2):521–523, 2018.

[8] Z. Dyson, E. Klemm, S. Palmer, and G. Dougan. Antibiotic resistance and typhoid. *Clinical Infectious Diseases*, 68(2):S165–S170, 2019.

[9] J. Vázquez, R Enriquez, R. Abad, B. Alcalá, C. Salcedo, and L. Arreaza. Antibiotic resistant meningococci in Europe: Any need to act? *Fœderation of European Microbiology Societies Microbiology Reviews*, 31(1):64–70, 2007.

[10] M. Kitaoka, S. Miyata, D. Unterweger, and S. Pukatzki. Antibiotic resistance mechanisms of *Vibrio cholerae*. *Journal of Medical Microbiology*, 60(Pt 4):397–407, 2011.

[11] T. Foster. Antibiotic resistance in *Staphylococcus aureus*. Current status and future prospects. *FEMS Microbiology Reviews*, 41(3):430–449, 2017.

[12] G. Rossolini, F. Arena, P. Pecile, and S. Pollini. Update on the antibiotic resistance crisis. *Current Opinion in Pharmacology*, 18:56 – 60, 2014.

[13] Faststats - deaths and mortality, 2017. *https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_09-508.pdf*

[14] Antibiotic resistance threats in the United States, 2019. *https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf*

[15] P. E. Reynolds. Structure, biochemistry and mechanism of action of glycopeptide antibiotics. *European Journal of Clinical Microbiology and Infectious Dis-eases*, 8(11):943–950, 1989.

[16] S. Dzidic, J. Suskovic, and B. Kos. Antibiotic resistance mechanisms in bacteria: Biochemical and genetic aspects. *Food Technology and Biotechnology*, 46(1):11–21, 2008.

[17] P. Vannuffel and C. Cocito. Mechanism of action of streptogramins and macrolides. *Drugs*, 51(1):20–30, 1996.

[18] H. Yoneyama and R. Katsumata. Antibiotic resistance in bacteria and its future for novel antibiotic development. *Bioscience, Biotechnology, and Bio-chemistry*, 70(5):1060–1075, 2006.

[19] P. Lambert. Bacterial resistance to antibiotics: Modified target sites. *Advanced Drug Delivery Reviews*, 57(10):1471–1485, 2005.

[20] M. E. Rupp and Fey P. D. Extended spectrum β-lactamase (ESBL)-producing *Enterobacteriaceae. Drugs*, 63(4):353–365, 2003.

[21] A. Gratia. Sur un remarquable exemple d'antagonisme entre deux souches de coilbacille. *Comptes rendus des séances de la Société de Biologie*, 93:1040-1042, 1925.

[22] E. Cascales, S. Buchanan, D. Duché, C. Kleanthous, R. Lloubès, K. Postle, M. Riley, S. Slatin, and D. Cavard. Colicin biology. *Microbiology and Molecular Biology Reviews, 71(1):158–229, 2007*

[23] A. Fleming. On the antibacterial action of cultures of a *Penicillium*, with special reference to their use in the isolation of *B. influenzæ. British Journal of Experimental Pathology*, 10(3):226–236, 1929.

[24] R. Perez, T. Zendo, and K. Sonomoto. Novel bacteriocins from lactic acid bacteria (LAB): various structures and applications. *Microbial Cell Factories*, 13(1):3, 2014.

[25] X. Zhao and O. P. Kuipers. Identification and classification of known and putative antimicrobial compounds produced by a wide variety of *Bacillales species. BioMed Central Genomics*, 17(1):882, 2016.

[26] N. Heng, P. Wescombe, J. Burton, R. Jack, and J. Tagg. 'The diversity of bacteriocins in gram-positive bacteria'. In *Bacteriocins: Ecology and Evolution*, pages 45–92. Margaret A. Riley and Milind A. Chavan, editors. Springer, Heidelberg, Germany, 2007.

[27] M. Rea, R. Ross, P. Cotter, and C. Hill. 'Classification of Bacteriocins from Gram-positive Bacteria'. in *Biotechnology of Lactic Acid Bacteria*. Fernanda Mozzi, Raul R. Raya and Graciela M. Vignolo editors. Springer, New York, NY, 2011.

[28] H. Kaya, B. Özel, and Ö. Şimşek. '*A Natural Way of Food Preservation: Bacteriocins and Their Applications.*' in Health and Safety Aspects of Food Processing Technologies. Malik, A., Erginkaya Z., Erten H. (editors) Springer Cham, Switzerland. 2019.

[29] P. D. Cotter, R. P. Ross, and C. Hill. Bacteriocins - a viable alternative to antibiotics? *Nature Reviews Microbiology*, 11(2):95–105, 2013.

[30] P. D. Cotter, C. Hill, and R. P. Ross. Bacteriocins: developing innate immunity for food. *Nature Reviews Microbiology*, 3(10):777–88, 2005.

[31] P. Alvarez-Sieiro, M. Montalban-Lopez, D. Mu, and O. P. Kuipers. Bacteriocins of lactic acid bacteria: extending the family. *Applied Microbiology and Biotechnology*, 100(7):2939–51, 2016.

[32] L. Lopetuso, M. Giorgio, A. Saviano, F. Scaldaferri, A. Gasbarrini, and G. Cammarota. Bacteriocins and bacteriophages: Therapeutic weapons for gastrointestinal diseases? *International Journal of Molecular Sciences*, 20(1):183, 2019.

[33] P. Hols, L. Ledesma-García, P. Gabant, and J. Mignolet. Mobilization of microbiota commensals and their bacteriocins for therapeutics. *Trends in Microbiology*, 27(8):690–702, 2019.

[34] J. M. Shin, J. W. Gwak, P. Kamarajan, J. C. Fenno, A. H. Rickard, and Y. L. Kapila. Biomedical applications of nisin. *Journal of Applied Microbiology*, 120(6):1449–1465, 2016.

[35] F. Bourdichon, S. Casaregola, C. Farrokh, J. Frisvad, M. Gerds, W. Hammes, J. Harnett, G. Huys, S. Laulund, A. Ouwehand, I. Powell, J. Prajapati, Y. Seto, E. Schure, A. Boven, V. Vankerckhoven, A. Zgoda, S. Tuijtelaars, and E. Hansen. Food fermentations: Microorganisms with technological beneficial use. *International Journal of Food Microbiology*, 154(3):87 – 97, 2012.

[36] V. L. Cavera, T. D. Arthur, D. Kashtanov, and M. L. Chikindas. Bacteriocins and their position in the next wave of conventional antibiotics. *International Journal of Antimicrobial Agents*, 46(5):494–501, 2015.

[37] N. Ríos Colombo, M. Chalón, S. Navarro, and A. Bellomio. Pediocin-like bacteriocins: new perspectives on mechanism of action and immunity. *Current Genetics*, 64(2):345–351, 2018.

[38] T. R. Klaenhammer. Bacteriocins of lactic acid bacteria. *Biochimie*, 70(3):337–49, 1988.

[39] M. C. Rea, A. Dobson, O. O'Sullivan, F. Crispie, F. Fouhy, P. D. Cotter, F. Shanahan, B. Kiely, C. Hill, and R. P. Ross. Effect of broad- and narrow-spectrum antimicrobials on *Clostridium difficile* and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl 1(1):4639–44, 2011.

[40] H. Majeed, A. Lampert, L. Ghazaryan, and O. Gillor. The weak shall inherit: bacteriocin-mediated interactions in bacterial populations. *PLoS One*, 8(5):63837, 2013.

[41] M. E. Hibbing, C. Fuqua, M. R. Parsek, and S. B. Peterson. Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, 8(1):15–25, 2010.

[42] B. C. Kirkup and M. A. Riley. Antibiotic-mediated antagonism leads to a bacterial game of rock-paper-scissors *in vivo*. *Nature*, 428(6981):412–4, 2004.

[43] A. Gravesen, M. Ramnath, K. B. Rechinger, N. Andersen, L. Jansch, Y. Hechard, J. W. Hastings, and S. Knochel. High-level resistance to class iia bacteriocins is associated with one general mechanism in *Listeria monocytogenes*. *Microbiology*, 148(8):2361–9, 2002.

[44] C. Bastos Mdo, M. L. Coelho, and O. C. Santos. Resistance to bacteriocins produced by Gram-positive bacteria. *Microbiology*, 161(4):683–700, 2015.

[45] K. Geldart and Y. N. Kaznessis. Characterization of class iia bacteriocin resistance in *Enterococcus faecium*. *Antimicrobial Agents and Chemotherapy*, 61(4), 2017.

[46] L. A. Draper, P. D. Cotter, C. Hill, and R. P. Ross. Lantibiotic resistance. *Microbiology and Molecular Biology Reviews*, 79(2):171–91, 2015.

[47] D. Raoult. Probiotics and obesity: a link? *Nature Reviews Microbiology*, 7(9):616, 2009.

[48] S. Y. Shaw, J. F. Blanchard, and C. N. Bernstein. Association between the use of antibiotics and new diagnoses of Crohn's disease and ulcerative colitis. *American Journal of Gastroenterology*, 106(12):2133–42, 2011.

[49] J. A. Madden and J. O. Hunter. A review of the role of the gut microflora in irritable bowel syndrome and the effects of probiotics. *British Journal of Nutrition*, 88(1):67–72, 2002.

[50] A. E. Livanos, T. U. Greiner, P. Vangay, W. Pathmasiri, D. Stewart, S. McRitchie, H. Li, J. Chung, J. Sohn, S. Kim, Z. Gao, C. Barber, J. Kim, S. Ng, A. B. Rogers, S. Sumner, X. S. Zhang, K. Cadwell, D. Knights, A. Alekseyenko, F. Backhed, and M. J. Blaser. Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. *Nature Microbiology*, 1(11):16140, 2016.

[51] T. Van Boeckel, C. Brower, M. Gilbert, B. Grenfell, S. Levin, T. Robinson, A. Teillant, and R. Laxminarayan. Global trends in antimicrobial use in food animals. *Proceedings of the National Academy of Sciences*, 112(18):5649–5654, 2015.

[52] S. Kraemer, A. Ramachandran, and G. Perron. Antibiotic pollution in the environment: From microbial ecology to public policy. *Microorganisms*, 7(6):180, 2019.

[53] G. E. Norris and M. L. Patchett. The glycocins: in a class of their own. *Current Opinion in Structural Biology*, 40(2016):112–119, 2016.

[54] T. J. Oman, J. M. Boettcher, H. Wang, X. N. Okalibe, and W. A. van der Donk. Sublancin is not a lantibiotic but an *S*-linked glycopeptide. *Nature Chemical Biology*, 7(2):78–80, 2011.

[55] J. Stepper, S. Shastri, T. S. Loo, J. C. Preston, P. Novak, P. Man, C. H. Moore, V. Havlicek, M. L. Patchett, and G. E. Norris. Cysteine *S*-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *Federation of European Biochemical Societies Letters*, 585(4):645–50, 2011.

[56] T. Hata, R. Tanaka, and S. Ohmomo. Isolation and characterization of plantaricin ASM1: a new bacteriocin produced by *Lactobacillus plantarum* A-1. *International Journal of Food Microbiology*, 137(1):94–9, 2010.

[57] H. Venugopal, P. J. Edwards, M. Schwalbe, J. K. Claridge, D. S. Libich, J. Stepper, T. Loo, M. L. Patchett, G. E. Norris, and S. M. Pascal. Structural, dynamic, and chemical characterization of a novel *S*-glycosylated bacteriocin. *Biochemistry*, 50(14):2748–55, 2011.

[58] M. A. Maky, N. Ishibashi, T. Zendo, R. H. Perez, J. R. Doud, M. Karmi, and K. Sonomoto. Enterocin F4-9, a novel *O*-linked glycosylated bacteriocin. *Applied and Environmental Microbiology*, 81(14):4819–26, 2015.

[59] H. Hanchi, R. Hammami, B. Fernandez, R. Kourda, J. Ben Hamida, and I. Fliss. Simultaneous production of formylated and nonformylated enterocins

L50A and L50B as well as 61A, a new glycosylated durancin, by *Enterococcus durans* 61A, a strain isolated from artisanal fermented milk in Tunisia. *Journal of Agricultural and Food Chemistry*, 64(18):3584–90, 2016.

[60] H. Wang, T. J. Oman, R. Zhang, C. V. Garcia De Gonzalo, Q. Zhang, and W. A. van der Donk. The glycosyltransferase involved in thurandacin biosynthesis catalyzes both *O*- and *S*-glycosylation. *Journal of the American Chemical Society*, 136(1):84–7, 2014.

[61] A. Kaunietis, A. Buivydas, D. Čitavičius, and O. Kuipers. Heterologous biosynthesis and characterization of a glycocin from a thermophilic bacterium. *Nature Communications*, 10(1):1115, 2019.

[62] H. Ren, S. Biswas, S. Ho, W. A. van der Donk, and H. Zhao. Rapid discovery of glycocins through pathway refactoring in *Escherichia coli*. *American Chemical Society Chemical Biology*, 13(10):2966–2972, 2018.

[63] A. Kerr. *The bacteriostatic spectrum and inhibitory mechanism of glycocin F, a bacteriocin from Lactobacillus plantarum KW30 : a thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Microbiology at Massey University, Palmerston North, New Zealand.* Thesis, Massey University, 2013.

[64] P. Main. *Investigating the bacteriocin library Lactobacillus plantarum A-1 : presented in partial fulfilment of the requirements for the degree of Master of Science in Microbiology at Massey University, Manuwatū Campus, Palmerston North, New Zealand.* Thesis, Massey University, 2014.

[65] R. Dorenbos, T. Stein, J. Kabel, C. Bruand, A. Bolhuis, S. Bron, W. J. Quax, and J. M. Van Dijl. Thiol-disulfide oxidoreductases are essential for the pro-

duction of the lantibiotic sublancin 168. *The Journal of Biological Chemistry*, 277(19):16682–8, 2002.

[66] J. Stepper. *The molecular and cellular characterisation of the first glycocin : plantaricin KW30 : a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Biochemistry at Massey University, Palmerston North, New Zealand*. Thesis, Massey University, 2009.

[67] K. Drower. *The bacteriostatic diglycocylated bacteriocin glycocin F targets a sugar-specific transporter :* Thesis, Massey University, 2014.

[68] C. V. Garcia De Gonzalo, E. L. Denham, R. A. T. Mars, J. Stülke, W. A. van der Donk, and J. M. van Dijl. The phosphoenolpyruvate:sugar phosphotransferase system is involved in sensitivity to the glucosylated bacteriocin sublancin. *Antimicrobial Agents and Chemotherapy*, 59(11):6844, 2015.

[69] M. Saier. The bacterial phosphotransferase system: New frontiers 50 years after its discovery. *Journal of Molecular Microbiology and Biotechnology*, 25(2-3):73–78, 2015.

[70] M. Bailie. *The role of the N-acetylglucosamine phosphoenolpyruvate phosphotransferase system from Lactobacillus plantarum 8014 in the mechanism of ac-tion of glycocin F :* Thesis, Massey University, 2017.

[71] P. van Langevelde, J. T. van Dissel, C. J. Meurs, J. Renz, and P. H. Groeneveld. Combination of flucloxacillin and gentamicin inhibits toxic shock syndrome toxin 1 production by *Staphylococcus aureus* in both logarithmic

and stationary phases of growth. *Antimicrobial Agents and Chemotherapy*, 41(8):1682–5, 1997.

[72] W. Kelly, R.V. Asmundson, and C.M. Huang. Characterization of plantaricin KW30, a bacteriocin produced by *Lactobacillus plantarum*. *Journal Of Applied Bacteriology*, 81(6):657 – 662, 1996.

[73] S. W. Bisset, S. H. Yang, Z. Amso, P. W. R. Harris, M. L. Patchett, M. A. Brimble, and G. E. Norris. Using chemical synthesis to probe structure-activity relationships of the glycoactive bacteriocin glycocin F. *American Chemical Society Chemical Biology*, 13(5):1270–1278, 2018.

[74] Z. Amso, S. W. Bisset, S. H. Yang, P. W. R. Harris, T. H. Wright, C. D. Navo, M. L. Patchett, G. E. Norris, and M. A. Brimble. Total chemical synthesis of glycocin F and analogues: *S*-glycosylation confers improved antimicrobial activity. *Chemical Science*, 9(6):1686–1691, 2018.

[75] M. I. Gibson, G. J. Hunt, and N. R. Cameron. Improved synthesis of *O*-linked, and first synthesis of *S*-linked, carbohydrate functionalised N-carboxyanhydrides (glyconcas)-linked, and first synthesis of *S*-linked, carbohydrate functionalised N-carboxyanhydrides (glycoNCAs). *Organic and Biomolecular Chemistry*, 5(17):2756–7, 2007.

[76] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, R. Breitling, E. Takano, and M. H. Medema. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, 43(1):237–43, 2015.

[77] D. Hoover and L. Steenson. *Bacteriocins of lactic acid bacteria*. Academic Press, 2014.

[78] S. Ahn, J. Stepper, T. S. Loo, S. W. Bisset, M. L. Patchett, and G. E. Norris. Expression of *Lactobacillus plantarum* KW30 *gcc* genes correlates with the production of glycocin F in late log phase. *Federation of European Microbiological Societies Microbiology Letters*, 365(23):261–261, 2018.

[79] M. Y. Galperin. Telling bacteria: do not LytTR. *Structure*, 16(5):657–9, 2008.

[80] T. Okinaga, G. Niu, Z. Xie, F. Qi, and J. Merritt. The *hdrRM* operon of *Streptococcus mutans* encodes a novel regulatory system for coordinated competence development and bacteriocin production. *Journal of Bacteriology*, 192(7):1844, 2010.

[81] Z. Xie, T. Okinaga, G. Niu, F. Qi, and J. Merritt. Identification of a novel bacteriocin regulatory system in *Streptococcus mutans*. *Molecular Microbiology*, 78(6):1431–47, 2010.

[82] P. A. Risoen, O. Johnsborg, D. B. Diep, L. Hamoen, G. Venema, and I. F. Nes. Regulation of bacteriocin production in *Lactobacillus plantarum* depends on a conserved promoter arrangement with consensus binding sequence. *Molecular Genetics and Genomics*, 265(1):198–206, 2001.

[83] G. Engelke, Z. Gutowski-Eckel, P. Kiesau, K. Siegers, M. Hammelmann, and K. D. Entian. Regulation of nisin biosynthesis and immunity in *Lactococcus lactis* 6F3. *Applied and Environmental Microbiology*, 60(3):814–25, 1994.

[84] D. B. Diep, R. Myhre, O. Johnsborg, A. Aakra, and I. F. Nes. Inducible bacteriocin production in *Lactobacillus* is regulated by differential expression of the pln operons and by two antagonizing response regulators, the activity of which is enhanced upon phosphorylation. *Molecular Microbiology*, 47(2):483, 2003.

[85] D. Sidote, C. Barbieri, T. Wu, and A. Stock. Structure of the *Staphylococcus aureus* AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. *Structure*, 16(5):727–735, 2008.

[86] S. Behr, R. Heermann, and K. Jung. Insights into the DNA-binding mechanism of a LytTR-type transcription regulator. *Bioscience Reports*, 36(2), 2016.

[87] A. Nikolskaya and M. Galperin. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Research*, 30(11):2453–2459, 2002.

[88] M. Boudes, D. Sanchez, M. Graille, H. van Tilbeurgh, D. Durand, and S. Quevillon-Cheruel. Structural insights into the dimerization of the response regulator ComE from *Streptococcus pneumoniae*. *Nucleic Acids Research*, 42(8):5302–13, 2014.

[89] B. Cantarel, P. Coutinho, C. Rancurel, Th. Bernard, V. Lombard, and B. Henrissat. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research*, 37:D233–8, 2009.

[90] C. Breton, L. Šnajdrová, C. Jeanneau, J. Koča, and A. Imberty. Structures and mechanisms of glycosyltransferases. *Glycobiology*, 16(2):29R–37R, 2005.

[91] L.L. Lairson, B. Henrissat, G.J. Davies, and S.G. Withers. Glycosyltransferases: Structures, functions, and mechanisms. *Annual Review of Biochemistry*, 77(1):521–555, 2008. PMID: 18518825.

[92] R. Nagar and A. Rao. An iterative glycosyltransferase EntS catalyzes transfer and extension of *O*- and *S*-linked monosaccharide in enterocin 96. *Glycobiology*, 27(8):766–776, 2017.

[93] H. Wang and W. A. van der Donk. Substrate selectivity of the sublancin *S*-glycosyltransferase. *Journal of the American Chemical Society*, 133(41):16394–7, 2011.

[94] G. De Gonzalo. *Studies on the mechanism of action of the antimicrobial S-linked glycopeptide sublancin*. Thesis, University of Illinois, 2015.

[95] K. Hashimoto, T. Madej, S. Bryant, and A. Panchenko. Functional states of homooligomers: Insights from the evolution of glycosyltransferases. *Journal of Molecular Biology*, 399(1):196–206, 2010.

[96] S. Bobeica, S. Dong, L. Huo, N. Mazo, M. McLaughlin, G. Jiménez-Osés, S. Nair, and W. van der Donk. Insights into AMS/PCAT transporters from biochemical and structural characterization of a double glycine motif protease. *eLife*, 8:42305, 2019.

[97] D. Lin, S. Huang, and J. Chen. Crystal structures of a polypeptide processing and secretion transporter. *Nature*, 523(7561):425–430, 2015.

[98] L. S. Havarstein, D. B. Diep, and I. F. Nes. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Molecular Microbiology*, 16(2):229–240, 1995.

[99] W. Aucher, C. Lacombe, A. Héquet, J. Frère, and J. Berjeaud. Influence of amino acid substitutions in the leader peptide on maturation and secretion of mesentericin Y105 by *Leuconostoc mesenteroides*. *Journal of Bacteriology*, 187(6):2218–2223, 2005.

[100] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita,

D. Zhang, C. Zheng, L. Y. Geer, and S. H. Bryant. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45(1):200–203, 2017.

[101] C. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223, 1973.

[102] T. Kouwen, A. Van Der Goot, R. Dorenbos, T. Winter, H. Antelmann, M. Plaisier, W. Quax, J. Van Dijl, and J. Dubois. Thiol-disulphide oxidoreductase modules in the low-GC Gram-positive bacteria. *Molecular Microbiology*, 64(4):984–999, 2007.

[103] K. Inaba, S. Murakami, M. Suzuki, A. Nakagawa, E. Yamashita, K. Okada, and K. Ito. Crystal structure of the DsbB-DsbA complex reveals a mechanism of disulfide bond generation. *Cell*, 127(4):789–801, 2006.

[104] A. Bolhuis, G. Venema, W. J. Quax, S. Bron, and J. M. van Dijl. Functional analysis of paralogous thiol-disulfide oxidoreductases in *Bacillus subtilis*. *Journal of Biological Chemistry*, 274(35):24531–24538, 1999.

[105] S. W. Bisset. *Understanding the mechanism of action of the glycosylated bacteriocin glycocin F*. PhD thesis, Massey University, 2019.

[106] D. B. Diep, M. Skaugen, Z. Salehian, H. Holo, and I. F. Nes. Common mechanisms of target cell recognition and immunity for class ii bacteriocins. *Proceedings of the National Academy of Sciences*, 104(7):2384–9, 2007.

[107] S. Wang, Q. Wang, X. Zeng, Q. Ye, S. Huang, H. Yu, T. Yang, and S. Qiao. Use of the antimicrobial peptide sublancin with combined antibacterial and immunomodulatory activities to protect against methicillin-resistant *Staphylococcus aureus* infection in mice. *Journal of Agricultural and Food Chemistry*, 65(39):8595–8605, 2017.

[108] A. E. Jacob and S. J. Hobbs. Conjugal transfer of plasmid-borne multiple antibiotic resistance in *Streptococcus faecalis* var. zymogenes. *Journal of Bacteriology*, 117(2):360–372, 1974.

[109] O. Kuipers, P. de Ruyter, M. Kleerebezem, and W. de Vos. Quorum sensing-controlled gene expression in lactic acid bacteria. *Journal of Biotechnology*, 64(1):15 – 21, 1998.

[110] T. Aukrust and H. Blom. Transformation of *Lactobacillus* strains used in meat and vegetable fermentations. *Food Research International*, 25(4):253–261, 1992.

[111] R. N. Costilow and T. W. Humphreys. Nitrate reduction by certain strains of *Lactobacillus plantarum*. *Science*, 121(3136):168, 1955.

[112] U. Schillinger and F. K. Lücke. Antibacterial activity of *Lactobacillus sake* isolated from meat. *Applied and Environmental Microbiology*, 55(8):1901–1906, 1989.

[113] A. M. Crutz-Le Coq and M. Zagorec. Vectors for *Lactobacilli* and other Gram-positive bacteria based on the minimal replicon of pRV500 from *Lactobacillus sakei*. *Plasmid*, 60(3):212–20, 2008.

[114] F. Yao, X. Y. Xu, and Q. Pan. A modified method for plasmid extraction from *Lactobacillus plantarum* contained lysozyme removal step. *Analytical Biochemistry*, 566(1):37–39, 2018.

[115] J. Chiu, P. E. March, R. Lee, and D. Tillett. Site-directed, ligase-independent mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Research*, 32(21):e174, 2004.

[116] J. Chiu, D. Tillett, I. W. Dawes, and P. E. March. Site-directed, ligase-independent mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *Journal of Microbiological Methods*, 73(2):195–8, 2008.

[117] M. Bradford. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1):248–254, 1976.

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

[119] R. W. Li. *Metagenomics and Its Applications in Agriculture, Biomedicine, and Environmental Studies*. Nova Science Publishers, Inc, Hauppauge, N.Y., 2011.

[120] L. Axelsson, I. Rud, K. Naterstad, H. Blom, B. Renckens, J. Boekhorst, M. Kleerebezem, S. van Hijum, and R. J. Siezen. Genome sequence of the naturally plasmid-free *Lactobacillus plantarum* strain NC8. *Journal of Bacteriology*, 194(9):2391–2, 2012.

[121] A. Maldonado, R. Jimenez-Diaz, and J. L. Ruiz-Barba. Induction of plantaricin production in *Lactobacillus plantarum* NC8 after coculture with specific Gram-positive bacteria is mediated by an autoinduction mechanism. *Journal of Bacteriology*, 186(5):1556–1564, 2004.

[122] A. Hemsley, N. Arnheim, M. Toney, G. Cortopassi, and D. Galas. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Research*, 17(16):6545–6551, 1989.

[123] C. J. Lohff and K. B. Cease. Pcr using a thermostable polymerase with 3' to 5' exonuclease activity generates blunt products suitable for direct cloning. *Nucleic Acids Research*, 20(1):144, 1992.

[124] D. Silva, G. Santos, M. Barroca, and T. Collins. *Inverse PCR for Point Mutation Introduction*, volume 1620, pages 87–100. Springer, New York, NY, 2017.

[125] L. Slabinski, L. Jaroszewski, L. Rychlewski, I. A. Wilson, S. A. Lesley, and A. Godzik. XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics*, 23(24):3403–5, 2007.

[126] B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta*, 1804(4):996–1010, 2010.

[127] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, 1990.

[128] T. Thongaram, J. Hoeflinger, J. Chow, and M. Miller. Human milk oligosaccharide consumption by probiotic and human-associated *bifidobacteria* and *lactobacilli*. *Journal of Dairy Science*, 100(10):7825 – 7833, 2017.

[129] S. Meydan, J. Marks, D. Klepacki, V. Sharma, P. Baranov, A. Firth, T. Margus, A. Kefi, N. Vázquez-Laslop, and A. Mankin. Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome. *Molecular Cell*, 74(3):481–493.e6, 2019.

[130] L Kelley, S Mezulis, C Yates, M Wass, and M Sternberg. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10:845, 2015.

[131] A. Guerler, B. Govindarajoo, and Y. Zhang. Mapping monomeric threading to protein-protein structure prediction. *Journal of Chemical Information and Modeling*, 53(3):717–725, 2013.

[132] V. Saunders. *Microbial genetics applied to biotechnology : principles and techniques of gene transfer and manipulation*. Springer, 2012.

[133] D Gibson. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Research*, 37(20):6984–6990, 2009.

[134] J.J Almagro Armenteros, K. D. Tsirigos, C. K Sønderby, T. N Petersen, O Winther, S Brunak, G von Heijne, and H Nielsen. Signalp 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4):420–423, 2019.

[135] C Herranz and A. J. M. Driessen. Sec-mediated secretion of bacteriocin enterocin p by *Lactococcus lactis*. *Applied and environmental microbiology*, 71(4):1959–1963, 2005.

6 Appendix

Activity assay

Figure 6.1: **GccF standards tube activity assay extended.** 20 μL of purified GccF diluted in MRS or *L. plantarum* NC8 pRV610*gcc* CFS was applied to each tube.

Estimation of protein size on SDS-PAGE

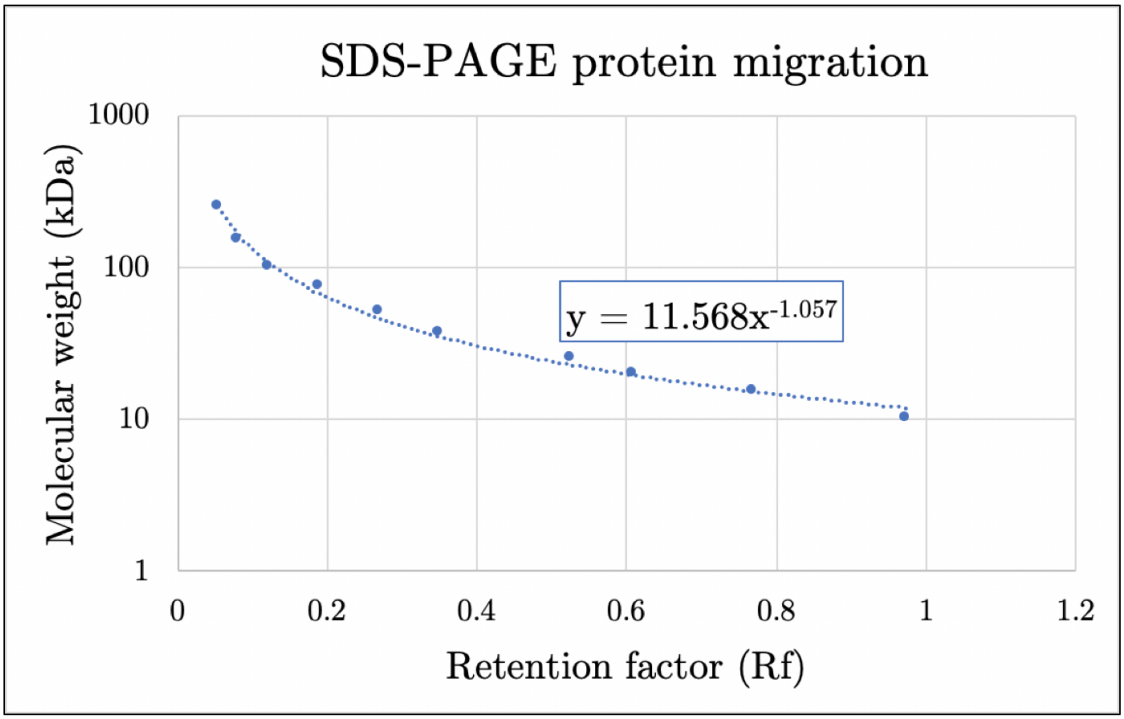


Figure 6.2: **Retention factors used for estimating protein size** belonging to the SDS-PAGE *E. coli* BL21 (DE3) pProEx_HTB_*gccH* induction trial.

IC_{50} of $GccF_{S18C}$

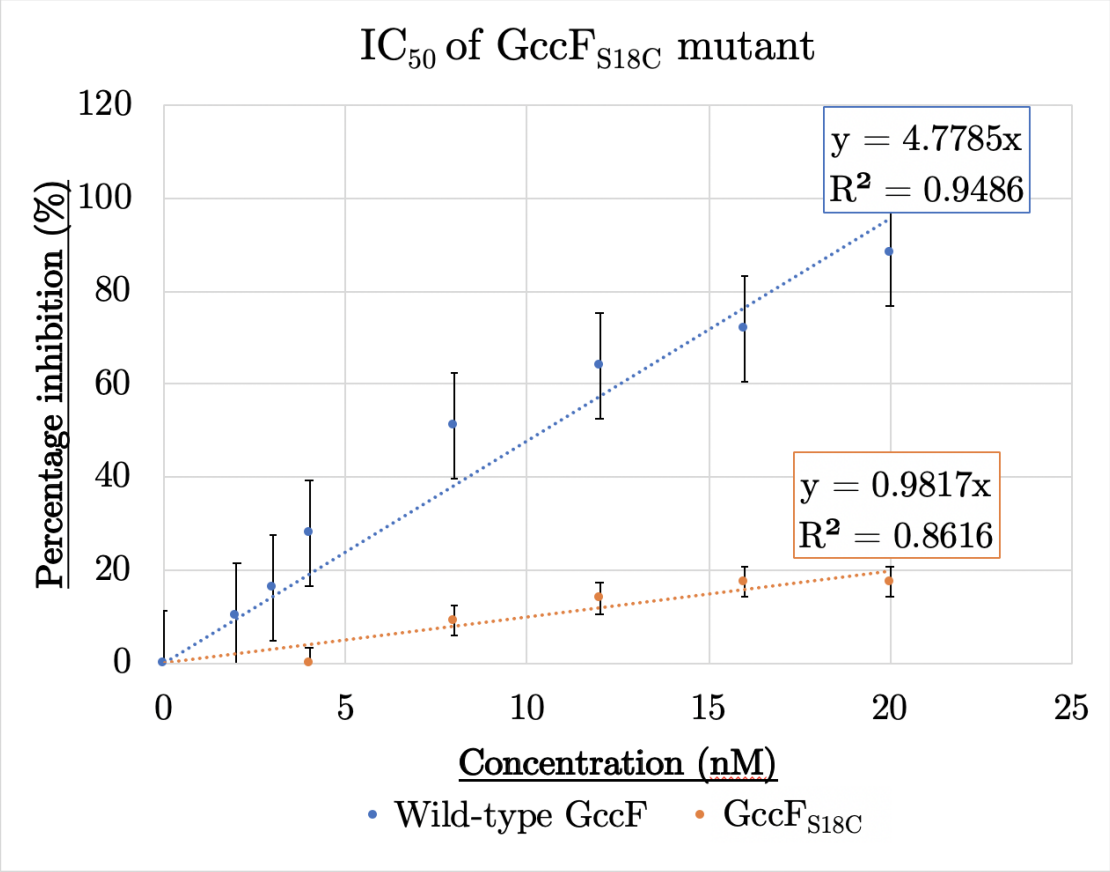


Figure 6.3: **IC_{50} of native and GccF purified from *L. plantarum* NC8 pRV610*gcc*_$GccF_{S18C}$**. The bars on each points represent standard error.

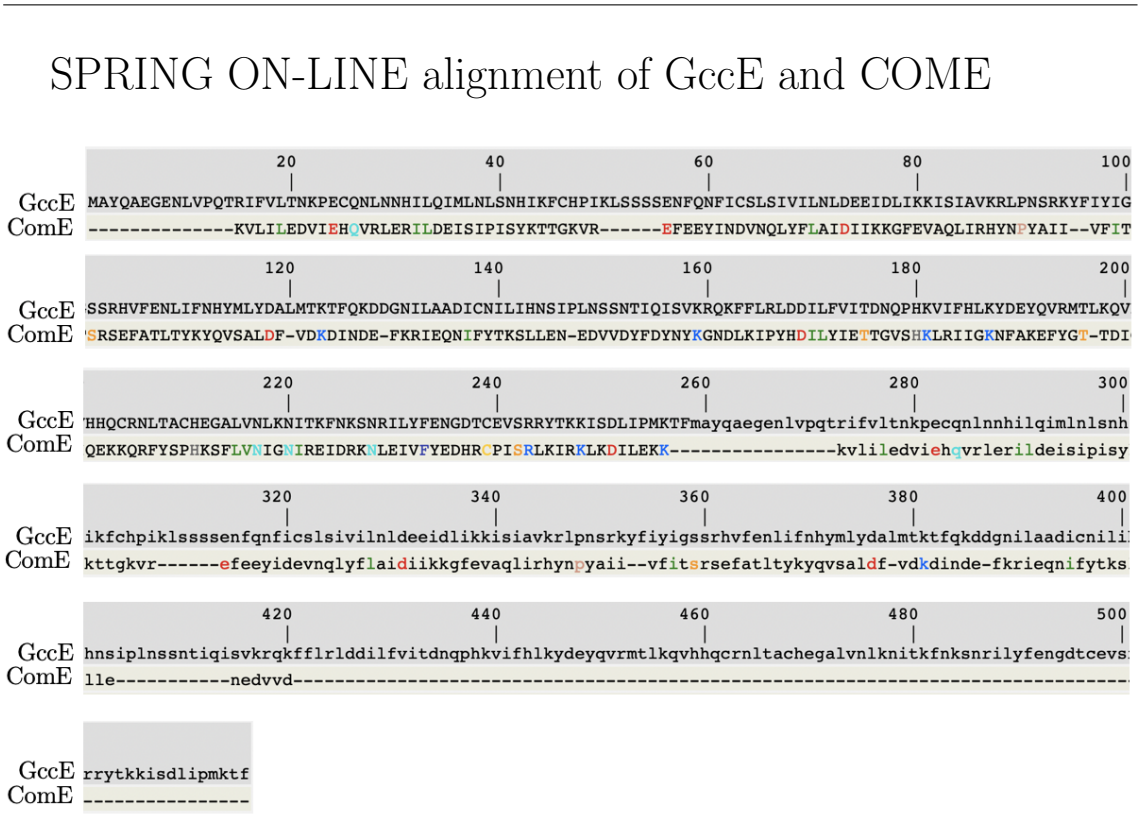


Figure 6.4: **SPRING ON-LINE alignment of GccE and COME used to predict a dimeric structure of GccE.**

Primers

Table 6.1: Primers used in this study

Name	**Sequence**
gccA_d123,125n_fs	GGATGTTGATGATGGAGC
gccA_d123,125n_rl	AACCACTCATTACTATTAACATAGAACACCCAGTCTTTAGA
gccA_d123,125n_fl	GTTAATAGTAATGAGTGGTTGGATGTTGATGATGGAGC
gccA_d123,125n_rs	ATAGAACACCCAGTCTTTAGA
gccC_stop_fl	GTTATGAAATAAACCATTATCACTATTATTGCGCTTAGTGTT ATC
gccC_stop_fs	CACTATTATTGCGCTTAGTGTTATC
gccC_stop_rl	ATAATGGTTTATTTCATAACAATGCCTCCCTCAATGATTC
gccC_stop_rs	AATGCCTCCCTCAATGATTC
gccD_stop_fs	AAAACTAGTATTAATTTTGCTAATACTTCTC
gccD_stop_fl	GATGAATCTATAAAAAAAAAGAAAACTAGTATTAATTTTGCT AATACTTCTC
gccD_stop_rs	TAAAGATCTCCCGTGTACC
gccD_stop_rl	CTTTTTTTTTATAGATTCATCTAAAGATCTCCCGTGTACC
gccE_stop_fs	GGAATGCCAAAATTTAAATAACCA
gccE_stop_fl	ATATTTGTTTAAACAAATAAACCGGAATGCCAAAATTTAAAT AACCA
gccE_stop_rs	TCTAGTCTGTGGAACTAAATTTTC
gccE_stop_rl	GGTTTATTTGTTTAAACAAATATTCTAGTCTGTGGAACTAA ATTTTC
gccE_148stop_fs	CACTATACAAATTTCTGTCAAAAGACA
gccE_148stop_rl	TTTGAACTATTTTAGGGAATAGAATTGTGTATTAGAATGTTA CAAATATC
gccE_148stop_fl	ATTCCCTAAAATAGTTCAAACACTATACAAATTTCTGTCAAA AGACA
gccE_148stop_rs	AGAATTGTGTATTAGAATGTTACAAATATC
gccE_148_ha_fl	ATTCCCTACCCATACGATGTTCCAGATTACGCTTTAAATAGTT CAAACACTATACAAATTTCTG

gccE_148_ha_rl	TTTGAACTATTTAAAGCGTAATCTGGAACATCGTATGGGTA GGGAATAGAATTGTGTATTAGAATGT
gccF_HA_F_T	TCCCATACGATGTTCCAGATTACGCTAGTTATCATTGTTAG TTTTGTGAATGTTTT
gccF_HA_F_NT	AGTTATCATTGTTAGTTTTGTGAATGTTTT
gccF_HA_R_T	AGCCTAATCTGGAACATCGTATGGGTAACCAAAACAATGC GAATACATATAATC
gccF_HA_R_NT	ACCAAAACAATGCGAATACATATAATC
gccF_his6_fl	CATCATCATCATCATAGCAGTTATCATTGTTAGTTTTGTGAA TGTTTTAGAC
gccF_his6_fs	CATTGTTAGTTTTGTGAATGTTTTAGAC
gccF_his6_rl	ATAACTGCTATGATGATGATGATGATGCTTTATACCAAAA CAATGC
gccF_his6_rs	ATGCTTTATACCAAAACAATGCG
gccF_tt_fl	CAACACTACTAAACCTGCATGGTGTTGGTATACTTTAGCA
gccF_tt_fs	ATGGTGTTGGTATACTTTAGCA
gccF_tt_rl	GCAGGTTTAGTAGTGTTGTTTTGAGCCTTAGAAATTTCAC TTATAG
gccF_tt_rs	TTTTGAGCCTTAGAAATTTCACTTATAG
gccF_s18c_fs	TGATTATATGTATTCGCATTGTTTTGG
gccF_s18c_fl	GTTATGATTGTGGAACCTGTGATTATATGTATTCGCATTGTT TTGG
gccF_s18c_rs	CAGCACCACACATTGCTA
gccF_s18c_rl	CAGGTTCCACAATCATAACCAGCACCACACATTGCTA
gccF_y16f_fl	GTTTTGATTCGGGAACCTGTGATTATATGTATTCGCATTGT TTTGG
gccF_y16f_rl	CAGGTTCCCGAATCAAAACCAGCACCACACATTGCTA
gccF_y16q_fl	GTCAAGATTCGGGAACCTGTGATTATATGTATTCGCATTGTT TTGG
gccF_y16q_rl	CAGGTTCCCGAATCTTGACCAGCACCACACATTGCTA

gccF_d17n_fl	GTTATAACTCGGGAACCTGTGATTATATGTATTCGCATTGTTTTGG
gccF_d17n_rl	CAGGTTCCCGAGTTATAACCAGCACCACACATTGCTA
gccF_gloop_rs	ACACATTGCTAAAGTATACCAAC
gccF_gloop_rl	CAGGTTCCTCCCGAATCATAAGCACCACACATTGCTAAAGTATACCAAC
gccF_gloop_fl	GGTGCTTATGATTCGGGAGGAACCTGTGATTATATGTATTCGCATTGTTTTGG
gccF_c43s_fs	AATGTTTTAGACTTATTAAGTTACGT
gccF_c43s_rl	CACAAAACTACGAATGATAACTGCTACTACCACTACTATGATGCTTTATA
gccF_c43s_fl	GCAGTTATCATTCGTAGTTTTGTGAATGTTTTAGACTTATTAAGTTACGT
gccF_c43s_rs	TACTACCACTACTATGATGCTTTATA
gccF_NarI_r	CATGGCGCCTACCCTCCTTTTAATAAGAAAATTATAAACC
gccF_NaeI_f	TAGGCCGGCGAATGTTTTAGACTTATTAAGTTACGTTGA
613_NarI_f	TGAGGCGCCGTGGAAGTTACTGACGTAAGATTA
613_NaeI_r	GCCGCCGGCTTATTTATTTTGACACCAGACCAAC
gccH_NdeI	GGCCATATGGTTACTAATATTTGTATTATTCCTTCT
gccH_XhoI	ATGCCTCGAGCTAAGCATC
pepN_NaeI_r	GCGCGCCGGCTTACAATTTTTCAGCAATATCAGTAATTGC
pepN_KasI_f	GCGCGGCGCCATGGCTGTAAAACGTTTAATTGAAAC
GccF_MscI_f	TAGTGGCCAGAATGTTTTAGACTTATTAAGTTACGTTGA
MscI_pepn_r	GCGCTGGCCATTACAATTTTTCAGCAATATCAGTAATTGC

List of pRV610*gcc* plasmids

Table 6.2: pRV610 mutant constructs created during this study

pRV610*gcc*
pRV610*gcc*_GccA$_{D123N,D125N}$
pRV610*gcc*_GccA$_{D123N,D125N}$_GccF$_{C43S}$
pRV610*gcc*_GccC$_{KX3}$
pRV610*gcc*_GccD$_{L4X}$
pRV610*gcc*_GccC$_{KX3}$_GccD$_{L4X}$
pRV610*gcc*_GccE$_{L19X}$
pRV610*gcc*_GccE$_{L148X}$
pRV610*gcc*_GccF$_{C43S}$
pRV610*gcc*_GccF$_{S18C}$
pRV610*gcc*_GccF$_{S18C,C43S}$
pRV610*gcc*_GccF$_{Y16F}$
pRV610*gcc*_GccF$_{D17N}$
pRV610*gcc*_GccF$_{G16del,\ S18_G19insG}$
pRV610*gcc*_GccF$_{His}$
pRV610*gcc*_GccF$_{HA}$
pRV610*gcc*$_{_TT}$GccF$_{His}$
pRV610*gcc*$_{_TT}$GccF

pRV610*gcc* Circular DNA sequence of 11199 bp

CTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGC

GTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCT

TTCTCGCCACGTTCGCCATCCATGGAAGCTGTCAGTAGTATACCTAATAATTTATCT

ACATTCCCTTTAGTAACGTGTAACTTTCCAAATTTACAAAAGCGACTCATAGAATTA

TTTCCTCCCGTTAAATAATAGATAACTATTAAAAATAGACAATACTTGCTCATAAGT

AACGGTACTTAAATTGTTTACTTTGGCGTGTTTCATTGCTTGATGAAACTGATTTTT

AGTAAACAGTTGACGATATTCTCGATTGACCCATTTTGAAACAAAGTACGTATATAG

CTTCCAATATTTATCTGGAACATCTGTGGTATGGCGGGTAAGTTTTATTAAGACACT

GTTTACTTTTGGTTTAGGATGAAAGCATTCCGCTGGCAGCTTAAGCAATTGCTGAAT

CGAGACTTGAGTGTGCAAGAGCAACCCTAGTGTTCGGTGAATATCCAAGGTACGCTT

GTAGAATCCTTCTTCAACAATCAGATAGATGTCAGACGCATGGCTTTCAAAAACCAC

TTTTTTAATAATTTGTGTGCTTAAATGGTAAGGAATACTCCCAACAATTTTATACCT

CTGTTTGTTAGGGAATTGAAACTGTAGAATATCTTGGTGAATTAAAGTGACACGAGT

ATTCAGTTTTAATTTTTCTGACGATAAGTTGAATAGATGACTGTCTAATTCAATAGA

CGTTACCTGTTTACTTATTTTAGCCAGTTTCGTCGTTAAATGCCCTTTACCTGTTCCA

ATTTCGTAAACGGTATCGGTTTCTTTTAAATTCAATTGTTTTATTATTTGGTTGAGT

ACTTTTTCACTCGTTAAAAAGTTTTGAGAATATTTTATATTTTTGTTCATGTAATCA

CTCCTTCTTAATTACAAATTTTTAGCATCTAATTTAACTTCAATTCCTATTATACAAA

ATTTTAAGATACTGCACTATCAACACACTCTTAAGTTTGCTTCTAAGTCTTATTTCCA

TAACTTCTTTTACGTTTCCGCCATTCTTTGCTGTTTCGATATGATATCAATTCATCAG

CAAGTTCTCTAATAGTTTTAGGCATGTTTTCCAAACCCCTTTCGGAGTTCAGCAAGC

ATTTCTTGTCTTGTCTGATCTCTTATTTTTTTATCATGCTCAATCTGTCGTTCGGCTA

AAATCTGTTTTTCAGTTGAGCCTAGTGGTAGGCCTTTAACTTTATCGGTAGCGCGCC

ATTTTTCTTGTTCTGTTAATTCACCATTATTCTGGATATTAAAAAGTTTTTGACGCT

CGTCTTGGAACTTACCTTGGGAAAAGTCGTTTGCGTCTTTCTTTTCAGGTTTCCAAG

TAAAGGTATAACCAATCACTGGTTTTCCACGACCTTTACCATATTTTTTCCTAACCG

TTAACCCTCTAAATAAGGGAGTTAATTCTTCTTTAATGGGTTTTATAACAAATTTAT

CAACGTTAGATGGACTACTCCAGTAACTTTTAGGCATATCGAGTAATTCAAAAAAAT

CTTCTTTAGAAAAATAAGCATATCCAGTGGTTCGGAACTGTTTTAGTAATCGAAACA

TAGTTTTTGCGTAACTACTCTTCAAATCTCTGAACTCTGCTAGAGCATAACGAACCC

AACTTTCAAGCTTGTTTAAAAGGGGTAAAGCACGTTCATAAACTTTTACGTCTACAT

AAGGTTCCTTTGCATCTCCATCGATTTTAAACTCTGTAAACAAGACAAAAAACTCTC

GAGTTAATCCACTTTTACTTCGCCTACCAAAATGTAATCCCATCATTTTTTCATAAGT

TCTCTGAATATCATCTTCAAAACGGTTATTTGCAGTTGGTTTATAAGCACTTAATTC

TTTTAATTGATCAAAAGTAAAGCGAATGGTCTGATTACTCTTATCACGCATACGAGA

AATAATTGAAAAAAATAAATTCATTTCAACCGGAGTAAATTTTCGAAGTGGAATAGT

ATTCAACTCTGGATCATATTTAATAATTTCGTTTCCCATGCTGCACCTCCTAATCATT

ACAAATTTAATATAAGTTAATTTTATCAAGTAGTCAACTTTATAATACTACTTGATA

AACAGACAAAACCTACTCGATAAACAGACAAAACCTACTTGATAAACAGACAAAACC

TACTTGATAAACAGACAAAACCTACTTGATAAAGTTCTGTAAGTCTTGGGGGAGTAA

GGCTCAAGAGGGGTCTAAAAGAGGTTTAAAAGAGGTTTATAAAAGAGGTATATAAG

AGGCACCACTGTACGAGATCAAAACGGGCCCATATCATGGCTTTCCCCGTCAAGCTC

TAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA

AAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTT

TTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTG

GAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGA

TTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTA

ACAAAATATTAACGCTTACAATTTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGA

AGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTG

CTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAA

CGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACCGAA

TTCCTGCAGCCCGGGCATCAGCGCTAACTGCTAATTATTCACTTAGAGAAATTTAAA

GGAAGACAATAATTATTATGTTATTGTCTTCCTTATTTGCGTCATAAATTTTCCAAA

TAACTCTTTATTGTGGAAATTACAAAATCAAAGAATAAATACTCTATCAACGTAACT

TAATAAGTCTAAAACATTCACAAAACTAACAATGATAACTGCTACTACCACTACTAT

GATGCTTTATACCAAAACAATGCGAATACATATAATCACAGGTTCCCGAATCATAAC

CAGCACCACACATTGCTAAAGTATACCAACACCATGCAGGTTTTCCACCGTTGTTTT

GAGCCTTAGAAATTTCACTTATAGTAAGTGTCTTAACCAATTTACTCATTCCTAATA

CCCTCCTTTTAATAAGAAAATTATAAACCAAAAACAAACATATTGGAGAGCAACTAT

CTTATTCGGTAACTAAAATGTCTTCATCGGGATTAAATCACTAATCTTTTTTGTATA

TCTTCTTGAAACCTCGCAAGTATCACCATTTTCAAAATATAGGATCCTATTGCTCTT

GTTAAATTTAGTGATGTTTTTTAGGTTAACTAAGGCTCCTTCATGGCAAGCAGTAAG

ATTACGACATTGATGATGAACTTGTTTTAAAGTCATACGGACTTGATATTCATCATA

TTTCAAATGAAATATTACCTTATGCGGTTGATTGTCAGTTATTACAAATAATATATC

ATCCAATCTTAGAAAAAACTTCTGTCTTTTGACAGAAATTTGTATAGTGTTTGAACT

ATTTAAGGGAATAGAATTGTGTATTAGAATGTTACAAATATCAGCAGCCAATATATT

TCCATCATCTTTTTGAAAAGTTTTGGTCATTAGTGCATCATAAAGCATATAATGATT

GAAAATTAAGTTTTCAAAAACATGTCTACTTGATCCAATGTAAATGAAGTATTTCCG

TGAATTAGGAAGCCTCTTGACAGCAATACTAATTTTTTTTATTAAGTCTATTTCTTC

ATCTAAATTTAAAATTACGATAGAGAGTGAACATATAAAGTTTTGGAAATTTTCAGA

GCTTGATGATAGTTTTATTGGATGACAAAACTTGATATGATTCGATAAATTGAGCAT

TATTTGTAATATATGGTTATTTAAATTTTGGCATTCCGGTTTATTTGTTAAAACAAA

TATTCTAGTCTGTGGAACTAAATTTTCTCCTTCTGCTTGATAAGCCATAAGTGTAAA

TCTCCCTTACTTATCCCTTTATTATTCCCCCACTGTATAATACTTAATACTTTGCCAT

TGTGATAGTGAATTATTGCAGGTGTCTGCTTGAATCCATATTTTTTCTTAAACTGTA

CCCATGAAGTCCATCCCTTTTGGCGTTCCCATGCTACGTTGAGAAAGACAACATTGC

TGGTCATATCTTCTTTTTTCAATTCATTAACTAAAATAGGATCAAATAAATTACAAT

CTGAACAAGTAGGCCGTCCAATATATACTAGAAAATCTCTTTTATCCGAAATTTCCT

TTTCAAAAGTTGGAATTGTTATTGAATTTAATTCCTTGTAAACCATAGTGTTTGTCT

TTGATTCTGAAATAATTTCAGCCCGAAGTACGTTCTCATGTTCATAAGAATTCCGTT

TCGTAATAATCACCGAAAATAATCCAATTACTAATAAAATAATGAGAAGTATTAGCA

AAATTAATACTAGTTTTCTTTTTTTTAATAGATTCATCTAAAGATCTCCCGTGTACCC

CTTATATTTAGTTGTTAAGTCCTCATATCCAAAGACTTGATTTATCCTTTGAACAAT

TTTTCCATGGTGTATAAACAATATGGTTGGTGTTGATGATATTCCCAATGTTCCCAA

AATAAAATCTTTGCTTGGATCTGCATTTATATCGATTGCTTTTACAGTGACACGTTT

ATTTCGAATAAAGCGGTTCAATATGGGAGTACTATGTTCACATCCTTCACAACCCTT

TTGGTAGAAGTAAACTACATCATCCTTCTTATTTTTTATGTCTGCTTGCAATTGTGT

CGCCTTTATTGAGTGATATAATGGCTCTTGTACCAAAGTGAGCCTTTGCCAAGAATA

TACTCCTATTATTGAGATAACACTAAGCGCAATAATAGTGATAATGGTTTTTTTCAT

AACAATGCCTCCCTCAATGATTCAGTAGAGTTTTTGTTAGTTTCTATTTGGTGCATC

TCACCATTTTTAACAGTGAAAATATCATCACATCTTATAGCCAATGTTAAATCGTGA

GTAATAAATATTAATTTAGGTTTGTTTTGTGATAATAAATAATTTTCTACTTTCAAT

TTTGTATTGATGTCCAGCCCATTCGTTATTTCATCAAAAATGATAAACCTACTTGGA

ATTACCATAGACCTTATAATATTTAATAATTGTTTTTGACCTCCAGATAATGAATAT

CCAGAGGATCCTAATTGTGTATATATTCCGTTTGGTAAATGATTCAAACATTCGTCA

AAACCAATACTACTAGCAATCCTATTAATTGATTTAACTGTGATATTTTTCCTACCA

AGAGTTATATTATTCATAACTGTATCCGAAAATATTTGTGGTGATTGTTCAACATAG

GTAATTGTATTACCTAAAGATTCTGCCCCTATTGCATAATAAGGTATGTTTCCTAAT

GATAGACCCTCTGCAGGATAAAAACCTGCCATTAGTTTAGCAATAGTTGACTTGCCA

CTCCCACTATTGCCAATTATTGCAGTTGGAAACTTTGAGAAAACACCGTTCATGTTT

TTAATTATTGGTGTAACCCCATCATATGAAAATGCTACCTTTCTTATTTCAATAGAA

TTTGAAAGTGTTAGTTTGCTTTTATTGTTATTGTCTTCAGGAGAAAGTAATAGATCT

AAGTAGCGTTCTGTTGCTACTTTTCCTTGAGACAGATTTACTTGCATAGTTATTATT

TTCACAAACGGATTTAGAACATTTCCCGTTAAAGCATAAAAACTAAGCAATGTTCCA

AGCACAATATTGCCCTTTATTACTAGAAAAGCCCCCACCGAGAATAAAATAATTCCA

AACAGTTGTGAAACACCATTTCTTGCTATCTGCTGTAATCTATCATATAAGCCAAAA

CTACGGTTTGCTGCAGTATACTTTTCCATACTTTTTGTTATTTTGGAATTGATGTAG

CTACCAGCATGAAAAGTTTTTATAGTTGAAAAATTTTTTATTTGTTCAATAATATCT

TGGCTAAATGTTTCCTGTGATTTAAACAAATTTGTACTTCTTTTTTTAATTGCGTCT

GAACTTAGATAAAATAATAATGCCATTAATATCATTGGAATAAACAATAAAATCGTC

AAAAATGGATTGATTCTAGCAAGAAGAAAAAGGGTTATCCCCATAGTAATAATGTCA

ATTGGTAAGTCACCTAAAAGCACAAGTTGTCTACCTGCAATTTGATAAATCGCATTA

AATCTTGTAATAATTTCACCATCTTCAAATGAATCTACTACTAACCGCTTCTTTTTGG

GAAAAGCCTTAAAGATGTCTTCATATAATTCTTTACTGATTGAATTACCAGTATAGG

CAATAAGGATTGCATTAATAACTTCCAAAAAAATTGATGCTACTTGAACAGTCAAAT

ATACGAGAGTTATAGTAGGAATCGCTGAAATAAACTTATTGGGAACAATGCGGTTAA

AATAGGCAGAATACATAGTTGCTAAGAATAGTGTTGTAGCAAATATAATTATTGAAA

GCATCCAATATATTAACAGGCATGCCCAGTGTTTCCTAAAACTAAACCACAAAATAG

GGACTTTTAAGCTTTCAACTTTTGTTTTTAGAATTTGAAAATTTTGTGATCTAGGTA

TCGTTAAGATAAGCGGCGTCCAATGCATCATAAATTTAGCCTTGTCAATACTTTCTA

TTTTTCCACTACCTGGGTCTGCCCAATATAATTTTGAACCAGAACACTTCGTGAGAA

CTACAAAGTGTAAATAGCCATTTTGCATAATTTGTGTCAGCACTGGAAATTCAATTT

CATCAAACACATCTGGATTCTTTACACACTTTTCTACAACAGACTCAACCTGAACTG

ACTTTAATCCGTTTATTATACTGATAAAATTAGTCCCATAATTTGTTGTTGTGATAA

TTTGCCTTATCTGCCAGTCATCAATTTTAATTTTACTAATCATACCAATGATTGTTG

CTACTGCTGCTGGTCCACAATCATTTTGGTCAATTTGTTTAATTATTCTCATTTTTG

CTGTCACCTTGTAATTTTGAATACAGTTTCACTGCAGATTTGACACTACTTTTTATA

TCGCCTGAAAAGTTTTTAGGAATTTCTGATATTAAAGGGAAAAGATGGCTGTAGTTG

GCACTTATTATGTCACATTCTAAAATAACCTGAGCTATATGATAGTAGTTTCCAGAT

ATATCGCTATTAACATCAATCATTCCTTTGTGTTTATTTAAATATCTCAACATTTTTA

CTTCTATAGACTTAGCTTCCGCTATAATTTCATTAATTTTAAATGATTCAATATAGT

ATATTGCATCACTGTCTTCCCCACCTGTTATCTGTAAAACATCTTTAAAAGATAATA

CTGCTTGAGTGGTTTTACCTTCTCTTAATAAAATACGGCCAAGCAATTTTTTTGCAA

ATGGCGAATATTTTTCTTGAAGACTGTCGGAAGCAACTAAGTCTAAGGTCCTTTTTA

CTAACTGTTTGAGTTTGTCCTGAGGCAGTACATCAAATCCATCACGTGCTAGTAAAA

ACGGCCATCTAGCATTTTGAGGCTCCTCACAAGTCATCTCTTGTAATAAACGAATGT

TTCTCTTTATTTTATCTTTATCTCGTAATACTTCTTTATCATATCCATCATGATAAAG

AATAATATCATCACATGCGAAGTGCCTAACATCGTAACCTAGTTTTTGATCTTCTTT

TCTTACTTCTTCGTGAATTTTTGCATAATAATGAAAGGACGACTTTTTAGGGAAGAT

TCTCCCTACTGTTTGGTATATCTGTCCTGAATGATCAGAAAATGTTGGGTTAATTAC

GAACTTAAAATTTTTTGCTTGAACTTTAAATAAAATTTTCTTTAGCTGAGCTCCATC

ATCAACATCCAACCACTCATCACTATCAACATAGAACACCCAGTCTTTAGATGCTAA

TTTGAGTGCCTTATTTCTAACCTCTGAGAAATCATTTTTCCAATTTGTTACTGATAT

TTTTGCTTGTGGCATATTCTTTTTAACCAGATGCACAGTTTCATCAGTTGAACCAGT

GTCTAAAACAATCAATTCATCTCCATTTTCCATATTTTGAGCAATTGAATTCACACA

TCTTTCAATAGTTGCTTGAGAATTTTTACATATTACAATATATGATACATAAATATG

ATGTCTAAATTGATCAATGTTTGTTAAATTTCCAAAATCAAATGATTTATGGACAGG

CCTTAAATGAAGGTTCAAATAACTGTCAATTTCATTTTGTCTATTTTTCATTCTCATT

CCTGCTTTCGTGGTATCATGTATTTAAGAAAAGAGGTAACAAAATGGTTACTAATAT

TTGTATTATTCCTTCTGCTGTCGTTGGTATTCAAGAAAGTAGTATATCAGGCAGTAT

GATAATAGACTCAGCAATCTCAGCACATTTTACTCACCGGCTAAAAAAGTCGTTGCA

AGGAGTTCCGAACATCTCAAAAATTTTTACTCTATCAGAATTCGAGTCAAATTGGGA

GCCAAATCCAAATACTATATATATTTTCCCAGATAACGTGTATCATGTTTTACCCGC

CTCTTTTTTAAACAAAGCTAAGTACGTGAAGTTTCCTAACAGTAAGATATTTTCGTC

AGATCTCAAGGAATTATCTCATGCAATTAAAATAAAATTATTTCAATATGATGCTTA

GTTACATCGTACATCAGAAGATATTATTTTAGTATTTACATGCACAGCATGCTTCAA

TTCAATCCTAGATATTAAGATCCATAATTTCTGTAACAGTTCTTTTGTTCCTTGTTT

TCTGGATTGTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTA

GTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA

TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC

CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC

TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG

GAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCG

CTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTT

ATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA

AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC

CTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC

TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGA

CCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTT

CTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGG

GCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC

GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA

ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG

CCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA

GTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGT

AGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAA

GAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGT

TAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAAT

TAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGT

TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCC

ATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCT

GGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCA

GCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCC

GCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT

AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCG

TTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCC

CCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGT

AAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACT

GTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC

TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAT

ACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGG

CGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGT

GCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAA

ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT

ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCAT

GAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCAC

ATTTCCCCGAAAAGTGCCAC

pET28a_*gccH* Circular DNA sequence of 5650 bp

ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCC

CAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTC

GGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGCTAAGCA

TCATATTGAAATAATTTTATTTTAATTGCATGAGATAATTCCTTGAGATCTGACGAA

AATATCTTACTGTTAGGAAACTTCACGTACTTAGCTTTGTTTAAAAAAGAGGCGGGT

AAAACATGATACACGTTATCTGGGAAAATATATATAGTATTTGGATTTGGCTCCCAA

TTTGACTCGAATTCTGATAGAGTAAAAATTTTTGAGATGTTCGGAACTCCTTGCAAC

GACTTTTTTAGCCGGTGAGTAAAATGTGCTGAGATTGCTGAGTCTATTATCATACTG

CCTGATATACTACTTTCTTGAATACCAACGACAGCAGAAGGAATAATACAAATATTA

GTAACCATATGGCTGCCGCGCGGCACCAGGCCGCTGCTGTGATGATGATGATGATG

GCTGCTGCCCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGG

AATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGA

GATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGC

GGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTT

CGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGG

GACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACG

GCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTC

GAGATCCCGGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGC

CCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGT

CGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAG

CCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATT

ACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCG

TTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAAT

CTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGC

GTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTG

ATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACT

AATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATT

TTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCAC

CAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTG

GCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAA

GGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGC

ATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGC

GCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGA

CGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTT

TCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGC

GGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGC

GCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGC

ACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGT

TAGCTCACTCATTAGGCACCGGGATCTCGACCGATGCCCTTGAGAGCCTTCAACCCA

GTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTC

TTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGC

GAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGG

AATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGG

CGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCCCACGGGTGCGCATGATCGTGC

TCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAAT

GAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACC

TGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG

AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTA

CCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATT

TTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGT

AACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATC

GGTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAA

CAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTT

CTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATCTGTGAATCGCT

TCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGG

TGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGA

TGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGG

GCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGC

GGCATCAGAGCAGATTGTACTGAGAGTGCACCATATATGCGGTGTGAAATACCGCAC

AGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGAC

TCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTA

ATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGG

CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT

CCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCC

GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC

TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT

GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC

CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGG

TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGC

CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGA

AGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGC

TGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA

CCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAG

GATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAA

ACTCACGTTAAGGGATTTTGGTCATGAACAATAAAACTGTCTGCTTACATAAACAGT

AATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGA

TTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTC

GGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTT

GTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCA

GACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTA

CTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGG

TATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCC

TGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTAT

TTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATT

TTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAA

CTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAAC

CTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATC

GCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCT

TCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAA

TTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGAATTAATTCATGAGCGGATA

CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCG

AAAAGTGCCACCTGAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTT

TTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAA

ATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTC

CACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCG

ATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTA

AAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAG

CCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGC

GCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGC

GCCGCTACAGGGCGCGTCCCATTCGCCA

Mass spectrometry methods continued

In all experiments, MS1 scans were acquired over a mass range of 500-2,800 m/z with detection in the Orbitrap mass analyser at a resolution setting of 70,000 and collected in profile mode. The raw data was visualised by Xcalibur Qual Browser (ThemoFisher Scientific, USA) and an average of the spectrums with baseline subtraction was used for analysis (Table 6.5).

Table 6.3: Chromatography instrument configuration

LC system	Dionex Ultimate 3000 Quaternary RSLC system (ThermoFisher Scientific)
Mass spectrometer	Q Exactive^{TM} Focus Hybrid Quadrupole-Orbitrap^{TM} (ThermoFisher Scientific)
Ionization source	Heated Electrospray Ionization (HESI-II)(ThermoFisher Scientific)
Analytical column	N/A (bypass mode)
Column oven	temperature 25^{o}C
Gradient	Isocratic 50:50 (A:B) for 1 minute
Flow rate	0.1 ml/min
Mobile phase	A: 0.1% Formic acid/ 2% MeCN/ water B: 0.1% Formic acid/ 98% MeCN/ water
Injection volume	2 μL

Table 6.4: Mass Spectrometer Settings

Capillary temperature [°C]	320
Aux gas heater temperature [°C]	250
S-Lens RF level [%]	55
Source voltage [kV]	3.5
Polarity	Positive
Shealth gas flow rate	15
Aux gas flow rate	2
AGC target	3e6
Max. injection times [ms]	Auto
Full MS mass range	500 - 2800 [m/z]
Resolution	70,000
No. of microscans	1
Spectrum data type	Profile

Table 6.5: Xcalibur Qual Browser Settings

Averaging peak spectrums	Across 50% peak width
Baseine subtraction:	
- Polynomial order	2
- Below curve [%]	10
- Tolerance	0.01
Mass tolerance [ppm]	10