University of Birmingham

School of Biosciences

Study into chromosome structure and gene expression

A research project report submitted by

Stephen Bevan

as part of the requirement for the degree of MRes in
Molecular and Cellular Biology

This project was carried out at: School of Biosciences, University of Birmingham

Under the supervision of: Professor S. Busby Date: Oct ‘11 – Feb ‘12

Contents

List of Figures and Tables

List of Abbreviations

CIP – Calf Intestinal Phosphatase
DAPI- 4,6 diamidino-2-phenylindole
dNTP's – Deoxynucleotide Triphosphates
***Dps* –** DNA binding Protein from Starved cells
EHEC – Enterohemorrhagic *Escherichia coli*
EPEC – Enteropathogenic *Escherichia coli*
ETEC -Enterotoxigenic *Escherichia coli*
***Fis* -** Factor-for-inversion stimulation protein
FITC – Fluorescein Isothiocyanate
FLP – Flippase Recombinase
FtsK – Filamenting Temperature Sensitive mutant K
GFP- Green Fluorescent Protein
HF – High Fidelity
H-NS – Histone-like nucleoid-structuring protein
HU – *Escherichia coli* protein that mediates chromatin compaction through DNA bending
IHF – Integration Host Factor
IPTG- Isopropyl β-D-1-thiogalactopyranoside
Kan – Kanamycin
***Lac* –** Lactose
LB – Luria Bertani broth
LRP - Leucine regulatory protein
LysE- Lysis Gene E
***Mal* –** Maltose
mntH – Gene encoding for the principal Manganese/divalent cation transporter
mntP – Gene encoding a putative efflux pump for manganese
mntR – Transcriptional repressor acting in response to changes in cellular levels of manganese ions
mntS – Gene member of the mntR transcriptional miniregulon
MukB- Chromosome partition protein MukB
NAP's- Nucleoid Associated Proteins
OD – Optical Density
OxyR -Bifunctional regulatory protein sensor for oxidative stress [
PBS- Phosphate buffered saline
PCR- Polymerase Chain Reaction
***RhtA* –** Resistance to Homoserine and Threonine
SacB – A native *Bacillus subtilis* gene encoding for a levansucrase enzyme
SMC – Structural Maintenance of Chromosomes
TAE- Tris/Acetic acid/EDTA
TRITC – Tetramethyl Rhodamine Isothiocyanate
ypeC – Uncharacterised protein, the structural gene of which is adjacent to the structural gene of mntH

Abstract

This study focuses on the mntR miniregulon, the associated genes of which are involved in manganese homeostasis. A lacI-GFP fusion tagged to *dps*, a gene known to be repressed by mntR upon increased manganese levels, has been used to image the gene's position within the cell using *lac* operators upstream of the relevant gene. The number of GFP foci exhibited also gives an indication as to the accessibility of the gene under varying growth conditions. Results presented in this study indicate that *dps* during exponential growth is a gene with reduced accessibility when compared to the *araC* gene, a gene responsible for transcriptional regulation of the arabinose operon. The accessibility of the gene is itself also altered slightly upon cellular growth with manganese and iron and thus confirms *dps* induction by iron in conditions of low manganese as hypothesised by Yamamoto et al. Finally this study also confirms a model proposed by Espeli et al with regards to the cellular positions of specific genes dependent on their macrodomain location.

Introduction

1.1 Escherichia coli: a model bacterium

Escherichia coli is a gram negative, rod shaped bacterium found to populate, in particular, the large intestine of endotherms. The facultative anaerobic nature of the bacterium lends itself to growth in these highly specialised conditions. The bacterium itself was discovered in 1885 by Theodor Escherich, from whom the genus takes its name (emedicine.medscape.com)
To this day, *Escherichia coli* is still one of the most infectious bacterium affecting the intestinal region of the host. The urinary tract is the most frequently affected site with over 90% of such problems a result of *Escherichia coli* infection; however, six varieties, such as EPEC, EHEC, and ETEC are also known to promote enteric infections commonly leading to diarrhoea. Prognosis of such infections is usually good; however in worst case scenarios (particularly in immuno-comprimised patients) the onset of pneumonia may be encouraged. In the neonatal population, *Escherichia coli* mediated meningitis currently possesses an 8% mortality rate (Robins-Browne and Hartland, 2002 and emedicine.medscape.com).
Non pathogenic *Escherichia coli* strains are used in biological research and biotechnology. The bacterium is one of many model organisms experimentally used to further knowledge and understanding of biologically crucial processes and pathways. Many advantages for using *Escherichia coli* in this fashion centre on its genomic attributes, however, the increased growth rate of bacteria over other potential model organisms also adds to its suitability. *Escherichia coli* possess a relatively small genome and therefore mapping and sequencing of specific regions is possible. Bacterial competency to transformation, transfection and conjugation, along with the possibility of gene doctoring protocols discussed below, make the transfer and incorporation of genetic constructs into the host genome possible with the potential of integrating highly defined mechanisms of expression control (Lodish et al, 2007).

1.2 The bacterial nucleoid and Escherichia coli chromosome

The *Escherichia coli* genome comprises of a single circular DNA chromosome consisting of 4,639,221 base pairs coding for 4,279 genes (figures correct for *Escherichia coli* laboratory strain K-12, differing strains may vary slightly) (http://ecocyc.org/). Unlike in eukaryotic cells, prokaryotic chromatin is not restricted to a membrane bound nuclear structure; instead, due mainly to dramatic improvements in electron microscopy, it is believed highly dense prokaryotic chromatin is packaged in a pseudo-structure in the cytoplasm known as the nucleoid. The structure of this nucleoid is usually described as 'quite irregular' and will vary between bacterial species. Despite the lack of integrity provided by membranes, the nucleoid adopts a relatively distinct central region in the cell with chromatin highly compacted within its confines. An absence of ribosomes is another feature that also characterises the spatial boundaries of the nucleoid (Kleppe et al, 1979 and Thanbichler et al, 2005).
As previously mentioned, the *Escherichia coli* genome is over 4.6 million base pairs in length, a chromosome that linearly would span 1.6 mm. As the average dimensions of an *Escherichia coli* cell are limited to 1-2 μm in length and 1 μm in width, the spatial constraints required to fit the chromosome within the cell, let alone the nucleoid, are considerable (Luijsterburg et al, 2006). It is estimated that chromosome compaction to a scale of 10^3 would be required; it is however important that DNA compaction does not affect any critical associated cellular processes such as transcription, replication and segregation (Travers and Muskhelishvili, 2005). Several methods of compaction and chromosomal organisation have been proposed.

1.2.1 Macrodomains

Throughout the *Escherichia coli* genome there have been four distinct regions identified, known as macrodomains. These macrodomains, ORI, TER, LEFT and RIGHT, are defined as regions within the chromosome that possess intra-domain interactions but do not have the ability to interact with other distinct domains. Two further macrodomains, NS, whose regions flank either side of the ORI macrodomain, also exist however, these regions are believed to have a higher interactive ability than the four principal macrodomains mentioned earlier (Figure 2) (Valens et al, 2004). Fluorescent tagging of all defined macrodomains confirm that these regions occupy distinct regions of the cell throughout the cell cycle (Figure 1). Closer

inspection of the two macrodomains flanking the ORI indicates a greater mobility, an observation that could well explain their high interactive abilities (Espeli et al, 2008).

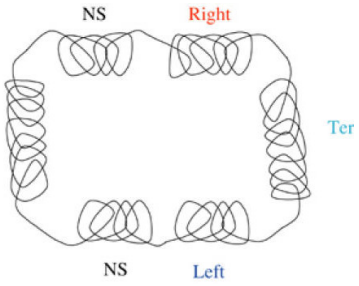


Figure 1 Potential model for macrodomain organisation of the *Escherichia coli* genome (Valens et al 2004)

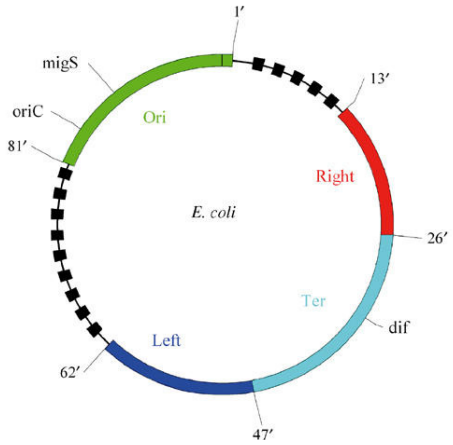


Figure 2 Representation of *Escherichia coli* macrodomains and there relative position. The replication origin, oriC, and termination site, dif, are indicated. (Valens et al 2004)

1.2.2 Supercoiling

In both eukaryotes and prokaryotes, double stranded DNA is coiled, the helical turns of which arise in a structurally optimal manner. To compensate for the significant spatial constraints of the *Escherichia coli* genome, supercoiling exists; a process that contributes significantly to reducing the volume the chromatin occupies within the cell. Two variants exist, positive and negative supercoiling, which result in either an increase, in the case of positive supercoiling, or a reduction in the number of optimal helical turns exhibited by the DNA chromosome. The resulting torsional stress induced on the DNA causes supercoiling whereby the double helix is described as being folded back down upon itself or coiled around itself to form ‘coiled coil’ sections of DNA (Travers and Muskhelishvili, 2005 and Peter et al, 2004).

Supercoiling is an enzyme mediated process brought about through the actions of varying topoisomerases. This enzyme group works antagonistically in controlling the distribution of

supercoiling through the DNA; gyrase and topoisomerase IV in particular, favour negative supercoiling (Travers and Muskhelishvili, 2005 and Peter et al, 2004).
At a local level, RNA polymerase can affect the stability and distribution of supercoiling during transcription. The introduction of positive and negative supercoiling, upstream and downstream of RNA polymerase respectively, is crucial for the functioning of RNA polymerase (Peter et al, 2004). Given the high activity of RNA polymerase within the cell, these constant changes in local supercoiling would not be energetically advantageous to the cell, therefore, to compensate, approximately 400 topologically independent supercoiled domains exist (Luijsterburg et al 2006). These domains were confirmed through the discovery that it would take many 'nicks' of the DNA to completely disrupt the chromatin structure (Delius et al, 1974). The formation of these domains can not only control supercoiling at a local level, particularly given the influence of RNA polymerase, but also contribute to compacting distant regions of the genome. Environmental changes, such as oxygen stress and temperature alter the levels of gene expression throughout the genome and, therefore, also have a significant global effect on the distribution of supercoiling (Peter et al, 2004).

1.2.3 Nucleoid Associated Proteins (NAPs)

Nucleoid associated proteins are often closely compared to eukaryotic histones as a result of their similar functions in the organisation of chromatin. There are different subtypes of NAP; they all function to provide increased compaction and stabilisation of prokaryotic supercoiled loops. In general there are two distinct sets, those responsible for DNA bending and those responsible for bridging separate DNA supercoiled loops.
Of those responsible for bridging DNA supercoiled loops, Histone-like nucleoid-structuring protein (H-NS), Chromosome partition protein MukB and a leucine regulatory protein (LRP), are the principal mediators. A dimeric protein, H-NS, projects two DNA-binding C termini in opposite directions (Luijsterburg et al, 2006). The ability of these dimers to bridge adjacent DNA loops throughout the chromosome, preferably those regions that are natively curved, will therefore facilitate significantly increased compaction (Figure 3D). Proof of this hypothesis is observed in H-NS over-expressed mutants, resulting in nucleoid compaction above and beyond that witnessed in wild type chromosomes (Thanbichler and Shapiro, 2006). In a similar manner to H-NS, MukB and LRP also bridge topologically unique supercoiled loops and therefore further compact the chromatin. Although the mechanism for DNA binding

in MukB is unclear, it is believed to once again form a dimeric Structural maintenance of chromosome (SMC) complexes that forms a central core structure allowing the bridged loops to emanate radially (Figure 3H). LRP protein, through its native octamer or hexadecamer structure, has the ability to bridge DNA on a much larger scale. Although DNA binding is generally mediated through a specific consensus sequence, it is believed that leucine residues are able to bind with a higher affinity. In much the same way that MukB forms a core structure, LRP octamers form a central disk region that mediates both the 'wrapping' of DNA around itself and also bridging events discussed earlier (Figure 3L). Both of these compaction methods are themselves promoted by upstream DNA supercoiling. Although present in low levels throughout the cell, their role cannot be understated, 10 % of all *Escherichia coli* genes are affected by LRP through its binding affinity for leucine residues alone (Luijsterburg et al, 2006).

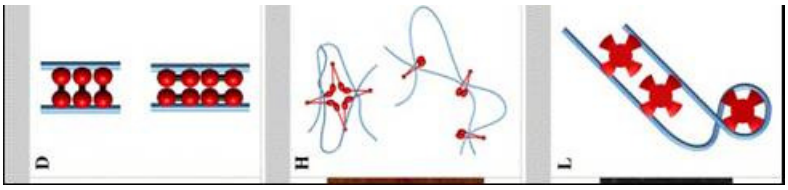


Figure 3 : D) Model illustrating the bridging mechanisms of H-NS, H) Gathering of DNA coils using SMC and L) LRP mediated DNA wrapping (Luijsterberg et al 2006)

In *Escherichia coli* there are three main DNA bending proteins associated with chromosomal organisation: Integration Host factor (IHF), HU and Factor for inversion stimulation protein (Fis). Integration host factor (IHF) protein binds DNA through a specific consensus sequence recognised by two flexible ß-sheets inserted into the minor groove of DNA. The binding of these flexible ß-sheets, in particular apical proline residues, to the DNA induces a hydrophobic effect causing sharp bends in the structure of the DNA (Figure 4G). Manipulation experiments in *Escherichia coli* have revealed that IHF mediated bending has the ability to reduce the length of DNA by up to 30 %. HU functions in a very similar manner to IHF, apical proline residues on HU bind to the minor groove of DNA, once again inducing bending of DNA (Luijsterburg et al, 2006). Unlike IHF bending, HU promotes flexible bends that causes negative supercoiling of the double helix (Figure 4G) (Thanbichler and Shapiro, 2006). Compaction via this method is estimated to reduce the length of DNA by up to 50 %. The factor for inversion stimulation (Fis), although classified as a DNA bending protein, has a secondary mode of action through the formation and stabilisation of DNA micro loops beyond those established as a result of supercoiling. The formation of these micro loops is

however, believed to counteract the effect Fis has on a reduction in negative supercoiling on a global level. Fis imparts steric hindrance to bound DNA preventing increased supercoiling, whilst the protein also acts as a repressor towards DNA gyrase activity and expression (Travers and Muskhelishvili 2000).

In terms of DNA bending, compaction is mediated through the interactions of the DNA flanking Fis binding sites and the peripheries of the Fis dimer. For Fis to function effectively as a bending protein, initial bending of the DNA is primarily required and therefore it is likely Fis is unable to act alone in facilitating compaction via bending (Figure 4K) (Luijsterburg et al, 2006).

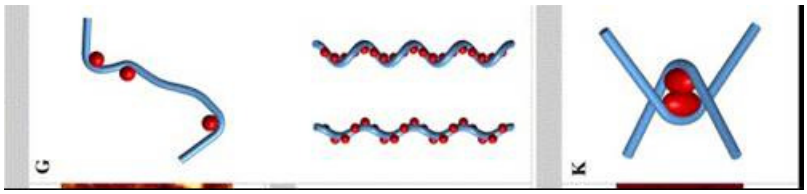


Figure 4 G) Models for DNA compaction mediated by HU and IHF, K) DNA Compaction model mediated by Fis (Luijsterburg et al, 2006)

1.2.4 Macromolecular crowding

In terms of DNA organisation, macromolecules are defined as non-DNA binding proteins and cytoplasmic RNA. During varying stages of the cell cycle, the levels of these macromolecules can be high. Entropy- driven organisation between these sizeable macromolecules and prokaryotic chromatin results in the aggregation of nucleic acids and therefore compaction of the chromosome. The afore mentioned process also aids the binding affinity of NAPs to the chromosome promoting further downstream compaction of the nucleoid (Thanbichler and Shapiro, 2006)

There is a clear relationship between all the compaction processes mentioned thus far. This suggests co-operative effects that combine to allow overall compaction to take place. The relationships between macromolecular crowding, NAP recruitment and Fis and primary bending via IHF/HU illustrate this feature. The key consideration regarding compaction however is ensuring that within the global chromatin organisation, particular genes and operons are accessible to RNA polymerase and transcription regulators during periods of transcriptional activity.

1.3 Bacterial transcription

1.3.1 RNA polymerase and sigma factors

Replication is initiated at a specific point on the chromosome, aptly named the origin of replication. As the name suggests, this site is present in the ORI macrodomain mentioned previously. Replication continues in a bidirectional manner until a termination site, defined as the dif site, within the TER macrodomain is reached.

Prokaryotic transcription, in concordance with the eukaryotic process is catalysed by RNA polymerase. Initiation of transcription is brought about through the binding of a sigma factor associated with RNA polymerase to the -10 and -35 promoter elements on DNA. Prokaryotic RNA polymerase has the ability to bind to a range of sigma factors; each individual factor itself is responsible for initiating transcription of gene clusters with associated functions. Examples of these include those encoding heat shock proteins, stationary phase proteins and the 'housekeeping' genes, transcription initiation of which is carried out by sigma factor 70 (Browning and Busby, 2004)

1.3.2 Regulation of bacterial transcription

The majority of bacterial transcription is regulated through transcription factors. Transcription factors possess an ability to either repress or activate gene transcription. Transcriptional repressors bind to DNA and inhibit transcription through steric hindrance, either through direct binding to the gene's promoter or through conformational structural changes to the coiling of DNA. Either mechanism inhibits polymerase binding or the formation of the DNA open complex. Transcriptional activators function by increasing the affinity of RNA polymerase binding to DNA or facilitating the interchange between closed and open complexes required for transcription (Browning and Busby, 2004). Binding of these activators usually takes place at a site upstream of the gene's promoter (Balleza et al, 2009).

1.3.3 The *Lac* Operon

Gene expression in response to environmental stimuli is fundamental to the propagation of bacteria. In response to these stimuli, genes of related functions are grouped together on the prokaryotic genome in structures known as operons. These operons are the targets for the majority of transcriptional repression or activation. The Lactose (*lac)* operon illustrates the balance between repression and activation based upon variable environmental stimuli. Expression of the *lac* operon is responsible for lactose metabolism within the cell. The operon is constitutively repressed through lacI repressor bound with high specificity to two operators. Only through induction via lactose, or a related homologue Isopropyl-β-D thiogalactopyranoside (IPTG) for example, does a conformational change in the structure of the repressor cause its dissociation and expression of the promoter (Lodish et al, 2007).
In this study, the Green Fluorescent Protein (GFP)- linked LacI induction/repression system coupled to gene constructs allows the formation of foci within the cell dependent on accessibility of the repressor to the operators. The formation of foci will not only give an indication as to the native location of the gene within the genome but also give an idea as to its accessibility on a local scale.

1.4 Current issues

The background regarding genomic compaction and prokaryotic transcription raises important issues such as:

- Does supercoiling and other compaction methods affect levels of transcription?
- Do genomic locations and thus position within the nucleoid have a consequence on transcription?

It has already been mentioned that the distribution of positive and negative supercoils in relation to RNA polymerase has an effect on transcription. Hypotheses to suggest that the level of supercoiling at a local level is also responsive to changes in the external environment have also been made (Peter et al, 2004). As a result it could be that supercoiling functions as a regulatory process by promoting or repressing the initiation of transcription following changes to environmental stimuli such as osmotic pressure, temperature and pH.

In terms of gene expression being affected by genomic position, it is important to consider the fact that not all genes are constantly expressed, if transcribed at all. As a result it is conceivable that the genomic structure is flexible to meet the demands of RNA polymerase accessibility. At a global level it would be energetically favorable for highly expressed genes to be located at sites readily accessible to RNA polymerase. As accessibility to RNA polymerase is a crucial factor in gene expression, postulations that genes expressed at low levels may be buried within the supercoiled loops, while highly expressed genes are collectively located on individual loops, potentially near the nucleoid periphery, could be made. It may also be the case that the native locations of a gene may be different upon gene induction, allowing increased exposure to RNA polymerase. Clearly all hypotheses are viable; however without further research using techniques such as the one employed in this study, and others such as 3C, no definitive answer to these problems can be published at this time.

1.5 Aim of the project

The aim of this project is to fluorescently tag binding sites of mntR protein and analyse the number and position of resulting foci within the cell under varying growth conditions. Through this, an indication whether transcription could be affected by a genes' location on the chromosome can be elucidated.

1.6 *Escherichia coli mntR regulon*

Manganese is an important micronutrient in, not only *Escherichia coli,* but also other species of bacteria. It functions as a catalytic enzyme cofactor as well as playing a role in protecting the cell from oxidative stress. This is achieved through OxyR-mediated neutralisation of reactive oxygen species and/or through displacement/replacement of iron in iron-cluster proteins, thus preventing downstream oxidative damage to the proteins. Despite these crucial functions, excess cellular manganese has been proven to be toxic and potentially lethal, with the most likely mechanism a disadvantageous result of its close relationship to iron and iron metabolism. Given this, the cell must control both the levels and distribution of manganese within the cell satisfying its advantageous attributes against potential toxicity (Waters et al 2011). Until recently very little information regarding manganese homeostasis has been published. It has now been suggested that mntR protein acts as transcriptional repressor in response to cellular levels of manganese ions. Following chromatin immunoprecipitation, four binding sites of mntR have been confirmed on the *Escherichia coli* chromosome, leading to the postulation of the mntR miniregulon. The four binding sites are adjacent to *mntH*, *mntR*, *mntP* (formerly known as yebN) and *dps* genes, the expression of which is affected by manganese and mntR repression (Yamamoto et al 2011).

The first of these transcription factor binding sites, *mntH* is located within the LEFT macrodomain and is believed to encode the principal bacterial manganese transporter (http://ecocyc.org/). Upon mntR binding to the *mntH* promoter, repression of *mntH* expression is triggered, preventing further toxic manganese accumulation. Variants of this system controlling manganese transport are exhibited not only in *Escherichia coli* but other bacteria such as *Bacillus subtilis* and *Staphylococcus aureus* (Yamamoto et al 2011). The *mntP* gene has recently been identified as a putative efflux pump for manganese. Sequence homology to Lysis gene E (LysE) efflux pumps, along with the mapping of mntR binding to particular *mntP* promoter elements and sequences, indicated a potential link to a mntR positively regulated efflux pump upon increased levels of manganese. In fact upon deletion of *mntP*, *Escherichia coli* cells exhibited a significant increase in sensitivity to manganese and over a two fold increase in intracellular manganese, results consistent with an inability to remove excess manganese (Waters et al 2011).

The primary function of DNA binding protein from starved cells (*dps)* is as a stationary-phase nucleoid associated protein that, mediated by Fis and H-NS repression, facilitates increased supercoiling during oxidative and starvation stress. During a 2011 study into the mntR regulon it was discussed that *dps* in *Escherichia coli* was repressed by manganese via mntR (Yamamoto et al 2011 and Grainger et al, 2008). *Dps* protein is coded for as part of the RIGHT macrodomain and is known to protect cells from oxidative stress through its ability to bind and seclude excess cellular Fe^{2+} (http://ecocyc.org/). It is therefore clear that the role *dps* occupies in reducing oxidative stress is coupled to the same role Mn^{2+} is known to fulfil. A hypothesis suggesting that elevated expression of *dps* should compensate for low levels of manganese in reducing the possibility of Fe^{2+}-associated oxidative stress would be plausible (Yamamoto et al 2011).

A paper published last year has also identified a potential fifth member to the proposed mntR miniregulon. *MntS*, formerly known as *rybA*, encodes a small 42 amino acid protein that has been proven to have altered expression levels dependent on manganese and mntR. The function of this small protein is still unclear, however results from the study indicate that *mntS* is repressed by manganese via mntR and similarly to mntH, both genes are expressed during periods of low intracellular manganese. Indeed in mntS mutants, over expression of mntS causes elevated sensitivity to high concentrations of manganese; whilst in mntS deficient cells, manganese mediated repression of mntH is disturbed, suggesting that mntS is involved in maintaining the availability and distribution of manganese within the bacterial cell. The most interesting result suggested that repression of mntS is not complete, residual mntS is always left in the cell at a sufficient concentration to mediate complete mntH repression. Given this result, one can postulate that mntS does not confer a definitive manganese import/export process but does, in fact, play a role in mediating manganese availability and distribution. A two function hypothesis regarding mntS can be proposed. During increasing intracellular manganese, mntS may facilitate the binding of mntR to the promoter DNA of mntH and therefore trigger the complete repression process discussed earlier. During basal levels of intracellular manganese however, mntS may act as a manganese chaperone, mediating correct distribution throughout the cell depending on demand (Waters et al 2011).

1.7 Fluorescent protein: gene fusions

In order to visualise the positions of the targeted genes, a construct of fluorescent proteins coupled to well defined control systems was utilised. Two distinguishable fluorescent proteins, Green Fluorescent protein (GFP) and mCherry, linked to LacI and MalI repression systems respectively, were used for separate mntR binding sites on the chromosome. In order to ensure that fluorescent foci are witnessed solely at the desired locations, the relevant endogenous operators for LacI and MalI have been removed from the host chromosome. Given this, LacI-GFP and MalI-mCherry repression is only permitted to take place at the operator repeats that will be incorporated in the donor plasmid constructs. LacI- GFP fusion has been incorporated as part of the host genome, whilst the MalI-mCherry fusion was transformed into the cell on a plasmid.

1.8 Gene doctoring

Gene doctoring, a method published in 2009 by Lee et al, is a highly controlled protocol adapted from gene gorging as published by Herring et al. Gene gorging functions by inserting specific DNA modifications onto a host genome making use of the λ recombinase operon. The λ red proteins incorporate a linear fragment of DNA into a region of the host chromosome targeted by suitable homology regions flanking the insertion position. Three proteins; *gam, exo* and *bet* are coded for within the operon, each responsible for a separate role within the recombination process. The *gam* protein protects the plasmid DNA from the host RecBCD degradation complex; *exo* generates the DNA overhangs, key in successful homologous recombination, whilst *bet* protein itself acts as a catalyst for recombination to take place (Lee et al, 2009). Unlike early recombination methods based on the λ red recombinase operon, gene gorging allows linearization of a plasmid through the presence of a Sce-I nuclease that itself is incapable of cleaving the host genome. Sce-I recognition sites, either side of the desired insertion sequence, allow cleavage of this region by the nuclease from the plasmid into a linear fragment (Herring et al, 2003).

The λ recombinase operon along with Sce-I nuclease gene and chloramphenicol resistance is encoded for on a mutagenesis plasmid, pACBSR, transformed into the host cell along with the donor plasmid containing the homologous insert and kanamycin resistance. The inclusion of an arabinose inducible promoter on the mutagenesis plasmid allows ease of controlled

induction whilst successful transformants are selected through the presence of chloramphenicol and kanamycin resistance respectively (Lee et al, 2009 and Herring et al, 2003).

Gene doctoring, as carried out in this investigation, incorporates these fundamental principles whilst conferring increased selection as a result of the *SacB* gene. The gene, itself present on the pACBSR mutagenesis plasmid and donor plasmids, encodes for a levansucrase enzyme catalysing the hydrolysis of sucrose and the synthesis of high molecular weight levan polymers. In *Escherichia coli*, the expression of SacB in the presence of sucrose is lethal and therefore provides definitive selection for pACBSR expression and donor plasmid linearization (Pelicic et al, 1995 and Lee et al, 2009). In combination with kanamycin resistance, transferred from the inserted fragment, successful recombination can be clearly indicated.

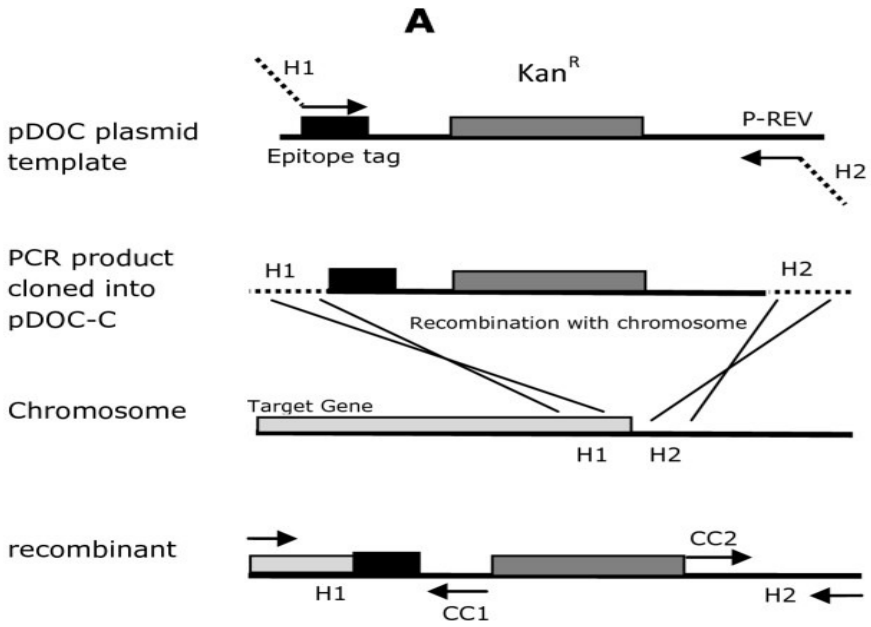


Figure 5: Schematic outlining the principles of Gene Doctoring (Lee et al 2009). Diagrams illustrate the incorporation of the two homology regions in combination with, in this study, several *lac* operators indicated by the region flanking CC1 and CC2

2 Materials and Methods

Strain	Description	Origin
MG 1655	*recA56, araD139 (ara-leu)7697, lacX74, galU, galK, hsdR, strA*	Cherepanov et al, 1995
RLG 221	*F^-, λ^-, ilvG, rfb-50, rph-1*	R. Gourse
pDOC-G	*LacI GFP – Kan*	Lee et al, 2009
SXB1	20 Maltose (Mal) operators inserted on the chromosome of MG 1655 in between ypeC and mntH genes.	This study
SXB2	20 Lac operators inserted on the chromosome of LacI – GFP – Kan strain in between dps and rhtA genes.	This study
LR06	22 Lac operators inserted between yabI and thiQ genes indicating the location of araC	L. Rowley Unpublished

Table 1: *Escherichia coli* strains utilised in this study

Plasmid	Description	Origin
pLR8	Plasmid containing multiple Lac operator repeats	L. Rowley Unpublished
pJB32	Template plasmid used for cloning in homology regions	J. Bryant Unpublished
pLR17	Plasmid containing multiple Mal operator repeats	L. Rowley Unpublished
pACBSR	Mutagenesis plasmid containing the lambda red recombinase operon and Sce1 meganuclease as well as both Ampicillin and Chloramphenicol resistance	Scarab (Lee et al 2009)
pCp20	Plasmid encoding the FLP recombinase protein under the influence of a temperature sensitive origin and a phage promoter	Cherepanov et al 1995
pSB1	ypeC homology region in pJB32	This study
pSB2	dps homology region in pJB32	This study
pSB3	mntH homology region added to pSB1	This study
pSB4	rhtA homology region added to pSB2	This study
pSB5	pSB3 with inserted Mal Operator repeats	This study
pSB6	pSB4 with inserted Lac Operator repeats	This study
pLER106	Plasmid encoding the MalI: mCherry fusion	L. Rowley Unpublished

Table 2: List of plasmids utilised in this study

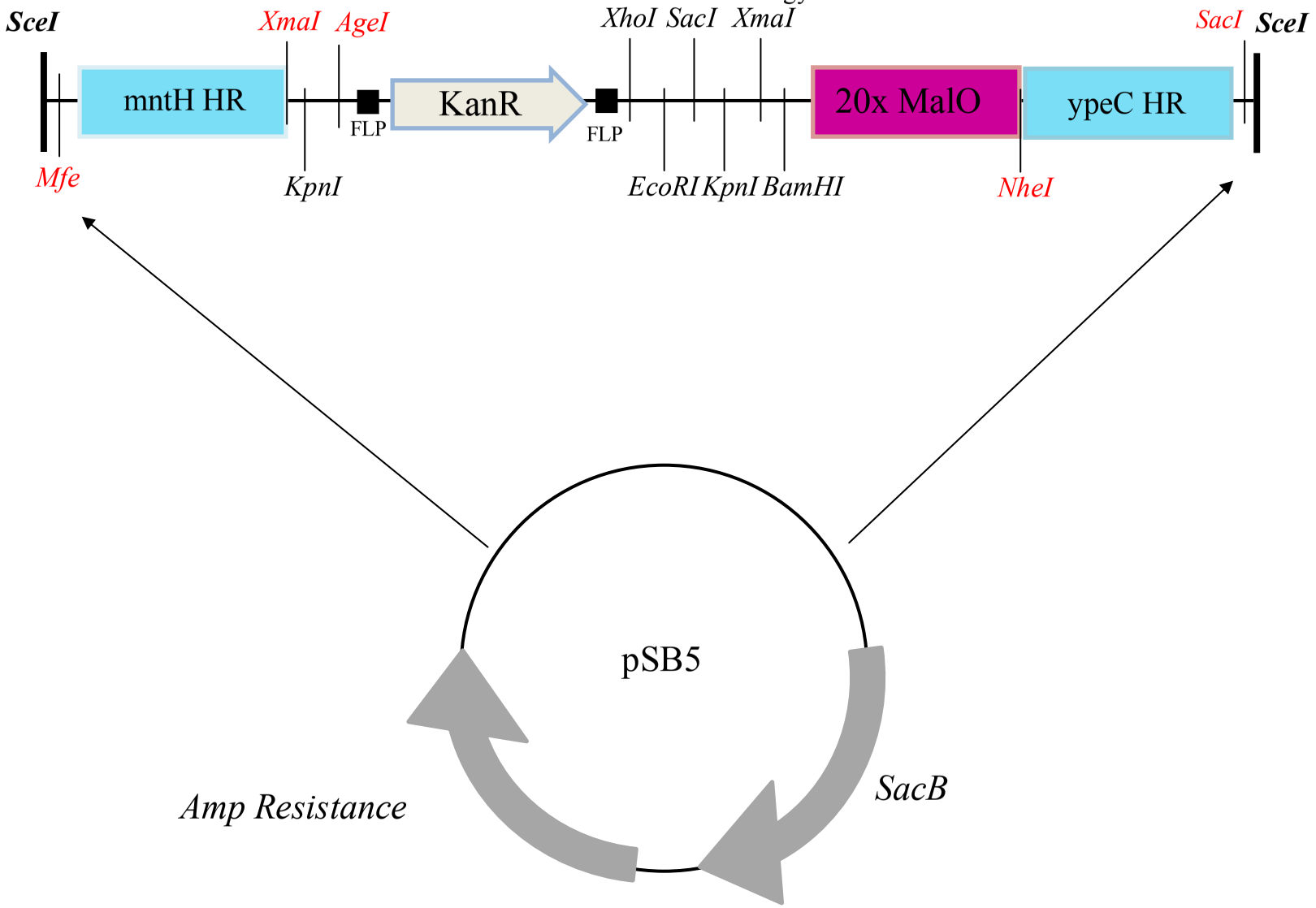


Figure 6: pSB5 plasmid map

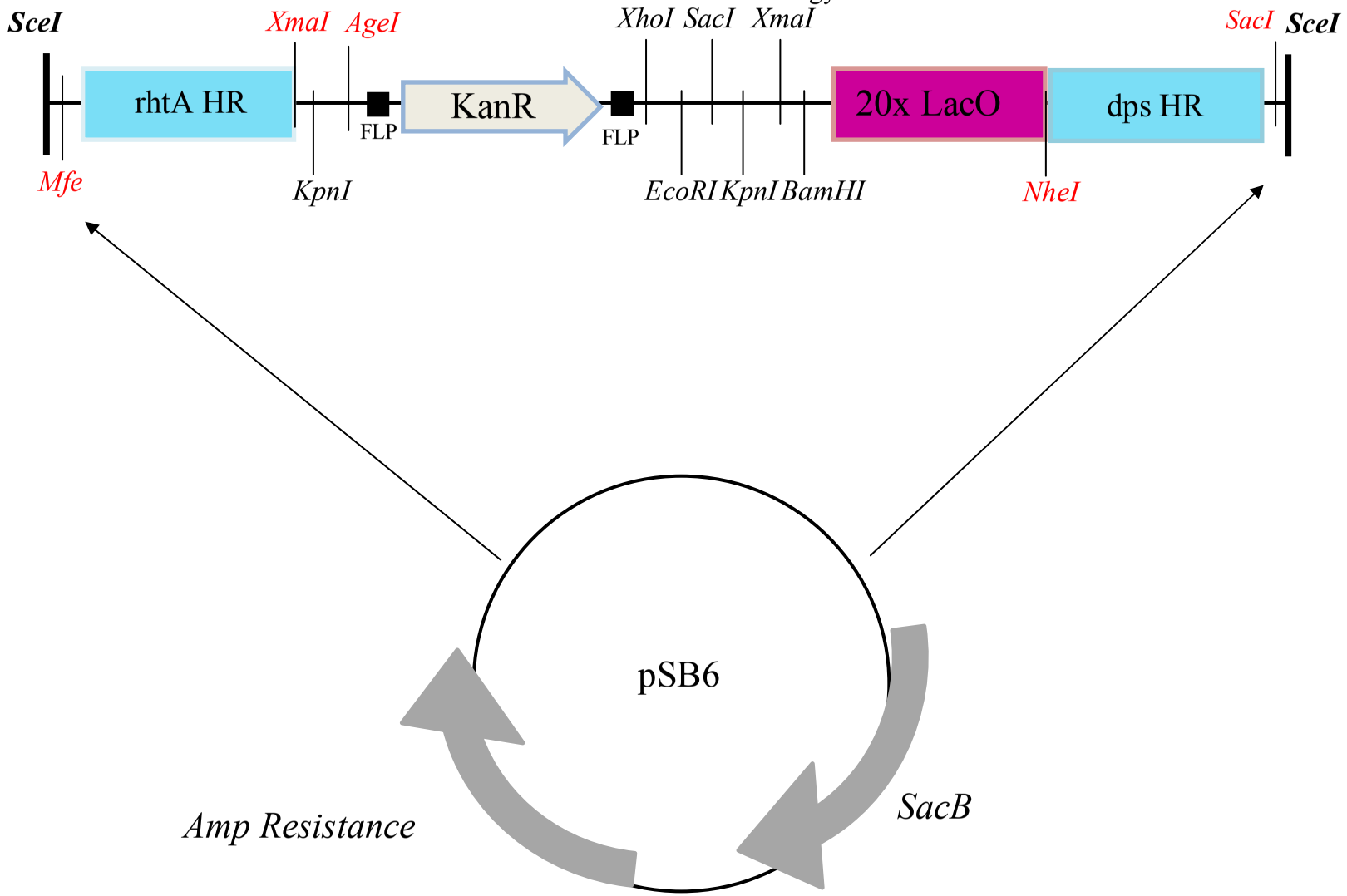


Figure 7: pSB6 plasmid map

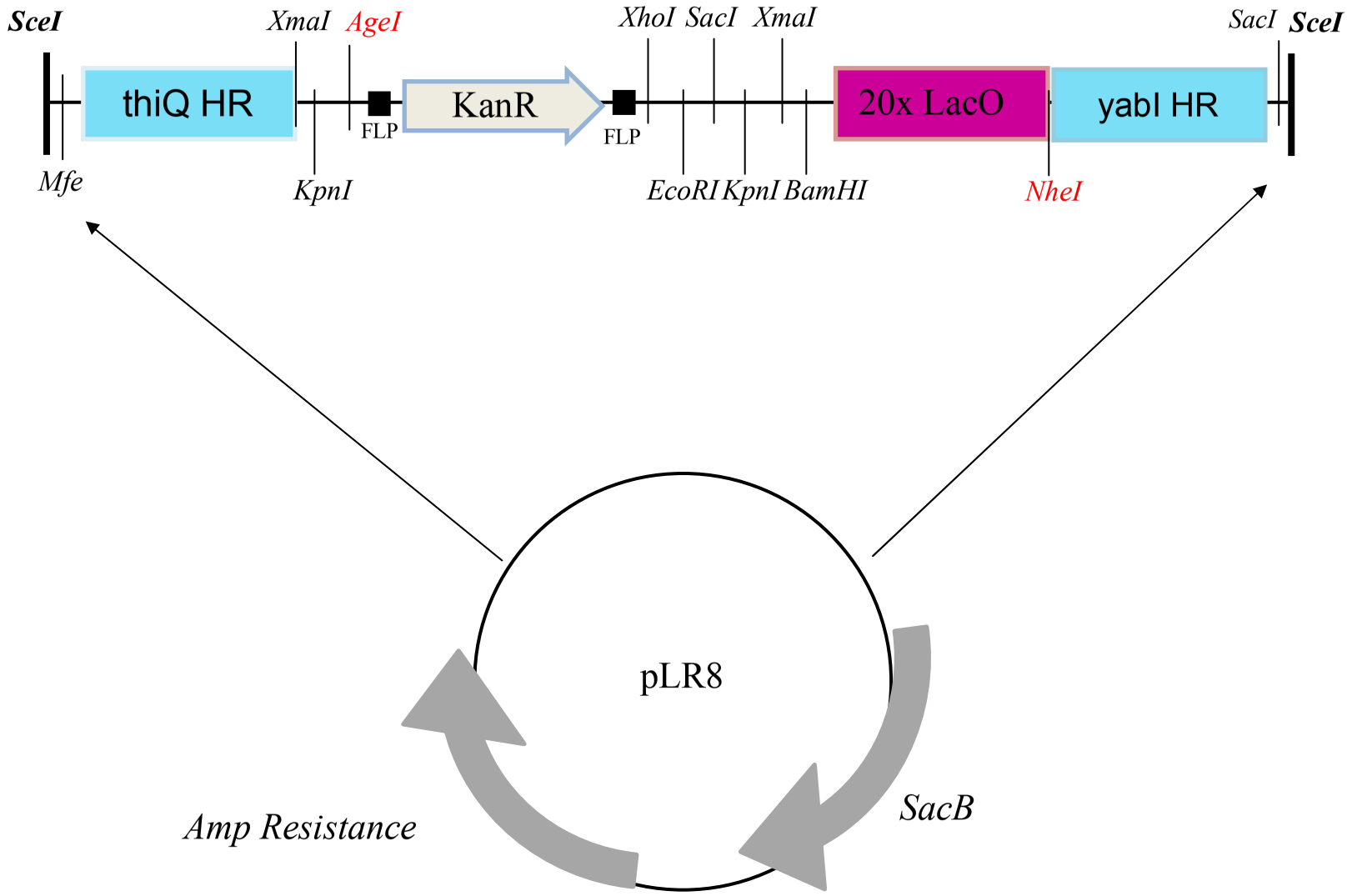


Figure 8: pLR8 plasmid map

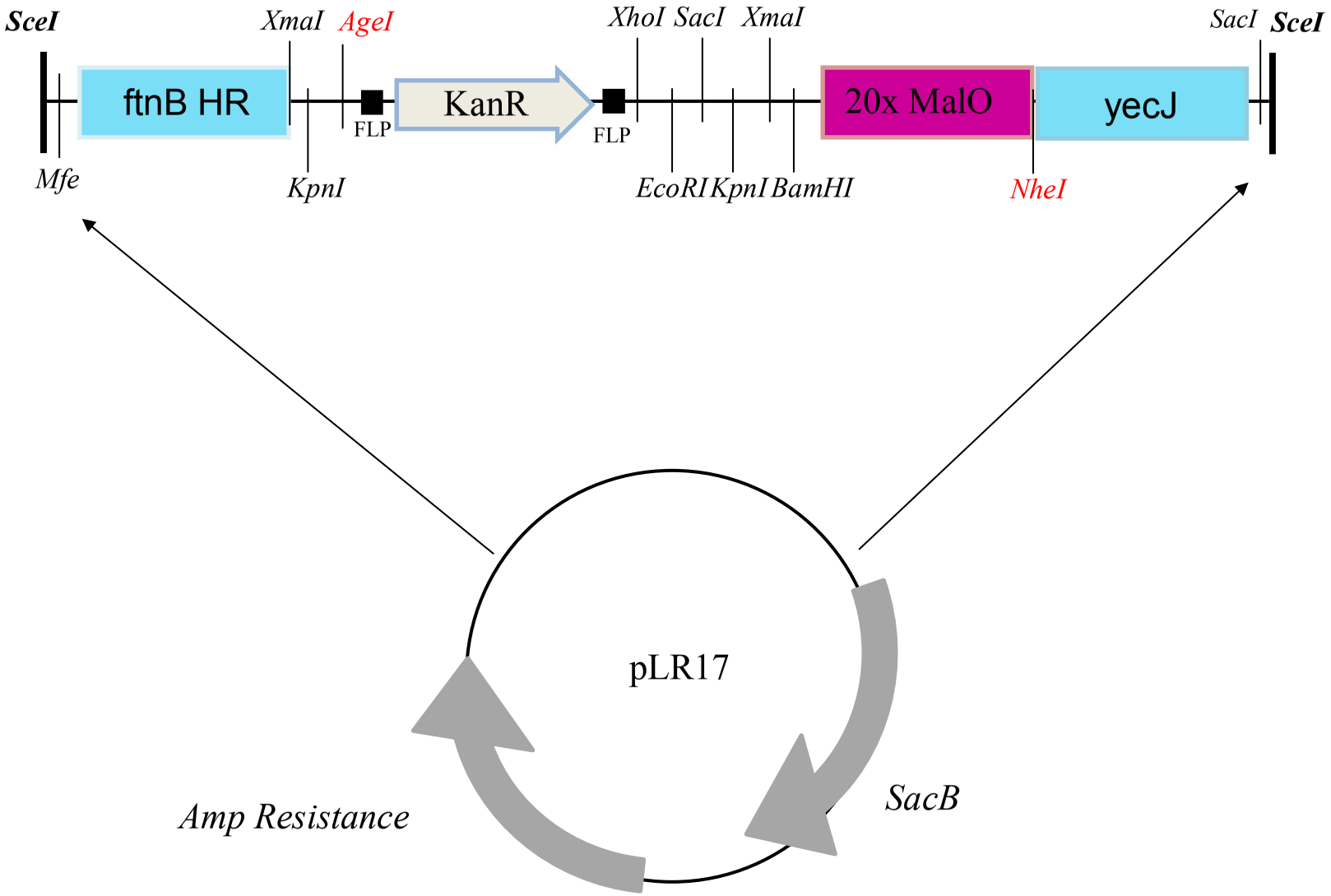


Figure 9: pLR17 plasmid map

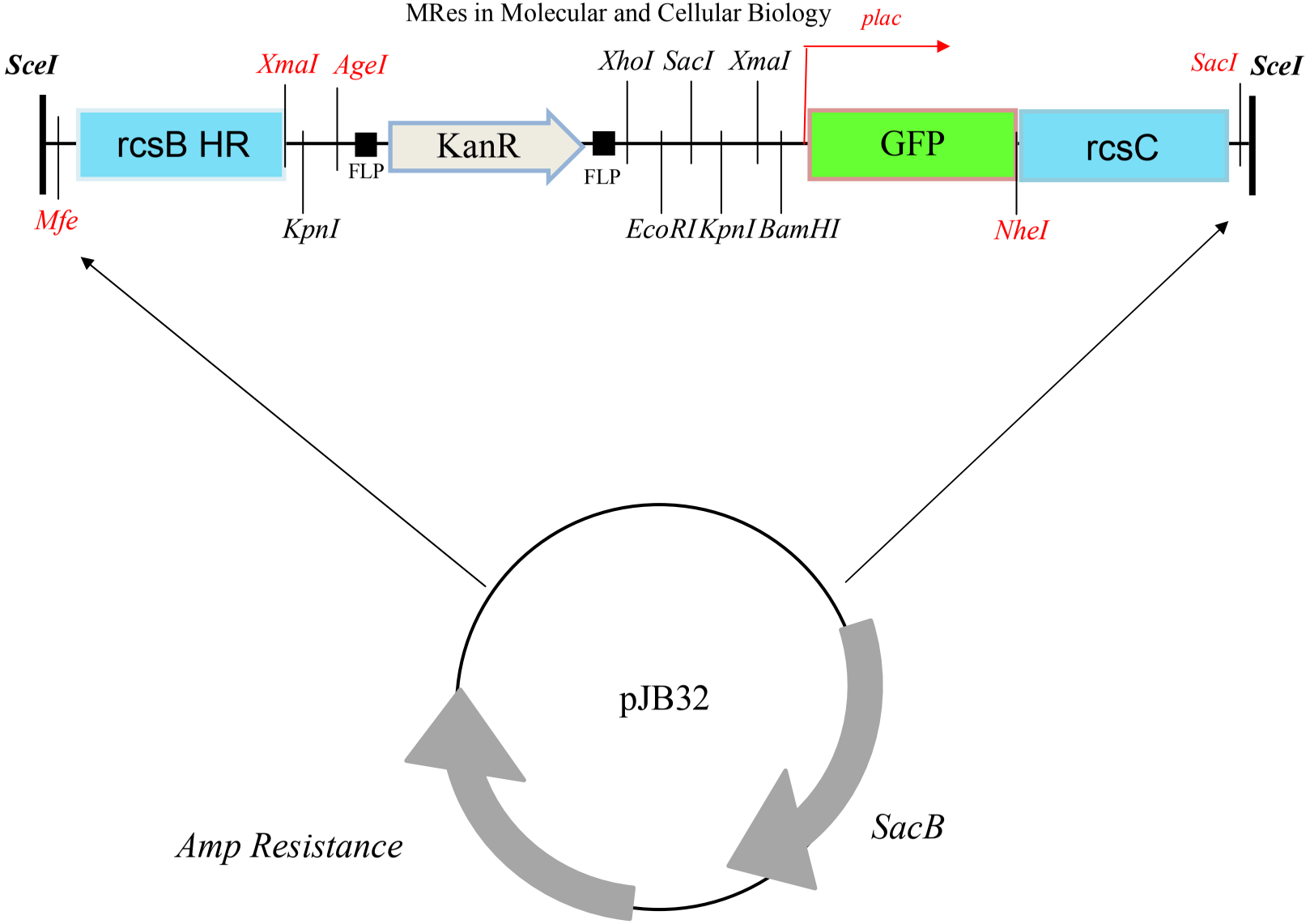


Figure 10: pJB32 plasmid map

Primer	Primer Sequence (5'-- 3')	Binding specificity
D58793	GGA-TGT-GCT-GCA-AGG	Anneals downstream of the second SceI site of pJB32, sequencing back through the insert beginning with the second homology region.
D58794	TAT-GCT-TCC-GGC-TCG	Anneals upstream of the first SceI site of pJB32, sequencing through the insert beginning with the first homology region.
D75238	AGC-GTG-AGG-GGA-TCT-TG	Anneals within the sequence of the Kan cassette, sequencing towards the first homology region.
D69988	TCA-AGA-TCT-GAT-CAA-GAG-ACA-GGA-TGA-GG	Anneals within the Kan cassette, sequencing towards the second homology region.
D76207		Anneals within the Kan cassette, sequencing towards the first homology region.
D68556	TTT-ACG-TCG-CCG-TCC-AG	Anneals within GFP, sequencing through Kan cassette towards the first homology region.
D69232	GCCG-CCC-GGG-CGA-CGC-TTG-CCG-CGT-CTT-ATC	XmaI and Sac1 primers for *pLR8* homology regions
D69233	GCCG-GAG-CTC-CTG-AAC-ATG-CGT-TGC-ATC-AAC	
D71881	GGAC-CAA-TTG-ATT-GTT-CTG-ACT-ATG-AAC-AAC-AAC-C	MfeI and XmaI primers for *rcsB* homology regions on pJB32
D71882	GGAC-CCC-GGG-ATG-GGA-ATC-GTA-GGC-CG	
mntH forward	GCT-CAA-TTG-ATG-AGG-CTT-ATC-TGA-CGC	MfeI primer for the *mntH* homology region
mntH reverse	GCT-CCC-GGG-GCC-AAT-GGA-GCA-CAA-TGC	XmaI primer for the *mntH* homology region
ypeC forward	GCT-GCT-AGC-GGA-CGC-GTT-TAA-TGG-CG	NheI primer for the *ypeC* homology region
ypeC	GCT-GAG-CTC-GTG-CTG-	SacI primer for the *ypeC* homology region

reverse	GTG-GTA-ACA-CG	
Dps forward	GCT-GCT-AGC-GCT-ACT-TTT-CCT-CTA-CAC-CG	NheI primer for the *dps* homology region
Dps reverse	GCT-GAG-CTC-CCC-CAG-AGC-TAC-ACC-G	SacI primer for the *dps* homology region
rhtA forward	GCT-CAA-TTG-TGT-GGT-TCC-TGC-TAC-CG	MfeI primer for the *rhtA* homology region
rhtA reverse	GCT-CCC-GGG-GAG-AAA-TTC-TGC-ATG-GTT-ATG-C	XmaI primer for the *rhtA* homology region
mntH check primer	GCT-ACA-GCT-GCG-GCG-GC	Anneals within the *mntH* gene upstream of the homology region. Sequencing through the first homology region towards the Kan cassette. Used to PCR potential gene doctoring recombinants
ypeC check primer	GCG-GCA-ATA-ACC-GTT-TCT-TGC-G	Anneals within the *ypeC* gene downstream of the homology region. Sequences back through the second homology region. Used to PCR potential gene doctoring recombinants
Dps check primer	CGT-TGT-GGA-TGT-CCA-GCG	Anneals within the *dps* gene downstream of the homology region. Sequences back through the second homology region. Used to PCR potential gene doctoring recombinants
rhtA check primer	TTT-CGT-CTG-GGT-TGT-GCT-GGC	Anneals within the *rhtA* gene upstream of the homology region. Sequences through the first homology region towards the Kan cassette. Used to PCR potential gene doctoring recombinants

Table 3: List of primers utilised as part of this study. Sequences highlighted indicate MfeI, XmaI, NheI and SacI restriction sites juxtaposing primer sequences.

Buffers and Solutions

All buffers and solutions were purchased from Sigma Aldrich, BDH or Fisher scientific unless stated. Prior to use, all solutions used in bacterial growth were autoclaved at 120 °C and 15 psi for 20 minutes.

DNA loading dye

0.025 % Bromophenol Blue, 0.025 % Xylene Cyanol F, 20 % Glycerol, 10 mM Tris, 1 mM EDTA

Ethidium Bromide

10 mg/ml (Biorad)

DNA markers

100 bp and 1 Kb ladders (NEB)

40 x Tris/Acetic acid/EDTA (TAE) buffer

2 M Tris acetate, 100 mM Na_2EDTA (National Diagnostics)

Agarose gel

0.8 % Agarose in 1 x TAE buffer. Heated in the microwave prior to use.

Deoxynucleotide Triphosphates (dNTPs) (Bioline)

Diluted from 0.25 mM to a concentration of 1 µM each

Primers (Alta bioscience)

Diluted to a final concentration of 1 µM

Calf Intestinal Phosphatase (CIP) (New England Biolabs)

DNA Sequencing

Functional Genomics and Proteomics Lab, University of Birmingham

3.2 picomole of primer added to a 10 µl sample

CasAmino Acids

10 ml 20 % solution made using sterile water

Fructose

10 ml 20 % solution made using sterile water

10x M9 minimal salts (Values as for 100 ml)

6 g Na_2 HPO_4

3 g KH_2PO_4

0.5 g NaCl

1 g NH_4Cl

Phosphate Buffered Saline (PBS) (Values made up to 100 ml using sterile water)

2.8 ml 5 M NaCl

50 µl 4 M KCl

8 ml 0.1 M Na_2HPO_4

60 µl 1 M KH_2PO_4

Manganese Chloride ($MnCl_2$)

10 mM solution in sterile water

Iron (III) Chloride ($FeCl_3$)

17 mM solution in sterile water

Growth Media

Luria Bertani Broth: 20 g Tryptone (peptone), 10 g Yeast extract, 10 g NaCl in 1 Litre of distilled water. All solutions were autoclaved at 120 °C and 15 psi for 20 minutes and then stored at room temperature prior to use.

Nutrient Agar: 23 g Nutrient Agar powder (Nifco) in 1 litre distilled water. Prior to use, the agar solution is autoclaved at a temperature of 120 °C and a pressure of 15 psi for 20 minutes. Cooled Nutrient Agar plates were stored at 4 °C before use.

M9 minimal salts media:
10 ml 10 x M9 salts
90 ml Sterile distilled water
200 µl 1 M $MgSO_4$
100 µl 0.1 M $CaCl_2$
0.5 ml 20 % Casamino acids
1.5 ml 20 % Fructose or Glucose

Bacterial Growth with Mn^{2+} and Fe^{3+} was carried out in minimal salts media with the addition of the relevant Iron or Manganese containing substance taking place 30 minutes before slide preparation.

Antibiotics

All antibiotics created were filter sterilised prior to use and stored at -20 °C

Kanamycin (Kan)
Made at 50 mg/ml with sterile water and used at a 1000 x dilution of 50 µg/ml

Ampicillin
Made at 40 mg/ml with sterile water and used at 80 µg/ml

Chloramphenicol
Made at 35 mg/ml with ethanol and used at a 1000 x dilution of 35 µg/ml

Competent Cells

1 ml of the overnight culture was added to 50 ml of LB broth and incubated at 37 °C until mid-logarithmic growth phase, usually established through an optical density of 0.3-0.5 at 650 nm. Upon reaching the correct optical density (OD), the sample was subjected to ice for 10 minutes and harvested through centrifugation at 4000 rpm at 4 °C for 10 minutes. Pelleted cells were then re-suspended in 25 ml ice cold $CaCl_2$ and incubated on ice for 20 minutes. Following a further identical centrifugation step, the pelleted cells were re-suspended in a solution of 2.3 ml $CaCl_2$ and 1 ml 50 % glycerol and stored at -80 °C.

Digestions

3 separate double digestions were carried out in this study using three pairs of restriction enzymes; *MfeI* HF and *XmaI*, *NheI* HF and *SacI* HF, *NheI* and *AgeI* HF. A 50 µl reaction volume was made using the following ratios;

- 5 µl Buffer – appropriate to restriction enzymes as recommended by NEB
- 0.5 µl Bovine Serum Albumin (BSA) – if required
- 0.5 µl High Fidelity (HF) enzymes
- 1 µl Non- HF enzymes
- Volume of DNA required to bring total reaction volume up to 50 µl

In each digestion, the reaction volume was left to incubate at 37 °C for three hours.
If the digested plasmid was to be used as a vector for subsequent ligations, a Calf Intestinal Phosphatase (CIP) reaction was carried out using 2.5 µl CIP and 5 µl buffer 3, in a 50 µl reaction. This reaction prevents further uncontrolled ligation of the vector by removing 5' phosphates and was once again left at 37 °C for a further hour.

Ligations

Various ligation ratios of insert and vector volumes were calculated and assembled based on the concentration of DNA purified. A total reaction volume of 20 µl was reached using sterile water, 1 µl of T4 DNA ligase and 2 µl T4 DNA ligase buffer. The reaction was left resting horizontally on ice for at least 10 minutes prior to incubation overnight (8 hours) at 16 °C. Upon completion of the ligation period, the sample was kept at 4 °C until use.

Transformations

1-2 µl of the plasmid was added to 100 µl of appropriate *Escherichia coli* competent cells and placed on ice for at least 30 minutes. Following ice incubation, cells were subjected to heat shock at 42 °C for 90 seconds, before being placed back on ice for a further 5 minutes. 500 µl LB broth was added to the cells and incubated at 37 °C for an hour before centrifugation, re-suspension and plating out on appropriately antibiotic treated nutrient agar plates. Incubation of the plates took place at 37 °C overnight.

Plasmid DNA extraction

Overnight cultures of plasmid carrying cells were grown in LB, supplemented by appropriate antibiotics. Plasmid DNA extraction was carried out following manufacturers instructions of the QIAgen Spin Miniprep kit. Elution was carried out in 50 µl of sterile distilled water.

PCR used to amplify insert fragments

A 50 µl reaction volume consisting of 25 µl Biomix Red mixture, 1 µl of each primer specific to the insert being amplified and 5 µl of DNA miniprep template were made up to volume using distilled sterile water. A typical PCR cycle as stated below was then carried out;

95 °C	------ 5 minutes	
95 °C	------ 30 seconds	X 35
Annealing Temperature	------ 1 minute	
71 °C	------ 15/30 seconds per kb of amplified fragment	
71 °C	------ 7 minutes	

Colony PCR including boil preparation

A Biomix Red reaction volume, as stated previously, was set up, replacing the sterile water and the 5 µl DNA template. An *Escherichia coli* colony was re-suspended in 50 µl sterile water, from which a 23 µl sample was added to the primer pair/ biomix red solution. The PCR cycle was similar to that mentioned previously, however, an increased initial boiling time of 10 minutes was implemented to ensure complete cell lysis prior to polymerase chain reactions occurring.

Where required, a gradient of annealing temperatures, suitable to the melting temperatures of different primer combinations was implemented.

Gel Electrophoresis

Liquid agarose gel was cooled in a casting plate coupled with sufficiently sized gel comb until solid. 5 µl DNA samples combined with loading dye were loaded and the gel ran at 100 V in 1 X TAE buffer for approximately 40-45 minutes. Upon completion of the run, gels were stained in 10 mg/ml ethidium bromide solution or SYBRSafe and viewed using a UV trans-illuminator or blue light box depending on the purpose of the gel.

Agarose gel extraction

Upon completion of the run, agarose gels were stained in SYBRSafe for 30 minutes and the required bands physically extracted with the aid of a blue light box. Upon extraction the required DNA was purified using the recommended Bioline protocol and Bioline reagents. Elution of the DNA was carried out into 50 µl of sterile distilled water.

DNA purification

Purification was carried out using the Bioline purification kit, following the instructions accompanying the kit. The purified product was eluted into 20 µl distilled sterile water.

Gene Doctoring Protocol

MG 1655 and LacI – GFP competent cells were co transformed with the appropriate donor plasmid, pSB5 and 6 respectively, along with the pACBSR mutagenesis plasmid. Transformed single colonies were then patched onto two plates, one containing kanamycin and sucrose, the other containing kanamycin and chloramphenicol. A small sample of cells exhibiting sucrose sensitivity were re-suspended in 0.5 ml of LB media, kanamycin and chloramphenicol and grown at 37 °C for between 2 and 3 hours. Cells were then centrifuged and washed in 0.1 X LB to ensure removal of the antibiotics. Following further centrifugation, the cells were re-suspended once again in 0.5 ml of 0.1 X LB containing 0.5 % arabinose, inducing the expression of the SceI and λ red genes. Incubation at 37 °C for a further 3 hours followed, before the cells were plated out onto kanamycin and sucrose nutrient agar plates. These plates were incubated overnight at 30 °C and then at 37 °C for further colony growth. Appropriate Polymerase Chain Reactions (PCR) and sequencing, using the check primers listed in table 3, were carried out to ensure successful gene doctoring. Patching of successful colonies onto Kanamycin/ Chloramphenicol and Kanamycin/ Sucrose agar plates confirmed the complete removal of the mutagenesis plasmid from the cell.

Removal of Kanamycin cassette

The presence of Flippase recombinase (FLP) recognition sequences both upstream and downstream of the Kan cassette in both gene doctoring donor plasmids allows relatively easy removal through the use of flippase recombinase (FLP). Successfully gene doctored cells were transformed with pCp20, a temperature sensitive plasmid containing the FLP recombinase gene, and grown on ampicillin nutrient agar plates at 30 °C. Colonies were then re-streaked on non-selective agar plates and grown at 37 °C. Single colonies from this plate were patched on agar plates alone, as well as plates supplemented with kanamycin and ampicillin antibiotics, and once again grown at 37 °C overnight. Appropriate PCR reactions and sequencing were carried out to confirm successful removal of the Kanamycin resistance cassette.

Microscopy

A 1 in 50 dilution of an overnight sample was grown to an OD of 0.1 in 5 ml minimal media at 23 °C. Growth with 10 µM metal ions was carried out for 30 minutes prior to slide preparation. Upon reaching the desired OD, 1 ml of the sample was spun down and re-suspended in PBS wash buffer, twice. During this process 5 µl of poly-L-lysine solution was spread across the centre of the slides and left to dry. The PBS supernatant was removed from the sample and the cell pellet re-suspended in 5 µl Hoechst 33258 and 5 µl FM 4-64. Incubation for 5 minutes followed before 5 µl was added to the poly-L-lysine treated slides and the cover slip placed firmly on top.

Microscopy was carried out on a Nikon Eclipse 90i, with cell visualisation achieved using bright field, whilst 4,6 diamidino-2-phenylindole (DAPI), Fluorescein Isothiocyanate (FITC) and Tetramethyl Rhodamine Isothiocyanate (TRITC) filters were used to visualise the Hoechst 33258, GFP and FM 4-64 fluorescence respectively.

Growth curve protocol

Relevant dilutions of the overnights cultures were carried out to achieve an OD of approximately 0.05 at 650 nm in M9 minimal media. 1 ml samples every hour were subjected to spectrophotometer analysis again at 650 nm.

Microscope Analysis

Positions of foci were measured from the nearest apical pole of the cell. In cells exhibiting more than one focus, all foci measurements were taken from the same apical pole.

3 Results

3.1 Cloning

The construction of both donor plasmids, used in the subsequent gene doctoring procedures, was based on the same fundamental principles. The first homology regions of pSB5 and pSB6, representative of the final 500 bp of the *mntH* and *rhtA* (Resistance to Homoserine and Threonine A) genes respectively, were PCR amplified from MG 1655. These PCR products were then digested using an MfeI and XmaI restriction enzyme combination. Ligation of these inserts to a previously identically restricted pJB32 vector followed (Figure 9). The second 500 bp homology regions of pSB5 and pSB6 again representative of the final 500 bp of the *ypeC* and *dps* genes respectively were constructed in a similar manner. This restriction took place using a SacI and NheI enzyme pair and ligation was into the vector of pJB32 already containing the appropriate first homology region.

Completion of the donor plasmid constructs came about through insertion of the 20 Mal or Lac operators in combination with Kanamycin resistance flanked by FLP recombinase sites into the appropriate plasmid vector (Figures 5 and 6). The Mal operators were directly digested from pLR17 (Figure 8) and the Lac operators from pLR8 (Figure 7) using the NheI and AgeI restriction enzymes.

Successful digestion and ligation of the homology regions and operator repeats were verified at every stage through PCR amplification, an example of which is illustrated in figure 10, and sequencing. Transformants that exhibited positive verification via both methods were considered for gene doctoring.

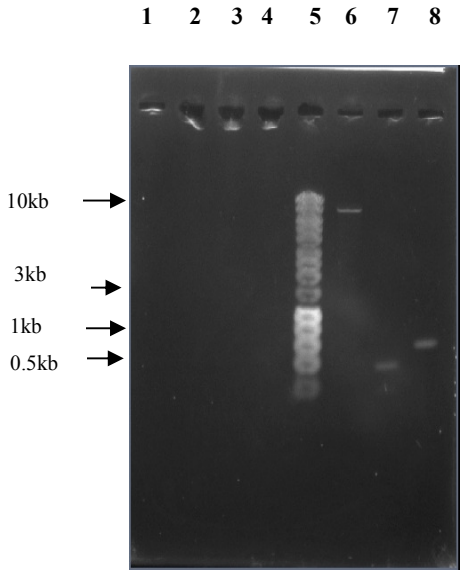


Figure 11: Successful digestion of pJB32 vector and mntH and rhtA inserts. A single band of approximately 9.5kb for the pJB32 vector (lane 6) and of approximately 500bp for the mntH insert (lane 8) confirms digestion in accordance with the relevant restriction sites. A decrease in size of rhtA insert (lane 7) was due a second downstream MfeI restriction site not previously noticed. Despite this size reduction, further ligation to the vector and gene doctoring should have been unaffected.

3.2 Gene doctoring

Gene doctoring, using the two previously constructed donor plasmids and the pACBSR mutagenesis plasmid, was carried out using the protocol described earlier (materials and methods). Two host cell genomes; MG1655 and pDOC-G, containing lacI-GFP, were doctored with pSB5 and pSB6 donor plasmids respectively. Potential recombinants were analysed through PCR amplification of the relevant host genome section as shown in figures 11 and 12. Candidates that conform to the expected 3 kb single band were sent for sequencing to truly validate incorporation of the donor plasmid insert onto the chromosome at the desired position. Prior to the formation of downstream SXB competent cells, verification of pACBSR linearization and removal from the cell was carried out via ampicillin and chloramphenicol sensitivity. Upon verification, removal of the Kanamycin resistance cassette was carried out using an FLP recombinase enzyme present on pCp20 (Figure 13).

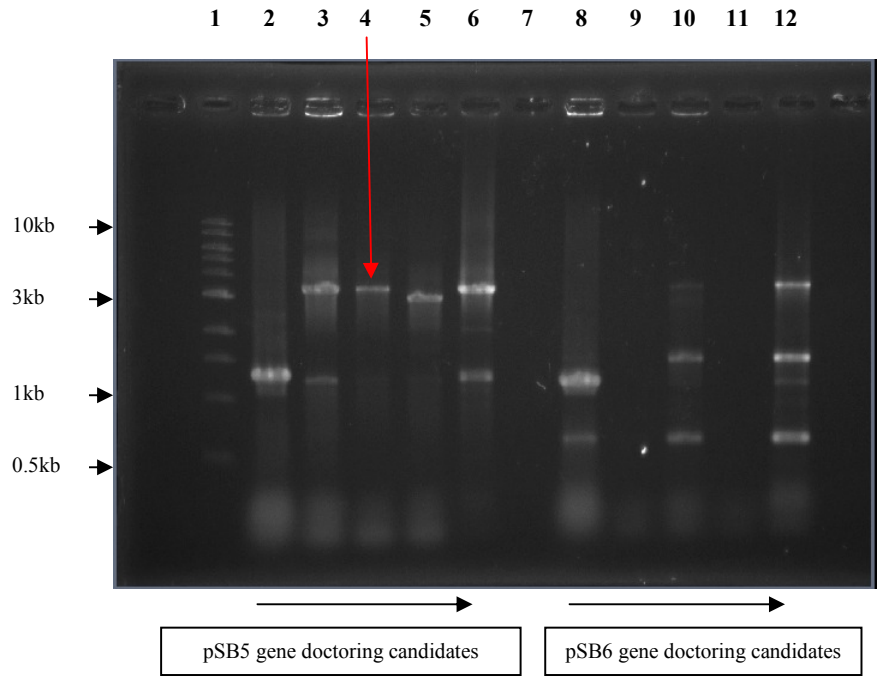


Figure 12: Check PCR using the relevant check primers for potential gene doctoring recombinants of both donor plasmids. From Left to Right, Lane 1 = Marker Lanes 2-6 = pSB5 candidates, Lanes 8-12 = pSB6 candidates. The presence of single 3 kb band from lane 4 indicated potential successful gene doctoring of pSB5. This was completely confirmed through sequencing.

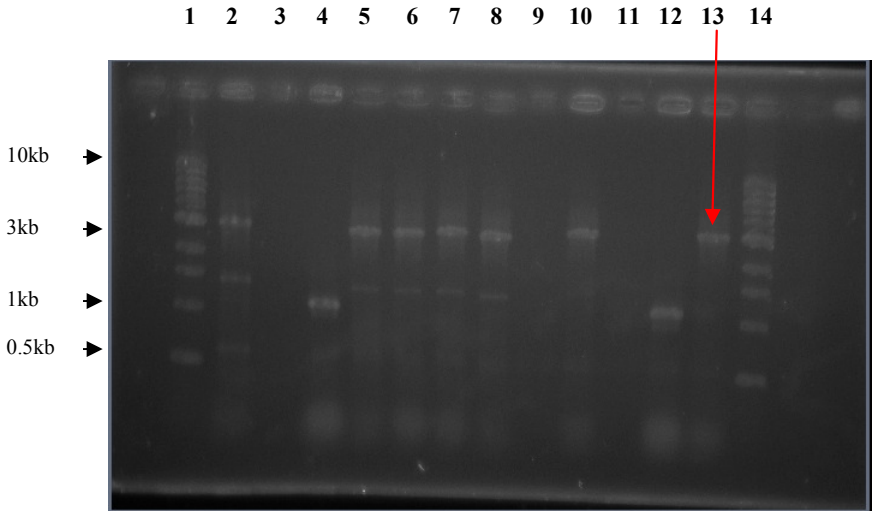


Figure 13: Check PCR using the relevant check primers for potential gene doctoring recombinants of donor plasmid pSB6. Once again, the presence of a single 3 kb band in lane 13 (lanes numbers running from left to right) suggested successful gene doctoring of pSB6, confirmed later through sequencing

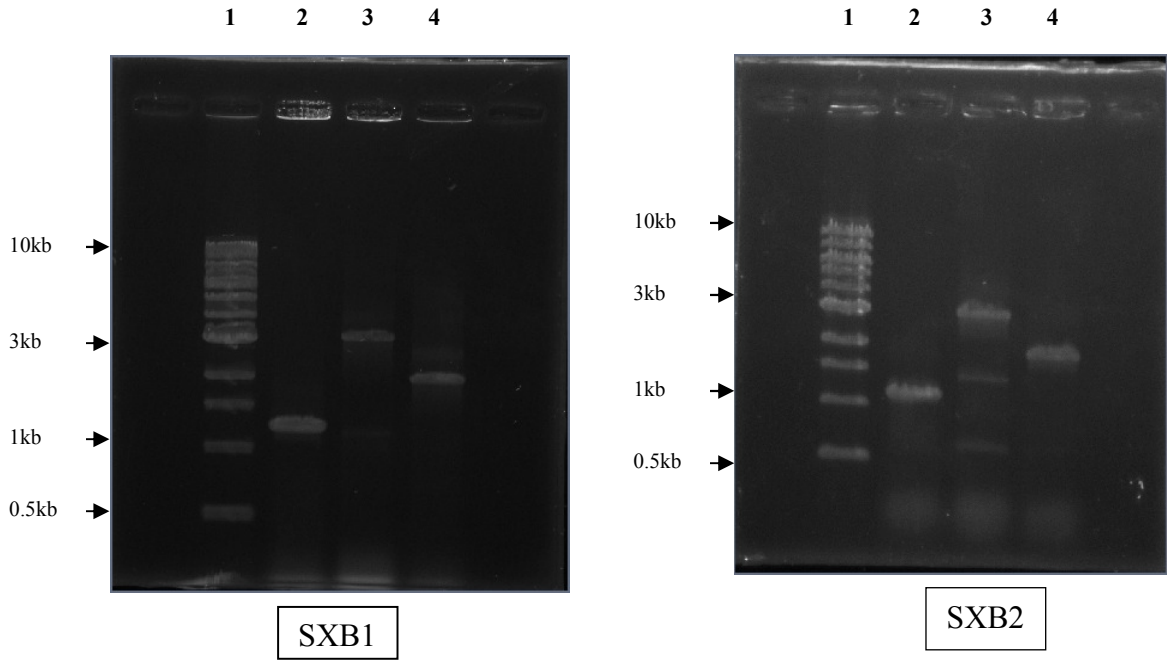


Figure 14: Successful removal of the Kanamycin cassette from SXB1 and SXB2 strains. From left to right Lane 1 = Marker, Lane 2 = Donor Plasmid, Lane 3 = The relevant SXB strain Lane 4 = The relevant SXB strain minus the Kan cassette. The reduction in size between the third and fourth lanes is consistent with that estimated for the loss of the Kanamycin cassette from the insertion construct.

3.3 Growth Curves

Given the relationship between chromosomal positioning, transcription and cell proliferation, it was important to confirm that the insertion of the constructs did not have a disadvantageous effect on culture growth which could possibly falsify any results. Several growth curve measurements in M9 minimal growth media, a representation of which is illustrated in figure 14, were undertaken in an attempt to prove that this was not the case. Comparison of the relevant SXB strain to the initial gene doctoring host strain shows little difference between the growth rates over the entire experimental period.

Fundamentally the construction of both SXB strains was identical, apart the additional requirement of MalI: mCherry in the case of SXB1. Transformation with the relevant pLER106 plasmid, however, coincided with significant growth problems that affected any planned microscopic analysis. Growth in both LB media and M9 minimal salts yielded no difference in the growth rate of cultures, whilst growth at 37 °C and without chloramphenicol produced problems regarding plasmid retention. It is clear from the growth curves that the construction of SXB1 had no effect on the growth rate of cultures, therefore it was postulated that the presence of pLER106 plasmid reduced growth rates in the culture. Indeed transformation of MG 1655 with pLER106 confirmed this hypothesis, through exhibiting a growth rate significantly reduced from that witnessed in not only native MG 1655 but in other MG 1655 transformant cultures. The necessity of the pLER106 plasmid for successful imaging of the SXB1 strain led to the postponement of future investigations using this strain until the growth problems were resolved.

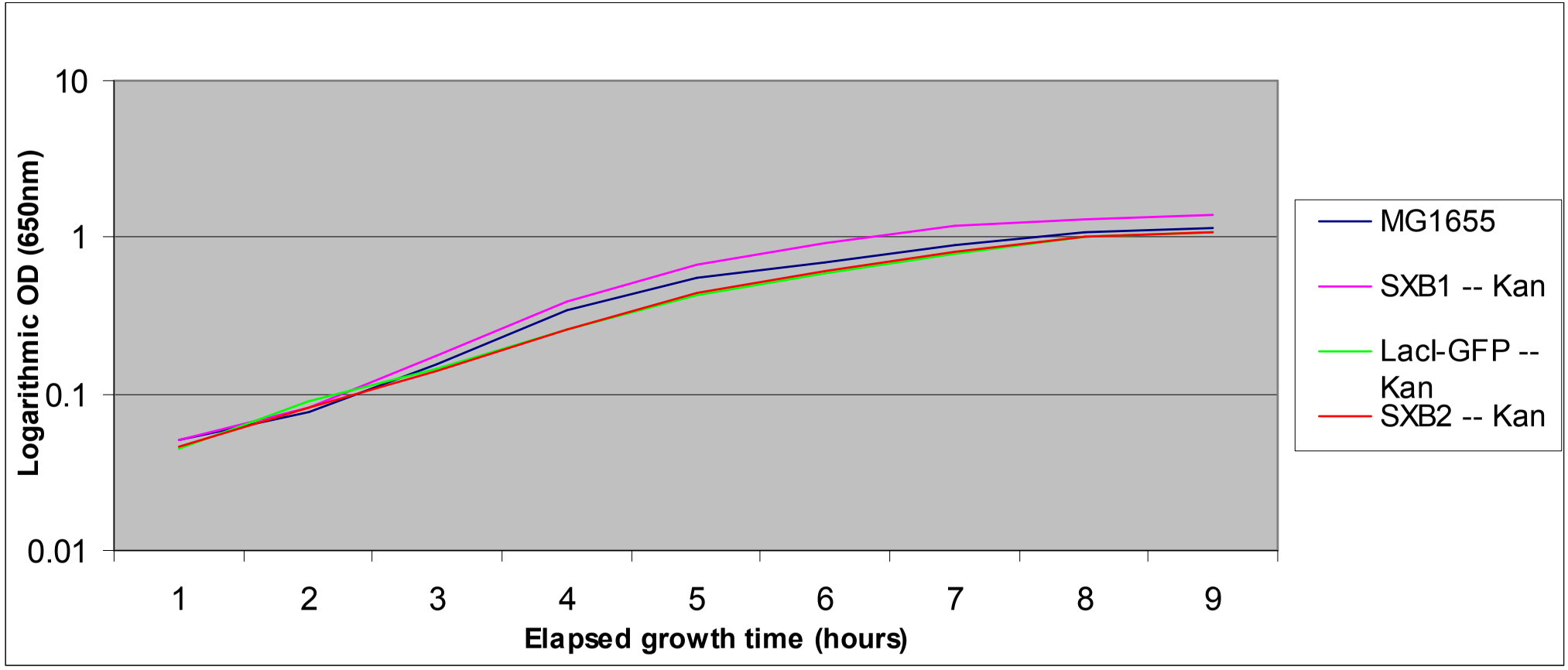


Figure 15: Growth curve in M9 minimal media for each of the four strains experimentally tested in this study

3.4 Visualisation of pDOC-G strain control

To ensure that the foci observed in all samples were localised to the targeted *lac* operators, controls of the pDOC-G strain were visualised in combination with both the experimental LR06 and SXB2 strains under every experimentally tested growth conditions. Through the 50 control cells examined, the complete lack of foci exhibited in the pDOC-G strain, vindicated the assumption that foci production was located at the gene of interest as opposed to any other area on the genome.

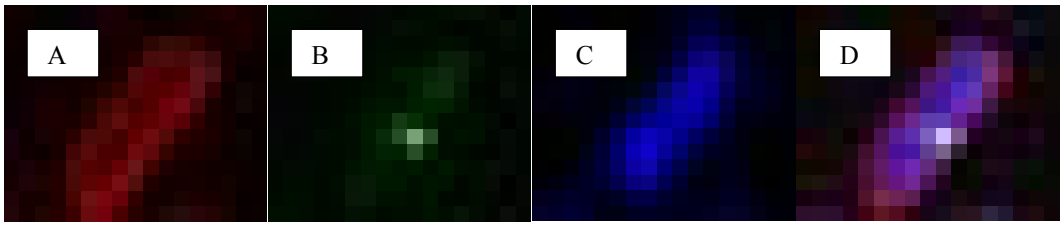


Figure 16: Visualisation of an example cell exhibiting a single focus through separate individual filters. A) Visualisation of FM 4-64 labelled membrane under TRITC filter B) Visualisation of a LacI- GFP focus under a FITC filter C) Visualisation of Hoechst 33258 labelled DNA under a DAPI filter D) Composite image of the cell

3.5 LR06 vs. SXB2

The LR06 strain shows GFP foci in close vicinity of the *araC* gene as opposed to juxtaposing the *dps* gene in the case of SXB2. Given the two distinctly separate locations, the number of cells with foci within a culture gives an indication as to the accessibility of the gene to transcription factors and/or RNA polymerase, and therefore the effect on transcriptional regulation and gene expression. As illustrated in figure 16, there is considerable overall reduction in the number of cells with foci in the SXB2 strain over the LR06 strain. The 23 % reduction in cells possessing more than one focus in combination with an increase of 12 % in cells possessing no foci illustrates this trend. The reduction of single foci exhibited indicates a difference in gene accessibility between *dps* and *araC*.

Given the relative distances of both genes from the origin of replication, chromosome replication will yield differing ratios of one foci displaying cells to those exhibiting two foci. The increase in cells showing one focus at the *dps* locus is indicative of this principle and is not a product of increased gene accessibility.

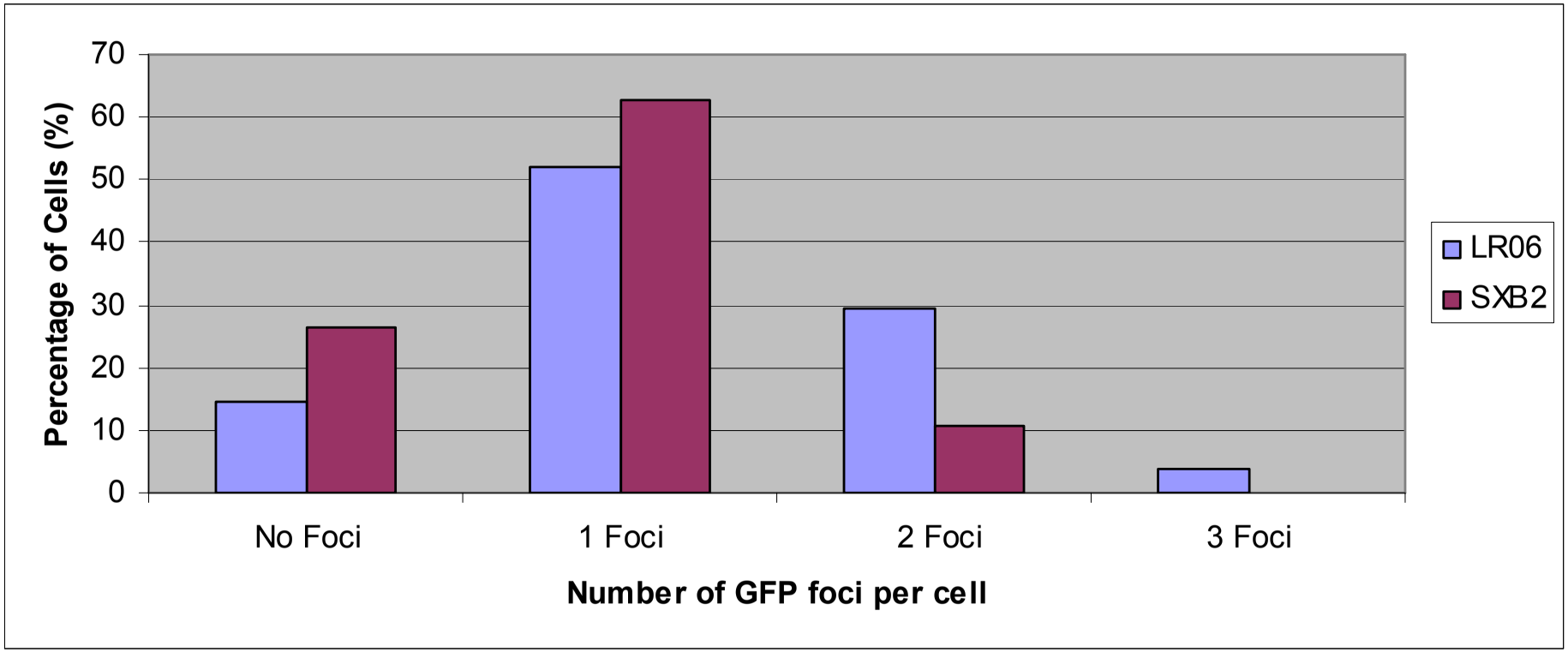
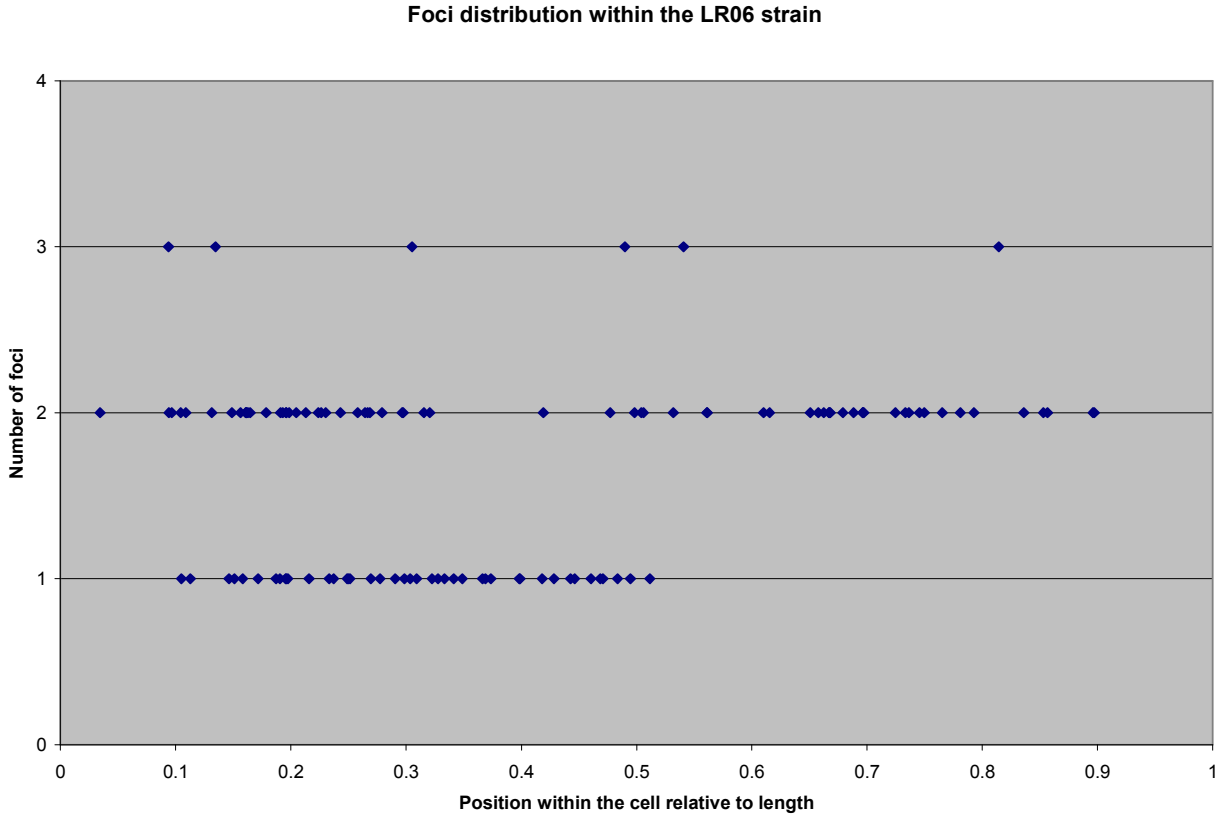


Figure 17: Comparison between the number of cellular foci present in an LR06 (LacI-GFP bound to the *araC* locus) *Escherichia coli* sample and an SXB2 (LacI-GFP bound to the *dps* locus) *Escherichia coli* sample. Sample size in each case was 102 cells and bacterial growth was carried out in minimal media/

In terms of the position of the majority of foci within the cell, an increased polar nature within the LR06 strain, especially portrayed in cells possessing a single focus, could be argued (Figure 17). Indeed statistical analysis, based upon the presence of a normal distribution of foci, indicate a shift in the mean position of 0.08 in the direction of the cell poles against that witnessed in SXB2 (Table 4). In LR06 cells displaying two foci, the large concentration between 0.15 - 0.35 supports the previously discussed observation, whilst in combination with a foci concentration between 0.6 – 0.8 signifies two distinct regions within the cell that the *araC* gene may occupy within the cell (Figure 17). Unfortunately comparison with SXB2 cells displaying two foci, on this occasion is unable to take place through a deficiency in data points.

Strain	Mean position of single focus relative to cell length	Variance based upon a normal distribution
LR06	0.306	0.013
SXB2	0.372	0.008
SXB2 + $MnCl_2$	0.350	0.009
SXB2 + $FeCl_3$	0.382	0.005

Table 4: Mean position of single focus and Variance for each microscopically analysed sample. The mean position and variance figures were based a normal distribution of data points across the data range.



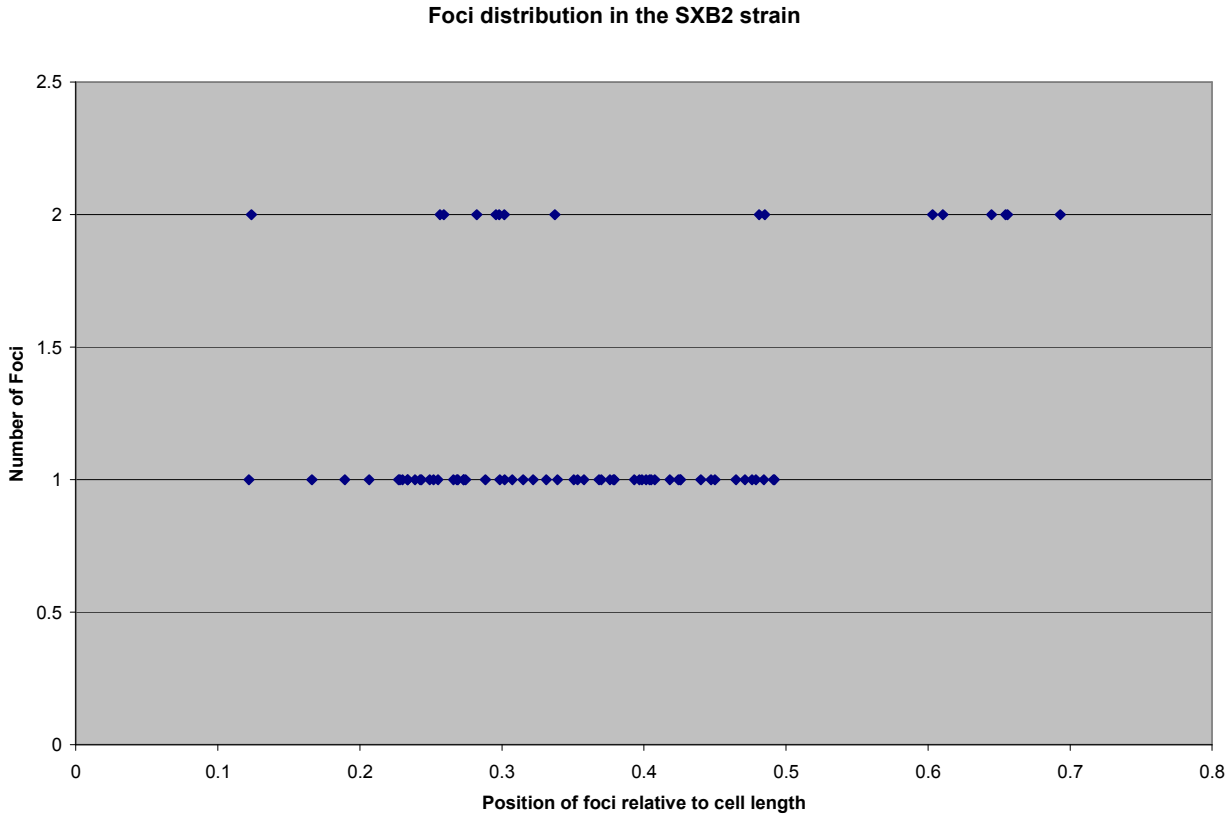


Figure 18: Distribution of witnessed foci relative to the cells' length in top) LR06 and bottom) SXB2 *Escherichia coli* strains.

3.6 Comparison of the SXB2 strain under Manganese and Iron supplemented growth

The antagonistic relationship between manganese and iron, particularly in mediating oxidative stress and metal toxicity, is well documented. Increased levels of intracellular iron triggers the uptake of extracellular manganese providing the protection against iron-induced oxidative stress. As previously mentioned, *dps* is repressed by manganese during stationary phase growth, yet it is utilised to bind Fe^{2+} as a replacement for manganese in protecting the cell from excess iron toxicity (Yamamoto et al, 2011). It may be concluded that growth conditions supplemented with 10 μM manganese or iron may alter the distribution of foci originally witnessed in the SXB2 strain.

Manganese or Iron supplemented growth had the effect of increasing the number of foci displayed at the *dps* location. Although not a considerable increase, 89 % of cells under manganese supplemented growth and 94 % under iron supplemented growth exhibited at least one focus compared to the 81 % witnessed under no metal supplementation. The levels of cells possessing two foci or three foci remain approximately the same throughout the three differing growth conditions; the increase in foci shown under iron growth is attributed to a 10% increase in cells containing one focus (Figure 18). The increase in number of cells containing one focus prompts a conclusion of increased gene accessibility under conditions of both increased manganese and iron to varying degrees.

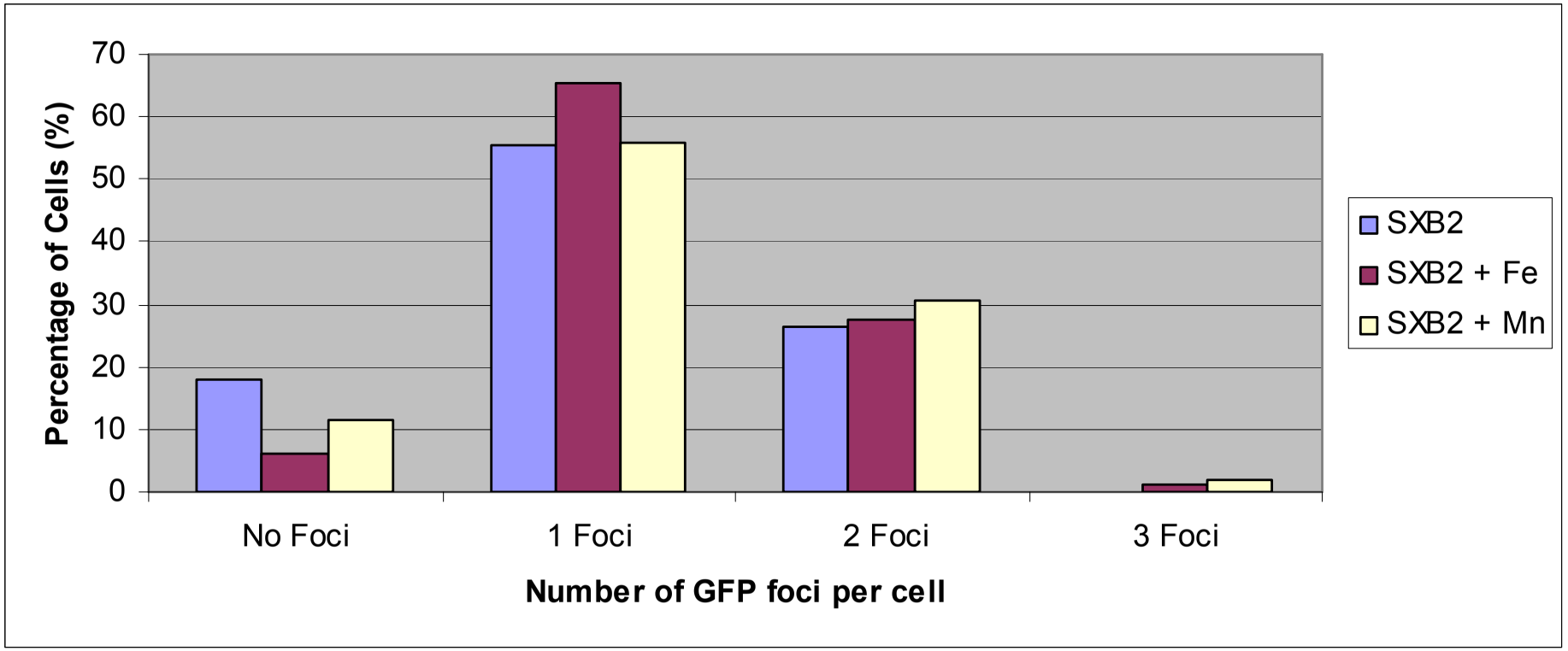


Figure 19: Comparison between the numbers of cellular foci exhibited in an SXB2 *Escherichia coli* sample under varying growth conditions. SXB2 was grown in sole M9 minimal media, SXB2 + Mn had M9 minimal media supplemented growth for 30 minutes with 10 µM $MnCl_2$ and SXB2 + Fe had M9 minimal media supplemented growth for 30 minutes with 10 µM $FeCl_3$. Minimum sample size was 250 cells.

Analysis of the foci position with respect to the cell length shows no appreciable difference between SXB2 and SXB2 + Fe strains. However, a small polar shift in the manganese positive sample was identified. In concordance with previous one focus SXB2 cells, the largest concentration of foci for each growth condition was witnessed between a 0.3-0.5 region of the length of the individual cell (Figure 19). Indeed the similarity in mean foci positions, once again calculated based upon a normal distribution of data points, provided no evidence for a change in gene position as a result of iron supplemented growth whilst also confirming the small polar shift witnessed with the addition of manganese (Table 4). Although not conclusive evidence, it could be argued that, in concert with results seen in the LR06 strain, cells exhibiting more than one focus have segregated regions either side of the cell midpoint, where the presence of both *araC* and *dps* genes are witnessed. Given that the copy number of a chromosome is strongly linked to cell division, it may be the case that these regions of the cell are designated for controlled cell separation rather than gene locations resulting from particular chromatin organisation.

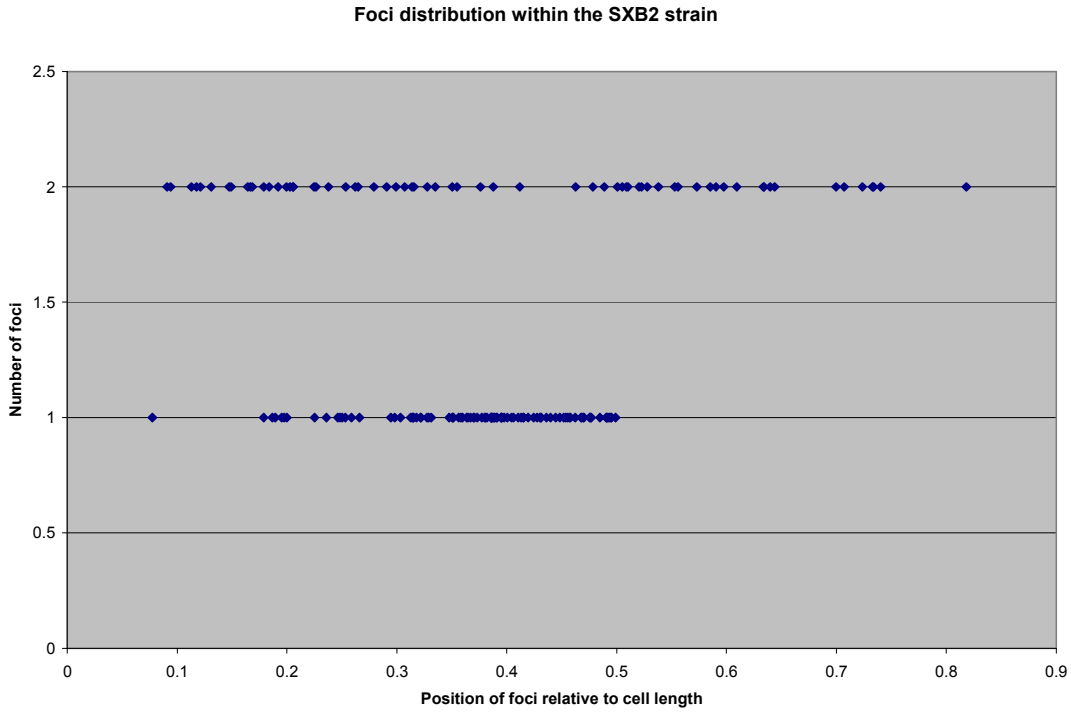


Figure 20 : Distribution of witnessed foci relative to cell length in SXB2 strains following growth in M9 x minimal salts media

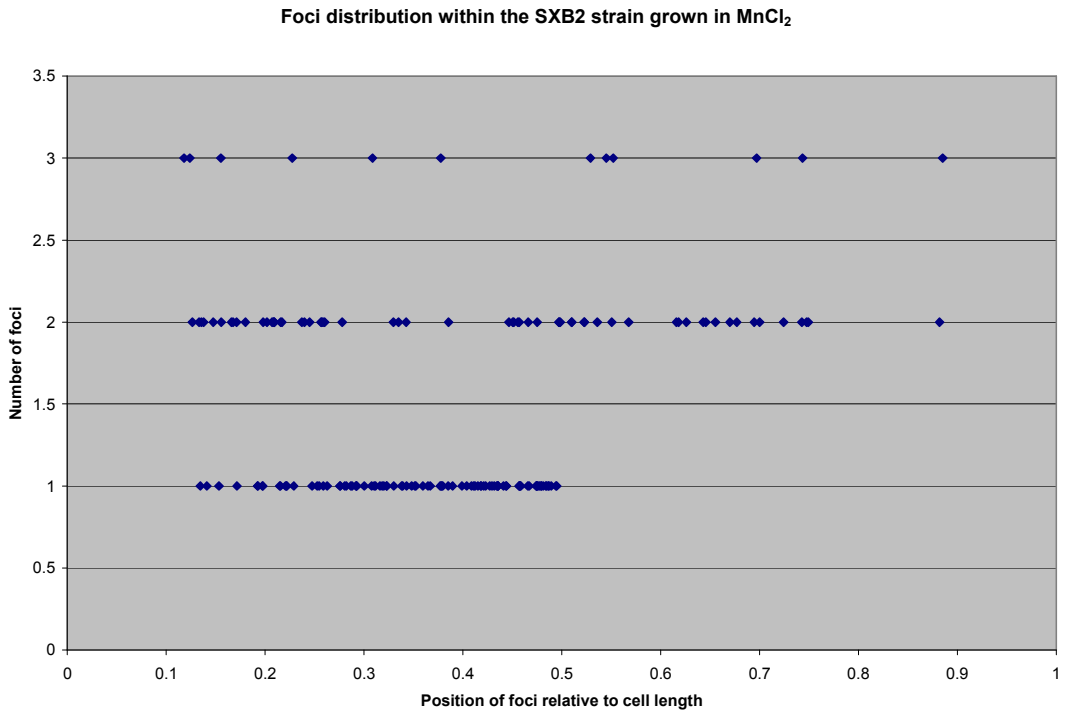


Figure 21 : Distribution of witnessed foci relative to the cell length in SXB2 strains following growth in Manganese supplemented M9 media

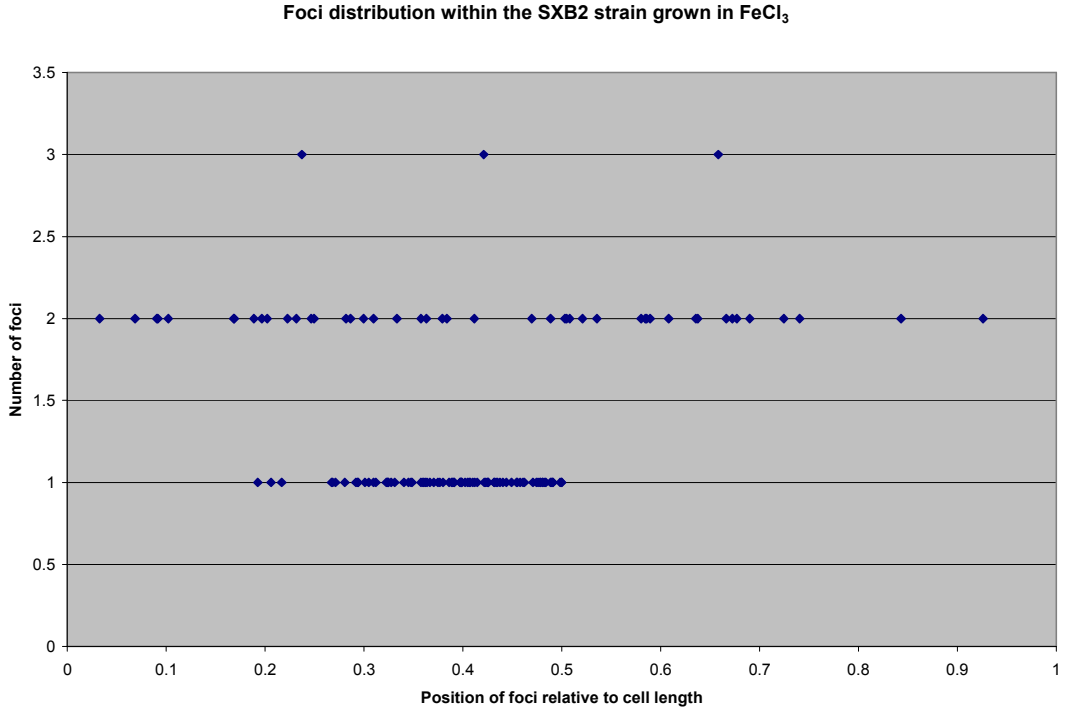


Figure 22 : Distribution of witnessed foci relative to the cell length in SXB2 strains following growth in Iron supplemented M9 media

4 Discussion

4.1 Difference in the number of foci exhibited at araC and dps loci

Escherichia coli araC protein is a transcriptional activator responsible for regulation of five operons associated with arabinose transport, metabolism and degradation. The *araC* gene is located in the NS Right macrodomain, approximately 1.5 centisomes from the ORI. In contrast to *araC*, *dps*, a stationary phase protein, is located 18.27 centisomes from the ORI in the RIGHT macrodomain (http://ecocyc.org/). Given the two distinct locations on the genome, it is conceivable that each gene could be incorporated in separate topologically unique supercoiled loops along with having a unique level of local chromatin organisation. If, as surmised, the difference in single focus number is as a result of varying levels of gene accessibility, it could be argued that local supercoiling and compaction at the two locations is the cause. This hypothesis, in combination with results presented previously, indicate that local chromatin organisation exhibited at the *araC* locus confers increased gene accessibility above that present at the *dps* locus. This analysis is a credible outcome when considering the specific functions of the genes; *araC*, a gene that is transcriptional regulated through all stages of the cell cycle, requires an increased degree of accessibility to RNA polymerase. *Dps*, would be repressed during the experimentally tested exponential growth phase and it is, therefore, logical that reduced gene accessibility, through a reduction in the number of cells exhibiting one focus, is witnessed at this locus.

4.2 Relative foci position at the araC and dps loci

Macrodomains are known to occupy distinct regions in the cell and are defined as regions of the genome that posses only intra-domain interactive capabilities. Indeed time lapse photography of fluorescently tagged loci across the genome mapped the specific regions macrodomains occupy through the cell cycle. *Espeli et al* published a model for their findings as illustrated in figure 22 (Espeli et al, 2008).

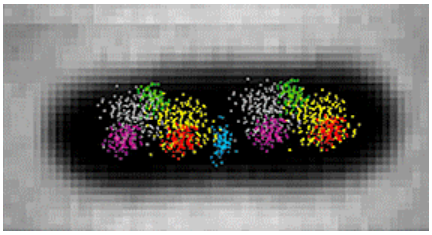


Figure 23: Fluorescent time-lapse map of *Escherichia coli* macrodomains in a mid-age cell.
ORI = Green, NSR = Grey, RIGHT = Red LEFT = Blue

Results presented in this study identified a polar shift of foci between the *araC* and dps loci. Comparison between these results and the model published by *Espeli et al* collaborate the notion that RIGHT macrodomain loci, such as *dps*, exhibit single foci nearer to the midpoint of the cell than those loci in closer proximity to the ORI macrodomain, such as *araC*.
Espeli et al also published data surrounding the positioning of multiple foci within the cell. These multiple foci are believed to be as a result of chromosome replication and gene copy number. Espeli et al postulated specific regions either side of the cellular midpoint that individual macrodomains occupy. For the ORI and TER macrodomains, these regions were strictly symmetrical either side of the midpoint. For the LEFT and RIGHT macrodomains, foci locations were described as asymmetrical; in fact measured foci between 0.38-0.44 and 0.7-0.76 of the cell length were identified. NS macrodomains have already been proven to have increased motility and interactive properties, an observation corroborated during the work of *Espeli et al*. NS macrodomains, in conjunction with the LEFT and RIGHT macrodomains possess an asymmetric distribution either side of the midpoint however it has been noted that the distribution appears less defined due to the NS domains' increased levels of motility (Espeli et al, 2008).
Lau et al also presented similar data however, it was suggested that the TER macrodomain was variable in the position it occupies within the cell with respect to stages of chromosome

replication. *Lau et al* agreed that the spread of ORI macrodomains from the mid point was symmetrical and that following the completion of chromosome replication, the replication factories and TER macrodomains migrated towards the midpoint of the newly formed sister cells. However it was proposed that prior to complete chromosome segregation, the single focus representing the TER macrodomain shifted asymmetrically to one side of the mid point. It was hypothesised that this asymmetric shift provided directionality for the chromosome segregation protein, Filamenting temperature sensitive mutant K (FtsK) and ensured that the midpoint of the cell was void of chromatin upon the onset of cell division. Given the fact that *Espeli et al* produced their model based upon mid –age cell imaging and that chromosome segregation is believed to take place approximately half way through the cell cycle, it is likely the process described by *Lau et al* was not taken into consideration. Considering the shift proposed by *Lau et al* it is unsurprising that macrodomains closer in proximity to the TER macrodomain also experience a similar level of movement. This process may provide an explanation for the asymmetrical distribution of two foci seen at loci within the LEFT, RIGHT and both NS macrodomains. Anchoring of the ORI at the midpoint of the newly formed sister cells could well be a requirement for the spatial regulation of chromatin upon complete cell division (Lau et al 2003).

Unfortunately, the deficiency in the number of two foci data points from this study allows no conclusion regarding two foci position to be made. However early analysis would suggest that the *araC* loci does possess a wider spread of foci either the side of the midpoint in concordance with the increased motility of NS-R macrodomain. It is also possible to suggest that *dps* foci exhibit an asymmetrical distribution of foci about the midpoint, again consistent with the observations published regarding LEFT and RIGHT macrodomain loci.

4.3 Foci position and Chromosome replication

In a study of this nature, it is essential that the locations of tagged genes are a true representation of native chromatin organisation and not a result of overriding cellular processes such as cell replication and division. Chromosome migration to the midpoint of a cell, a process fundamental to successful division, is a process independent from transcription regulation and gene expression and therefore will ‘skew’ the gene locations obtained through this experimental procedure. Although impossible to truly eradicate these results from those

obtained at interphase of the cell cycle, it is possible to statistically estimate the effect these results has on a data sample and therefore make allowances in any conclusions drawn. Figure 24 plots the single focus position relative to cell length (on the x-axis), the y-axis plots the probability that these data points describe a normal distribution. The line in this figure represents the ideal normal distribution. When the data points and the line deviate the data no longer can be described by a normal distribution. The data shows that, up to a relative single focus position of (up to) approximately 0.45 the data has a normal distribution and therefore it can be attributed to the gene native position within the cell's interphase chromatin organisation. Above a relative single focus position of 0.45, the data is 'skewed' away from this normal distribution and thus indicates, as postulated, that these results are independent from native organisation and are as a result of cell division events. It is also worth noting a reduction in confidence with increasing foci polarity (Figure 24).

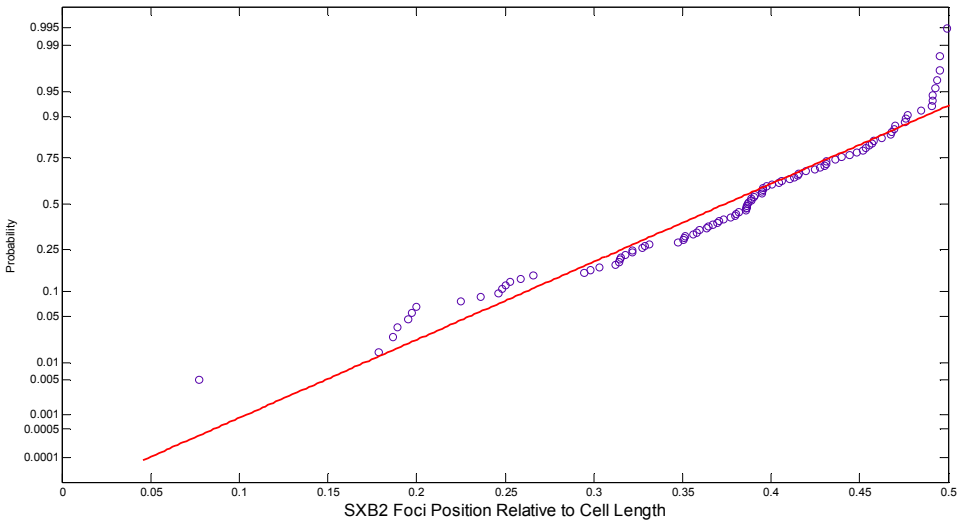


Figure 24: A probability plot outlining the distribution of SXB2 foci positions against a normal distribution model.

4.4 Foci dynamics at dps locus under Manganese and Iron supplemented growth

The numbers of foci displayed at the *dps* locus, in comparison to other foci such as *araC*, have already been discussed. Transcriptional regulation of the *dps* gene is a well studied topic; it is conceivable that an increase in cellular manganese or iron concentration would have an effect on the both the number and relative position of foci.

Yamamoto et al published evidence suggesting that *dps* expression in conjunction with an increase in cellular manganese was unaffected during exponential phase but was markedly repressed upon entry to stationary growth phase. It is also known that during exponential phase *dps* is repressed through the binding of Fis to its promoter, however as the growth rate reduces, the levels of Fis repression reduce rapidly (Yamamoto et al, 2011 and Grainger et al 2008). These findings would suggest that a decrease in gene accessibility following the addition of manganese would be more appropriate. Results presented in this study however do not support this. The small increase in the number of cells with one focus following the addition of manganese cannot be attributed to the rapid decline of Fis repression with a reduction in growth rate or the potential overlap between Fis and manganese mediated repression. It could be that this conflicting result will be eradicated through an increased sample size and/or repeats, however if proved to be correct, further investigation to explain the result is required.

Although not experimentally verified, *Yamamoto et al* also hypothesised that, in an effort to protect the cell from oxidative stress, *dps* expression may be induced under conditions of low manganese and increased iron (Yamamoto et al, 2011). If, as proposed, gene induction leads to an increase in gene accessibility, the 12 % increase in single focus cells witnessed following the addition of iron corroborates the hypothesis presented by *Yamamoto et al*.

In terms of foci position, the addition of iron had no considerable effect at the *dps* locus. This suggests that the increase in gene accessibility is not a result of structural changes to chromatin organisation such as supercoiling or NAP binding. Growth with 10 μM manganese however, identified a very small polar shift in foci. Indeed the small nature of the shift could well be eradicated with a larger sample size and increased repeats. However, if this shift proves to take place in vivo, changes in local supercoiling could possibly provide an explanation. As previously mentioned, *dps* during exponential growth is repressed by Fis

protein; a protein that itself functions through steric hindrance and prevention of negative supercoiling (Travers and Muskhelishvili, 2000). With a reduction in growth rate, Fis repression reduces rapidly, therefore allowing increased freedom for changes to local supercoiling topology. Manganese however, mediates mntR binding to the *dps* locus upon entry to a more stationary growth phase. Although mntR binding confers similar repression, the constraints placed upon supercoiling via Fis are not replaced. The conformational change in local supercoiling between these two *dps* repressive states could provide an explanation for the shift in foci position.

4.5 Future work

Clearly conclusions drawn as part of this study would need to be validated further. It would also be improper to suggest that results identified at the *dps* locus are representative in establishing how gene location and chromatin organisation is altered during transcription and required periods of gene expression.

An increased sample size, particularly for cells exhibiting more than one foci, would not only aid in confirming the presence of asymmetrical distributions, outlined in *Espeli et al*, but would also allow numerical comparisons with other loci from other macrodomains to be made. Time constraints unfortunately curtailed experimental repeats, an aspect that would require attention before reliable conclusions could be made. The mntR miniregulon contains five separate binding sites including the *dps* locus (Yamamoto et al, 2011 and Waters et al, 2011). Comparison of results from other binding sites, possibly within a single strain, would provide a more substantiated indication as to the variations in gene location and chromatin organisation in vivo. Resolving the growth issues confronted by this study when using the MalI: mCherry fusion plasmid would allow analysis of the *mntH* locus to take place.

Cell analysis during this study was carried out at exponential growth phase, a phase during which *dps* is known to be highly inactive. Analysis following growth up to stationary phase, and therefore increasing the possibility of *dps* expression, could provide different foci distributions to that witnessed as part of this study. In conjunction with this, analysis following increasing concentrations of metal ions may also provide different foci distributions.

5 References

- **Balleza E , Lopez-Bojorquez LN, Martinez-Antonio A, Resendis-Antonio O, Lozada-Chavez I, Balderas-Martinez YI, Encarnacion S, Collado-Vides J. (2009)** Regulation by transcription factors in bacteria: beyond description. FEMS Microbiol Rev Jan;33(1):133-51.

- **Browning. DF, Busby. SJ (2004)** The regulation of bacterial transcription initiation Nature Reviews Microbiology Jan;2(1):57-65

- **Delius. H, Worcel. A (1974)** Electron microscopic studies on the folded chromosome of *Escherichia coli* Cold Spring Harb

- **Espeli. O, Boccard. F (2006)** Organisation of the *Escherichia coli* chromosome into macrodomains and its possible functional implications. Struct Biol Nov;156(2):304-10

- **Espeli. O, Mercier R, Boccard F (2008)** DNA dynamics vary according to macrodomain topography in the E. coli chromosome. Mol Microbiol Jun;68(6):1418-27.

- **Grainger. DC**, **Goldberg MD, Lee DJ, Busby SJ. (2008)** Selective repression by Fis and H-NS at the *Escherichia coli* dps promoter. Mol Microbiol Jun;68(6):1366-77.

- **Herring. CD, Glasner JD, Blattner FR. (2003)** Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. Gene Jun 5;311:153-63

- **Kleppe. K**, **Ovrebö S, Lossius I**. **(1979)** The bacterial nucleoid. Journal of General Microbiology May;112(1):1-13.

- **Lau. IF, Filipe SR, Søballe B, Økstad OA, Barre FX, Sherratt DJ. (2003)** Spatial and temporal organisation of replicating *Escherichia coli* chromosomes. Mol Microbiol Aug;49(3):731-43.

- **Lee. DJ, Bingle LE, Heurlier K, Pallen MJ, Penn CW, Busby SJ, Hobman JL.(2009)** Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. BMC Microbiol Dec 9;9:252.

- **Lloyd. GS, Godfrey. RE, Busby. SJ (2010)** Targets for the MalI repressor at the divergent *Escherichia coli* K-12 malX-malI promoters. FEMS Microbiol Apr;305(1):28-34

- **Lodish. H, Berk. A, Kaiser. C.A, Krieger. M, Bretscher. A, Ploegh. H, Amon. A, Scott. M.P (2007)** Molecular Cell Biology Sixth Edition

- **Luijsterburg. MS, Noom MC, Wuite GJ, Dame RT (2006)** The architectural role of nucleoid-associated proteins in the organisation of bacterial chromatin: A molecular perspective. Journal of Structural Biology Nov;156(2):262-72

- **Pelicic V, Reyrat JM, Gicquel B (1996)** Expression of the Bacillus subtilis sacB Gene Confers Sucrose Sensitivity on Mycobacteria. Journal of Bacteriology Feb;178(4):1197-9

- **Peter. BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR (2004)** Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. Genome Biol ;5(11):R87

- **Robins – Browne RM, Hartland EL (2002)** *Escherichia coli* as a cause of diarrhea. Journal of Gastroenterol Hepatology Apr;17(4):467-75

- **Thanbichler. M, Wang. SC, Shapiro. L (2005)** The bacterial nucleoid: a highly organized and dynamic structure. J Cell Biochem Oct 15;96(3):506-21

- **Thanbichler. M, Shapiro. L (2006)** Chromosome organization and segregation in bacteria. Journal of Structural Biology Nov;156(2):292-303

- **Travers. A, Schneider. R, Muskhelishvili. G (2000)** DNA supercoiling and transcription in *Escherichia coli*. The FIS connection Feb;83(2):213-7.

- **Travers. A, Muskhelishvili. G (2005)** DNA supercoiling – a global transcriptional regulator for enterobacterial growth? Nature Reviews Microbiology. Feb;3(2):157-69

- **Travers. A, Muskhelishvili. G (2005)** Bacterial chromatin. Current opinions in Genetics and Development Oct;15(5):507-14

- **Valens. M, Penaud S, Rossignol M, Cornet F, Boccard F (2004)** Macrodomain organisation of the *Escherichia coli* chromosome. EMBO Journal 2004 Oct 27;23(21):4330-41

- **Waters. LS, Sandoval. M, Storz. G (2011)** The *Escherichia coli* MntR miniregulon includes genes encoding a small protein and an efflux pump required for manganese homeostasis. Journal of Bacteriology Nov;193(21):5887-97

- **Yamamoto K, Ishihama A, Busby SJ, Grainger DC (2011)** The Escherichia coli K-12 MntR miniregulon includes dps, which encodes the major stationary phase DNA-binding protein. Journal of Bacteriology Mar;193(6):1477-80

- http://emedicine.medscape.com/article/217485-overview#a0104
- http://ecocyc.org/

6 Appendix

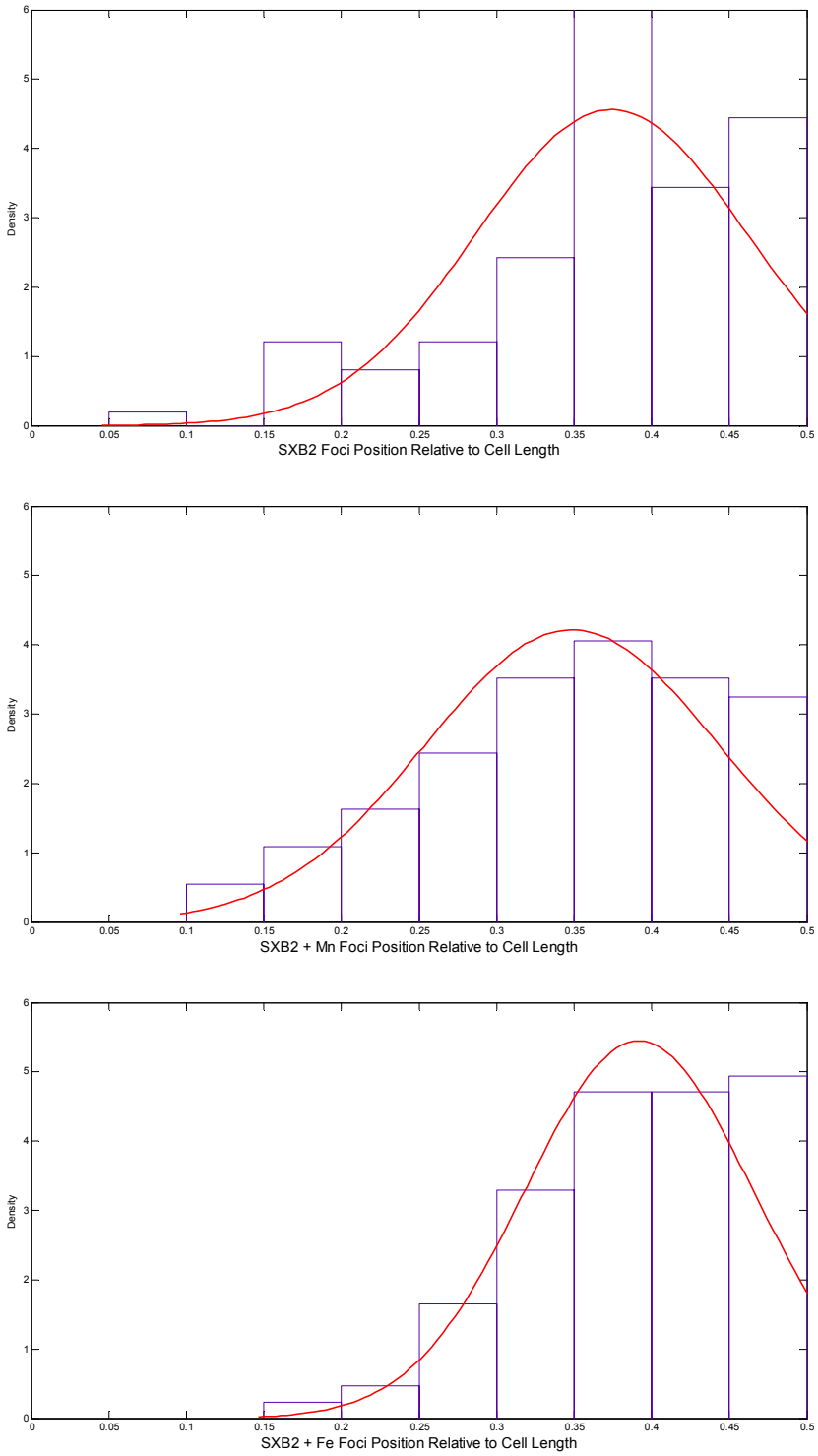


Figure 25: Distribution of witnessed foci relative to the cell length in SXB2 strains following growth in top) M9 x minimal salts media, middle) Manganese supplemented M9 media and bottom) Iron supplemented M9 media.
Line of mean foci position was based upon a normal distribution of data points.

University of Birmingham

School of Biosciences

Purification of transcriptional regulatory components of Group B *Streptococcus*

A research project report submitted by

Stephen Bevan

as part of the requirement for the degree of MRes in
Molecular and Cellular Biology

This project was carried out at: School of Biosciences, University of Birmingham

Under the supervision of: Dr David Lee Date: March ’12 – July ‘12

Contents

List of figures and tables

List of Abbreviations

APS- Ammonium Persulphate
BSA- Bovine Serum Albumin
CAMP factor – Group B Streptococcal pore-forming toxin whose name derives from its discovers; Christie, Atkins and Munch-Petersen
cfb – Structural gene responsible for transcribing the CAMP factor
CIP – Calf Intestinal Phosphatase
covR – Regulatory transcription factor of the covRS two component system
covR GuHCl – Insoluble covR protein extracted from the inclusion bodies before being homogenised into a Guanidine buffer and dialysed overnight into Buffer A.
covS – Sensor histidine kinase of the covRS two component system
cylE – Structural gene responsible for transcribing β hemolysin/cytolysin
cylX – Promoter region of the structural *cylE* gene
dNTP's – Deoxynucleotide Triphosphates
DTT- Dithiothreitol
EDTA-Ethylenediaminetetraacetic acid
EOD- Early Onset Disease
IPTG - Isopropyl β-D-thiogalactopyranoside
LB- Luria Bertani broth
LOD- Late Onset Disease
Ni-NTA – The chelation resin of nitrilotriacetic acid loaded with divalent Nickel ions incorporated in Nickel purification columns
OD- Optical Density
PCR- Polymerase Chain Reaction
rgfA – Regulatory transcription factor of the RgfC/RgfA two component system
rogB – Group B Streptococcal transcriptional regulator
rovS – Group B Streptococcal transcriptional regulator
rpoD – Group B Streptococcal Sigma factor
***scpB* –** Structural gene responsible for transcribing the C5a peptidase
SDS-Page- Sodium Dodecyl Sulphate Polyacrylmide gel electrophoresis
Spx – A family of unique RNA polymerase-binding proteins
***Stk1* –** Serine/Threonine Kinase 1
TAE – Tris/Acetic acid/ EDTA
TBE- Tris/Borate/EDTA
TEMED- Tetramethylethylenediamine
TFB – Standard Transformation Buffer

Abstract

Group B *Streptococcus* is a gram positive bacterium that has significant health implications for a large section of the neonates as well as the immuno-compromised elderly population. The bacterium is able to alternate between dormant, asymptomatic and invasive, potentially lethal, states when confronted with a suitable host. Understanding the regulation of its virulence factors may develop the understanding of how Group B *Streptococcus* is able to alternate between these two contrasting 'life cycles'. This study focuses on the purification of several transcription factors, known to be members of two-component regulatory systems responsible for the regulation of expression of several virulence factors. Through purification of one particular transcriptional regulator, covR, this study has been able to confirm its DNA binding ability at sections of DNA upstream of the structural genes of two pore-forming toxins, whose expression is regulated by the covRS regulatory system. Band-shift analysis has also confirmed that phosphorylation of the transcriptional regulator covR, increases the affinity of DNA/protein interactions. In contrast to other studies, this study also suggests that covR, in both its phosphorylated and unphosphorylated forms, functions as a monomer. There is to date, no indication that covR oligomerises to a dimer or any other more complex structural conformation. During the course of this study a Group B Streptococcal sigma factor and one of the Spx family of proteins has also been purified.

1 Introduction

1.1 Group B Streptococcus and health implications

Bacterial infections have considerable health implications throughout the entire human population. The adaption of many bacterial strains to antibiotics has heightened the problems of infection containment and prevention to recently unseen levels. The current hype surrounding methicillin resistant Staphylococcus aureus (MRSA), the so called 'Super Bug', has placed considerable emphasis on the causes of bacterial infections affecting the human population.

One such bacterium identified by the scientific community, is Group B *Streptococcus*, a gram positive bacterium capable of forming long chain cocci within the vaginal region of a large population of women (Rajagopal 2009). It is already known that this bacterial population colonises up to 25 % of healthy women, however, it usually resides asymptomatically, that is to say it doesn't display any disease associated symptoms (Rajagopal 2009 and Maisey et al 2008). The importance of this bacterium arises however, when confronted with a host possessing a suboptimal immune system. Under these conditions, the bacteria are able to capitalise on the weakened immune system by promoting invasive disease symptoms and extensive tissue damage (Maisey et al 2008). Given this mode of action, alongside the native location of the bacteria, it is unsurprising that Group B Streptococcal infection has the highest incident rate of all life-threatening infections associated with the human neonatal population (Rajagopal 2009). Understanding not only the mode of transmission from mother to child, but also the regulation of its virulence from an asymptomatic to an invasive disease state, may illuminate potential areas of interest with regards to reduced pathogenicity.

Significant research into the transmission of the bacteria between the two demographic populations has already been carried out. In fact, it is believed that the transfer of bacteria can occur prior to birth, within the womb itself, or during birth as the newborn comes into contact with contaminated vaginal fluids (Rajagopal 2009). The fact that this bacteria is able to lie asymptomatic within the vaginal region, consequently means the majority of mothers are unaware they are carriers of the disease. It therefore only becomes evident once the disease manifests itself within the neonate, by which time the chance of successful medical treatment is significantly reduced. Recent campaigns promoting the awareness of Group B *Streptococcus,* along with advances in maternal screening and disease diagnosis and treatment

have attempted to reduce the levels of infection, however fatality levels remain high enough for continued extensive research (Rajagopal 2009).

Disease manifestation can be categorised into either early onset or late onset, based unsurprisingly on the process of manifestation and the age of the patient. Early onset disease (EOD) is generally defined as occurring within the first few hours or days of a newborn's life. Given the highly fragile nature of the host's immune system at such an age, infection often causes respiratory failure and pneumonia with continued progression causing bacteraemia (Rajagopal 2009 and Maisey et al 2008). The EOD fatality rate of approximately 30 % is attributed to the severity of the disease associated symptoms mentioned above. In comparison, late onset disease (LOD) arises in infants up to 3 months in age and is generally a result of bloodstream infection. LOD is highly linked with the onset of meningitis; however, the development of the immune system with age results in LOD possessing a fatality rate of approximately 5 % (Rajagopal 2009 and Maisey et al 2008). Although potentially less fatal than EOD, LOD is believed in 30 % of cases to cause long-term problems such as retardation and a loss of hearing (Maisey et al 2008). As previously mentioned, Group B *Streptococcus* also thrives in those patients that are said to be immuno-compromised. Given this fact, invasive infections are known to also affect the elderly population. Those patients over the age of 65 have the highest fatality rate of all patients affected by Group B *Streptococcus* infections (Rajagopal 2009 and Maisey et al 2008).

The fatality rates associated with Group B *Streptococcus* infections and the large section of the human population potentially effected reinforces the fact that Group B Streptococci is a significant problem for public health. Combine these facts with the development of antibiotic-resistant bacteria, the urgency for preventive techniques cannot be overstated.

1.2 Bacterial Virulence Factors

As previously mentioned, Group B Streptococci infections cause significant health problems, not only in the worldwide neonatal populations but those elderly patients who are believed to be more immuno-compromised than the average human population. The potency of these infections is currently attributed to the number of extracellular virulence factors accompanying the bacteria. Some of these virulence factors are responsible for the adherence of the bacterial cells to vaginal epithelial cells under the extreme environmental conditions exhibited. The adherence of the bacterial cells to host epithelial cells, although not strictly involved in bacterial pathogenicity, is crucial for transmission to the neonatal population (Liu and Nizet 2004). Other attributes possessed by virulence factors include mechanisms that are able to avoid the immunological clearance response mediated by neutrophils and macrophages, along with an increased ability to penetrate epithelial and endothelial cells particularly with regards to the Blood/Brain barrier and the lungs. Many bacteria also have increased pathogenicity due to the secretion of a number of toxins and the ability of the bacteria to impart an inflammatory response on the host organism (Liu and Nizet 2004). As previously mentioned, these attributes are exacerbated in the neonatal population due to their under developed immune system (Liu and Nizet 2004).

There are generally considered to be three main virulence factors associated with Group B *Streptococcus*; β hemolysin/cytolysin, CAMP factor and C5a peptidase. Host C5a is a crucial protein required for the recruitment of neutrophils to a site of infection. The peptidase possessed by some virulent bacteria cleaves this protein rendering neutrophil recruitment absent. This study essentially will not focus on C5a peptidase (Rajagopal 2009).

1.2.1 β hemolysin/cytolysin

Although there are other factors required for the progression of this infection, aside from the three outlined previously, this study will focus on two pore forming toxins: β hemolysin/cytolysin and the CAMP factor (Rajagopal 2009). The discovery of the haemolytic properties possessed by Group B *Streptococcus* came through laboratory growth on blood agar plates in the early 1930s (Liu and Nizet 2006). This fact had largely been neglected however until recently, when further research into the bacterial infection established that Group B *Streptococcus* variants have the ability to lyse not only red blood cells but a large variety of other eukaryotic cell types (Liu and Nizet 2006).

The structural gene responsible for β hemolysin/cytolysin, *cylE,* codes for a protein of approximately 78 kDa (Rajagopal 2009). The *cylE* gene is one of twelve comprising the *cyl* operon however, through mutant expression assays; it is believed that the protein produced by the *cylE* gene is not only necessary for β hemolysin/cytolysin but also sufficient for bacterial haemolytic properties (Liu and Nizet 2006). The remaining members of the operon are involved in the regulation of *cylE* expression, depending on environmental stimuli (Rajagopal 2009). Given the obvious importance this virulence factor possesses with regards to the potency of the bacteria, significant research has been carried out into clarifying not only the mode of action, but also the protein's structure and how transcription of the *cylE* gene is regulated. To date, little is known about the structure of the protein due to the inability to purify it in an active form (Liu and Nizet 2006). In terms of regulation, the *cylE* gene is constitutively expressed, however, the levels at which the gene is transcribed can be further regulated through environmental stimuli (Liu and Nizet 2006). A more in depth review of virulence factor expression will be discussed later.

β hemolysin/cytolysin is a virulence factor located on the surface of the cell, allowing it to be easily cleaved and transferred to host epithelial cells upon conversion from an asymptomatic to an invasive state (Liu and Nizet 2004 and Liu and Nizet 2006). Indeed, the protein is only believed to be active when bound to cell membranes, or in the presence of large stabilising molecules (Rajagopal 2009). Upon binding to host epithelial cells, the protein begins its invasive function through several modes of action. The protein's main invasive technique is causing the formation of pores within the epithelial and endothelial cells of the lungs and blood/brain barrier. This allows the bacteria, in combination with the other remaining virulence factors, to further convey its invasive response deeper into the target tissues and the bloodstream (Liu and Nizet 2004 and Liu and Nizet 2006). β hemolysin/cytolysin also imposes an increased inflammatory response within the host organism through encouraging the apoptosis of macrophages and the expression of interleukin – 8, a pro-inflammatory cytokine (Liu and Nizet 2004).

In combination with β hemolysin/cytolysin production, is the presence of a carotenoid pigment. The formation of this pigment is directly linked to β hemolysin/cytolysin expression from the *cylE* gene, as *cylE* knockout mutants are void of this pigment. The pigment possesses twelve unsaturated bonds, promoting a conclusion that it is involved in pathogen resistance to oxidative stress imparted by the host organism. To date the functional link

between β hemolysin/cytolysin and the carotenoid pigment has not fully be elucidated however, a direct link between pigment levels and the haemolytic activity of β hemolysin/cytolysin has been proven (Rajagopal 2009 and Liu and Nizet 2006).

1.2.2 CAMP factor

The CAMP factor, similar to β hemolysin/cytolysin, is a 23.5 kDa extracellular pore forming toxin transcribed from the *cfb* gene (Rajagopal 2009 and Liu and Nizet 2004). Its role in Group B *Streptococcus* virulence was first established through in vivo rabbit studies. The presence of the CAMP factor triggered the onset of septicaemia, which eventually became lethal (Rajagopal 2009). The CAMP factor is believed to cause cell lysis through its ability to bind glycophosphotidylinostol anchored proteins within the membranes of targeted cells. Oligomerisation of the protein aids in the formation of these pores that are so crucial to the protein's mode of action (Rajagopal 2009). Recent studies have postulated that the CAMP factor is involved in a compensatory relationship with β hemolysin/cytolysin (Rajagopal 2009). In conditions where β hemolysin/cytolysin has limited functionality, such as areas populated by the lung surfactant dipalmitoylphophatidylcholine, the expression of the CAMP factor may increase to maintain the virulence levels previously exhibited (Rajagopal 2009 and Liu and Nizet 2006). In general however, it is believed that the CAMP factor is a significant understudy to β hemolysin/cytolysin with regards to bacterial virulence. The CAMP factor has also been proven to bind sections of human antibodies, therefore it is currently up for debate as to whether its reduced cytolytic properties in comparison to β hemolysin/cytolysin, or the ability to bind immunoglobulin proteins causes the greatest pathogenetic problems (Liu and Nizet 2004).

1.3 Regulation of Prokaryotic Transcription

The majority of bacterial transcription is regulated through transcription factors. Transcription factors possess an ability to either repress or activate gene transcription. Transcriptional repressors bind to DNA and inhibit transcription through steric hindrance, either through direct binding to the genes promoter or through conformational structural changes to the coiling of DNA. Either mechanism inhibits polymerase binding or the formation of the DNA open complex. Transcriptional activators function by increasing the affinity of RNA polymerase binding to DNA or facilitating the interchange between closed and open

complexes required for continual transcription (Browning and Busby, 2004). Binding of these activators usually takes place at a site upstream of the genes' promoter (Balleza et al, 2009).

1.3.1 Two Component Systems

Group B *Streptococcus* alternates between an asymptomatic state, within the vaginal region of infected females, to a virulent state upon transfer to a neonatal patient (Maisey et al 2008). In order to mediate this increase in virulence with a change in environment successfully, regulation of expression of the appropriate genes and virulence factors is crucial. Group B *Streptococcus* tackles this problem through the use of two component regulatory systems (Jiang et al 2005). Fundamentally, changes in environmental stimuli are recognised by a membrane bound sensor histidine kinase. Upon recognition of the appropriate changes in environmental conditions, the sensor kinases autophosphorylate and then in turn, phosphorylate their related regulatory transcription factors at a specific conserved aspartate residue (Jiang et al 2005 and Lin et al 2009). It is believed that the phosphorylation of the regulator protein confers the necessary changes in DNA affinity and therefore, has a knock on effect on the transcription of target genes (Lin et al 2009).

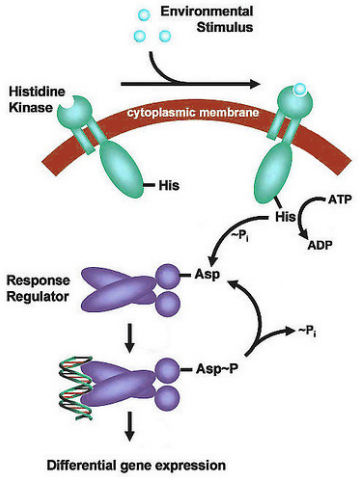


Figure 1 : Illustration portraying how prokaryotic two component systems function (Microbiology Bytes = http://www.microbiologybytes.com/blog/2010/02/15/two-component-signal-transduction-in-bacteria/)

Group B *Streptococcus* possesses approximately twenty of these two component regulatory systems, of which many are responsible for the regulation of the virulence factors mentioned previously (Rajagopal 2009). The major two component system investigated in this study is the covRS system that has been proven, through Quantitative Reverse Transcription

Polymerase Chain Reaction (QRT-PCR), to regulate the production of both β hemolysin/cytolysin and CAMP factor from their associated structural genes (Jiang et al 2008). covR is believed to repress transcription of the *cyl* operon, including that of *cylE*, whilst activating the *cfb* gene responsible for CAMP factor production (Jiang et al 2005). The antagonistic nature of covR gene regulation at these two sites lends further evidence to the compensatory relationship of the two virulence factors mentioned earlier. It is important to note that a serine/threonine kinase, Stk1 is also able to phosphorylate covR at a different site and regulates covR in an opposite manner to covS (Lin et al 2009). The reasons behind this antagonistic regulation has yet to be fully established, however it further illustrates the refined nature of gene regulation required to adapt to the various changes in environment.

Several other proteins targeted in this study are also members of two component systems responsible for the regulation of Group B *Streptococcus* virulence factors. RogB functions as transcriptional repressor affecting the production of the sialic acid capsular polysaccharide utilised in the immune evasion response. RovS is known to regulate the level of β hemolysin/cytolysin, as well as superoxide dismutase and fibrinogen binding protein A; virulence factors involved in bacterial immune evasion and host cell adherence respectively. Finally, rgfA represses transcription of the *scpB* gene, a gene responsible for the production of the C5a peptidase discussed previously (Rajagopal 2009).

1.4 Aims and Objectives

The initial aim of this project was to over-express and purify five Group B *Streptococcus* proteins; rgfA, rogB, rpoD, rovS and covR. As previously mentioned rgfA, rogB, rovS and covR are transcription factors involved in the regulation of many virulence genes, whilst rpoD is a Streptococcal sigma factor. The initial aim was to over-express and purify these proteins allowing downstream experiments such as band-shifts and DNA foot printing to be carried out. Through these experiments it was hoped an indication as to the binding regions of these transcriptional regulators with respect to their relevant associated virulence genes could be identified. A further series of proteins, including spxs 0090, 1031, 1617 and 2046, were added to the list as the project developed.

1.5 Spxs

As mentioned in the aims and objectives, a series of numbered proteins were investigated in combination with the transcriptional regulators. One protein, 1617, was of particular interest. The 1617 protein is known as a Spx, a family of RNA polymerase-binding proteins first discovered in *Bacillus subtilis*. Throughout further research, homologues of this protein family have been found in a number of low G + C content, gram-positive bacteria. The function of this family of proteins, is to bind the alpha subunit of RNA polymerase and disrupt the transcriptional activation mechanisms of two-component regulatory systems. To date, no specific DNA binding ability has been identified for spx proteins, it is therefore suggested that disruption of transcriptional activation must arise through an unproven mechanism. It is conceivable that the steric hindrance applied by binding of the spx to the alpha subunit is sufficient to impart its regulatory effects. Research carried out by Zuber has also elucidated a role for spxs in the activation of transcription in response to oxidative stress (Zuber 2004).

Throughout the genome of Group B *Streptococcus*, three spx proteins are coded for. Usually many species of bacteria possess only one, leading to the obvious question, why does Group B *Streptococcus* have inflated numbers? Clearly, within the regulation system of Group B *Streptococcus*, spxs are linked to two component systems in regulating gene expression in response to environmental conditions. The ability to promote and repress transcription, as seen before with covR, further highlights the complex nature of gene regulation taking place within the bacteria. The ability to activate certain genes in response to oxidative stress, suggests that these spx proteins may have a role in controlling the virulence of bacterial cells. Through purifying one of these proteins therefore, it is hoped that downstream experiments can further develop our knowledge of how these proteins function, do they have a role to play in the virulence of the bacteria? and why does Group B *Streptococcus* require three separate spxs?

2 Materials and Methods

Strain	Description	Origin
RLG 221	*F⁻, λ⁻, ilvG, rfb-50, rph-1*	R. Gourse
T7 Express	*Escherichia coli* strain used for over expressing recombinant protein	New England Biolabs

Table 1: *Escherichia coli* strains utilised in this study

Plasmid	Description	Origin
pET28	Protein over-expression vector	Novagen

Table 2: List of plasmids utilised in this study

Primer	Primer Sequence (5'—3')	Binding specificity
D76562	GGG-GTT-ATG-CTA-GTT-ATT-GCT-CAG-C	Reverse sequencing primer for pET28a plasmid
D44965	GGG-GAA-TTG-TGA-GCG-GAT-AAC	Forward sequencing primer for pET28a plasmid
D77115DL	AGC-TCC-ATG-GTT-GAA-AAT-TAT-TTA-GAA-AAA-GAC	NcoI primer for the Group B Strep gene rogB
D77100DL	TAG-TCT-CGA-GAG-AGT-TAT-AGG-AGC-AGC-TAC-GGG	XhoI primer for the Group B Strep gene rogB
D77114DL	AGC-TCC-ATG-GAT-ATT-TTT-ATA-CTT-GAG-GAT-G	NcoI primer for the Group B Strep gene rgfA
D77098DL	TAG-TCT-CGA-GTC-TTT-TTC-TAT-CCT-TTT-GAT-ATT-TAT-CC	XhoI primer for the Group B Strep gene rgfA
D77093DL	AGC-TCC-ATG-GGT-AAA-AAG-ATC-TTA-ATA-ATC-G	NcoI primer for the Group B Strep gene covR
D77094DL	TAG-TCT-CGA-GTT-TTT-CAC-GAA-TCA-CAT-AGC-C	XhoI primer for the Group B Strep gene covR

D77091DL	GAT-TCC-ATG-GCA-GAG-AAA-AAA-GGA-AAT-ACA-AC	NcoI primer for the Group B Strep gene rpoD
D77092DL	TAG-TCT-CGA-GAT-CTT-CCA-TGA-AAT-CTT-TAA-GTT-GC	XhoI primer for the Group B Strep gene rpoD
D77095DL	AGC-TCC-ATG-GAA-AAA-GAA-TTA-GGA-AAA-ACA-C	NcoI primer for the Group B Strep gene rovS
D77096DL	TAG-TCT-CGA-GGC-ATT-CTT-TAT-TAT-TGC-CAA-GTA-C	XhoI primer for the Group B Strep gene rovS
D77284DL	AGC-TCC-ATG-GTT-ACC-TTA-TTT-TTA-CCC	NcoI primer for the Group B Strep gene 1031
D77285DL	TAG-TCT-CGA-GAT-CAT-GTT-TTC-CCT-CAA-TTT-C	XhoI primer for the Group B Strep gene 1031
D77288DL	AGC-TCC-ATG-GTT-AAG-ATT-TAT-ACT-ATT-TCA-AG	NcoI primer for the Group B Strep gene 2046
D77289DL	TAG-TCT-CGA-GAA-GTG-CTG-CAC-GAA-GAC-GAG	XhoI primer for the Group B Strep gene 2046
D77290DL	AGC-TCC-ATG-GCT-TAT-ACG-TTT-TAT-GAA-TAC-CCC	NcoI primer for the Group B Strep gene 1617
D77291DL	TAG-TCT-CGA-GTA-CTA-AGT-TGA-GGT-CTT-TAT-AC	XhoI primer for the Group B Strep gene 1617
D77292DL	AGC-TCC-ATG-GCA-AGA-GAT-TTT-GAG-GAG-TTG	NcoI primer for the Group B Strep gene 0090
D77293DL	TAG-TCT-CGA-GAT-CCG-CTT-TAA-AAG-CAT-TGA-AAA-AAG-G	XhoI primer for the Group B Strep gene 0090
D77294DL	AGC-TCC-ATG-GCT-TCA-ATT-TTA-GAT-GAC-TAT-GAA	NcoI primer for the Group B Strep gene 0249
D77295DL	TAG-TCT-CGA-GTA-TTT-TAA-AAT-TAT-CCT-TTT-CTA-A	XhoI primer for the Group B Strep gene 0249

D77780DL	GAA-TAT-GAA-TTC-TTT-ATC-TAA-AAT-AGT-ACG-CTT-C	Anneals approximately 400 bp upstream of the cfb gene running towards the gene itself.
D77781DL	GAA-TAT-AAG-CTT-AGA-TAG-ATA-CAT-CAT-ATG-TTT-ATC-G	Anneals at the beginning of the cfb gene, running back towards the genes' upstream sequence
D77782DL	GAA-TAT-GAA-TTC-GGC-GCA-AGT-ATA-TTA-TAG-C	Anneals approximately 200 bp upstream of the cfb gene, running towards the beginning of the gene.
D77783DL	GAA-TAT-AAG-CTT-GCT-ATA-ATA-TAC-TTG-CGC-C	Anneals approximately 200 bp upstream of the cfb gene and is a reverse complimented sequence of D77782DL.
D77777DL	GAA-TAT-CAA-TTG-TAG-ATG-TGC-TTT-CTA-AGA-GCG	Anneals approximately 400 bp upstream of the *cylX* gene running towards the gene sequence
D77778DL	GAA-TAT-AAG-CTT-CAC-ACC-TAC-TCA-AAA-TAT-TAG-AAC-G	Anneals at the beginning of the *cylX* gene running back towards the gene's upstream sequence
D77779DL	GAA-TAT-AAG-CTT-ACA-AAA-TCA-TGA-GTA-AAA-CAC-TGC	Anneals approximately 200 bp upstream of the *cylX* gene running towards the gene sequence

Table 3: List of primers utilised as part of this study. Sequences highlighted indicate NcoI, XhoI, MfeI, HindIII and EcoRI restriction sites juxtaposing primer sequences.

Promoter sequence	**Primers used in PCR amplification**	**Location of the sequence fragment**
1	D77780DL and D77781DL	400 base pairs immediately upstream of the cfb promoter
2	D77781DL and D77782DL	200 base pairs immediately upstream of the cfb promoter
3	D77780DL and D77783DL	First 200 base pairs of promoter sequence 1
4M	D77777DL and D77778DL	400 base pairs immediately upstream of the *cylX* gene
4E	*Digestion of 4M fragments with EcoRI and HindIII restriction enzymes	200 base pair region immediately upstream of the *cylX* gene
5	D77777DL and D77779DL	First 200 base pair region of promoter sequence 4M

Table 4: List of promoter sequences used in band-shift experiments.

Buffers and Solutions

All buffers and solutions were purchased from Sigma Aldrich, BDH or Fisher scientific unless stated. Prior to use, all solutions used in bacterial growth were autoclaved at 120 °C and 15 psi for 20 minutes.

Required for Polymerase Chain Reactions (PCR) and PCR purifications;

Primers (Alta bioscience)

Diluted to a final concentration of 1 μM

Deoxynucleotide Triphosphates (dNTPs) (Bioline)

Diluted from 0.25 mM to a concentration of 1 μM each

Sodium Acetate

Concentration of 3 M Na^+ and pH 5.3

Required for running DNA agarose gels;

Agarose gel

0.8 % Agarose in 1 x TAE buffer. Heated in the microwave prior to use.

40 x Tris/Acetic acid/EDTA (TAE) buffer

2 M Tris acetate, 100 mM Na_2EDTA (National Diagnostics)

DNA markers

100 bp and 1 Kb ladders (NEB)

DNA loading dye

0.025 % Bromophenol Blue, 0.025 % Xylene Cyanol F, 20 % Glycerol, 10 mM Tris, 1 mM EDTA

Ethidium Bromide

10 mg/ml (Biorad)

Required for digestions and vector preparation;

Calf Intestinal Phosphatase (CIP) (New England Biolabs)

Required for production of competent cells;

Standard Transformation Buffer (TFB) I

30 mM KOAc

50 mM $MnCl_2$

100 mM RbCl

10 mM $CaCl_2$

15 % Glycerol

TFB II

10 mM NaMOPS pH 7

7.5 mM $CaCl_2$

10 mM RbCl

15 % Glycerol

Required for induction of cell cultures;

Isopropyl β-D-thiogalactopyranoside (IPTG) (Bioline)

Diluted in sterile water to a stock concentration of 1 M

Buffers used in cell lysis, protein extraction and protein purification;

Re-suspension Buffer

20 mM Tris

150 mM NaCl

5 % Glycerol

Lysis Buffer

40 mM Tris HCl pH 8

300 mM KCl

10 mM Ethylenediaminetetraacetic acid (EDTA)

0.2 % Sodium Deoxycholate

Denaturing Buffer

8 M Urea

50 mM Tris pH 7.5

500 mM NaCl

10 % Glycerol

GuHCl Denaturing Buffer

6 M GuHCl

50 mM Tris pH 7.9

1 mM EDTA

1 mM Dithiothreitol (DTT)

10 % Glycerol

Buffer A

50 mM Tris pH 7.5

100 mM NaCl

10 % Glycerol

10 mM Non-fluorescent blank imidazole

Buffer B

50 mM Tris pH 7.5

100 mM NaCl

10 % Glycerol

500 mM Fluorescent blank imidazole (Imidazole that has a significantly fluorescence as a result of the highest levels of purity possible, believed to be less than 0.005%)

No Salt Buffer

50 mM Tris pH 7.5

5 % Glycerol

0.1 mM EDTA

0.1 mM DTT

Low Salt Buffer

50 mM Tris pH 7.5

50 mM NaCl

5 % Glycerol

0.1 mM EDTA

0.1 mM DTT

High Salt Buffer

50 mM Tris pH 7.5

1 M NaCl

5 % Glycerol

0.1 mM EDTA

0.1 mM DTT

All used in the construction, running and staining of protein polyacrylamide gels;

Upper Buffer

15.15 g Tris

10 ml 10 % Sodium Dodecyl Sulphate (SDS)

500 µl Tetramethylethylenediamine (TEMED)

pH 6.8

Made up to 250 ml using distilled water

Lower Buffer

45.4 g Tris

10 ml 10 % SDS

250 µl TEMED

pH 8.8

Made up to 250 ml using distilled water

Lower Gel component

25 % Lower Buffer

50 % Protogel

23 % H_2O

2 % Ammonium Persulphate (APS)

Upper stacking gel component

25 % Upper Buffer

10 % Protogel

63 % H_2O

2 % APS

All figures are appropriate for the formation of 15 % Gels

Butan-1-ol

Water saturated 50 % - 50 %

10 well/15 well Bis-tris protein gels (Invitrogen – pre cast)

Sodium Dodecyl Sulphate- Polyacrylamide gel electrophoresis (SDS-Page) Loading Buffer

2 g SDS

20 ml Glycerol

5 mg Bromophenol

Made up to 92 ml with a 1 in 10 dilution of stacking buffer

SDS-Page Loading Dye

1 ml SDS-Page loading buffer

87 µl β- mercaptoethanol

See Blue Plus 2 Prestained standard (Invitrogen)

Broad range marker (Bio-Rad)

Kaleidoscope polypeptide standards (Bio-Rad)

Comassie Blue stain

50 % Methanol

10 % Acetic acid

2 g Brilliant Blue R

Made up to I litre with distilled water

Fast De-stain

40 % Methanol

10 % Acetic acid

Made up to 1 litre with distilled water

Required for Bradford Assays;

Protein Assay Dye Reagent concentrates (Bio-Rad)

Required for DNA Acrylamide gels for band-shifts;

Gel buffer

7.5 % Acrylamide

7.5 % Glycerol

0.5 X Tris/Borate/EDTA (TBE)

Acrylamide Gels

20 ml Gel Buffer

30 µl TEMED

200 µl 10 % APS

DNA Sequencing

Functional Genomics and Proteomics Lab, University of Birmingham

3.2 picomole of primer added to a 10 µl sample

Growth Media

Luria Bertani Broth: 20 g Tryptone (peptone), 10 g Yeast extract 10 g NaCl in 1 Litre of distilled water. All solutions were autoclaved at 120 °C and 15 psi for 20 minutes and then stored at room temperature prior to use.

Nutrient Agar: 23 g Nutrient Agar powder (Nifco) in 1 litre distilled water. Prior to use, the agar solution is autoclaved at a temperature of 120 °C and a pressure of 15 psi for 20 minutes. Cooled Nutrient Agar plates were stored at 4 °C before use.

Antibiotics

All antibiotics created were filter sterilised prior to use and stored at -20 °C

Kanamycin

Made at 50 mg/ml with sterile water and used at a 1000 x dilution of 50 µg/ml

Ampicillin

Made at 40 mg/ml with sterile water and used at 80 µg/ml

Competent Cells

1 ml of the overnight culture was added to 50 ml of LB broth and incubated at 37 °C until mid-logarithmic growth phase, optical density 0.3-0.5 at 650 nm. Upon reaching the correct OD, the sample was cooled on ice for 20 minutes and harvested through centrifugation at 4000 rpm at 4 °C for 10-15 minutes. Pelleted cells were then re-suspended in 15 ml ice cold TFB I buffer and incubated on ice for a further 20 minutes. Following a further identical centrifugation step, the pelleted cells were re-suspended in 2 ml TFB II and incubated on ice for 30+ minutes. Cells not required immediately for use were stored at -80 °C.

PCR used to amplify insert fragments

Typical 100 µl reaction volumes consisting of the following were constructed;

10 µl GC phusion buffer

2 µl 12.5 mM each dNTPs

1 µl each of the appropriate 10 µM oligos

1.5 µl phusion enzyme

Each reaction volume was made up to 100 µl using an appropriate volume of DNA template, depending on concentration, and sterile distilled water. Typically 1-2 µl of genomic Group B *Streptococcus* DNA template was implemented.

A typical PCR cycle as stated below was then carried out;

94 °C	------ 5 minutes	
94 °C	------ 30 seconds	X 32
Annealing Temperature	------ 30 seconds	
71 °C	------ 15/30 seconds per kb of amplified fragment	
71 °C	------ 7 minutes	

DNA purification

Purification was carried out using either the QIAgen or Bioline purification kit, following the instructions accompanying the kit. The purified product was eluted into 20 µl distilled sterile water.

Digestions

2 separate double digestions were carried out in this study using three pairs of restriction enzymes; NcoI and XhoI, HindIII and EcoRI. A 50 µl reaction volume was made using the following ratios;

- 5 µl Buffer – appropriate to restriction enzymes as recommended by NEB
- 0.5 µl BSA – if required
- 2 µl HF enzymes
- Volume of DNA required to bring total reaction volume up to 50 µl

In each digestion, the reaction volume was left to incubate at 37 °C for three hours.
If the digested plasmid was to be used as a vector for subsequent ligations, a Calf Intestinal Phosphatase (CIP) reaction was carried out using 2.5 µl CIP and 5 µl buffer 3, in a 50 µl reaction. This reaction prevents further uncontrolled ligation of the vector by removing 5' phosphates and was once again left at 37 °C for a further hour.

Gel Electrophoresis

Liquid agarose gel was cooled in a casting plate coupled with sufficiently sized gel comb until solid. 5 µl DNA samples combined with loading dye were loaded and the gel ran at 90 V in 1 X TAE buffer for approximately 40-45 minutes. Upon completion of the run, gels were stained in 10 mg/ml ethidium bromide solution or SYBRSafe and viewed using a UV trans-illuminator or blue light box depending on the purpose of the gel.

Agarose gel extraction

Upon completion of the run, agarose gels were stained in SYBRSafe for 30 minutes and the required bands physically extracted with the aid of a blue light box. Upon extraction the required DNA was purified using the recommended Bioline protocol and Bioline reagents. Elution of the DNA was carried out into 50 µl of sterile distilled water.

Ligations

Various ligation ratios of insert and vector volumes were calculated and assembled. A total reaction volume of 20 µl was reached using sterile water, 1 µl of T4 DNA ligase and 2 µl T4 DNA ligase buffer. The reaction was left resting horizontally on ice for at least 10 minutes prior to incubation overnight (8 hours) at 16 °C. Upon completion of the ligation period, the sample was kept at 4 °C until use.

Transformations

1-2 µl of the plasmid was added to 100 µl of appropriate *Escherichia coli* competent cells and placed on ice for at least 30 minutes. Following ice incubation, cells were subjected to heat shock at 42 °C for 90 seconds, before being placed back on ice for a further 5 minutes. 500 µl LB broth was added to the cells and incubated at 37 °C for an hour before centrifugation, re-suspension and plating out on appropriately antibiotic treated nutrient agar plates. Incubation of the plates took place at 37 °C overnight. For the transformation of ligations, a larger initial volume of 5 µl was added to the volume of competent cells.

Plasmid DNA extraction

Overnight cultures of plasmid carrying cells were grown in LB, supplemented by appropriate antibiotics. Plasmid DNA extraction was carried out following manufacturers instructions of the QIAgen Spin Miniprep kit. Elution was carried out in 50 µl of sterile distilled water.

Protein Over-expression

A large cell culture consisting of 5 ml relevant overnight culture in 500 ml LB media was grown at 37 °C until OD_{650} of approximately 0.5. Upon reaching the desired optical density (OD), the cell cultures were induced to a final concentration of 1 mM IPTG and grown for a minimum of a further 3 hours. Prior to centrifugation at 10,000 rpm, a 1 ml sample was taken for gel examination to ensure that over-expression was successful. The pelleted cells were stored at -20 °C until required for protein extraction.

Cell lysis and protein extraction

The relevant stored pelleted cells were initially re suspended in 20 ml lysis buffer along with 250 µl 50 mg/ml lysozyme and incubated on ice for 10-15 minutes. The re suspended cells were then subjected to four separate cycles of sonication, each for one minute. To complete cell lysis, the solution was centrifuged for 25 minutes at 18,000 rpm, after which the cellular supernatant was separated from the pelleted inclusion bodies.

The pelleted inclusion bodies, along with remaining cell residue, was subjected to two further rounds of homogenisation in lysis buffer, incubation, sonication and centrifugation as previously outlined. After each centrifugation step, only the pelleted inclusion bodies were carried forward through the extraction process. To complete insoluble protein extraction, the pelleted inclusion bodies were re suspended in 15 ml of guanidine buffer and sealed in a sufficiently sized membrane dialysis bag and left to dialyse in buffer A overnight.

Purification steps

Nickel Affinity column

Initial purification was facilitated through the use of the histidine tag present on all proteins of interest over-expressed using the pET28a vector. Upon establishing the location, with regards to its solubility, the protein was extracted from the relevant cellular fraction i.e. the supernatant or inclusion bodies, as outlined in the materials and methods section. For each of the proteins, the appropriate cell fraction was applied to a 1 ml GE Healthcare Ni-NTA column at a rate of 1 ml per minute. The high affinity of the Ni-NTA resin for histidine residues underlines the specificity of this purification technique. Prior to the application of the sample, the entire purification system was primed with Buffer A; elution of bound protein from the column was brought about through increasing levels of imidazole through increasing volumes of buffer B. Gradient or step wise elution could have been carried out depending on preference.

MonoQ and Heparin columns

Where necessary, further protein purification was carried out through the use of either a GE healthcare MonoQ or Heparin purification column. Although both rely on the principle of ion exchange, they are antagonistic in the characteristics of proteins they purify. Prior to applying the protein sample to the column, it was vitally important to appropriately dilute the salt concentration within the sample in order to prevent the protein falling out of solution. Dilution utilised appropriate volumes of no salt buffer whilst initial injection of the protein was carried out into a system filled with the low salt buffer. Elution of bound protein was brought about through an increase in salt concentration as a result of subjecting the system to increased volumes of high salt buffer.

Given that the pH of buffers used in these systems was 7.5, any protein carrying a pI above that would require the use of a Heparin column, whilst proteins possessing a pI lower than 7.5 would carry a negative charge and therefore would be purified using a monoQ purification column.

Concentrating purified protein samples

Pooled protein fractions were applied to a 20 ml Vivaspin concentrator with a minimum molecular weight limit of 5,000 kDa. The sample was then subjected to several rounds of centrifugation at 4500 rpm for 15 minutes. If necessary, further concentration can be carried out in 0.5 ml vivaspin concentrator with centrifugation again at 4500 rpm.

Bradford Assays

In to every assay sample, 1 ml of a 1 in 5 dilution of Bio Rad concentrate reagent with distilled water was added. A series of standards using Bovine Serum Albumin (BSA) were carried out, ranging from the addition of 0-8 µl of a 1 mg/ml BSA dilution. Optical Density readings at A595 were obtained in order to obtain a standard curve of OD with a known increase in protein concentrations. Testing the concentrations of unknown protein samples was carried out in a very similar fashion, a given volume of protein was added to a 1 ml sample of reagent solution and the OD595 was taken. As a general rule, two or three different protein volumes were tested for every sample, this allows an average of the results to be taken. Following the ODs of all the samples being taken, calculations based upon the standard curve elucidated the sample protein concentrations.

Band-shift reactions

A 15 µl reaction volume consisting of 3 µl 50% Glycerol, 1.5 µl 10 x Transcription Buffer and 1 µl appropriate PCR DNA sequence was created. The remaining reaction volume was comprised of varying ratios of protein and distilled water, giving a range of protein concentrations. Where phosphorylation of the protein was required, a relevant volume of phosophate donor, acetyl phosphate was added to the overall band-shft reaction volumes in place of a volume of distilled water. Reactions were incubated at 37 °C for 20 minutes prior to being run on a 7.5 % Acrylamide gel at 100 volts in 0.5 x TBE and staining in ethidium bromide.

Glutaraldehyde crosslinking

An initial series of 15 µl reactions consisting of a protein amount equivalent to 10 µg, Hepes buffer to a final concentration of 50 mM and DTT to a final concentration of 0.1 mM. The remaining reaction volume was occupied by varying volumes of a 0.2 % Glutaraldehyde solution appropriate to the individual reaction.

Each reaction volume was left to incubate at room temperature for 20 minutes before 2 µl of ethanolamine and 200 µl of acetone was added to end the crosslinking reaction. Following this, each reaction was stored at -70 °C for 30 minutes before being centrifuged at 13,000 rpm for 15 minutes. The pelleted protein was re-suspended in 10 µl SDS- Page loading dye and run on Invitrogen pre cast Bis-tris gels in 1 X MES running buffer at 150 volts.

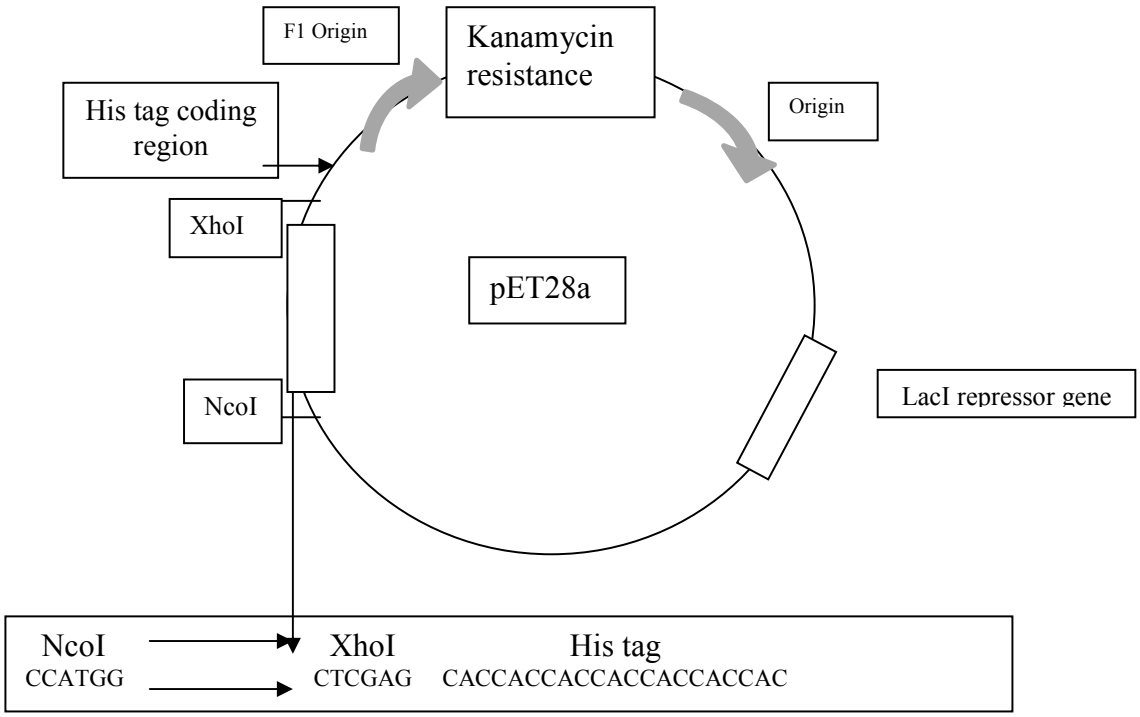


Figure 2: Plasmid map for pET28a, the plasmid responsible for protein over-expression through IPTG mediated induction of the LacI repressor. Each protein sequence was ligated to the vector using in frame XhoI and NcoI restriction sites. Providing the protein sequences were inserted in frame at the outlined region, the coding of a Histidine tag juxtaposing the protein sequence was present. The presence of the Histidine tag facilitated the primary nickel affinity purification step.

3 Results

3.1 Cloning

The individual gene fragments were PCR amplified from genomic Group B *Streptococcus* DNA using the appropriate combination of XhoI and NcoI primers. Following PCR purification, the fragments, along with a mini prep stock of pET28a, were digested using a combination of XhoI and NcoI restriction enzymes. The individual digested gene fragments were ligated into the identically digested pET28a vector and then transformed into RLG 221 competent cells prior to being sequenced.

Alterations in the protein sequence could potentially have a significant effect to the proteins' function. Given that downstream results rely on mimicking the functionality of the protein within the host cell, it is vitally important to verify that the correct protein sequence was amplified and transformed. Once satisfied of the above criteria, the expression plasmids complete with the protein sequences were transformed to T7 Express competent cells.

3.2 Over-expression test cultures

In order to establish that over-expression within the constructs was successful, small test cultures were created for all the designated genes. The purpose of these test cultures was to not only indicate successful over-expression, but also to give an indication whether the protein was soluble and therefore present in the cellular supernatant, or insoluble and therefore likely to be localised in inclusion bodies. 10 ml cultures induced in an identical manner to the large cultures, outlined in the materials and methods, were grown for each protein for a variety of times and under different growth conditions. Samples from before and after IPTG induction, along with a sample from the cell supernatant and the inclusion bodies, were run on a SDS-PAGE gel to confirm both over-expression and the solubility of the over-expressed protein sample. Unlike the extraction process outlined for the larger over-expression cultures, cell lysis of the test cultures did not require the use of lysozyme or the sodium deoxycholate lysis buffer. Instead, re-suspension was carried out in buffer A and only one round of sonication and centrifugation was used before the samples were applied to the gel.

The results of these test cultures are indicated in the table below;

Protein	**Approximate protein size (kDa)**	**Proposed over-expression conditions based on results of the test cultures**
covR	25.8	3 hour induction at 37 °C (overnight induction possible)
rpoD	42.1	3 hour induction at 37 °C (overnight induction possible)
rogB	60	Initial cell growth at either 30 °C or 37 °C prior to overnight induction at room temperature
rovS	33.1	Overnight induction at room temperature
rgfA	29.7	Overnight induction at 37 °C
1031	16	Overnight induction at 37 °C
1617	13.5	Overnight induction at room temperature
2046	15	Overnight induction at room temperature
0090	19.5	Early indications suggested overnight induction at room temperature however over-expression was not as successful as other proteins so further optimisation may be required

Table 5: Over-expression conditions for each of the outlined proteins, as indicated by the test cultures conducted previously

In order to extract the levels of protein required for downstream experiments, one litre cell cultures were grown. These cultures were then subjected to cell lysis and protein extraction prior to the purification process. The following figures indicate the purification process carried out for the indicated proteins. After each sample had been applied to any purification column, a selection of fractions identified as potentially protein positive were run on a gel for not only, qualitative confirmation of the proteins' presence but also to give an indication as to the concentration present in the fraction.

3.3 covR purification procedure

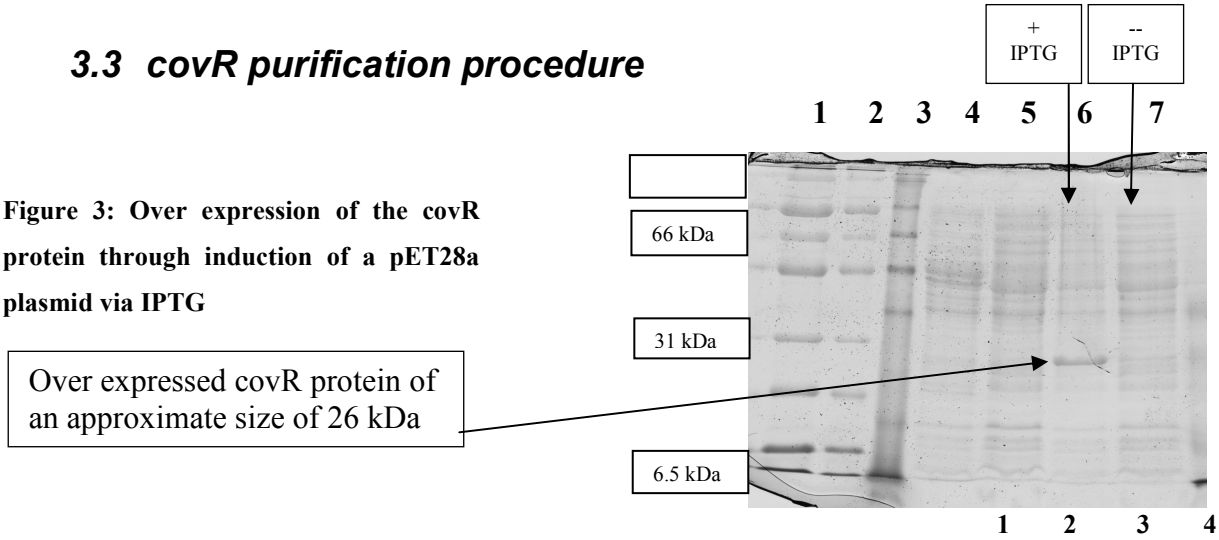


Figure 3: Over expression of the covR protein through induction of a pET28a plasmid via IPTG

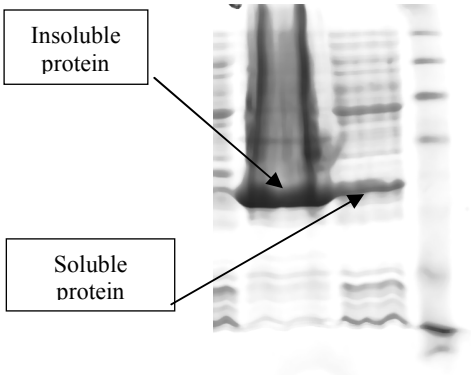


Figure 4: Analysis of the protein sample to determine its solubility and location within the cell

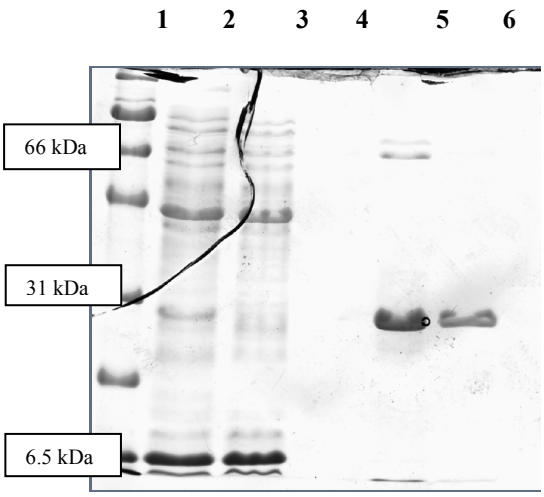


Figure 5: Application of soluble covR on a Nickel Affinity column following cell lysis and protein extraction.
Elution of bound protein required an imidazole concentration of approximately 500 mM
Lane 1 = Broad Range protein marker, Lane 2 = Pre loaded sample, Lane 3 = Flow through, Lanes 4 + = Fractions obtained following nickel affinity mediated purification.

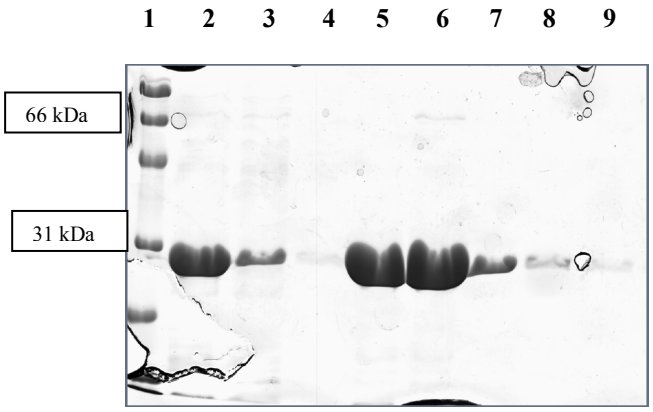


Figure 6: Application of the insoluble covR protein sample on a Nickel Affinity column following cell lysis, protein extraction and dialysis.
Complete elution of bound protein required an imidazole concentration of above 250 mM imidazole
(Left to Right) Lane 1 = Broad Range protein marker, Lane 2 = Pre loaded sample, Lane 3 = Flow through, Lanes 4 + = Fractions obtained following nickel affinity mediated purification.

The lack of complete purity following nickel affinity purification of both the insoluble and soluble protein samples requires further purification using a monoQ purification column.

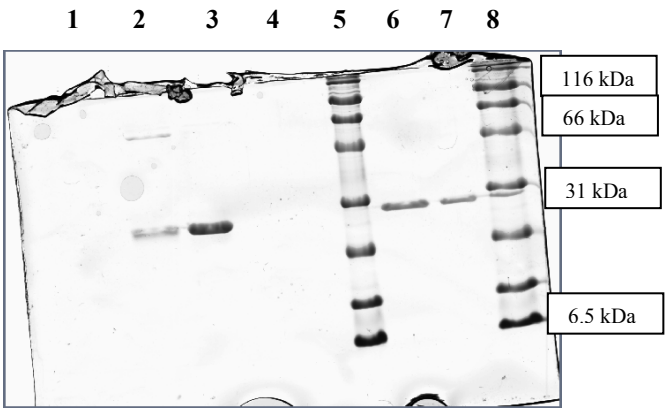


Figure 7: Application of the nickel column fractions (positive for the presence of covR) to further purification via a monoQ column. Elution of bound protein began at a salt concentration of 250 mM and above.
(Left to Right) Lanes 1-4 = Fractions obtained following purification Lanes 5 and 8 = Broad Range protein marker Lane 6 = Pre loaded sample Lane 7 = Flow through

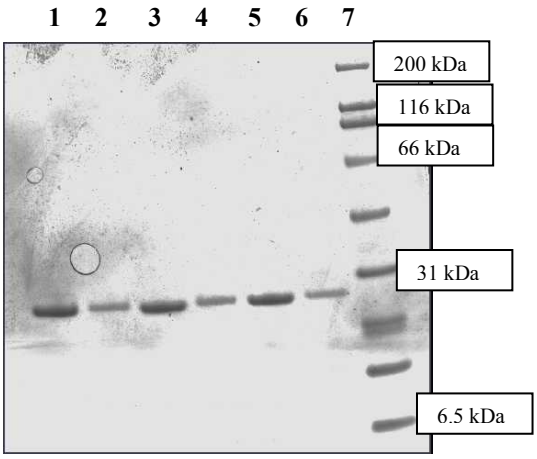
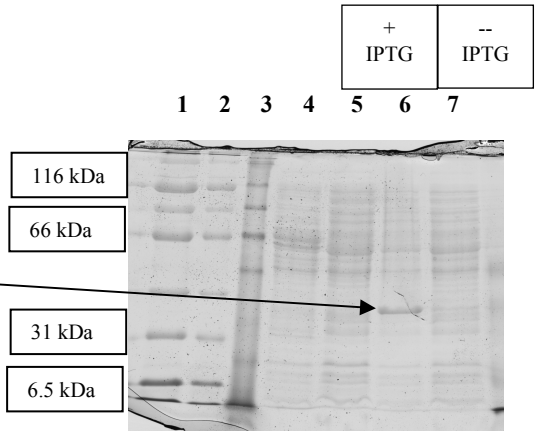


Figure 8: Following complete purification, all relevant protein fractions were pooled and concentrated. The following image indicates the purity of the final protein samples
(Left to Right) Lanes 1 and 3 = 5 mg dialysed covR protein Lanes 2 and 4 = 1 mg dialysed covR protein Lane 5 = 5 mg covR soluble protein Lane 6 = 1 mg covR soluble protein Lane 7 = Broad Range protein marker

3.4 rpoD purification procedure

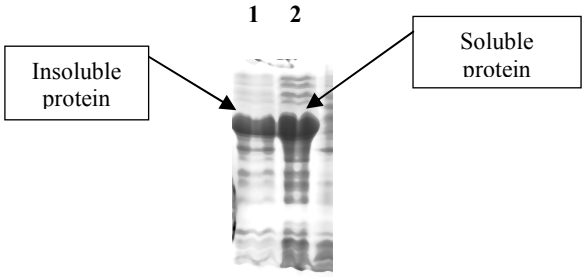
Over-expression of the rpoD protein through induction of the pET28-a plasmid via IPTG

Over expressed rpoD protein with an approximate size of 42 kDa



⇩

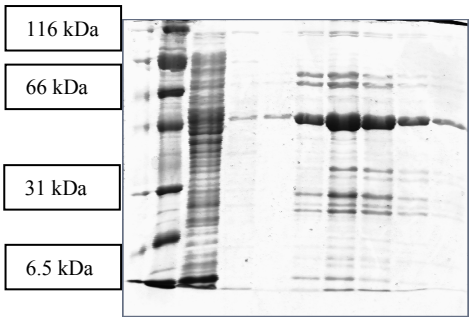
Analysis of the protein sample to determine its solubility and location within the cell



⇩

Application of protein sample on a Nickel Affinity column. Elution of bound protein required an imidazole concentration of above 250 mM

Lane 1 = Broad Range protein marker, Lane 2 = Pre loaded sample, Lane 3 = Flow through, Lanes 4 + = Fractions obtained following nickel affinity mediated purification



⇩

Application of the Nickel column fractions positive for covR to further purification via a MonoQ Column. Elution of bound protein began at a salt concentration of 150 mM and above

Lane 1 = Broad Range protein marker, Lane 2 = Pre loaded sample, Lane 3 = Flow through, Lanes 4 + = Fractions obtained following purification.

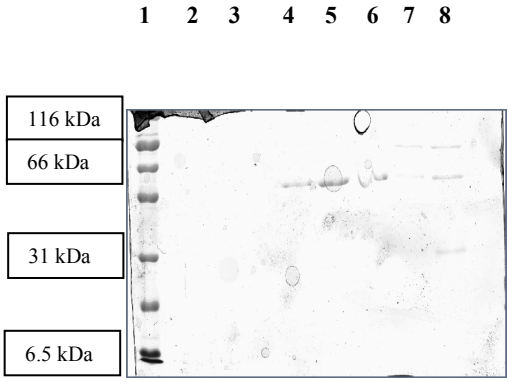
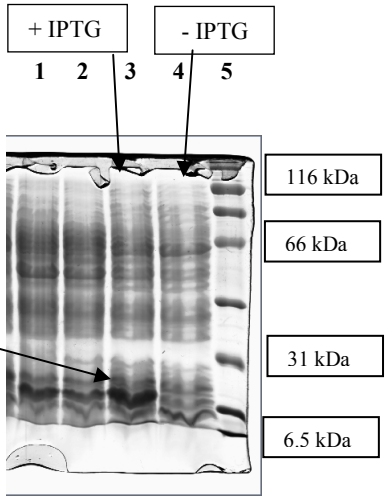


Figure 9: Flow chart outlining the several purification steps involved in purifying the rpoD protein

3.5 1617 protein purification

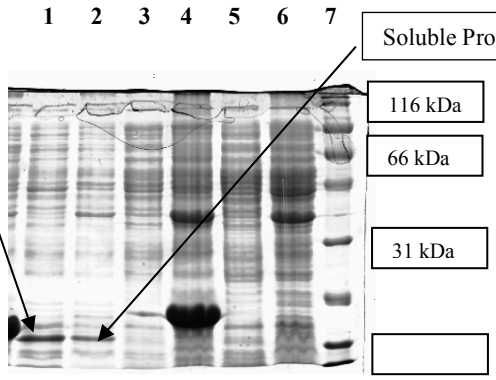
Over-expression of the 1617 protein through induction of the pET28-a plasmid via IPTG

Over expressed 1617 protein of an approximate size of 13.5 kDa



Insoluble protein

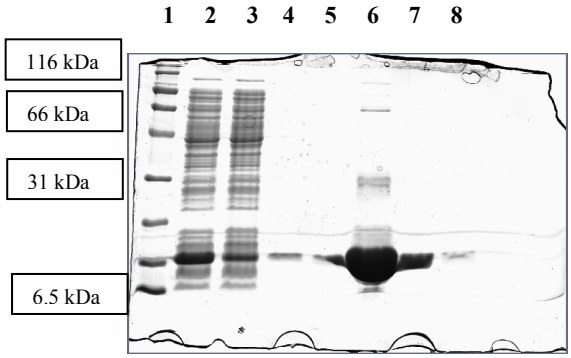
Soluble Protein



Analysis of the protein sample to determine its solubility and location within the cell

Application of the protein on a Nickel Affinity column. Elution of bound protein required an imidazole concentration of approximately 500 mM.

Lane 1 = Broad Range protein marker, Lane 2 = Pre loaded sample, Lane 3 = Flow through, Lanes 4 + = Purification fractions



Application of the Nickel column fractions positive for 1617 to further purification via a Heparin Column. Elution of bound protein began at a salt concentration of 500 mM

Lane 1 = Broad Range marker, Lane 2 = Pre loaded sample, Lane 3 = Flow through, Lanes 4 + = Purification fractions

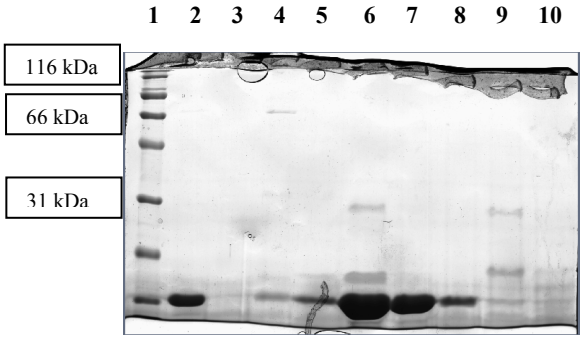


Figure 10: Flow chart outlining the purification steps involved in purifying protein 1617

3.6 Protein Concentration

For the majority of experiments involving purified protein, it was fundamental that the concentrations of the proteins were established. Establishing the concentrations relied on the use of the spectrometry based technique known as a Bradford assay, the results of which are shown below;

Protein	Stock protein (Average mg/ml)	Average molarity (µM)
covR soluble	7.53	292
covR GuHCl	4.89	190
covR GuHCl fraction 15	7.95	300
covR GuHCl fraction 18	7.38	286
rpoD soluble	6.95	208
1617 fractions 20-23	13.01	957
1617 fractions 24-26	4.73	348

Table 6: Bradford assay results for the purified protein samples. All figures rounded to three significant figures

3.7 Band shift Analysis

Recent studies into the relationship of β haemolysin/cytolysin, CAMP factor expression and the two component control system involving covR, have identified that the covR protein binds to DNA (Lamy et al 2004 and Jiang et al 2008). In fact, the study identified that the protein binds to DNA sequences of both the *cylX* promoter region and *cfb* gene. These binding regions were identified as being within a 400 base pair stretch, upstream of the translation initiation site (Lamy et al 2004). As previously mentioned these suspected binding regions were initially used to confirm whether the purified covR protein functioned similarly to other purified covR protein samples from other studies and to that in-vivo.

3.7.1 CAMP promoter band-shifts

Initially, a series of band-shifts utilising both the soluble covR protein sample and that produced following dialysis of the inclusion bodies, in combination with DNA sequences upstream of the *cfb* gene, were analysed.

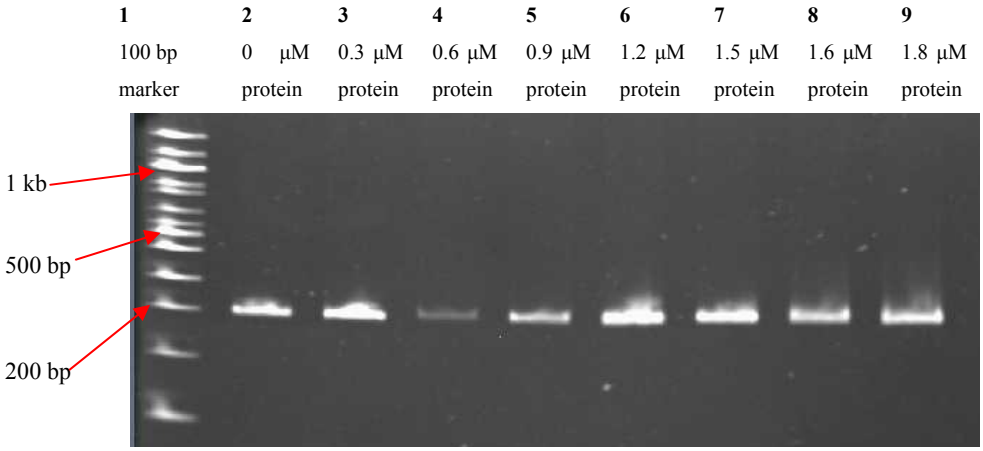


Figure 11: Band-shift results following incubation of the indicated levels of covR protein produced under dialysis with DNA representative of 200 base pairs immediately upstream of *cfb* promoter

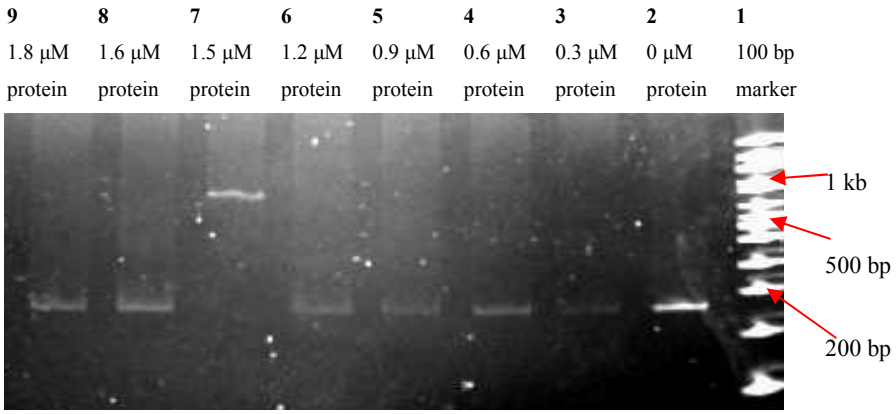
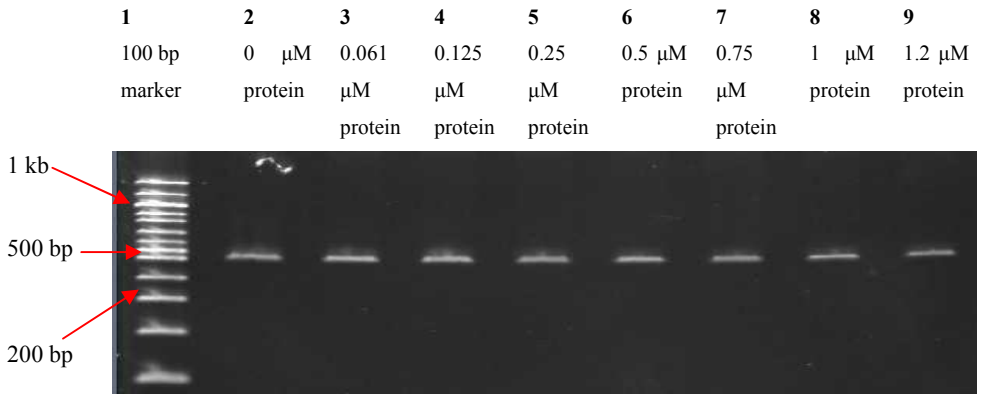


Figure 12: Band-shift results following incubation of the indicated levels of covR protein produced under dialysis with DNA spanning 400- 200 base pairs upstream of the *cfb* promoter

The band-shift results illustrated in figure 12 indicate that the insoluble dialysed protein does indeed bind to DNA within the sequence of promoter fragment 3. The single shift in DNA shown at a concentration of 1.5 µM, indicate that there is a single binding site present within a 200 base pair sequence upstream of the CAMP factor translation initiation site. The major discussion however is the nature of the shift exhibited. Common paradigm indicates that a shift in DNA occurs at a given protein concentration and that any given concentration above that will also exhibit the same shift. It is also commonly documented that band-shifts of this nature portray a gradual appearance, that is to say that signs of a shift are seen at lower protein concentrations than the critical concentration required for complete band-shift. In this example however, neither of these common themes exist. Firstly, the shift appears abrupt and clean, but also at protein concentrations above the 1.5 µM mentioned previously, the DNA appears to return to state more consistent with it being unbound/ 'free' of protein.



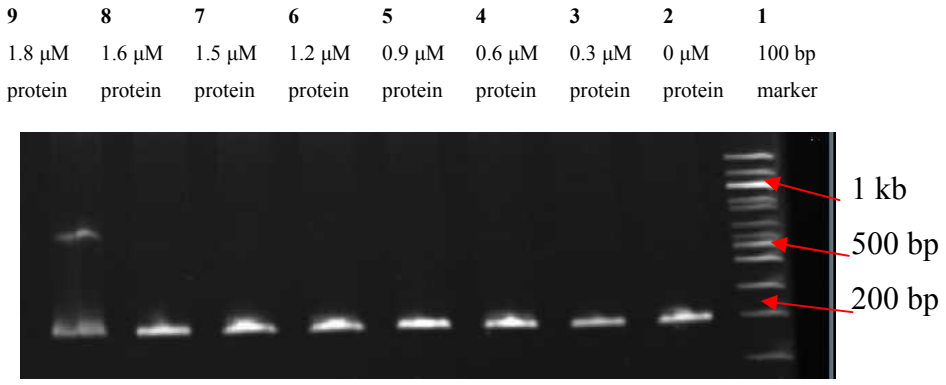


Figure 14: Band-shift results following incubation of covR soluble protein with DNA representative of 200 base pairs immediately upstream of the *cfb* promoter

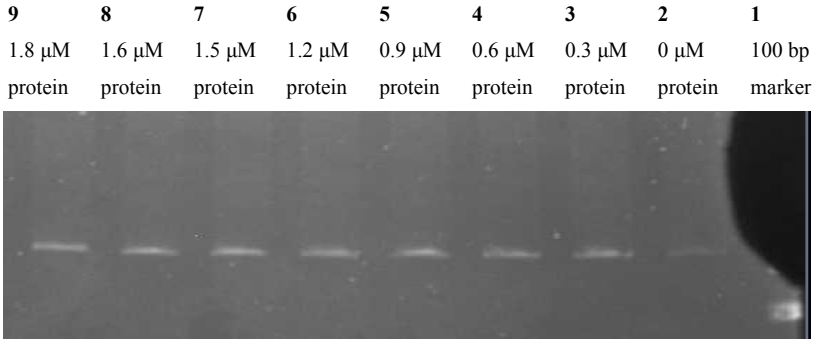


Figure 15: Band-shift results following incubation of covR soluble protein with DNA representative of bases 400-200 upstream of the *cfb* promoter

When incubated with the cellular soluble protein, a difference in DNA shift is exhibited. In contrast to the single shift presented using dialysed protein, the only shift witnessed using the soluble protein occurs in the 200 base pairs immediately upstream of the gene sequence as shown in figure 14. The interesting point to note is the appearance of the single shift upon incubation with promoter sequence 2, but not when the protein was incubated with fragment 1, a sequence that encompasses the DNA of fragment 2 and beyond. Again, in contrast to the band-shift of the insoluble dialysed covR, the soluble mediated shift portrays the hallmarks of a more traditional shift, i.e. the shift is more gradual. However, it cannot be explained whether the shift occurs only at a narrow protein concentration range as shown in the guanidine dialysed protein shift. The concentration of protein required for the soluble shift, 1.8 µM, is

0.3 µM higher than the protein concentration required for a shift in fragment 3 using the protein extracted from the insoluble inclusion bodies.

In this study it has been suggested that there are two binding sites for the transcription factor covR within the 400 base pairs immediately prior to the *cfb* gene. The evidence of these binding sites, through the shifting of DNA upon incubation with protein, is however still debatable. It is well documented that phosphorylation of many proteins not only aids in the binding ability to DNA, but in many cases it is a prerequisite (Jiang et al 2008). In order to test whether such a situation applies to covR, the protein was incubated with a phosphate donor, acetyl phosphate. The band-shift results of this experiment are indicated in figures 16 and 17.

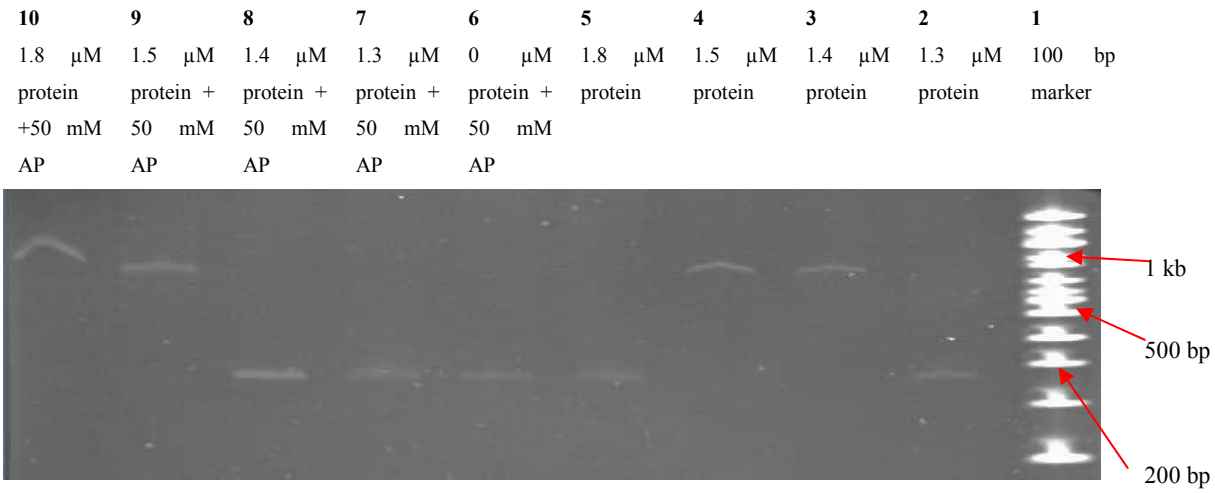


Figure 16: Band-shift results following incubation of covR protein produced under dialysis with DNA representative of DNA spanning 400 – 200 base pairs upstream of the *cfb* promoter and differing levels of Acetyl Phosphate

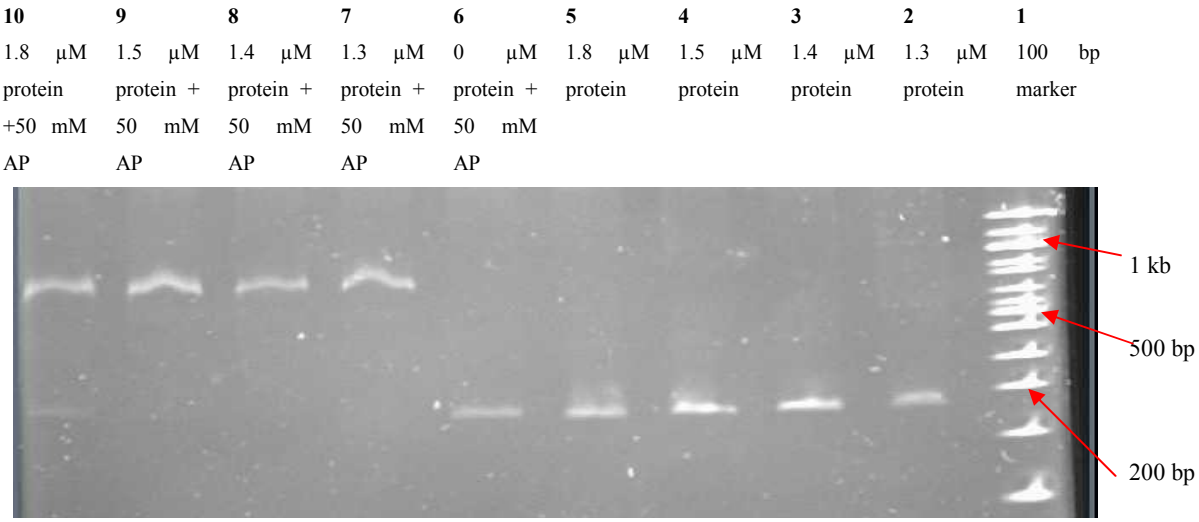


Figure 17: Band-shift results following incubation of covR soluble protein with DNA representative of DNA spanning 400 – 200 base pairs upstream of the *cfb* promoter and differing levels of Acetyl Phosphate

The initial results (lanes 2-5) using the protein extracted from the insoluble inclusion bodies in combination with promoter sequence 3 correlated with previous results shown in figure 12. Upon reaching a concentration of 1.4/1.5 µM, a sudden shift in the DNA was shown. This shift in the DNA then appeared to be lost with a further increase in protein returning to a conformation more consistent with free DNA as indicated by lane 6. With the addition of acetyl phosphate, the same shift in DNA is experienced at a concentration of 1.5 µM. It could be argued that the concentrations required for this shift are slightly raised upon the presence of acetyl phosphate however, until more research has been carried out it is suggested that this result is an anomaly and not a characteristic commonly portrayed. The other major point of interest surrounds the shift shown with a protein concentration of 1.8 µM. When unphosphorylated, no DNA shift at this concentration of insoluble protein extracted from the inclusion bodies has been witnessed. However, when phosphorylated we see a definite shift as shown in figure 16. It is possible that this is further evidence that the presence of phosphate does indeed raise the critical protein concentrations required for DNA binding; a hypothesis that would corroborate the lack of binding at a phosphorylated protein concentration of 1.4 µM. It is difficult to be entirely certain as to the nature of this individual protein shift due to its appearance on the gel. The conformation of the band could indicate that a second shift at a slightly different location is taking place. However, it is also highly conceivable that as the sample was loaded on an outer lane, the gel dynamics have forced the same shift to appear differently than its identical counterparts witnessed in previous experiments.

The results of phosphorylation on the soluble protein sample were a lot more decisive. No shift took place when unphosphorylated soluble protein was incubated with promoter sequence 3, a result replicated in figure 15. However, upon phosphorylation a definite single shift in the DNA occurred that appears to occur at a lower concentration than previous shifts outlined in this study thus far. This result suggests that phosphorylation, although not a necessary requirement, certainly aids in the binding ability of covR to specific DNA fragments.

3.7.2 cylX band-shifts

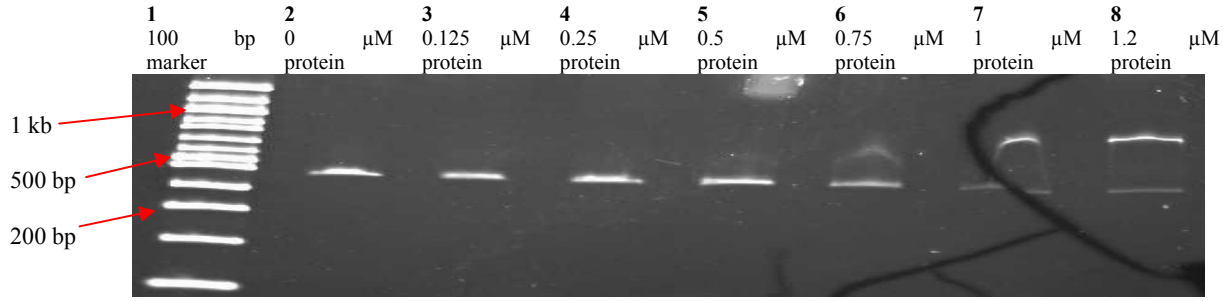


Figure 18: Band-shift results following incubation of covR protein produced through dialysis with DNA representative of 400 base pairs immediately upstream of the *cylX* promoter.

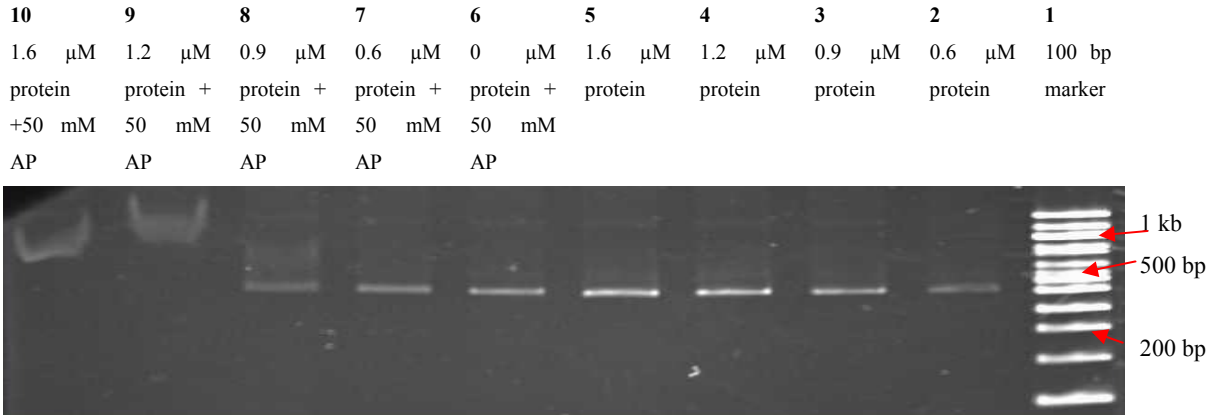


Figure 19: Band-shift results following incubation of covR protein produced under dialysis with DNA representative of 400 base pairs immediately upstream of the *cylX* promoter and differing levels of Acetyl Phosphate

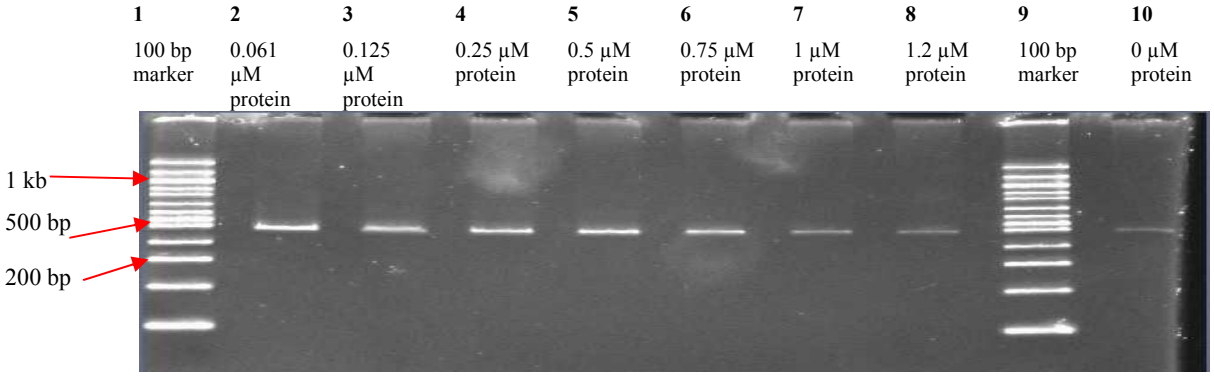


Figure 20: Band-shift results following incubation of covR soluble protein with DNA representative of 400 base pairs immediately upstream of the *cylX* promoter

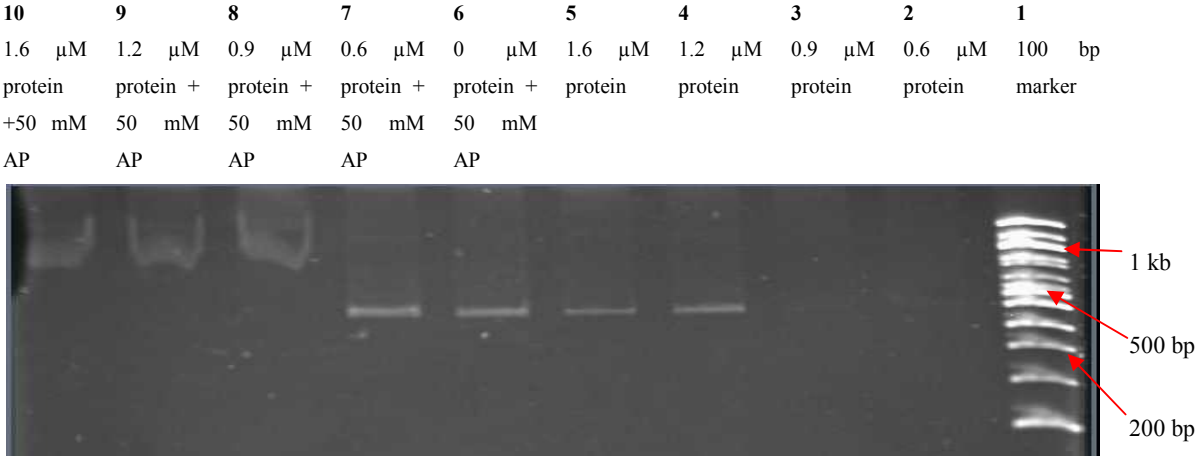


Figure 21: Band-shift results following incubation of covR soluble protein with DNA representative of 400 base pairs immediately upstream of the *cylX* and differing levels of Acetyl Phosphate

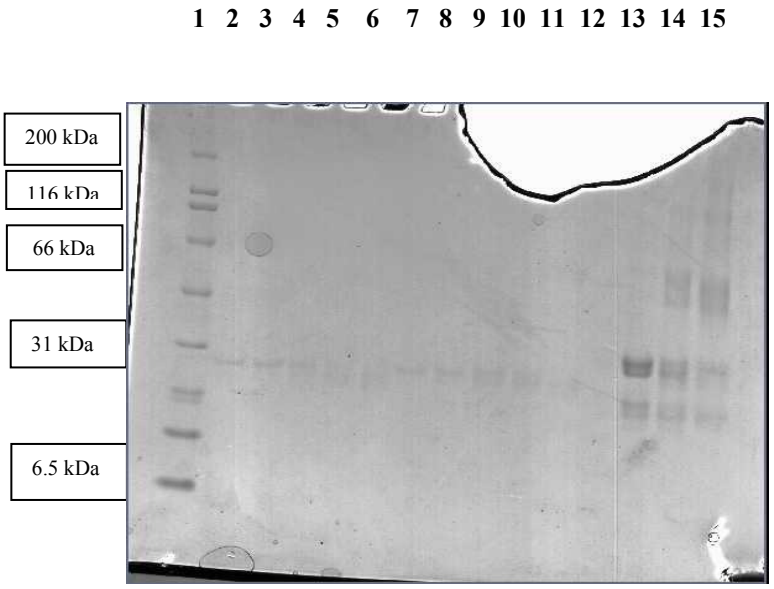
The other potential binding site for the covR protein is the *cylX* promoter region, a DNA sequence associated with the expression and regulation of β hemolysin/cytolysin (Rajagopal 2009). In order to determine whether the protein does in fact bind to potential promoter sequences, the same principle of band-shifting a 400 base pair region prior to the translation initiation site was examined. As shown in figures 18 and 20, soluble protein incubated with fragment 4 indicated no DNA shift however, using the insoluble protein extracted from the inclusion bodies a single shift appeared with a protein concentration of 0.75 µM and above. The shift can be described as a classical shift in that a gradual increase in protein concentration, results in an increased level of shift in the DNA sample. The critical concentration required for the binding of DNA at the *cylX* location is considerably lower than that seen at the CAMP factor location.

Once again the protein samples were subjected to acetyl phosphate incubation to ascertain the effect that phosphorylation may have at the *cylX* location, results shown in figures 19 and 21. With the soluble protein, it is clear that phosphorylation promotes the binding to DNA. It, however, is too early too confirm whether it is a requirement in-vivo. In a identical manner to the soluble protein results, those shown using the protein produced via dialysis promoted the conclusion that phosphorylation does indeed aid in DNA binding at the *cylX* location. The critical concentrations seen using both protein samples were of a similar level, however, the result at a concentration of 0.9 μM with the guanidine protein suggests that this protein sample could have a less efficient response to phosphorylation than the soluble protein. Also, during the same band-shift, the result at a concentration of 1.6 μM is debatable. The slightly lower nature of the band could well be a result of three factors. Simply, as the sample is located on the outer lane of the gel, the running dynamics in this lane could once again be responsible for the drop in DNA complex size and it is in fact an identical shift to that seen previously. Secondly, as witnessed before, it is possible that above a certain concentration the protein has an inhibitory effect on DNA binding. Finally, and potentially most unlikely, a second separate shift aside from the one seen at a concentration of 1.2 μM has taken place.

3.8 Oligomerisation state of the covR protein

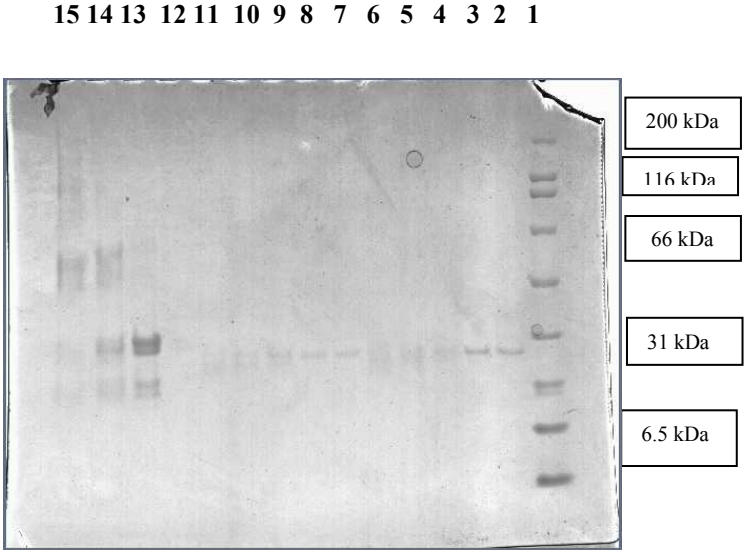
As well understanding the sites at which covR binds within sections of the Group B *Streptococcus* genome, it is also useful to ascertain the potential conformation a protein possesses in-vivo. A number of glutaraldehyde dilutions were carried out for both the soluble protein sample and that produced following dialysis, to examine if the protein exists as a monomer or does in fact dimerise under the correct conditions. As shown in figures 22 and 23 there is no clear evidence to suggest that, in the absence of a phosphorylating agent, covR exists in any other conformation than a monomer. The levels of glutaraldehyde reached in the more concentrated reaction volumes is believed to be more than sufficient to facilitate cross linking of protein dimers if indeed they exist. The presence of *yfH* dimers, in the absence of acetyl phosphate, ensures that the experimental procedure functioned correctly and also corroborates the belief that glutaraldehyde concentration was not a limiting factor in detecting covR dimerisation.

As previously illustrated during the series of band-shifts, phosphorylation of the protein appears to have a positive effect on its functionality. To test whether this increased functionality was due to dimerisation mediated by phosphorylation, each of the protein samples were incubated in acetyl phosphate prior to glutaraldehyde fixation, results shown in figure 19. Once again, the presence of *yfH* dimers following phosphorylation acted as a control. With regards to the soluble covR protein sample and that produced following dialysis, again there was no evidence to suggest that phosphorylation facilitates covR dimerisation or that covR dimerisation even takes place in vitro.



Lane	Protein	Volume of 0.2% Glutaraldehyde (µl)
2	covR soluble	0
3	covR soluble	2
4	covR soluble	4
5	covR soluble	8
6	covR soluble	16
7	covR soluble + AP	0
8	covR soluble + AP	2
9	covR soluble + AP	4
10	covR soluble + AP	8
11	covR soluble + AP	16
13	yfH	0
14	yfH	2
15	yfH	8

Figure 22: Results of glutaraldehyde crosslinking experiments for covR protein produced via dialysis. The contents of each lane (right to left) are described in the legend opposite the gel using a broad range protein marker in lane 1. The use of the yfH protein acts as a control ensuring that the crosslinking process was efficient and that the introduction of acetyl phosphate has no untoward effects on protein binding.



Lane	Protein	Volume of 0.2% Glutaraldehyde (µl)
2	covR GuHCl	0
3	covR GuHCl	2
4	covR GuHCl	4
5	covR GuHCl	8
6	covR GuHCl	16
7	covR GuHCl + AP	0
8	covR GuHCl + AP	2
9	covR GuHCl + AP	4
10	covR GuHCl + AP	8
11	covR GuHCl + AP	16
13	yfH + AP	0
14	yfH +AP	2
15	yfH + AP	8

Figure 23: Results of glutaraldehyde crosslinking experiments for soluble covR protein. The contents of each lane (left to right) are described in the legend opposite the gel using a broad range protein marker in lane 1. The use of the yfH protein acts as a control ensuring that the crosslinking process was efficient.

4 Discussion

4.1 Identification of covR binding sites at the cylX and cfb loci

It has been presented in this study that covR has the ability to directly bind to specific fragments of DNA. Particular DNA fragments, the *cylX* promoter region and sequences upstream of the *cfb* gene, coincide with regions that are believed to encompass gene promoters (Lamy et al 2004 and Jiang et al 2008).

With regards to the *cfb* gene, responsible for the CAMP factor, two regions of upstream DNA appear compatible for direct covR binding. Although the exact sequence or location of the binding regions cannot be defined, it is possible to confirm that one region is located in the 200 base pairs immediately upstream of the start codon, whilst the other is in a region between 200 and 400 base pairs upstream of the translation initiation site.

In terms of the *cylX* promoter region, the band-shift analysis carried out in this study indicated a single covR binding site. Due to temporal constraints, the band-shift analysis attributed to this gene location was less than extensive and therefore, it is impossible to narrow down the location of this binding site beyond it being located in the first 400 base pairs upstream of the genes translation initiation site. This result corroborates the findings of *Lamy et al*, where similar EMSA analysis demonstrated a single covR binding region within this promoter region. *Lamy et al*, however used DNase I foot printing assays to further define the location of the binding site. A 30 base pair region between base pairs -281 to -252 relative to the translation initiation start site was specifically protected by the protein. The high affinity of this region for covR binding suggests that this region, or a region encompassed by these 30 base pairs, is the region in which the protein binds (Lamy et al 2004). Only through further investigation and the use of narrower DNA fragments could the band-shift analysis of this study confirm *Lamy et al*'s findings, it is a firm recommendation that this should be carried out. In addition to these results, as part of the *Lamy et al* study, further DNase I foot- printing on a number of other Group B *Streptococcus* genes showed a potential binding motif for the covR protein. The sequence of this binding motif has been published as 5'-TATTTTAAT-3' and through bioinformatics it can be seen that this motif is conserved within the *cylX* promoter region. The location of this motif within the upstream *cylX* sequence is also consistent with the location of the potential covR binding site outlined, not only in *Lamy et al*, but also confirmed in this study (Lamy et al 2004).

4.2 Effect of protein phosphorylation on DNA binding

As outlined in the introduction, two component regulatory systems, such as covR and its partner sensor protein covS, rely heavily on a change of phosphorylation state to mediate a response to environmental stimuli (Jiang et al 2005). In general, it is widely accepted that phosphorylation of the regulator protein has a direct effect on its ability as a transcriptional regulator (Jiang et al 2008). Given this fact, it is vital to understand the effects of phosphorylation on covR binding at the *cylX* and *cfb* loci. In order to mimic the kinase activity of covS in vitro, the covR protein was incubated in acetyl phosphate, a proven phosphate donor.

When phosphorylated, the affinity for which the protein, from both the soluble and insoluble dialysed samples, binds to DNA appears to increase. At both the *cylX* and *cfb* loci, phosphorylation of the soluble protein sample leads to successful binding of the relevant DNA fragments, an attribute not seen previously prior to phosphorylation. With regards to the sample produced following dialysis, the positive effects of phosphorylation on the proteins' binding ability was not to the same degree as that witnessed with the soluble fraction. At both loci the shift in DNA exhibited with and without phosphorylation were approximately the same. The increase in binding affinity through phosphorylation witnessed in this study corroborates that seen by *Jiang et al*. During the Jiang study, a protein concentration of 0.75 μM was required to shift DNA at the *cylX* locus and a concentration of approximately 1.5 μM was required to shift DNA at the *cfb* locus (Jiang et al 2008). These figures approximately match those required to shift the relevant DNA in this study. Aside from the concentration of acetyl phosphate subjected to the protein sample, it is reasonable to suggest that the conditions created in the band-shifts of both studies are similar.

4.3 Differences between the soluble covR protein sample and that produced via Dialysis

The presence of covR protein, not only in the cellular supernatant but also in the inclusion bodies, allowed this study to purify two separate samples of the protein. Through the band-shift analysis it appears that the two independent protein samples show different properties. In general, the use of the soluble form of the protein produced more typical band-shift results particularly with regards to the effect of phosphorylation on the protein's affinity to bind DNA. When examining the results using protein from inclusion bodies, although the rough

principle of binding to DNA remains the same, there are several discrepancies in the mode of action not seen when the soluble protein was used.

The first discussion surrounds the binding of DNA at both examined loci at relatively low concentrations without the presence of acetyl phosphate. Results indicated by *Jiang et al* suggested that although binding without the aid of phosphorylation is possible, it generally requires a protein concentration in excess of 6 μM as opposed to concentrations of between 0.75 and 1.5 μM seen in this study (Jiang et al 2008). In combination with the similarity of band-shifts results with and without acetyl phosphate, this issue promotes the conclusion that the protein sample produced following dialysis is in fact, at least partially phosphorylated prior to incubation with acetyl phosphate. There are several explanations for this conclusion; firstly, it is possible that in-vivo some of the protein begins to be phosphorylated causing a decrease in solubility. The decrease in solubility leads to the build up of this phosphorylated protein in the inclusion bodies, a sample which is then extracted and purified via dialysis. The second, potentially more likely explanation, is that through the sheer volume of protein expression, a sample must be stored in inclusion bodies purely on the basis that the supernatant has reached a saturation point with regards to the protein. The protein stored in the inclusion bodies may well then have an increased susceptibility to phosphorylation, resulting in the properties exhibited.

Phosphorylation of the covR protein arises from the kinase activity of its relevant sensor protein and therefore phosphorylation is usually highly regulated, depending on environmental stimuli (Jiang et al 2004). Unfortunately this regulation system is absent in *Escherichia coli* and therefore it is distinctly possible that an unspecific *Escherichia coli* kinase is able to phosphorylate accessible covR protein producing the varied properties of the two independent covR samples. Contrary to these potential explanations however, is the likelihood that phosphorylation of the protein is unable to survive the extensive extraction and purification process utilised in this study. If this proves to be the case then an alternative explanation for why the insoluble dialysed protein samples appears partially natively phosphorylated must be investigated.

The most puzzling aspect of the results involving the insoluble dialysed protein, surrounds this apparent sudden shift in the DNA at protein concentrations above the critical concentration. As previously mentioned, it is generally accepted that upon reaching a concentration high enough to promote a shift in DNA, any further increases in protein

concentration will simply maintain the shift. At various points in this study however, this has not been the case. Indeed, further increases in protein concentration above the critical concentration have triggered the DNA to return to a conformation identical to that of unbound DNA. A simple explanation is that covR binding is highly concentration dependent and therefore, with too high a concentration, an inhibitory effect preventing further binding to DNA is conveyed. It is conceivable that this inhibitory feedback mechanism is a further addition to the complex regulation system imparted by the bacteria responsible for controlling its virulence at different stages of its life cycle. Unfortunately, without definitive evidence this hypothesis is circumstantial. Further experimental repeats using a broader protein concentration range may suggest that this is purely an anomaly rather than an attribute portrayed through other insoluble dialysed protein band-shifts. If however this phenomenon is endemic, it is hoped that investigation into the binding mode of action may elucidate potential advantageous reasons for the described inhibitory effect.

4.4 Structural conformation of covR

Throughout the scientific community, there are many examples of proteins that are required to be dimerised for effective binding to DNA. To date, little is known about the binding of covR to DNA, particularly in terms of the proteins' structural conformation and key binding residues. In order to begin to elucidate such information, this study attempted to use glutaraldehyde fixation as a method of identifying the structural conformation of covR in-vivo. Results presented in this study appear to indicate that both the soluble protein and that produced following dialysis does not naturally form a dimer. It has already been suggested that the binding of covR to DNA increases following protein phosphorylation. Given this fact, it is conceivable that the structural conformation of the protein alters upon phosphorylation, to mediate binding to DNA. Investigation into this hypothesis once again utilised incubation of the protein with acetyl phosphate prior to glutaraldehyde fixation. In opposition to the results of *Jiang et al* (Jiang et al 2008), this study again presents no evidence for protein oligomerisaion even upon protein phosphorylation. These results appear to suggest that covR binding to DNA requires a sole monomer of the protein.

4.5 Future Work

The results presented in this study are arguably promising and corroborate, to a certain extent, results illustrated in previous studies by *Lamy et al* and *Jiang et al*. The discrepancies between results involving the soluble covR protein and that produced following dialysis warrants further investigation. It is important to determine whether the insoluble dialysed protein has in fact, been phosphorylated prior to extraction and purification by a native *Escherichia coli* kinase or, whether the anomalies regarding this protein sample discussed earlier are a result of another unidentified phenomenon. It is also critical to further validate the results published in this report with the necessary experimental repeats. This was unfortunately an element of the investigation lacking due to temporal constraints. Further repeats of the glutaraldehyde fixation experiments, potentially with an even larger excess of glutaraldehyde will conclusively confirm whether or not covR does in fact dimerise, a result that had been previously postulated by *Jiang et al*.

In order to confirm the proposed binding site of covR at the *cylX* locus it seems logical to attempt band-shifts using the 30 base pair fragment identified by *Lamy et al*. It may also be appropriate to carry out similar EMSA analysis using the protein samples purified in this study with the proposed binding motif. Through this, a strong degree of confidence in both the functionality of the purified sample and also the proposed binding motif can be obtained. Given the fact that covR is a transcriptional regulator, it may also be useful to analyse the promoter sites at which RNA polymerase binds and how these sites relate to the covR binding sites previously established. An indication as to how covR affects the binding of RNA polymerase would also confirm its role as a transcriptional repressor at this locus.

With regards to the *cfb* locus, initial work should focus on narrowing the promoter fragments used in the band-shift experiments. Although early bioinformatic analysis of the upstream region failed to yield the presence of similar binding motifs published at the *cylX* locus, further work, in combination with a more refined band-shift analysis, may well be develop understanding of covR mediated regulation at this locus.

Once again due to temporal constraints, not all of the proteins outlined in this study were over-expressed or purified. For those that have over-expressed an appropriate purification process would need to be carried out. In the case of the remaining transcriptional regulators it would be logical to undertake a similar band-shift analysis to that carried out for covR, on DNA fragments appropriate for their target genes. For those proteins that had failed to over-

express previously, firstly an increased number of test cultures, altering induction temperatures, concentrations of IPTG etc could be analysed. A more complex method of over-expression that could be implemented, is that based on diauxie, a process where by a cell culture metabolises two potential sugar sources sequentially as opposed to simultaneously. Growth in both glucose and lactose media would allow the initial growth to take place preferentially with the aid of glucose sugar until the bacteria is required to switch to the remaining lactose. At this point, the presence of lactose will activate the pET28a plasmid, due to the lac repressor gene, consequently leading to over-expression of the coded protein.

As previously mentioned rpoD functions as a Group B *Streptococcus* sigma factor, therefore prior to any downstream experiments it is crucial to test that it functions correctly with respect to RNA polymerase. Similarly, 1617, being an spx, it would be interesting to investigate how it diverts RNA polymerase from promoters in Group B *Streptococcus* and to which promoters in particular it preferentially functions at. This initial analysis may give an indication as to why Group B *Streptococcus* possesses three spx proteins and if at all, these spx's play a role in the high degree of virulence shown by these bacteria.

5 References

- **Balleza E, López-Bojorquez LN, Martínez-Antonio A, Resendis-Antonio O, Lozada-Chávez I, Balderas-Martínez YI, Encarnación S, Collado-Vides J (2009)** Regulation by transcription factors in bacteria: beyond description. FEMS Microbiol Rev 33. 1. 133-151

- **Browning. DF, Busby. SJ (2004)** The regulation of bacterial transcription initiation Nature Reviews Microbiology 2.1 57-65

- **Cumley. N. J, Smith. L. M, Anthony. M and May. R. C (2012)** The CovS/CovR acid response regulator is required for intracellular survival of Group B *Streptococcus* in macrophages Infection and Immunity doi:10.1128/IAI.05443-11

- **Dmitriev. A, Mohapatra. S. S, Chong. P, Neely. M, Biswas.S, Biswas. I (2011)** CovR-controlled global regulation of gene expression in *Streptococcus* mutans PLoS ONE 6(5): e20127. doi:10.1371/journal.pone.0020127

- **Jiang. S, Cieslewicz. M. J, Kasper. D. L and Wessels. M. R (2004)** Regulation of virulence by a two-component system in Group B *Streptococcus* J. Bacteriol 187.3.1105-1113

- **Jiang. S, Ishmael N, Dunning Hotopp J, Puliti M, Tissi L, Kumar N, Cieslewicz MJ, Tettelin H, Wessels MR (2008)** Variation in the Group B *Streptococcus* csrRS regulon and effects on pathogenicity J. Bacteriol 190(6):1956

- **Lamy. M, Zouine M, Fert J, Vergassola M, Couve E, Pellegrini E, Glaser P, Kunst F, Msadek T, Trieu-Cuot P, Poyart C (2004)** CovS/CovR of Group B *Streptococcus*: a two-component global regulatory system involved in virulence 54.5 1250-1268

- **Lin. W, Walthers. D, Connelly. J. E, Burnside. K, Jewell. K. A, Kenney. L. J, Rajagopal. L (2009)** Threonine phosphorylation prevents promoter DNA binding of the Group B *Streptococcus* response regulator CovR Mol Micro 71(6), 1477–1495

- **Liu. G. Y and Nizet. V (2004)** Extracellular virulence factors of Group B Streptococci Frontier of Biosciences 1.9 1794-802

- **Liu. G.Y and Nizet. V (2006)** The Group B Streptococcal β hemolysin/cytolysin The Comprehensive Sourcebook of Bacterial Protein Toxins Chapter 43 735-745

- **Maisey. H. C, Doran. K. S and Nizet. V (2008)** Recent advances in understanding the molecular basis of Group B *Streptococcus* virulence Expert Rev Mol Med 10. e27

- **Rajagopal. L (2009)** Understanding the regulation of Group B Streptococcal virulence factors Future Microbiol 4.2 201-221

- **Taminato M , Fram D, Torloni MR, Belasco AG, Saconato H, Barbosa DA (2011)** Screening for Group B *Streptococcus* in pregnant women: a systematic review and meta-analysis Revista Latino-Americana de Enfermagem 19.6

- **Zuber. P (2004)** Spx-RNA Polymerase interaction and global transcriptional control during oxidative stress J Bacteriol April 2004 1911–1918

- *Group B Streptococcus sequenced gene accessed via* http://genolist.pasteur.fr/SagaList/

- *Initiative for vaccine research, World Health Organisation – Group B Streptococcus*

- *Columns and supporting information obtained from GE Healthcare*

- *pET28a vector supporting information including plasmid map obtained from Novagen*