H 24/3506

MONASH UNIVERSITY
THESIS ACCEPTED IN SATISFACTION OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
ON..... 6 December 2002 .....
..... [redacted] .....
Sec. Research Graduate School Committee

ERRATA

throughout thesis: "*Sau*3A 1" for "*Sau*3AI".

throughout thesis: "gentamicin" for "gentamycin".

ADDENDUM

p9 line 20: delete " (which are used as precursors for folic acid)".

p13 para 2, 3rd line: delete "distal" and read "distant".

p13 para 2, 4th line: delete "proximal" and read "proximate".

p46 para 3: Comment: For each RNA sample analysed, the relative level of 16S rRNA was measured. On the basis that the level of 16S rRNA is constant, regardless of environmental conditions, these values were used to normalise expression of other genes.

p130 line 5: delete "The ends were filled in" and read "Protruding nucleotides from the 3' termini were removed".

p185 line 22 and p189 Figure 5.1 legend line 10: delete "LWLNEGX" and read "LWLNEG".

throughout thesis: the term "homology" is meant to refer to "sequence identity".

Identification and Characterisation of Regulated Genes of *Corynebacterium pseudotuberculosis*

A thesis submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

by

Sandra Cameron McKean, BSc (Hons).

July, 2002

CSIRO, Livestock Industries

and

Department of Microbiology, Monash University

Table of Contents

Table of Figures

Table of Tables

Table of Tables

Abstract

Corynebacterium pseudotuberculosis is the etiological agent of two specific diseases: ulcerative lymphangitis and caseous lymphadenitis (CLA). CLA is a disease of sheep and goats that is characterised by caseous abscesses in the lymph nodes. Although the steps leading to the establishment of CLA are relatively well understood, little is known about how this is controlled at the molecular level. This in part stems from the lack of sequence information available for this pathogen and a relative paucity of molecular studies.

An empirical observation that has been made during the study of virulence gene expression in a wide number of bacterial pathogens is that genes that are intimately involved in the virulent phenotype are often up-regulated in the *in vivo* environment in comparison to when the organism is grown *in vitro*. This observation has led to the development of technologies aimed at identifying genes of bacterial pathogens that are up-regulated *in vivo* or in models that mimic aspects of *in vivo* infection.

The work described in this thesis aimed to increase our understanding of the molecular mechanisms of *C. pseudotuberculosis* pathogenesis by identifying regulated genes of *C. pseudotuberculosis*. Two independent approaches were developed. The first utilised a promoter probe vector containing green fluorescent protein as reporter in conjunction with Differential Fluorescence Induction (DFI) to identify up-regulated promoters. The second involved transcriptional profiling using a *C. pseudotuberculosis* DNA macroarray to identify both up and down regulated genes.

A *C. pseudotuberculosis* DNA array was constructed and used to identify a number of *C. pseudotuberculosis* genes whose expression changed upon heat shock. Most up-regulated genes coded for proteins belonging to the heat shock protein family. A number of down regulated genes were also identified including *fas*, encoding a fatty acid synthase, an iron permease component (*fagC*) and phospholipase D (*pld*).

Macrophage induced genes were identified using both the array and DFI technologies. Using the DFI approach the genes encoding a non-ribosomal peptide synthase and the beta chain of propionyl coA carboxylase were identified as up-regulated. Using the array approach the heat shock genes of *C. pseudotuberculosis* were identified as the most highly up-regulated genes following macrophage infection. A number of other regulated genes

were also identified, including those encoding a metallopeptidase, iron homeostasis proteins and a putative acetyl coA carboxylase subunit.

To ascertain the functional significance of regulated genes during macrophage infection three mutant strains were generated. Genes coding for a metallopeptidase, a Cu,Zn superoxide dismutase and FagC were targeted. The first two mutants were attenuated to varying degrees within the macrophage infection model, while the iron permease mutant demonstrated wildtype characteristics. An attempt to complement the metallopeptidase mutant was unsuccessful. At this stage the reason for this is not clear.

Pld is the most important virulence determinant of *C. pseudotuberculosis* that has been identified thus far. The observation that *pld* was down regulated following heat shock led to further studies of its regulation. *pld* expression was also found to be cell density dependent. However, expression of *pld* during macrophage infection was high and did not appear to be under density dependent control. While expression of *pld* by *C. pseudotuberculosis* had no effect on their ability to be phagocytosed by macrophages its expression did contribute to bacterial induced macrophage death. A preliminary analysis of the *pld* promoter was also performed.

In conclusion, the work described in this thesis has advanced our knowledge of gene regulation in *C. pseudotuberculosis*. The possible significance of regulated genes during the infection cycle is discussed.

omeostasis

phage infection
e, a Cu,Zn
re attenuated to
ermease mutant
etallopeptidase

s that has been
ng heat shock led
ell density
as high and did not
y *C.*
macrophages its
liminary analysis

edge of gene
ted genes during

Declaration

The material presented in this thesis has not been submitted elsewhere for the award of any other degree or diploma. To the best of my knowledge this thesis contains no material previously published or written by another person, expect where due reference is made in the text of the thesis.

[redacted]

Sandra McKean

Acknowledgements

I would like to sincerely thank Dr. Robert Moore for his advice, encouragement, discussion and enthusiasm over the past three years. It has been a lot of fun. I would also like to thank Dr. John Davies for his valuable supervision from afar.

I am grateful to the Australian Animal Health Laboratory (CSIRO, Livestock Industries) for both financial support and for providing a dynamic working environment. Thankyou especially to Dr. Andrew Bean for assistance with the flow cytometer, Mary Broadway for assistance with tissue culture techniques and the AAHL library staff. Additionally thankyou to everyone in Labs 4 and 11 for producing a conducive and happy work environment, in particular Tim Doran, Kerri Bruce, Scott Sheedy, Tony Davis, Volker Haring, Kathy Granger and Melissa Kowalski. Thanks also to the jugglers, soccer players and holders of AAHL girls nights for, at times needed stress release.

Finally I would like to thank my family for all their love and support. Ultimately you have made this possible for me. I would especially like to thank Rhys for being there every step of the way. Thankyou.

List of Abbreviations

Amp	ampicillin
AP	alkaline phosphatase
ASC	antibody secreting cell
BAP	bacterial alkaline phosphatase
BHI	brain heart infusion
bp	basepair
BSA	bovine serum albumin
CAT	chloramphenicol acetyltransferase
CDP-star	Disodium 4-chloro-3-(methoxyspiro {1,2-dioxetane-3,2'-(5'-chloro) tricyclo [3.3.1.1^{3,7}] decan}-4-yl)phenyl phosphate
CFU	colony forming units
CLA	caseous lymphadenitis
CM	conditioned medium
CSPD	Disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2'-(5'-chloro) tricyclo [3.3.1.1^{3,7}] decan}-4-yl)phenyl phosphate
Cu,ZnSOD	copper-zinc SOD
DAG	diacylglycerol
DD-PCR	differential display-polymerase chain reaction
DEPC	diethyl pyrocarbonate
DFI	Differential fluorescence induction
DIG	digoxigenin
DMEM	Dulbecco's modified Eagle's medium
dNTP	equimolar mixture of dATP, dCTP, dGTP, dTTP
EDTA	ethylenediaminetetraacetic acid
FACS	fluorescent activated cell sorter
FAS	fatty acid synthase
FCS	fetal calf serum
FITC	fluorescein isothiocyanate
GDP	genome directed primers
Gfp	green fluorescent protein
HAIR	HspR associated inverted repeat
hr	hour(s)
IFN-γ	interferon gamma
IHF	integration host factor
IPTG	isopropyl-β-D-thiogalactopyranoside
IVET	*in vivo* expression technology
IVIAT	*in vivo* induced antigen technology
Kan	kanamycin
kb	kilobase pairs
kDa	kilodaltons
LB	Luria Bertani medium
MBN	Mung bean nuclease
min	minute(s)
MOI	multiplicity of infection
MOPS	3-[N-Morpholino]propanesulfonic acid
mRNA	messenger RNA
MW	molecular weight
OD	optical density
PA	phosphatidic acid

PBS	phosphate buffered saline
PCC	propionyl coA carboxylase
PCR	polymerase chain reaction
PEG	polyethylene glycol
PI	propidium iodide
PKC	protein kinase C
Pld	phospholipase D
RAP-PCR	random arbitrarily primed PCR
RBS	ribosome binding site
RLB	rapid ligation buffer
RNI	reactive nitrogen intermediate
ROI	reactive oxygen intermediate
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	reverse transcriptase
SDS	sodium dodecyl sulfate
SOD	superoxide dismutase
srp	sterically repressed promoter
SSC	sodium chloride, sodium citrate buffer
STM	signature tagged mutagenesis
TAE	tris, acetate, EDTA buffer
TB	Terrific broth
TE	Tris-EDTA buffer
Tris	Tris(hydroxymethyl)methane
UC	uncut
UL	ulcerative lymphangitis
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-indolyl-ß-S-galactoside

Chapter 1

Literature Review

1.1 *Corynebacterium pseudotuberculosis*

C. pseudotuberculosis, which has previously been referred to as both *C. ovis* and *Preisz-Nocard bacillus* (Williamson, 2001), is a gram positive, mesophilic bacterium. These bacteria are facultative anaerobes, non motile, rod shaped and surrounded by a layer of electron dense surface lipid, which is external to the cell wall (Hard, 1969a). *Corynebacterium* is a heterogeneous genus ranging from pathogenic species such as *C. pseudotuberculosis*, *C. diphtheriae* and *C. ulcerans* to soil saprophytes such as the industrially important amino acid producing *C. glutamicum* and *C. ammoniagenes*. Analysis of 16s rRNA sequences indicate that within the *Corynebacterium* genus *C. pseudotuberculosis* is most closely phylogenetically related to *C. ulcerans* followed by *C. diphtheriae* (Takahashi *et al.*, 1997). Outside of the genus Corynebacteria are closely related to other mycolic acid containing genera such as *Dietzia*, *Gordona*, *Mycobacterium*, *Nocardia*, *Rhodococcus*, *Tsukamurella* and *Turicella*. *Corynebacterium* together with *Mycobacterium* and *Nocardia* make up the CMN group of bacteria. Within its hosts *C. pseudotuberculosis* is primarily an intracellular pathogen that resides and replicates within macrophages (Hard, 1972;Tashjian and Campbell, 1983).

C. pseudotuberculosis is distinguishable in the laboratory on the basis of the morphology of its hydrophobic colonies and biochemical analysis for the presence of urease and phospholipase D (Pld) activity. Two biotypes of *C. pseudotuberculosis* have been defined, a biovar ovis and a biovar equi (Songer *et al.*, 1988). In general equine strains can be distinguished from ovine strains on their ability to reduce nitrate. Biovars have also been characterised by Restriction endonuclease analysis (Songer *et al.*, 1988), or Pulsed-Field gel electrophoresis (Connor *et al.*, 2000).

1.2 Diseases caused by *C. pseudotuberculosis*

C. pseudotuberculosis is the etiological agent of two specific diseases, caseous lymphadenitis (CLA) which affects sheep and goats and ulcerative lymphangitis (UL) which affects horses, mules and occasionally cattle. *C. pseudotuberculosis* has also been observed to be pathogenic in a number of other mammalian species including camels, buffalo, deer and humans (Collier *et al.*, 1998;Mills *et al.*, 1997;Peel *et al.*, 1997).

UL, which is caused by the equi biovar, is characterised by inflammation of the lymphatic vessels, which most commonly occurs on the fetlocks of infected horses. Nodules form in the subcutaneous tissue surrounding the site of infection and may enlarge to 5-7 cm. Infection frequently spreads along the draining lymph nodes and further abscesses may arise. Abscesses often rupture to yield ulcers with a necrotic base that discharges caseous green pus (1999). In Australia only the ovis biotype appears to be present, hence cases of UL are not observed. The remainder of this thesis is therefore given to the description and study of ovine strains of *C. pseudotuberculosis* which cause CLA and are prevalent in Australia.

CLA is characterised by caseous abscesses primarily in the superficial lymph nodes of infected animals. Necrosis of affected lymph nodes is also observed. In some cases other organs may also be infected. This is most commonly seen in the visceral organs and in particular the lungs (Batey, 1986b).

Laboratory animals show different susceptibilities to infection with *C. pseudotuberculosis*. In the guinea pig, disease progression is rapid and typically fatal (Dickinson and Bull, 1931). The mouse model of infection has been well studied and much of the information regarding the immunological responses of the host to *C. pseudotuberculosis* infection have been derived from these studies. Jolly (1965a) showed that different levels of disease could be induced in the mouse by altering either the dose or route of administration. Although high doses are lethal to the mouse, lower doses result in a disease that may be contained by the animal (Batey, 1986a). There are however differences in disease pathogenesis, in particular in the organs affected. In sheep primary lesions develop at the site of infection (skin) followed by the establishment of secondary lesions in the lymph nodes and visceral organs. Conversely, in the mouse, the peritoneal cavity or skin (depending on route of inoculation) is considered the primary site and the liver, lung, spleen or kidney the secondary site (Batey, 1986a). Unlike in sheep the development of lesions in the lungs is not observed in the mouse.

1.3 Prevalence of CLA in Australia

CLA is a common bacterial disease of sheep in Australia. Within the years 1988-1990 almost half of adult sheep (from 412 flocks) culled in Western Australia showed signs of

CLA (Paton, 1997). More recent data suggest that although the rate of incidence has decreased, disease prevalence is still high (estimated at approximately 20% in Western Australian flocks in 1997 (Paton, 1997). When used correctly commercially available vaccines, which contain detoxified phospholipase D, are able to substantially reduce the incidence of CLA (Paton *et al.*, 1995;Piontkowski and Shivvers, 1998). There are no reliable data available on the prevalence of CLA in goats however it is generally thought that the disease level is probably lower than in sheep.

1.4 Economic significance of disease caused by *C. pseudotuberculosis*

Abscesses in the lymph nodes have little effect on the general health of the sheep, however the disease is of economic significance in terms of reduced wool production and condemnation of carcasses at slaughter. Sheep with signs of infection have on average a 4 to 7% reduction in clean wool production in the year they are infected (Paton *et al.*, 1994). In subsequent years of infection, wool production is observed to return to normal levels (Paton, 1997). The losses to the Australian wool industry are estimated to correspond to a monetary value of $15 million (Paton, 1997). Approximately 1% of adult carcasses are condemned at slaughter due to CLA infection. In Australia this corresponds to losses of $4 million annually (Paton, 1997). Additional losses of approximately $10 million per year occur through the cost incurred in the trimming of less severely infected carcases (Paton, 1997). Therefore the total annual loss to the Australian sheep industry is at least $30 million per annum.

1.5 CLA: route of infection and disease pathogenesis

The primary route of infection is via the skin. *C. pseudotuberculosis* has been shown to be able to cross intact skin of recently shorn sheep (Nairn and Robertson, 1974), however most infections probably occur at sites of skin wounding (Batey, 1986b). Infection is therefore most common following shearing as this is when sheep are most likely to present with skin lesions. It was initially thought that the primary source of bacteria for infection was from the ruptured abscesses of already infected sheep. However, more recent studies convincingly show that a major source of infective bacteria is from contaminated aerosols of sheep with discharging lung abscesses (Paton *et al.*, 1995;Ellis *et al.*, 1987). *C. pseudotuberculosis* can however survive for long periods of time in the environment, including for at least 24 hours in sheep dip (Nairn and Robertson, 1974), raising the possibility of transmission from environmental sources.

The exact mechanisms by which disease is then established are not well understood and are in general based on experimental studies in which large numbers of bacteria are administered by subcutaneous inoculation (Pepin *et al.*, 1991b;Pepin *et al.*, 1994). Experimental infection in sheep is typically associated with a transient temperature increase, localised inflammation and abscess formation at the site of infection (Pepin *et al.*, 1991b). Inflammation is caused at least in part by an infiltration of immune cells and possibly by increased prostaglandin biosynthesis as a result of activation of mammalian arachondic acid-producing pathways by the bacteria. It is likely that bacteria replicate extracellularly at the site of infection prior to lymph node localisation. Alternatively bacteria are taken up by phagocytic cells and then actively transported to the lymph nodes. Massive immune infiltration into the lymph node occurs. Neutrophils are predominant early, however around day 3 post infection macrophages become the predominant cell type (Pepin *et al.*, 1991a). Within the lymph node repetitive cycles of bacterial phagocytosis, intracellular bacterial replication and phagocytolysis play an essential role in abscess formation. Fully formed abscesses which may reach a diameter of 10 cm have been shown to consist of a central necrotic region surrounded by an internal layer of macrophages and epithelioid cells, which is in turn surrounded by a layer that consists predominantly of lymphocytes and a few fibroblasts (Pepin *et al.*, 1991b). This is enclosed in a fibrous capsule. The encapsulation of the infection limits bacterial dissemination but to the advantage of the bacteria it also provides a defence against the normal immunological defences of the host.

A number of studies investigating the interaction between *C. pseudotuberculosis* and phagocytes have been performed. Studies have been performed *ex vivo* using caprine mammary macrophages (Tashjian and Campbell, 1983) or *in vivo* using a mouse model in which *C. pseudotuberculosis* is injected into the peritoneum (Hard, 1969b;Hard, 1972). No studies have been performed using ovine models. All studies demonstrate that *C. pseudotuberculosis* is rapidly engulfed by macrophages and that fusion between phagosomes containing bacteria and primary and secondary lysosomes occurs to generate phagolysosomes. A decrease in macrophage viability also occurs within a few hours of infection. The *in vivo* study in the mouse peritoneum indicated that necrotic macrophages and their contents are phagocytosed by scavenging macrophages which are at least

partially successful in killing the bacteria. However infection was persistent with bacteria recoverable 10 days post infection (Hard, 1972).

1.6 Virulence determinants and their role in *C. pseudotuberculosis* pathogenesis

Virulence genes are those that are considered essential for successful infection of a host by a bacterial pathogen. Often virulence genes are not required for growth outside the host, hence their expression may be tightly regulated. Due to a lack of molecular information relatively little is known about the genes involved in *C. pseudotuberculosis* survival and replication *in vivo*. The major virulence determinant identified to date is a protein with phospholipase D activity, while a putative serine protease and toxic cell wall components have also been suggested to be important for disease progression. Regulation of expression has not been described for any of these genes.

1.6.1 Phospholipase D

1.6.1.1 Protein characteristics and enzyme activity

Pld is a secreted exotoxin with a molecular weight of 31.4 kDa which possesses sphingomyelinase activity (Hodgson *et al.*, 1990). Pld has been shown to increase vascular permeability *in vivo*, have dermonecrotic properties, exhibit synergistic haemolysis of sheep blood cells in the presence of products from *Rhodococcus equi* and reduce the viability of ovine neutrophils *ex vivo* (Batey, 1986b;Yozwiak and Songer, 1993). The chromosome contains a single copy of the gene and is transcribed as a monocistronic RNA of 1.1 kb (Hodgson *et al.*, 1990).

1.6.1.2 Mutant strains

The importance of the *pld* gene to pathogenesis is convincingly demonstrated by studies performed with strains of *C. pseudotuberculosis* in which the *pld* gene has been inactivated (Hodgson *et al.*, 1992;McNamara *et al.*, 1994;Hodgson *et al.*, 1994;Simmons *et al.*, 1998;Hodgson *et al.*, 1999). Studies with one such ovine strain called Toxminus found that in the absence of a functional *pld* gene no abscessation in the lymph node occurred, whereas animals infected with wildtype *C. pseudotuberculosis* demonstrated extensive abscessation (Hodgson *et al.*, 1992). Similar studies have been performed in a caprine strain of *C. pseudotuberculosis*. The *pld* gene of strain Whetten 1 was replaced with an allele containing a nonsense mutation to generate strain W1.31r1 (McNamara *et al.*, 1994). The mutant strain demonstrated the same growth rate *in vitro* but was unable to cause

abscesses in the lymph node at a 100-fold higher dose than wildtype. In both studies the *pld* mutant strains caused a small amount of abscessation at the site of infection in some animals. These two studies convincingly demonstrate the importance of Pld for virulence and the lack of bacteria in the lymph nodes supports the hypothesis that a major role of Pld, *in vivo*, is in aiding bacterial dissemination to the lymph nodes.

Additional evidence for the importance of Pld *in vivo* comes from the observation that vaccination with formulations in which Pld is the major component provides protection against subsequent disease challenge (Eggleton *et al.*, 1991b). Pld specific antibody responses are also associated with protection of sheep from CLA (Eggleton *et al.*, 1991a). In particular, immunization of sheep with Pld or passive transfer of serum from immunised to non immunised mice reduces the number of bacteria that localise to the lymph node following experimental infection (Jolly, 1965b;Hard, 1970).

1.6.1.3 *in vivo* effects

Pld, *in vivo*, is believed to increase vascular permeability thereby leading to increased transmission of the bacteria from the site of infection to the lymph node. Marked leakage of plasma from small blood vessels is observed at the site of experimental infection (Carne and Onon, 1978). This results in the flooding of lymphatic spaces thus increasing bacterial transport from the periphery to the draining lymph nodes. Other studies using supramaximal concentrations of bacteria or purified toxin have demonstrated that Pld degrades the sphingomyelin on sheep erythrocytes *in vivo* and that dramatic changes in erythrocyte morphology occur resulting in significant pitting (Brogden and Engen, 1990). Interestingly, *in vitro*, Pld is unable to independently lyse sheep erythrocytes, instead requiring the presence of a cholesterol oxidase from *Rhodococcus equi* (Linder and Bernheimer, 1997). However, *in vivo* haemolytic anaemia following inoculation of Pld can be observed in lambs (Hsu *et al.*, 1985). This suggests that additional components required for haemolysis are present in the host but not in ovine blood preparations used for *in vitro* assays.

Mammalian cells also possess phospholipases, including two isoforms of Pld (Colley *et al.*, 1997). It is likely that *C. pseudotuberculosis* Pld is able to activate mammalian signalling pathways. Hydrolysis of the membrane component phosphatidylcholine by mammalian Pld results in the formation of phosphatidic acid and choline. Phosphatidic acid (PA) which may act as a signalling molecule (e.g., by activating a PA-activated kinase), can be

hydrolysed by phosphatide phosphohydrolase to form diacylglycerol (DAG) or in a reaction involving phospholipase A_2 be converted to lysophosphatidic acid (LPA) (Exton, 1997). Both DAG and LPA may act as signalling molecules (Exton, 1997). DAG, which is embedded in the membrane, is able to activate the important signalling molecule protein kinase C (PKC). PKC when in its activated state is able to phosphorylate and thus activate a variety of proteins including those involved in cell functions such as actin assembly (Lennartz, 1999). DAG may be further degraded to generate arachidonic acid. Arachidonic acid is the primary substrate for the production of prostaglandins, prostacyclins, thromboxanes and leukotrienes. Prostaglandins are potent vasodilators, therefore some of the Pld mediated effects on vascular permeability may be through this type of pathway rather than simple destruction of mammalian cell membranes. Mammalian Pld has also been implicated to play a role in a number of immune cell functions including phagocytosis by macrophages (Lennartz, 1999), ATP-induced killing of *M. tuberculosis* by macrophages (Kusner and Adams, 2000) and production of superoxide and degranulation in neutrophils in response to chemotactic agents (Olson and Lambeth, 1996). However, whether activation of any of these pathways by bacterial Pld would be advantageous to *C. pseudotuberculosis* is debatable.

1.6.2 CP40 (serine protease)

A putative serine protease called CP40 has also been suggested to be a virulence factor (Walker *et al.*, 1994). CP40 was identified using an approach that utilises locally derived antibody-secreting cells (ASC) from sheep with CLA. In comparison to antibodies found in serum, ASC cells obtained locally (from the site of infection) produce antibodies of highly restricted specificity. Supernatant from cultured ASC cells was used to probe blots containing whole-cell antigens of *C. pseudotuberculosis*. The CP40 protein band was consistently recognised and in two of three sheep analysed was the major antigen. Like Pld, CP40 is a secreted protein (Wilson *et al.*, 1995). When CP40 protein was administered as a vaccine an 82% reduction in the proportion of infected sheep and a 98% reduction in lung lesions was observed 3 months after challenge (Walker *et al.*, 1994). The substrates for CP40 have not been identified. The observation that this protein is secreted does however raise the possibility that its targets are of host origin.

1.6.3 Cell wall components

Although other specific virulence determinants have not yet been identified it is likely that components of the cell wall may play a role in virulence. A direct correlation has been found between the percentage of bacterial lipid that is found in a floccular layer external to the cell wall and the virulence of the bacterium in a mouse model of infection (Muckle and Gyles, 1983). Interestingly this was observed as increased induction of chronic abscessation with increasing lipid content but increased mortality was not observed. It is likely that the surface lipid is at least partly responsible for clumping which contributes to *C. pseudotuberculosis*' pathogenic properties. The correlation between lipid content and disease severity suggests that genes involved in lipid biosynthesis may be considered virulence factors.

1.6.4 AroQ (encodes a type II 3-dehydroquinase)

C. pseudotuberculosis strains that show reduced ability to synthesise aromatic amino acids have also been studied *in vitro* and *in vivo*. The *aroQ* gene of *C. pseudotuberculosis* has been inactivated by the insertion of an erythromycin resistance gene (Simmons *et al.*, 1997). The resulting strain showed reduced growth *in vitro* and in a mouse model of infection a three-fold log increase in LD50 compared to the parental wildtype strain was observed (Simmons *et al.*, 1997). Observations from studies such as this raise the question of whether genes that are essential *in vitro* can be considered to be virulence factors. Pathogenic bacteria need to be able to synthesise aromatic amino acids (which are used as precursors for folic acid) as the availability (p-aminobenzoic acid) is limited in host tissues. Therefore the expression of this gene can be considered essential for virulence. The same could be said for many genes that would traditionally be described as housekeeping genes. It is therefore difficult to draw a line between what is and isn't a virulence factor, although if the bacteria can survive without the gene *in vitro* but not *in vivo* then it can probably be considered a virulence factor.

1.6.5 Other virulence determinants

Studies of *C. pseudotuberculosis* pathogenesis clearly demonstrate an important role for Pld in the initial establishment of disease, however whether Pld is essential for disease progression once abscesses have been established in lymph nodes is unclear. Given that *C. pseudotuberculosis* must survive in the hostile environment of the phagocytic cell it is likely that genes other than those described in this section are required, for example genes

involved in protecting bacteria from free radicals produced by macrophages. Additionally within phagocytes the nutrient supply is likely to be limited, therefore genes involved in the scavenging of nutrients such as iron and other metabolites are likely to be up regulated. It therefore seems very likely that other virulence genes will be identified in the future.

1.7 Aspects of the regulation of bacterial gene expression

For bacterial pathogens infection consists of a series of steps. For *C. pseudotuberculosis* the bacteria must first penetrate through the skin, then move from the site of infection to the lymph nodes. This is either preceded or followed by internalisation of the bacteria within phagocytic cells. Within the lymph node the bacteria must then undergo cycles of replication, phagocytolysis and reinfection of new phagocytes. The ultimate step is exit from the host thereby allowing for a new cycle of infection to begin. The external environment, blood, lymph nodes and intracellular environment of macrophages will each require adaptations from the bacteria to allow survival and the scavenging of nutrients required for viability and replication. To respond to such changes bacteria have evolved regulatory systems to control gene expression in response to the external environment thus leading to appropriate adaptation to it. Conversely such regulation may also prevent the expression of genes unnecessary for survival within a given environment. This could be considered to be a mechanism whereby "energy" is conserved. Empirical evidence shows that for many bacterial pathogens the expression of those genes that play a direct role in bacterial survival within the host is often tightly regulated, i.e. such that their expression is timed with certain aspects of infection. In particular, virulence genes have often been observed to be *in vivo* induced during specific stages of infection.

Bacteria have evolved mechanisms to allow coordinated regulation of expression of multiple genes. A single gene may have its own promoter that is regulated by given stimuli, alternatively genes may be arranged into an operon in which two or more usually related genes are cotranscribed as a polycistronic unit from a single promoter. Single genes or operons that are dispersed around the chromosome, but that have promoters that respond to the same regulators in a similar fashion are said to comprise a regulon. Finally genes or operons that respond to a stimulus but not necessarily in the same manner (i.e. not all up or down regulated) are said to comprise a stimulon.

At the most fundamental level, gene regulation occurs at the promoter which is the minimal sequence recognised by the RNA polymerase (RNAP) that gives rise to basal

promoter activity. Promoter strength is regulated by a number of factors including DNA structure, sigma factors, DNA binding proteins and transcriptional regulator proteins. Regulation at the promoter occurs in response to factors such as changes in temperature, environmental factors and nutrient limitation. To allow this to occur bacteria have evolved efficient signalling cascades and mechanisms that allow them to sense and then adapt appropriately to changes in environmental conditions.

1.7.1 Regulation of expression at the promoter

1.7.1.1 Sigma factors

In eubacteria RNAP consists of a core of two α, one β and one β' subunit in addition to a variable σ subunit. It is the σ subunit that defines the specificity of the enzyme by binding to the DNA promoter sequence and hence also defines the basal strength of the promoter. Most bacteria possess multiple σ factors that each recognise different consensus sequences. In eubacteria σ factors can be classed into two groups: those that show similarity to the σ^{70} of *E. coli* form the σ^{70}-family (Lonetto *et al.*, 1992) and those that show homology to *E. coli* σ^{54} form the σ^{54}-family (Merrick, 1993). σ^{70} typically bind to two regions of the DNA known as the –35 and –10 regions. These are found at –30 to –35 and about –7 to –12 (relative to the transcription start site) and in *E. coli* and many other bacteria the consensus sequences for these two regions are TTGACA and TATAAT respectively. In gram positive bacteria such as *B. subtilis* a conserved motif at –16 (TRTG (R = purine) is often also observed (Voskuil and Chambliss, 1998).

Within the σ^{70}-family a number of sigma factors are found which can be divided into three groups. Group I sigma factors generally control the expression of the majority of genes which are required during exponential cell growth and as such are usually genes that are essential for cell survival. Group II comprises sigma factors which control expression of genes that are generally nonessential for growth. Group III sigma factors generally control the transcription of specific regulons needed during special physiological or developmental conditions, i.e. flagellar σ-factors, extracytoplasmic function σ-factors (a class of environmentally responsive transcriptional regulators) and heat shock σ-factors.

A wide variety of σ^{54}-dependent genes in both gram positive and gram negative bacteria have been identified. Like Group II σ^{70} factors, σ^{54} usually control the expression of non-essential genes (Merrick, 1993). Additionally RNAP containing σ^{54} is unable to initiate transcription unless an enhancer binding protein is also present (Morett and Segovia,

1993). A consensus recognition sequence for σ^{54}-dependent promoters has been defined as –26(TGGCACGNNNNTTGC (-12) T/A NNANNN (Barrios *et al.*, 1999).

The expression of σ-factors is controlled by a variety of mechanisms. The system is regulated such that appropriate σ-factors are present in the cell for the given environmental conditions. Expression of some σ-factors is controlled at the transcriptional level by the binding of transcriptional regulators, for some others, anti-σ factors interact with specific σ-factors to negatively regulate their activity, for others such as the heat shock σ factors activity is controlled primarily by changes in protein stability (Wosten, 1998).

1.7.1.2 UP elements

Upstream of the –35 region of a number of promoters is an AT rich region that is able to interact with the α subunit of RNAP. This α recognition element is termed the Upstream or UP element, and an AT rich consensus sequence of NNAAAWWTWTTTTNNNAAANNN (where W = A or T and N = any base) optimally located at –59 to –38 has been defined (Estrem *et al.*, 1998). This consensus sequence contains two subsites each of which bind one α subunit. UP elements increase promoter activity by recruiting RNAP to the promoter DNA, thereby increasing the equilibrium constant for the initial binding step (Gourse *et al.*, 2000). Interaction between the α subunits of RNAP can play an important role in activated transcription as well as basal transcription, however not all promoter-α subunit interactions that occur have a positive effect on promoter function.

1.7.1.3 Repressors and activators

Repressors and activators are DNA binding proteins that bind to specific sequences in the region surrounding a promoter and cause a subsequent decrease or increase in the level of transcription. In *E. coli* the majority of promoters (75%) are controlled by a single type of regulator (Gralla and Collado-Vides, 1996). Additionally, modulation of promoter activity by repression is twice as common as by activator regulation (Gralla and Collado-Vides, 1996). Dual regulators are transcriptional regulators that may act as either an activator or repressor depending on where its binding site is located. The majority of repressor binding sites are found around –30 where it is believed that the presence of a repressor protein mechanistically interferes with the interaction between RNAP and the DNA. Other mechanisms of repression can occur, as for a number of promoters RNAP is able to bind to the promoter but subsequent steps required for RNA synthesis are blocked (Fassler and

Gussin, 1996). In *E. coli* activator sites are typically found between –80 and –30 (Gralla and Collado-Vides, 1996). Promoters that require activation are in general rate limited in one or more steps in the pathway to transcription. Activators typically bring about an increased rate constant for one or more of these steps thereby up-regulating the level of transcription. Alternatively activation can be mediated by a change in the rate of promoter clearance.

In addition to regulator binding sites found in close proximity to the promoter a minority of promoters contain binding sites remote to the promoter (i.e. greater than 80bp from the transcriptional start site). Distal sites are however almost always found paired with sites that are in a proximal position and in general it appears that their role is to modulate the primary interaction, which occurs between the polymerase and the regulator in the proximal region (Gralla and Collado-Vides, 1996).

1.7.1.4 DNA structure

The topology of the DNA at and surrounding a promoter influences the ability of RNAP and regulatory proteins to interact with it, and hence partially determines promoter strength. DNA topology is determined by the base composition of the DNA, template supercoiling and interactions with proteins that alter DNA structure through looping, bending and alteration of supercoiling (Dai and Rothman-Denes, 1999).

Many promoter regions have been shown to be AT rich compared to the overall content of the genome. Base composition determines local DNA structure (curvature) and can therefore influence promoter activity. For example some promoters contain phased A tracts which confer an intrinsic bend to the DNA and enhance the interaction between RNA polymerase and the promoter by extending the contact region (Dai and Rothman-Denes, 1999). This occurs in the promoter of the *plc* gene from *C. perfringens* which contains three phased A tracts. The bending angle of these phased tracts is temperature regulated such that the DNA bending is greater at lower temperatures (Katayama *et al.*, 1999). The increased DNA bending angle leads to a higher affinity interaction between RNAP and the DNA, leading to increased promoter activity (Katayama *et al.*, 2001). Thus *plc* expression is temperature regulated in a primitive manner that has no requirement for trans-acting factors.

In addition to transcriptional regulators that bind to specific sites at one or more promoters, bacteria contain DNA binding proteins that bind to multiple sites throughout the genome. The identity and function of this type of protein is better understood in gram-negative bacteria than gram-positive bacteria. Among the architectural proteins identified in gram-negative bacteria are HU, FIS, H-NS and integration host factor (IHF). These proteins may play a role in a variety of cellular processes including condensation of the bacterial nucleoid, and DNA transcription, replication and recombination. Some DNA binding proteins bind to DNA non-specifically whereas others such as H-NS and IHF bind at specific sites for which consensus motifs have been defined. Binding of the protein to the DNA may induce structural changes in the DNA. IHF, for example, introduces an approximately 180° bend into the DNA (Engelhorn and Geiselmann, 1998). When this occurs in a promoter region it may facilitate interactions between regulatory proteins bound at distal upstream sites and RNAP at downstream promoter sites.

The best characterised architectural DNA binding protein in gram-positive bacteria is the histone-like protein HBsu of *Bacillus subtilis*. HBsu is essential for normal growth of *B. subtilis* (Micka *et al.*, 1991;Kohler and Marahiel, 1997) and has been shown to play a role in DNA repair and homologous DNA recombination (Fernandez *et al.*, 1997). Interestingly, analysis of the *B. subtilis* genome did not reveal the presence of proteins with homology to sequence specific architectural proteins of gram-negative bacteria such as H-NS, IHF and Fis (Kunst *et al.*, 1997).

Histone-like proteins have been identified in a number of other gram-positive bacteria including *M. tuberculosis* (Prabhakar *et al.*, 1998), *M. smegmatis* (Shires and Steyn, 2001), *Clostridium pasteurianum* (Kothekar *et al.*, 1998) and *Staphylococcus aureus* (Viter *et al.*, 1999). The role of these proteins in processes such as transcriptional regulation is however much less clear than for the well studied gram-negative architectural proteins.

Binding of transcriptional activators may also influence DNA structure at the promoter, for example activators may increase transcription by unwinding the DNA at the site of interaction leading to better alignment of the –35 and –10 regions and subsequently a better interaction between the DNA and RNAP (Dai and Rothman-Denes, 1999).

1.7.2 Regulatory systems

The mechanisms of transcriptional control described above focus on how regulation occurs at the DNA level. This is frequently the last step in a regulatory signal cascade in which changes in gene expression are made in response to alterations in the internal and external environment. A major regulatory feature of prokaryotes is the two component regulatory system. This is the predominant system by which modification in gene expression occurs in response to changes in the environment. In general a two component system consists of two proteins, a histidine protein kinase and a response regulator protein (Foussard *et al.*, 2001). The histidine kinase is a transmembrane protein that detects signals external to the cells. Binding of the signal may either activate or inactivate the protein's kinase activity. An activated kinase hydrolyses ATP and autophosphorylates at a specific histidine residue. Transfer of the phosphoryl group from the phosphokinase to an aspartate residue of the response regulator protein then occurs. This phosphorylation brings about a change in the conformation of the response regulator's regulatory domain which in turn activates an associated domain which effects the response at the transcriptional level. A large but variable number of two component systems are found in most bacteria. For example, based on the genome sequence, 63 have been observed in *P. aeruginosa* (Rodrigue *et al.*, 2000), 70 in *B. subtilus* and 11 in *Helicobacter pylori* (Stock *et al.*, 2000). Two component signalling systems play important roles in the regulation of a large number of cellular functions including production of virulence determinants, nutrient acquisition, energy metabolism, quorum sensing and adaptation to the physical environment.

Other transcriptional factors are regulated by metabolic intermediates. Binding of the metabolic intermediate to the regulator alters its structure and hence its DNA binding activity. Metals are commonly found to bind to transcription factors. For example in *C. diphtheriae* the regulator DtxR acts as a repressor in the presence of iron. The *in vivo* environment is usually low in iron, hence genes usually repressed by DtxR become activated. This includes a number of virulence determinants including the *tox* gene.

Many other forms of regulation exist, in particular at the post transcriptional level. Both RNA and protein stability can be significantly altered by environmental factors such as temperature and osmolarity.

1.8 Methods for the identification of differentially regulated genes in bacterial species

In recent years a number of technologies have been developed for the identification of differentially expressed genes, in particular to identify genes that show up-regulation *in vivo* compared to control *in vitro* conditions. Some of these approaches, such as *in vitro* Expression Technology and Signature Tagged Mutagenesis, have been developed specifically for prokaryotic systems, whereas others such as differential display have first been developed for eukaryotic systems and subsequently adapted for prokaryotic systems. As eukaryotic technologies often use the presence of a polyA tail on mRNA as an enrichment tool, adaptations have been required to allow these protocols to be utilised with bacterial systems. This section summarises some of the major technologies currently available. Differential Fluorescence Induction and DNA Array approaches are described in detail as these were undertaken during this study.

1.8.1 Signature Tagged Mutagenesis (STM)

The process of STM was pioneered by David Holden's group in 1995, using a murine model of typhoid fever caused by *Salmonella typhimurium* (Hensel *et al.*, 1995). STM has significant advantages over previously existing mutation methods as it can be used to identify mutations that result in a loss of function. This is required for the identification of virulence determinants as they will result in attenuation or non survival of the bacteria in the host. STM has been used for a number of bacterial and fungal pathogens in animal models, but it can only be used in organisms for which transposon induced mutagenesis systems exist.

A disadvantage of this approach compared to the IVET technology discussed later is that essential genes which are also virulence factors will not be identified in this study. This is because it is not possible to make a mutant in which an essential gene has been inactivated. On the other hand the advantage of this approach is that the essentiality of identified genes for the virulent phenotype is directly determined. For genes shown to be up-regulated *in vivo* by other techniques it is subsequently necessary to generate a mutant strain and test for attenuation *in vivo*.

1.8.2 PCR amplification methods

Differential display (DD) and its derivative RNA arbitrarily primed PCR (RAP-PCR) are polymerase chain reaction (PCR) based approaches that can be used to identify both up and down regulated genes in two or more samples (Fislage, 1998;Matz and Lukyanov, 1998). In DD arbitrarily primed PCR is utilised to produce fingerprints from the first-strand cDNA of the samples of interest. Typically each fingerprint contains between 50 and 100 distinguishable bands. By using successive primer pairs different genes will be amplified such that eventually the whole genome may be covered. Bands that appear to be regulated are excised from the gel and the identity of the gene determined by sequencing.

DD was initially developed for eukaryotic systems but has been used to identify regulated genes in a number of bacterial species. It has been successfully used to identify heat regulated genes in *Treponema denticola* (Tsai and Shi, 2000), *in vivo* induced genes in *Shigella flexneri* (Robb *et al.*, 2001) and *Streptococcus pneumoniae* (Orihuela *et al.*, 2001), acid induced genes in *Helicobacter pylori* (Dong *et al.*, 2001) and *Streptococcus mutans* (Hanna *et al.*, 2001) and macrophage induced genes *of Legionella pneumophilia* (Abu Kwaik and Pederson, 1996). DD can however be dogged with technical difficulties, in particular the number of false positives detected (Nagel *et al.*, 1999).

1.8.3 Proteomics

A number of approaches also exist for looking at regulation at the protein level. For at least some genes there are disparities between the relative levels of mRNA and their corresponding protein (Dutt and Lee, 2000). Monitoring changes in protein level is a way in which this problem can be avoided. The major proteomic approach is 2-dimensional gel electrophoresis in which proteins are separated using SDS-PAGE on the basis of charge by isoelectric focusing and then on relative size. Currently, because of major shortcomings with this technology, it is limited in its application for identifying *in vivo* regulated genes (differential proteome analysis). Difficulties include the preparation of suitable samples, particularly *in vivo* ones in which it would be necessary to either remove all host proteins from the preparation or alternatively be able to discriminate between host and bacterial proteins. Secondly analytical 2D electrophoresis is complicated by the fact that differentially regulated proteins must be extracted from the gel and sequenced or analysed by mass spectroscopy before the originating gene can be identified. Finally access to this

type of approach can be somewhat limited by the type of equipment required and costs involved.

A recently developed methodology for identifying proteins that are induced in the host is *in vivo* induced antigen technology (IVIAT) (Handfield *et al.*, 2000). This technology can be applied even when an animal model of infection is not available making it particularly suitable for human pathogens. Sera obtained from an infected host are absorbed with whole bacteria and bacterial extracts from *in vitro* grown bacteria thus removing antibodies to constitutively expressed proteins. The absorbed sera are used as a probe against an expression library containing genes from the pathogen cloned into an appropriate expression vector. Reactive clones contain a DNA fragment from the pathogen that is *in vivo* induced. Issues with this type of approach include antibody cross reactivity leading to false positives, variability between the sera of different animals and the degree of quantitation that may be obtained. Additionally not all proteins are immunogenic so any screening experiment will not be comprehensive. However, this is not unlike the situation with the majority of screening technologies in which only a proportion of regulated genes are typically identified.

1.8.4 *In vitro* Expression Technology (IVET)

IVET is a promoter trap approach that is used to identify genes whose expression is up-regulated in the host animal. The first IVET experiment was performed by Mahan *et al* using a mouse model of *Salmonella typhimurium* infection (Mahan *et al.*, 1993). Transcriptional fusions between random fragments of genomic *S. typhimurium* DNA and a promoterless *purA* gene were created. The corresponding pool of fusions was then transformed into a *S. typhimurium* Δ*pur*A strain and selection for homologous recombination events performed. Mice were infected with the pooled fusion strains and surviving bacteria recovered after three days. As purines are limiting for growth of *S. typhimurium* in the mouse, only those strains that express *purA* from fused promoters would survive *in vivo*. To separate bacteria containing constitutive promoters from those with inducible promoters the recovered bacteria were analysed *in vitro* for those that have low promoter activity.

Modifications to this approach have been made such that the technology can be applied to bacterial systems in which auxotrophic selection is not possible. The auxotrophic gene is

replaced by an antibiotic resistance gene and the antibiotic is co-administered with the bacterial library to the animal (Mahan *et al.*, 1995). Promoter fragments that have activity *in vivo* will drive expression of the antibiotic resistance gene leading to survival of the bacteria. An advantage of antibiotic selection over the auxotrophic selection is that promoters with different levels of activity can be selected for by adjusting the amount of antibiotic administered. Additionally the timing of administration can be altered to allow the identification of genes that are expressed at a particular time during infection.

1.8.5 Differential fluorescence induction (DFI)

DFI is an enrichment strategy for the isolation of promoters that are activated by a given stimulus (Figure 1.1) and can be considered to be a further modification of IVET in which selection is replaced by an enrichment process. Genomic DNA fragments are inserted into a promoterless vector containing the green fluorescent protein (*gfp*) gene as a reporter. Following a stimulus, bacteria containing plasmids with an activated promoter express *gfp* and are enriched for by fluorescence activated cell sorting (FACS). The collected population will contain clones with both constitutive and inducible promoters. To remove those clones that express *gfp* constitutively a second round of sorting is performed on this population after growth under *in vitro* control conditions. This time the non-fluorescent population which will contain the population of inducible clones is collected by FACS. The original stimulus is applied to the bacteria and a final round of enrichment is performed resulting in a population of inducible clones. Inserts of clones of interest are then sequenced thus allowing identification of promoter sequence and downstream gene sequence information. DFI has several advantages when compared to the more traditional IVET approaches described in earlier sections. These include

1. DFI is not based on absolute levels of expression of the selectable marker. This means that weakly expressed clones can be sorted as efficiently as highly fluorescent ones within a single experiment.
2. As a flow cytometer is used to detect reporter expression, DFI has high sensitivity and can be used even when gene expression levels are relatively low.
3. Approximately 3000 bacteria can be sorted per second. This high processivity makes the analysis of large libraries relatively easy.
4. Gene induction is easily quantitated by flow cytometry.
5. Unlike IVET it is possible to apply DFI to *in vitro* screening, for example following an environmental shock such as heat shock.

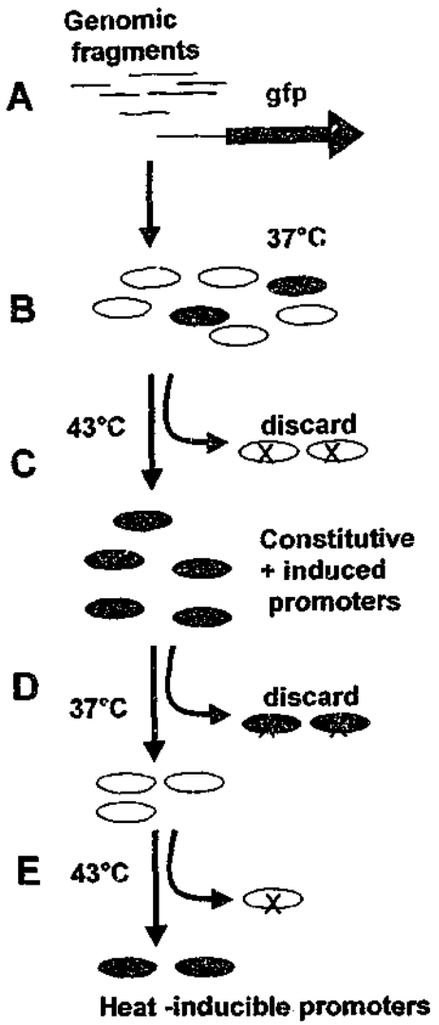

Figure 1.1: Differential Fluorescence Induction (DFI).

The process of DFI is outlined for the identification of heat regulated genes however heat shock may be substituted for other types of treatment, ie infection of macrophages. (A) Random fragments of chromosomal DNA are inserted upstream of the reporter gene (green fluorescent protein) in the promoter probe vector. (B) A proportion of these inserts will act as constitutive promoters giving rise to fluorescence at 37°C. (C) To select for promoters that are heat induced the culture is heat shocked. Fluorescent bacteria are then selected for by FACS analysis. This population will contain both heat induced and constitutive promoters. (D) To remove the constitutive promoters the selected bacteria are grown in nutrient rich conditions at 37°C and the non-fluorescing bacteria are collected by FACS while fluorescent bacteria (containing constitutive promoters) are discarded. (E) To remove false positives a second round of selection is performed at 43°C. Fluorescent bacteria isolated at this stage should contain heat activated promoters

1.8.5.1 Green fluorescent protein (Gfp) as a reporter gene

Since 1995 Gfp has been widely adopted by the scientific community as a tool for studying gene expression. Gfp was discovered by Shimomura *et al* (Shimomura *et al.*, 1962) while studying fluorescence in the bioluminescent jellyfish *Aequorea victoria*. In *Aequorea victoria* the photoprotein aequorin, upon binding calcium, emits blue light, that in turn excites Gfp which emits light at 508nm. The gene was cloned by Prasher *et al* (1992) and it was subsequently shown that expression of *gfp* in other organisms resulted in the production of a fluorescent protein (Chalfie *et al.*, 1994). This indicated that jellyfish specific enzymes were not required for fluorescence and led to the development of Gfp as a biological tool. Wildtype Gfp is a protein consisting of 720 amino acids and has peak emission and excitation wavelengths of 508nm and 395nm respectively (Tsien, 1998). Two major disadvantages exist in terms of using wild type *gfp* as a reporter gene. These are (i) that wildtype Gfp does not fold efficiently at temperatures above room temperature and (ii) that the maximal emission observed at 508nm is not entirely suitable for detection by existing laboratory equipment. For the second reason a number of red shifted Gfp proteins were developed that could be detected with fluorescein isothiocyanate (FITC) filter sets or the 488nm line (for excitation) of an argon laser which is used in a flow cytometer.

One such set of FACS optimised mutants were developed by Cormack *et al*, 1996 (Cormack *et al.*, 1996). Random mutations were made in a 20 amino acid region surrounding the chromophore to generate a library of 6×10^6 mutant genes in *E. coli*. Fluorescent activated cell sorting (FACS) was used to identify mutants that demonstrated increased fluorescence when excited at 488nm. Three types of mutants with shifted maxima were identified that fluoresced approximately 100-fold more than wild type Gfp at 488nm. Increased fluorescence was as a result of a red shifted absorption maxima (i.e. from 395nm for wtGfp to between 480 and 501nm for the three mutants) and increased amount of soluble protein as a result of more efficient protein folding.

A number of properties make *gfp* particularly suitable as a tool for biological studies. These include;

1. Stable fluorescence with almost no photobleaching observed

2. Unlike the majority of reporter proteins no co-factors (other than O_2) or substrates are required. It is therefore possible to perform expression studies in live cells in real time. This is quite different to the majority of reporter genes (i.e. CAT, luciferase, β-galactosidase) in which the cells need to be permeabilised for detection of the reporter protein or its reaction product.
3. Gfp maintains its fluorescence properties when it is part of a fusion protein, thus allowing the study of protein localisation without the need for generating antibodies.
4. Gfp fluorescence persists after formaldehyde fixation making it suitable for fluorescence microscopy
5. When using UV-optimised Gfp proteins it is possible to detect Gfp with a number of pieces of laboratory equipment including fluorometers, flow cytometers and fluorescence microscopes

1.8.5.2 Use of DFI approaches to identify differentially regulated genes.

DFI was first described in 1996. Since then the methodology has been used in 6 species of bacteria including 3 mycobacteria. Significant adaptations have been made from the original approach as outlined in Figure 1.1 to both the type of reporter plasmid that is used and the number of sorting steps utilised.

The first study to utilise DFI identified genes of *S. typhimurium* whose expression was up-regulated at pH 4.5 compared to pH 7 (Valdivia and Falkow, 1996). *S. typhimurium* is the causative agent of human gastroenteritis and murine typhoid fever. The rational for identifying acid induced genes is that *S. typhimurium* replicates within compartments of the mouse macrophage that are acidic. Therefore low pH may be a stimulus for the transcription of genes required for intracellular survival. Fluorescent bacteria were enriched from an acid treated population. The enriched population was then grown at neutral pH and the non fluorescent population enriched for. A final round of selection was performed on the acid treated population. A total of 8 acid inducible promoters were identified with induction varying between 2.6 and 17.1-fold, of which half also showed induction in a macrophage infection model. The genes of the identified promoters fell into three categories, (i) bacterial-cell surface structure and maintenance proteins, (ii) stress response proteins and (iii) generalised efflux pumps. A second related study by Valdivia *et al.*, (1997) utilised a mouse infection model to identify genes preferentially expressed when *S. typhimurium* associates with its host cell. Fourteen genes were identified of which

4 corresponded to known virulence factors. For 5 of the genes with unknown function, strains with disrupted genes were made, and tested in a mouse model. The mutations had varying effects on the bacteria's ability to colonise the spleen thus indicating that at least some genes regulated in such a manner may be important for virulence.

The DFI studies in mycobacteria are of particular interest because of the relatedness of corynebacteria to mycobacteria. Studies have been performed in *M. marinum*, *M. avium* and *M. tuberculosis* (Barker *et al.*, 1998;Triccas *et al.*, 1999;Ramakrishnan *et al.*, 2000). Tricas *et al.*, (1999) have modified the DFI procedure to identify macrophage induced genes of *M. tuberculosis*. In the first instance three rounds of enrichment inside the macrophage were performed. Secondly, instead of removing constitutive clones by a round of negative sorting a second reporter gene called *sacB* was utilised for *in vitro* selection. SacB is a secreted enzyme that confers sensitivity to sucrose, thus clones with constitutively expressing promoters could be removed in the presence of sucrose on agar plates. As in the standard DFI protocol a final round of positive selection in the macrophage was then performed. This enrichment was fairly ineffective as of 400 clones from the enriched collection only 7 showed reproducible induction ranging from 1.93 to 3.14 fold. Genes identified include those that may play a role in the metabolism of fatty acids and sulphur, a member of the PE-PGRS subfamily of the *M. tuberculosis* glycine/alanine-rich family and proteins with unknown function. A potential shortcoming of this and a number of other DFI studies is that regulation of gene expression has not been confirmed by a second method, for example northern analysis. Given that the fold induction observed was relatively small, confirmation of the data by a second method is important.

In a study of macrophage induced genes of *M. marinum* by Barker *et al* (1998) modifications were again made to the standard DFI protocol. Instead of sorting individual bacteria released from infected macrophages, phagosomal preparations were made and sorted. As phagosomes typically only contain 1 bacteria and its progeny this was seen as a way of avoiding false positives. Using the standard sorting process 4 of 400 individual clones identified showed reproducible regulation. In an attempt to increase the enrichment rate an additional sorting step was introduced at the start of the experiment that involved a negative sort of the entire library to remove constitutively expressing clones. Using such an approach the sorting efficiency was increased to 12 of 300 clones analysed individually. Thus the sorting efficiency of this experiment was similar to that observed in *M.*

tuberculosis. The fold induction ranged from 1.4 to 14.5 however most clones showed a 2-3 fold induction upon macrophage infection. Sequence analysis revealed a number of functions including drug efflux pumps, metallopeptidase and an arginyl-tRNA synthetase (Barker *et al.*, 1998).

The most recent use of DFI in a mycobacterial species demonstrates that DFI can successfully be used in an animal model. Genes of *M. avium* that show induction in macrophages and in frog granulomas were identified (Ramakrishnan *et al.*, 2000). Unlike the previously outlined *Mycobacterium* studies the sorting efficiency obtained was high, ranging from 100% to 17% depending on the sorting procedure. Clones containing homologues of PE-PGRS genes from *M. tuberculosis* were identified as being up regulated in both the macrophage and infection models. Interestingly a gene belonging to this family was also identified in the study performed by Triccas et al, (1999) looking at macrophage induced genes in *M. tuberculosis*. Mutant strains were generated for the PE-PGRS genes identified and were shown to be incapable of replication in macrophages and to demonstrate reduced persistence in frog granulomas (Ramakrishnan *et al.*, 2000). This study demonstrates that genes that show up-regulation inside macrophages may also be essential for growth *in vivo*.

1.8.6 Transcriptional profiling using DNA arrays

DNA arrays allow for the simultaneous measurement of relative gene expression under two or more sets of conditions for large numbers of genes. This has the potential to be an invaluable and powerful tool for molecular microbiology as it allows for the measurement of relative gene expression for the entire genome under any number of different growth conditions and time points. Additionally, array technology can be used to probe differences caused by changed genetic states, for example to determine how the inactivation of one gene affects expression of others in the genome. For all of the microbial genomes sequenced to date there are a large proportion of genes for which no function has been assigned. Array technology can be used to show which genes are regulated and this may in turn provide clues about their function.

1.8.6.1 Array technology

DNA arrays are generally termed macroarrays or microarrays, the term referring to the size of the DNA spots arrayed. Macroarrays contain relatively large DNA spots arrayed onto a nylon membrane, either robotically or by hand held array making devices (Schummer *et*

al., 1997). Typically a macroarray may contain between 200 and 5000 genes arrayed in an area of approximately 12x 8 cm. The classical microarray consists of up to 50,000 genes arrayed onto a glass microscope slide measuring 75x 25 mm(Lucchini *et al.*, 2001), although nylon membranes may also be used as the matrix for microarrays. A variation of the microarray is the high density oligonucleotide array in which oligonucleotides are arrayed instead of PCR products. Two types of oligonucleotide array exist. In the first oligonucleotides are first synthesised then arrayed in a manner analogous to PCR products; in the second oligonucleotides are synthesized directly onto the chip using a combination of photolithographic techniques and oligonucleotide chemistry (Lockhart *et al.*, 1996). This second type of oligonucleotide chip may have up to 64,000 "features" on a 1.28 cm^2 chip.

In the terminology suggested by Phimister (1999) the DNA arrayed onto the membrane or glass slide is considered the probe while the DNA or RNA used to probe the array is considered the target. A schematic of a DNA array experiment is outlined in Figure 1.2. The first requirement when performing a transcriptional profiling experiment is for an array. Partial or complete arrays for a number of microbial organisms are now commercially available. When all or part of the genome sequence for the organism of interest is known an ordered array may be prepared from PCR products or long oligonucleotides. In this type of array each spot corresponds to a single gene, hence the identity of each spot on the array is known. When the genome sequence is not available a random array may be prepared. For prokaryotic organisms this typically involves the preparation of a genomic library by shotgun cloning of fragmented genomic DNA into a suitable vector. DNA from clones is then spotted out onto the array or the inserts are amplified by PCR and subsequently arrayed. Important issues that must be considered with the arraying of random libraries include the size of insert cloned and the generation of a representative library. Large inserts are likely to contain several genes hence regulatory patterns may be obscured. If an array is to represent the entire genome, it may be necessary to have several times coverage meaning that there will be a sizable amount of redundancy in the system.

Arrays are typically used to identify changes in the expression profile of bacteria exposed to different conditions. This requires the isolation of RNA from the appropriate biological samples. Typically these are a control or reference sample and one or more experimental samples. Relatively large amounts of target are required (in the order of 25 µg total

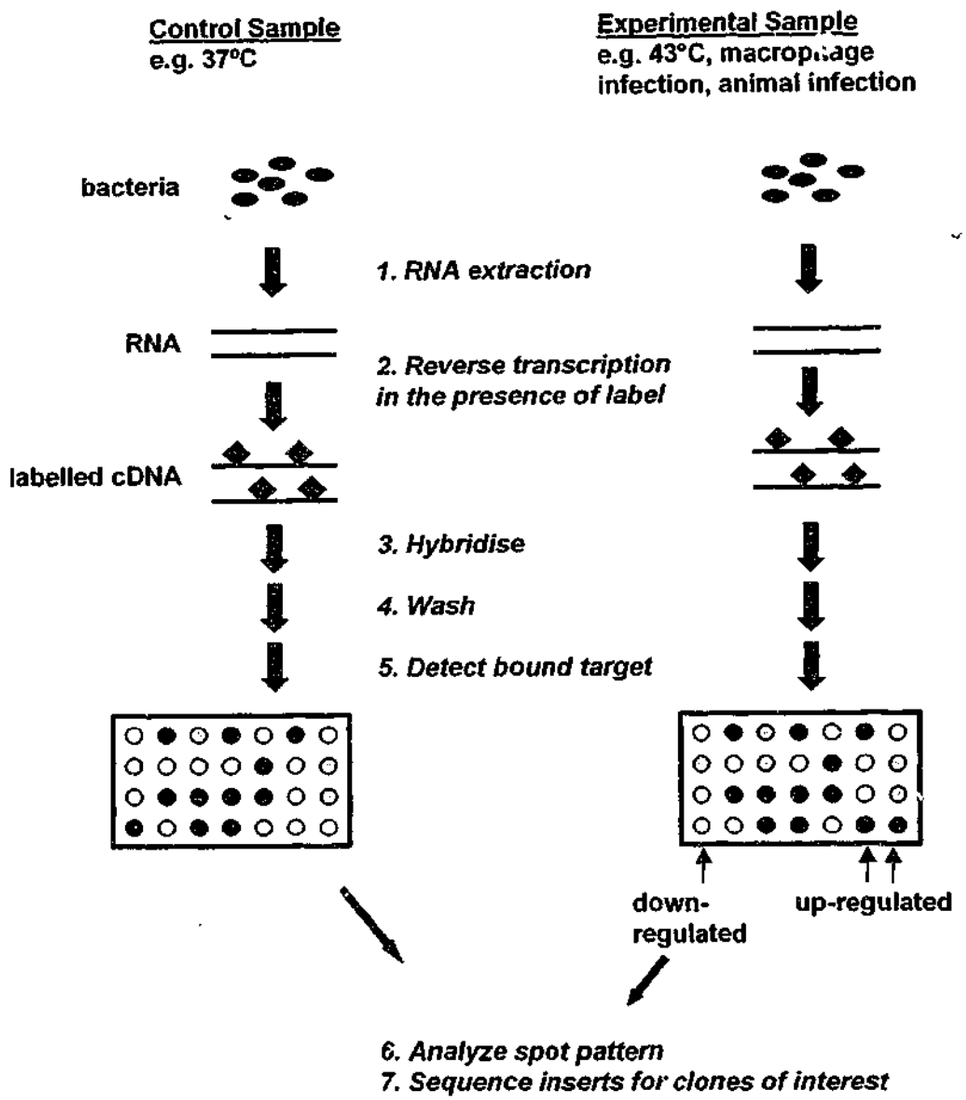

Figure 1.2: Schematic of transcriptional profiling using a DNA array approach.
(1) In an expression analysis experiment RNA is first isolated from control and experimental samples. (2) The RNA is reverse transcribed in the presence of an appropriate label to generate the target. (3) The labelled target is hybridised to the array(s). For macroarrays each target is hybridised to separate arrays (as shown) whereas for microarrays differently labelled targets are hybridised to the same array. (4) Following hybridisation the arrays are washed to remove non specifically bound target and (5) the remaining bound target detected with an appropriate system. (6) The relative intensities of the spots on the different membranes is measured and genes that are regulated identified. (7) For arrays of random libraries the inserts of clones of interest are then sequenced to provide functional information.

RNA). A number of alternatives exist for labelling of the target. Labels may be radioisotopic (usually ^{32}P or ^{33}P), fluorescent (i.e. Cy3 or Cy5 labelled dUTP) or non-isotopic alternatives such as DIG or biotin which have specialised detection systems (usually resulting in a colour, fluorescent or chemiluminescentt product). The media on which the DNA is arrayed in part determines what type of labelling system will be utilised. For example, radioactive targets are rarely used with glass chip arrays. In some cases the RNA may be labelled directly i.e. by using a cis-platinum coupled digoxigenin derivative (Hoevel *et al.*, 1999). However in most cases the RNA is reverse transcribed in the presence of a labelled dNTP to yield cDNA which has the added advantage of greater stability than RNA. Two types of primers are commonly used to reverse transcribe prokaryotic RNA. Random hexamers are commonly used but these will reverse transcribe ribosomal RNA as well as mRNA. When the genome sequence is known a collection of primers corresponding to each gene in the genome can also be utilised. The advantage of such an approach is that far less labelling reagent is required as rRNA is not reverse transcribed. A third alternative is genome-directed primers (GDP). Talaat *et al* (2000) have developed an algorithm that predicts the minimum number of primers to specifically anneal to all genes in a given genome. 37 primers were predicted for the *M. tuberculosis* genome. When compared to an array hybridised with cDNA generated by priming with random hexamers, the array probed GDP-derived target was more specific and more sensitive (Talaat *et al.*, 2000).

The manner in which hybridisation and washing steps are performed is again dependent on the type of array system being utilised. For macroarrays the methods are fundamentally the same as for traditional membrane protocols such as those used during northern or southern analysis. Macroarrays are typically incubated with hybridisation solution containing the target in a bottle, with rolling to ensure even dispersal of the target. Hybridisation is then followed by sequentially more stringent washes that serve to minimise background. Given that arrays for comparison are hybridised and washed in separate bottles it is important to ensure that identical conditions are applied to each filter to avoid inconsistent, poor quality data.

For gene expression analysis using microarrays a single array is probed with both the control and experimental samples. Because of the small area of the array, hybridisation can be performed in volumes less than 50 µL. As with traditional approaches, a variety of

hybridisation buffers, temperatures and wash conditions are utilised. Microarrays may be placed in hybridisation chambers that in turn are incubated in a water bath, otherwise hybridisation and washes may be performed in a flow cell in which the chips are anchored to the bottom of a sealed temperature controlled chamber through which solutions can be directed.

Data from the arrays is collected using appropriate systems, i.e. film for radioactive or chemiluminescent probes, confocal scanning microscopes or dedicated scanners for fluorescence detection on chip arrays. In the first instance data analysis is primarily concerned with determining fold induction or repression of genes from the experimental conditions compared to the control conditions. For arrays containing random libraries it is necessary to sequence the inserts of clones of interest at this stage. Analysis becomes more complex when there are two or more experimental conditions, for example multiple timepoints or treatments. In such cases computer aided clustering analysis can be perf rmed to identify groups of genes or "clusters" that demonstrate similar regulatory patterns across the experimental conditions. Correlation of cluster data with already obtained functional and protein data may provide insigh ts into why these genes are co-ordinately regulated. This could especially be the case for genes for which a function has not yet been assigned. This ability to identify groups of genes that belong to a given stimulon may be particularly beneficial during studies to identify virulence genes. Given that virulence gene expression is often co-ordinately regulated, it is likely that looking for genes that are regulated in a manner similar to already identified virulence genes may lead to the identification of new virulence genes.

1.8.6.2 Global transcriptional profiling and prokaryotes

DNA arrays have now been used to monitor changes in transcription for a number of bacterial species. As is the case for most molecular microbiology studies early experiments concentrated on *E. coli*. Global transcriptional analysis was first performed in *E. coli* by Chuang *et al*, (1993) using a macroarray containing overlapping lambda clones that spanned the entire genome of *E. coli* K-12. Arrays were used to study relative mRNA levels under a variety of growth conditions including heat shock, osmotic shock and nutrient starvation. Most genes that had been mapped to a certain phenotype were identified. Additionally new genes not previously assigned to a stimulon were also identified. With this type of approach it is likely that some regulated genes would not be

identified as the lambda clones contained large DNA inserts such that the 92% of the genome analysed was contained in approximately 450 clones. Given that *E. coli* has approximately 4,300 genes (Blattner *et al.*, 1997) this means that on average each clone contained 9 genes.

The first prokaryotes in which a global transcription profile was obtained at the single gene level were *E. coli* and *M. tuberculosis*. Early studies on *E. coli* used both macro and microarrays to study the heat shock response, response to IPTG (Richmond *et al.*, 1999) and the effect of growth media on gene expression (Tao *et al.*, 1999). In general the data obtained correlated very well with data generated through years of study of these responses. For example, of 119 genes identified as being heat regulated in *E. coli* all but 35 have been previously recognised as part of the heat shock stimulon (Richmond *et al.*, 1999). Microarray analysis has subsequently been used in *E. coli* to monitor changes in gene expression in response to changes in media composition (Khodursky *et al.*, 2000;Wei *et al.*, 2001;Oh *et al.*, 2002)and growth phase (Wei *et al.*, 2001), treatment with hydrogen peroxide (Zheng *et al.*, 2001), during superoxide stress (Pomposiello *et al.*, 2001) and to monitor changes in gene expression in mutant strains of bacteria (Arfin *et al.*, 2000;Campbell and Cronan, 2001).

A microarray approach has been used to identify genes of *M. tuberculosis* that have altered expression following treatment with the anti tuberculosis drug isoniazid (Wilson *et al.*, 1999). Isoniazid blocks the mycolic acid biosynthetic pathway by inactivating a type II fatty acid synthase (FAS-II) complex that is required for full-extension of the meromycolate chain. Using biochemical and genetic approaches it has been shown that some of the genes of the FAS-II complex are up-regulated in isoniazid treated *M. tuberculosis*. Using the array approach a cluster of five genes encoding type II fatty acid synthase enzymes and a trehalose dimycolyl transferase were up-regulated confirming previously obtained proteomic data and indicating that for these genes regulation occurs at the transcriptional level. A number of other genes not previously associated with sensitivity to isoniazid were also up-regulated. Given that the development of isoniazid resistant *M. tuberculosis* is a problem currently being faced by the medical community studies such as this one may provide other suitable targets for future drug development. This study also demonstrates how the targets of novel therapeutics could potentially be determined in a relatively easy manner.

Since 1999 global transcription studies have been performed for a number of other bacterial pathogens including *Vibrio cholerae* (Chakrabortty *et al.*, 2000). *Helicobacter pylori* (Ang *et al.*, 2001), *Bacillus subtilis* (Ye *et al.*, 2000;Yoshida *et al.*, 2001;Wiegert *et al.*, 2001;Helmann *et al.*, 2001), *Pasteurella multocida* (Paustian *et al.*, 2001), Group A *Streptococcus* (Smoot *et al.*, 2001), *Lactococcus lactis* (Even *et al.*, 2001) and *Streptococcus pneumoniae* (Rimini *et al.*, 2000). A variety of array approaches have been utilised in these studies including microarrays with fluorescent probes, macroarrays with radioactive probes and macroarrays with biotin labelled probes. Additionally arrays have been used to study genetic diversity in a number of species including *Mycobacterium* (Behr *et al.*, 1999), *H. pylori* (Israel *et al.*, 2001) and *Campylobacter jejuni* (Dorrell *et al.*, 2001).

1.8.6.3 Macroarrays versus microarrays

A comparison between nylon macroarrays and glass microarrays has been made using heat shock or IPTG treatment of *E. coli* K-12 as a model system (Richmond *et al.*, 1999). Nylon macroarrays were hybridised with radioactively labelled cDNA while microarrays were hybridised with fluorescently labelled cDNA. Both arrays contained PCR products corresponding to all 4290 annotated ORFs of *E. coli* K-12. Using the macroarray approach 119 genes were shown to be significantly up or down regulated by heat shock from 37°C to 50°C for 5 min. Of the 77 genes that were identified as being up-regulated 62 were also identified as up-regulated in the microarray approach. Fold induction was generally in the same order of magnitude but was somewhat variable between the two systems. Generally greater differences in gene expression were detected with the macroarrays. The microarray approach was, however, the more reproducible technique. One suggested explanation for the higher variability in the macroarray experiments is that separate membranes are used for each of the targets, therefore variability in spot deposition may affect hybridisation. This could potentially be overcome by performing sequential hybridisations to the same array.

A systematic study of different array approaches (Granjeaud *et al.*, 1999) (Baldwin *et al.*, 1999) has shown that a nylon microarray probed with a radioactive target provides the greatest sensitivity. Macroarrays with a radioactive probe and microarrays show a slightly reduced sensitivity while the use of enzymatic systems with colorimetric detection were

the least sensitive (Granjeaud *et al.*, 1999). The advantage of the nylon systems comes from their high capacity to bind probe DNA and hence target, in addition to the greater sensitivity associated with radioactive detection (Granjeaud *et al.*, 1999).

The other major difference between macroarray and microarray technology is the costs involved. Whilst the full implementation of a microarray system involves the purchase of expensive robotic and scanning systems for the manufacture and analysis of the chips, macroarray analysis can be performed relatively cheaply. This stems from the fact that arrays can be prepared with the aid of simple hand held arraying devices and pre-existing laboratory technologies can be used for array probing and analysis.

1.8.6.4 Future directions for array technology

Array technology is still in its infancy and as such most experiments that have been performed have studied already well understood phenomena. In general experiments have looked at gene regulation *in vitro* when one environmental condition is perturbed, for example heat shock or transfer from aerobic to anaerobic conditions. The application of array technology to *in vivo* derived bacteria should provide a wealth of data that will aid in the understanding of disease pathogenesis and has the potential to identify new vaccine candidates and therapeutic approaches. However, before such experiments can be performed, technology aimed at recovering relatively pure bacteria from infected animals in a manner that is quick and does not perturb gene expression needs to be further developed. Given the complex environment of the mammalian host this is likely to be more easily achieved for some bacteria than others. In the meantime useful data is likely to arise from the study of bacterial infection using tissue culture cell models.

An advantage of DNA array technology is that it can be used to look at gene expression in both the host and the pathogen using the same experimental approaches. Changes in mammalian host cell gene expression has been monitored following infection with a number of bacterial pathogens including *L. monocytogenes* (Cohen *et al.*, 2000), *Salmonella enterica* (Eckmann *et al.*, 2000) and *Pseudomonas aeruginosa* (Ichikawa *et al.*, 2000). The effect of a virulence factor of *S. typhimurium* on macrophage cell function has also been determined by comparing gene expression following infection with a wildtype strain and a strain in which the *phoP* gene is inactivated. Data demonstrated that PhoP plays a role in *Salmonella* induced human macrophage cell death (Detweiler *et al.*, 2001).

Studies such as this that lead to increased understanding of how the host modulates its response to infection may identify more effective ways to treat disease.

1.9 Project aims

Although the pathogenesis of CLA is relatively well understood, little is known about how this is controlled at the molecular level. On the basis of observations made in other bacterial species it is likely that at least some genes that are regulated *in vivo*, or in models that mimic aspects of infection, are intimately involved in the virulent phenotype. In order to increase the understanding of *C. pseudotuberculosis* pathogenicity the principal aim of this study was to identify genes of *C. pseudotuberculosis* that were regulated *in vivo* or in models that mimic aspects of the infection process.

The DNA sequence of 5 genes from the *C. pseudotuberculosis* genome was available in the Genbank database at the commencement of this study. Thus approaches that were based on knowledge of genome sequence could not be utilised. The first aim was therefore to develop approaches for the identification of differential gene expression that were not sequence based. The first was a DNA macroarray approach (Chapter 3) using an arrayed shotgun library. Given that complementary approaches for the identification of differentially regulated genes typically identify different subsets of genes, a second technique was also established. DFI was chosen as a contrasting methodology as it looks directly at promoter activity rather than at RNA levels (Chapter 4). Having established these technologies the second aim was to validate them in an environmental shock model. Models that mimic some aspects of *in vivo* infection are often used in studies of virulence gene regulation. This is done in part because it avoids the difficulties encountered in animal experimentation. Such studies have yielded a lot of useful information regarding the identification and study of virulence genes. Heat shock was chosen for this study as this response is very well characterised in a large number of bacteria and the response is likely to be similar in *C. pseudotuberculosis*.

The third principal aim was to identify genes that showed regulation following infection of macrophages. Given that there were likely to be enormous technical difficulties in extracting a viable single cell population of *C. pseudotuberculosis* from an infected sheep for flow cytometry or RNA extraction, a mouse macrophage model of infection was established instead. This is likely to closely mimic the interaction between macrophages

and *C. pseudotuberculosis in vivo*. The third aim was therefore to identify genes of *C. pseudotuberculosis* that show regulation upon infection of macrophages (Chapters 3 and 4).

The fourth aim was to determine whether genes identified as being up regulated during macrophage infection play a significant role in pathogenesis. It was anticipated that some of the genes that were up-regulated in this model would be essential for survival and replication within the macrophage. The approach taken to address this aim was to generate mutant strains of *C. pseudotuberculosis* in which genes of interest were functionally inactivated by insertion of an antibiotic resistance gene. Mutant strains could be tested in the macrophage infection model or animal model (Chapter 5).

The final aim arose during the study of the heat shock response using the macroarray approach. This study identified *pld* as being thermoregulated. This is the major virulence determinant of *C. pseudotuberculosis* identified to date, and no reports of its regulation have previously been made. In order to increase our understanding of the role of Pld in pathogenesis the final aim was to further investigate how *pld* was regulated and the significance of *pld* expression and regulation on macrophage function (Chapter 6).

Chapter 2

Materials and Methods

2.1 Chemical reagents

All aqueous samples were prepared using either double distilled or milliQ system purified water. Analytical grade reagents were used throughout. Prepared solutions were sterilised by autoclaving at 121°C for 20 min or by filtering through a 0.2 µM disposable filter (Sartorius AG, Gottingen, Germany).

2.2 Bacterial strains and plasmids

The *C. pseudotuberculosis* and *E. coli* strains used in this study are described in Table 2.1. and 2.2. respectively. Plasmids used in this study are listed in Table 2.3 (located at the end of Chapter 2).

2.3 Oligonucleotides

Oligonucleotides used in this study are described in Tables 2.4 and 2.5 (located at the end of Chapter 2). Oligonucleotides were custom synthesised by a number of sources including Invitrogen Corporation (Carlsbad, CA, USA), Geneworks (Adelaide, SA, Australia) and Applied Biosystems (Foster City, CA, USA).

2.4 Bacterial culture medium

All bacterial culture media were obtained from Oxoid (Basingstoke, Hampshire, England) or Difco (Detroit, MI, USA). Media was sterilised by autoclaving for 20 min at 121°C. When solid media was required, liquid media were supplemented with 1.5% (w/v) agar

Table 2.1: *C. pseudotuberculosis* strains used in this study

***C. pseudotuberculosis* strains**	**relevant characteristics**	**source (reference)**
C231	virulent wild type	Burrell, (1978)
Toxminus	*pld mutant* of C231	Hodgson, *et al*, (1992)
Toxminus +pTB111	complemented *pld* mutant of C231	Tachedjian, *et al*, (1995)
CPmet	*met* mutant of C231	this study
CPsodC	*sodC* mutant of C231	this study
CPfagC	*fagC* mutant of C231	this study

Table 2.2: *E. coli* strains used in this study

E. coli strains	Genotype	source (reference)
DH5-α	*deo*R, *end*A1, *gyr*A96, *hsd*R17(r_k- m_k+), *rec*A1, *rel*A1, *sup*E44, *thi*-1, Δ(*lac*ZYA-*arg*F)U169, Φ80*dlac*Δ(*lac*Z)M15, F-, λ-	Clontech
Top10 F'	F' [*lacI*^{q} Tn*10* (Tet^R)] *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *deo*R *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Stratagene
XL2 Blue MRF'	Δ(*mcr*A)183 Δ (*mcr*CB-*hsd*SMR-*mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F' *proAB lacI*^{q}Z*ΔM15* Tn*10* (Tet^r) Amy Cam^r]	Stratagene

prior to autoclaving. Routinely used media included Brain Heart Infusion (BHI) broth or agar (3.7% (w/v) BHI in H_2O), Luria-Bertani (LB) broth and agar (0.5% (w/v) yeast extract, 1% (w/v) tryptone, 1% (w/v) NaCl, pH 7.4) and Terrific Broth (TB) (1.2% (w/v) bacto-tryptone, 2.4% (w/v) bacto-yeast extract, 0.4% (v/v) glycerol, 17mM KH_2PO_4 and 72mM K_2HPO_4). Blood agar plates for assessment of phospholipase activity were prepared by the addition of defibrinated sheep blood and filtered culture supernatant of *Rhodococcus Equi* to LB broth to a final concentration of 10% (v/v) each.

When required, antibiotics were added to media at the following final concentrations: ampicillin 50 μg/ml, chloramphenicol 20 μg/ml and kanamycin 50 μg/ml. For blue-white selection solid media was supplemented with 200 μM isopropyl-β-D-thiogalactopyranoside (IPTG) and 10 μM 5-bromo-4-chloro-3-indolyl-β-galactopyranoside (X-gal).

2.5 DNA transformation techniques

2.5.1 Preparation and transformation of electrocompetent *E. coli*

Overnight cultures of *E. coli* were diluted 1/100 in 1 l of LB broth in a 2 l flask and incubated at 37°C in a shaking incubator until an OD_{600} of between 0.7 and 1 was reached. The bacterial suspension was cooled on ice for 30 min and then centrifuged at 4000 x *g* for 15 min at 4°C. The bacteria were washed sequentially in 1 l, 0.5 l then 30 ml of ice-cold

10% (v/v) glycerol. The resulting bacterial pellet was resuspended in 2-3 ml 10% (v/v) glycerol. 40 µl aliquots were frozen in liquid nitrogen and stored at -80°C.

10-100 ng of plasmid DNA was added to a thawed aliquot of electrocompetent cells. The mixture was transferred to a cold, 0.2 cm electroporation cuvette (BioRad, Hercules, CA, USA) and then electroporated using a BioRad gene Pulser set at 25 µF capacitance, 2.5 kVolts and 200 Ω. Following electroporation, 1 ml of LB broth was added to the cells and the bacterial suspension incubated with shaking for 1 hr at 37°C (to allow for recovery of the cells and expression of appropriate antibiotic resistance genes). Aliquots of the cells were then plated on LB agar containing the appropriate antibiotics.

2.5.2 Preparation and transformation of electrocompetent *C. pseudotuberculosis*

24 hr cultures of *C. pseudotuberculosis* were diluted 1/50 in 500 ml of BHI broth and incubated at 37°C in a shaking incubator until an OD_{600} of approximately 0.6 was reached. The bacterial suspension was cooled on ice for 30 min and then centrifuged at 8000 x *g* for 10 min at 4°C. The bacteria were washed twice with 250 ml ice-cold distilled H_2O and then washed once with 50 ml of 10% (v/v) glycerol. The resulting pellet was resuspended in 2-3 ml 10% (v/v) glycerol, aliquoted in 50 µl volumes and frozen in liquid nitrogen before storage at -80°C.

1-2 µg plasmid DNA was incubated with a thawed aliquot of electrocompetent *C. pseudotuberculosis* for 10 min on ice. The mixture was transferred to a cold, 0.2 cm electroporation cuvette (BioRad) and then electroporated using a BioRad gene Pulser set at 25 µF capacitance, 2.5 kVolts and 400 Ω. Following electroporation, 1 ml of BHI broth was added to the cells and the bacterial suspension incubated with shaking for 2 hr at 37°C (to allow for recovery of the cells and expression of appropriate antibiotic resistance genes). Aliquots of the cells were then plated on BHI agar containing appropriate antibiotics.

2.6 DNA purification techniques

2.6.1 Small scale purification of plasmid DNA from *E. coli*

The QIAprep® Spin Miniprep kit (Qiagen, Hilden, Germany) was used to prepare sequencing quality plasmid DNA preparations. The following protocol was used for

isolation of plasmid DNA for cloning, transformation and diagnostic purposes. Bacteria from 1.5-3 ml of overnight culture were collected by centrifugation. The bacterial pellet was resuspended in 0.5 ml 0.5 M NaCl, centrifuged and resuspended in 300 µl TE (50 mM Tris(hydroxymethyl)methane (Tris), 10 mM ethylenediaminetetraacetic acid (EDTA), pH 8) containing 50 µg/ml RNase A. 300 µl of Solution II (0.2 M NaOH, 0.1% SDS) was added and the suspension mixed gently by inverting the tubes several times causing lysis of the bacteria. After 2 min 300 µl of Solution III (2.55 M potassium acetate, pH 5.2) was added, mixed by inverting the tubes several times and then incubated on ice for 10 min. Following centrifugation the supernatant was transferred to a fresh tube. DNA was precipitated by the addition of 0.5 ml propan-2-ol and collected by centrifugation. The DNA pellet was washed in 0.5 ml 70% ethanol, air dried and then resuspended in 40 µl water.

2.6.2 Large scale purification of plasmid DNA from *E. coli*

For large-scale plasmid DNA preparations Qiagen plasmid Maxi and Qiagen plasmid Mega Kits (Qiagen, Hilden, Germany) were utilised as per the manufacturer's instructions.

2.6.3 Isolation of plasmid DNA from *E. coli* (96 well plate format)

2.6.3.1 Preparation of Beaut Broth

6 g tryptone, 12 g yeast extract, 6.28 g 3-(N-morpholino)-propanesulfonic acid (MOPS)-free acid, 1.5 g RNA (Type VI from Torula Yeast, Sigma) and 3 g DNA (sodium salt, from fish sperm, ICN, Aurora, Ohio, USA) were mixed together in a beaker. H_2O to 200 ml was gradually added to the powder stirring constantly such that a smooth paste was first formed followed by dissolution of the components. 2.1 g glycerol and 300 µl of 10 mg/ml RNase A was then added and the pH adjusted to 7.6 with 10 M NaOH. The volume was made up to 280 ml and the solution sterilized by autoclaving. Prior to use the media was supplemented with 15 ml 8% (w/v) KNO_3 (filter sterilized) and 5 ml 40% (w/v) glucose (filter sterilized).

2.6.3.2 Overnight cultures and DNA isolation

Overnight cultures were grown in 2.2 ml 96 Deep Well Plates (Advanced Biotechnologies Ltd, Epsom, Surrey, UK). A sodium lime glass ball (3 mm) was placed in the bottom of each well to aid in aeration of the cultures. 1.2 ml of Beaut Broth supplemented with

ampicillin (100 µg/ml) was added to each well and then inoculated from 96-well plate glycerol stocks using a flame sterilised 96-pin replicator. Plates were covered with an AirPore™ tape (Qiagen) and incubated at 37°C for 24 hr with shaking (300 rpm). Following centrifugation of the plates at 1250 xg for 5 min the supernatant was discarded and bacterial pellets were resuspended in 0.5 ml 0.5 M NaCl. The plates were covered with a deep well plate lid (Advanced Biotechnologies Ltd.) and vortexed until the bacteria were resuspended. The bacteria were again collected by centrifugation and then resuspended in 0.3 ml Solution I (TE supplemented with 50 µg/ml RNase A) by vortexing. To lyse the bacteria 0.3 ml Solution II (0.2 M NaOH, 1% (w/v) SDS) was added to each well, the plates were covered and then inverted several times to mix the contents. Following incubation for 10 min at room temperature 0.3 ml of Solution III (3 M potassium acetate, pH 4.8) was added. The plate contents were mixed by inversion and vigorous shaking. The plates were incubated at -20°C for 20 min and then centrifuged as above. A filtration plate (96 deep-well plate with a 2mm hole in the bottom of each well and a 1.5 cm x 1.5 cm piece of gauze wedged in the bottom of each well) was placed on top of a clean 96 deep-well plate. The filtration plate was pre-wet by the addition of 200 µl of 0.33 x SolutionIII to each well. The supernatants from the centrifuged plates were transferred to the filtration plate and allowed to drain through to the new plate underneath. A further 200 µl of 0.33 x SolutionIII was added to each well to ensure that the solutions were flushed through. 650 µl propan-2-ol was added to each well of the plate containing the filtered supernatant. A Deep well plate lid was placed over the plate and the plate contents mixed by rapid inversion. To pellet the DNA the plate was centrifuged as before. The supernatant was removed and 600 µl of 70% ethanol was added to each well. The Deep well plate lid was replaced, the contents mixed by inverting the plate several times and then centrifuged as before. The supernatant was removed, the pellets air dried and then resuspended in 100µl TE containing 125 U/ml RNase T1 (Life Technologies). The well contents were then transferred to shallow round bottomed 96 well plates (Costar, Cat# 3365) and stored at 4°C.

2.6.4 Isolation of plasmid DNA from *C. pseudotuberculosis*

Bacteria from 10 ml 24 hr cultures were collected by centrifugation and resuspended in STET buffer (8% (w/v) sucrose, 0.5% (v/v) Triton X-100, 4 mM EDTA, 10 mg/ml lysozyme, 0.75 mg/ml RNase, 1.6 mM Tris-HCl, pH 8). Following incubation at 37°C for 2 hr, 0.5 ml of Solution II (0.2M NaOH, 0.1%SDS) was added and the suspension mixed

gently by inverting the tubes several times allowing lysis of the bacteria. 0.5 ml of Solution III was then added and the resulting suspension centrifuged. The supernatant was transferred to a fresh tube and the DNA precipitated by the addition of 0.5 vol propan-2-ol. The DNA was collected by centrifugation, the resulting pellet washed once in 1 ml 75% (v/v) ethanol, air-dried and then resuspended in 20-40 µl water. As this protocol yields only small amounts of plasmid DNA, the DNA was always transformed into *E. coli*. Plasmid DNA was then isolated from *E. coli* cultures as described above.

2.6.5 Isolation of genomic DNA from *C. pseudotuberculosis*

This method is adapted from Zhang *et al.*, 1994. Bacteria from a 250 ml culture were collected by centrifugation and washed once in TE buffer (10 mM Tris, 10 mM EDTA, pH 8.7). The pelleted bacteria were then resuspended in 16 ml TE buffer supplemented with 10.3% (w/v) glucose and 10 mg/ml lysozyme. Following incubation at 37°C for 8 hr, 2 ml of 10% (w/v) SDS, 400 µl 5 M NaCl and 260 µl 20% (w/v) proteinase K were added. The resulting solution was mixed by inverting the tube several times and then incubated overnight at 56°C. Following centrifugation the supernatant was collected and extracted with an equal volume of phenol: chloroform: isoamylalcohol (25:24:1), and then an equal volume of chloroform. To precipitate the DNA an equal volume of propan-2-ol was added to the aqueous phase, incubated for 15 min on ice and then collected by centrifugation. The DNA pellet was washed once in 75% ethanol, air-dried and then resuspended in 2 ml H_2O. DNA concentration was determined by measuring the optical density at 260 nm. This protocol was scaled down when less starting material was available.

2.7 DNA sequencing

All DNA sequencing was performed by SUPERMAC (Sydney, NSW, Australia) or Newcastle DNA (Newcastle, NSW, Australia).

2.8 Manipulation of DNA

Restriction endonucleases and DNA modifying enzymes were obtained from Roche (Mannheim, Germany), Promega Corporation (Madison, Wisconsin, USA), New England Biolabs (Beverly, MA, USA) or MBI Fermentas (Vilnius, Lithuania).

2.8.1 Restriction endonuclease digestion

DNA was typically digested with 2-4 units of enzyme per µg DNA in a 20 µl reaction volume. Reactions were performed in 1 x restriction buffer with or without BSA (100 µg/ml) as recommended by the manufacturer. Reactions were performed at 37°C for 1-2 hr unless the manufacturer recommended alternative reaction conditions. When required, reactions were inactivated by heating at 65°C for 20 min.

2.8.2 Agarose gel electrophoresis

DNA fragments were separated according to size by horizontal gel electrophoresis in 1-2% agarose gels using a Tris-Acetate-EDTA (TAE) buffer system (40 mM Tris, 18 mM acetic acid and 1 mM EDTA). Agarose gels contained 350 ng/ml ethidium bromide to allow visualisation of the DNA under UV light. 0.1 vol 10x Sample buffer (0.42 % (w/v) bromophenol blue, 0.42% (w/v) xylene cyanol FF and 50% glycerol (v/v) in water) was added to samples before loading and DNA markers of the appropriate size range were run alongside DNA samples to allow size estimation of the DNA fragments. Samples were electrophoresed at 80-100 volts for 40-60 min. DNA was visualised on a UV transilluminator and images captured using a UVP camera and associated Grab-IT Annotating Grabber software (UVP, Upland, CA, USA).

2.8.3 Recovery of DNA fragments from agarose gels.

DNA bands were visualised on a UV transilluminator and excised using a scalpel blade. DNA was recovered from the agarose using either the Geneclean® II Kit (BIO 101, Carlsbad, CA, USA) or the QIAquick Gel Extraction Kit (Qiagen).

2.8.4 Ligations

In general ligations were performed with 10-100 ng vector DNA and vector: insert molar ratios between 1:1 and 1:5. When required linearised vector was dephosphorylated by treatment with Shrimp Alkaline Phosphatase (MBI Fermentas) or Thermostable Alkaline Phosphatase (Life Technologies) as recommended by the manufacturer. Ligation reactions contained 1 x ligation buffer (66 mM Tris-HCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 1mM ATP, pH 7.5) and 1 unit T4 DNA ligase (Roche or Promega) in a reaction volume of 10 µl. Ligation of sticky ends was performed at room temperature for 3-4 hr while blunt end

ligations were performed at 15°C, overnight. Reactions were heat inactivated by incubation at 70°C for 15 min and then transformed into *E. coli* or *C. pseudotuberculosis* as described earlier.

Ligation of *Taq* DNA polymerase-generated PCR products into pGEM®-T easy (Promega) was performed as per the manufacturer's instructions.

2.8.5 T4 DNA polymerase

T4 DNA polymerase was used to remove protruding nucleotides from the 3' termini of digested DNA. At the end of a restriction endonuclease reaction, 1 µl of 2 µM dNTPs and 2 units of T4 DNA polymerase were added, incubated at room temperature for 15 min and then heat inactivated at 75°C for 10 min. The DNA was then gel purified and used in a ligation reaction.

2.8.6 Mung bean nuclease

Mung bean nuclease (MBN) was used to blunt overhanging ends generated during sonication of *C. pseudotuberculosis* genomic DNA. 10 units of MBN were used per µg of DNA in a reaction volume of 7 µl. The reaction was performed in 1x MBN buffer (30 mM sodium acetate pH 5 at 15°C, 50 mM NaCl, 1mM $ZnCl_2$) supplemented with 3.75% (w/v) glycerol. The reaction was incubated at 37°C for 20 to 30 min and then inactivated by adding SDS to a final concentration of 0.01% (w/v).

*Sau*3AI digested genomic fragments were also treated with MBN. 400 ng of DNA was first digested with *Sau*3AI in a 10 µl volume. The reaction was then supplemented with $ZnSO_4$ to 1 mM and 2 units of MBN was also added. The MBN reaction proceeded at 30°C for 30 min and was then inactivated by the addition of SDS as above.

2.9 Amplification of DNA by the polymerase chain reaction (PCR)

For non-cloning PCR reactions EasyStart™50 tubes (Molecular Bio-products, San Diego, CA, USA) were utilised. These tubes contain PCR buffer, $MgCl_2$ and dNTPs. To these tubes were added primers to give a final concentration of 1 µM, 1 unit of Red Taq DNA

polymerase (Sigma), DNA template (genomic DNA, plasmid DNA or cDNA) and water to give a final volume of 50 µl. PCR reactions were performed in an AB Applied Biosystems DNA Thermal Cycler 480 and standard reaction conditions were 35 cycles of 95°C for 1 min, 55°C for 1 min and 72°C for 1 min per 1000 base pairs to be amplified. A final cycle with an extension time of 15 minutes was usually included.

For generation of PCR products for cloning a proof reading DNA polymerase was used. PCR was performed as per the manufacturer's instructions for *Pfx* DNA polymerase (Life Technologies) using the same cycling conditions as above. Alternatively *Tli* DNA polymerase (Promega Corporation) was used in the EasyStart™50 tubes with the addition of Triton X-100 to 5% (v/v). The PCR product was either used directly for cloning or was A-tailed and then cloned into pGEM®-T easy. To A-tail a PCR product it was first gel purified. A 10 µl reaction was then set up that contained 2 µl PCR product, 1x Red Taq buffer ,1 mM dATP and 2 units of Red Taq. The reaction was incubated at 70°C for 30 min. A portion was then ligated into pGEM®-T easy.

2.10 RNA isolation, manipulation and detection

2.10.1 Preparation of RNA solutions

To exclude RNase activity from glassware and solutions, all glassware was washed extensively with a detergent solution then distilled water. All solutions except those containing Tris were treated with 0.01% (v/v) diethyl pyrocarbonate (DEPC) for 16 hr, then sterilised by autoclaving. Those solutions containing Tris were prepared in DEPC treated water and then autoclaved.

2.10.2 Extraction of RNA from *C. pseudotuberculosis*

Bacteria from up to 5 ml of culture were collected by centrifugation, resuspended in 1 ml RNAzol™ B (Tel-Test, Inc, Friendswood, TX, USA) and then transferred to a 2 ml screw-cap tube containing 1 ml of 0.1 mm diameter glass beads (Daintree Scientific, Australia). The sample was homogenised in a Mini-BeadBeater-8 (Biospec Products) at maximum speed for 3 min. Following sedimentation of the beads by gravity the homogenate (approximately 0.5 ml) was transferred to a new tube. RNA isolation was then performed exactly as per the RNAzol™ B protocol. RNA was resuspended in DEPC treated water.

RNA concentration and purity was determined by measuring OD_{260} and OD_{280}. RNA quality was assessed by running 2 µg RNA on a 1.2% (w/v) agaose/TAE gel.

2.10.3 DNase treatment of RNA.

RNA was treated with RNase free DNase to remove trace DNA contamination. DNaseI was purchased from either Promega or Ambion and used as per manufacturers instructions.

2.10.4 Northern Analysis

2.10.4.1 Agarose gel electrophoresis of RNA

Agarose gel electrophoresis was performed under denaturing conditions as described in the Qiagen RNeasy Midi Handbook. This protocol is a modified version of that described by Sambrook *et al* (1989). A 1.2% (w/v) agarose gel was prepared in 1 x FA Gel buffer (20 mM MOPS, 5 mM NaAcetate, 1 mM EDTA, pH 7) supplemented with 220 mM formaldehyde and 100 ng/ml ethidium bromide and then equilibrated in 1x FA gel running buffer (1x FA Gel buffer, 246 mM formaldehyde). RNA samples were prepared by the addition of 0.2 vol of 5 x RNA loading buffer (4 mM EDTA, 886 mM formaldehyde, 20% (v/v)glycerol, 31% (v/v) formamide, 4 x FA gel buffer, bromophenol blue) and incubation at 65°C for 5 min. The samples were then cooled on ice and loaded onto the gel. The gel was run at 80V for 1.5-3 hr and then photographed.

2.10.4.2 Transfer of RNA to nylon membrane

To remove formaldehyde from the gel it was washed for 15 min in DEPC treated water, followed by two 15 min washes in 10x SSC (150 mM NaCl, 15 mM tri-sodium citrate). Transfer to Hybond N+ (Amersham Pharmacia Biotech, Buckinghamshire, UK) was performed in 20x SSC using capillary action as previously described (Sambrook, 1989). The RNA was fixed to the membrane by microwaving at 650 watt for 2 min.

2.10.4.3 Hybridisation and probe detection

Membranes were pre-hybridised in DIG Easy Hyb (Roche) for 30 min to 1 hr at 50°C in a rotary hybridisation oven. Up to 2 µg of gel purified PCR product was labelled with DIG-Chem Link from the Roche DIG-Chem Link Labelling and Detection Set. The labelled PCR product was denatured by incubating at 100°C for 10 min and then added directly to

the pre-hybridisation solution. The hybridisation reaction was allowed to proceed overnight under the same conditions used for pre-hybridisation.

The membrane was washed at low stringency in 2x SSC/0.1% (w/v) SDS at room temperature (2x 5 min) and then at high stringency 0.1x SSC/0.1% SDS at 68°C (2x 15 min). The probe was then detected as described in Section 2.11.4.3 for array probing.

2.10.5 Real-Time RT-PCR

2.10.5.1 Generation of cDNA by reverse transcription

Reverse transcription was performed using a TaqMan Reverse Transcription Reagents Kit (Applied Biosystems). 1 µg of DNaseI treated RNA was reverse transcribed in a 50 µl reaction containing 1x RT buffer (Applied Biosystems), 5.5 mM $MgCl_2$, 445 µM dNTPs, 2.5 µM Random hexamers, 20 U RNase inhibitor and 75 U Multiscribe reverse transcriptase (RT). The reaction was incubated for 10 min at 25°C, 45 min at 48°C and 5 min at 95°C. cDNA was diluted 1 in 50 in water prior to use in PCR reactions.

2.10.5.2 Real-time PCR

Real time PCR reactions were performed in a 96 well plate format in an ABI PRISM® 7700 Sequence Detector under universal cycling conditions (2 min at 50°C followed by 10 min at 95°C to activate the DNA polymerase, followed by 40 cycles of 15 sec at 95°C and 1 min at 60°C . Primers for real time PCR were designed using Primer Express® v1.5 software (Applied Biosystems) to have the following characteristics; melting temperatures of 58-60°C, G+C content of 20- 80%, no runs of more than 4 G, no more than 2 G or C bases in the 5 most 3' bases of the primer and an amplicon of between 50 and 150 base pairs. SYBR green 2x Master Mix (Applied Biosystems) which contains all the required components for a PCR reaction apart from primers and a DNA template was utilised. PCR reactions of 25 or 50 µl were performed in a 96 well plate format. Reactions contained 1x SYBR green master mix, 50 nM of forward and reverse primers and 5 or 10 µl diluted cDNA and were set up in duplicate or triplicate. Two negative controls were routinely performed. DNA contamination was determined by performing the PCR on negative control RT reactions (in which no enzyme had been added to the reverse transcription reaction). Environmental contamination was monitored in no template controls in which water was added to the reaction instead of DNA. Data were analysed using Sequence Detector v 1.7 software (Applied Biosystems). After completion of the PCR reactions a

dissociation curve was generated by melting samples at 60°C for 15 min, equilibrating them at 60°C for 20 min then slowly re-heating them to 95°C over 20 min. Dissociation curve data were analysed using Dissociation Curves Software (Applied Biosystems).

2.11 DNA Macroarray Techniques

2.11.1 Library array (colonies)

A library of genomic fragments ligated into pUC18 was transformed into *E. coli* and plated on LB agar supplemented with IPTG and X-gal to allow for selection of clones with inserts. White colonies were picked from the plates then patched into 96 well round bottomed plates containing 100 µl Beaut broth. The plates were covered with AirPore™ tapes (Qiagen) and incubated for 24 hr at 37°C. 50 µl of 60% glycerol (v/v) was added to each well. Plates were covered with a sealing tape, vortexed to distribute the glycerol and stored at -80°C. These plates were used to inoculate deep well plates as described in Section 2.6.3.2.

2.11.2 *Pvu*II digests of array DNA

*Pvu*II digests were performed on all array DNA samples to check for quality, quantity and insert size. Digests were performed in 96 well plates with each well containing 1x Buffer, 5 µl DNA and 2 units of *Pvu*II. Following digestion at 37°C for 2 hr samples were analysed on a 1% (w/v) agarose gel.

2.11.3 Array construction

60 µl denaturation solution (0.4 M NaOH) was placed into each well of a 384 well plate (120 µl/well, Nunc, International, Naperville, IL, USA). A 96 solid pin multi-blot™ replicator (V&P Scientific, Inc, San Diego, CA, USA) was washed in 70% ethanol and flame sterilised, dipped into a 96 well plate containing DNA and then transferred to the 384 well plate. The replicator was gently agitated to ensure dispersal of the DNA into the denaturation solution. This was repeated 3 times such that 4x 96 well plates were represented on each 384 well plate.

2.11.4 Arraying of DNA onto nylon membranes

A piece of nylon membrane was cut to the desired size and placed on a soft surface. A Multiprint perspex manifold (V&P Scientific, Inc) allowed for the alignment of 4x 384 spots within a space of 7x 11cm without overlapping of spots. The manifold was adhered to the membrane with double side tape. A 384 solid pin multi-blot™ replicator (V&P Scientific, Inc) was washed and flame sterilised then dipped into the wells of the 384 well plate containing DNA. Using the aligning pins (on the replicator) and holes in the manifold the replicator was aligned and then lowered onto the membrane. Pressure was gently applied across the top of the replicator to ensure that the DNA was transferred. This process was repeated for the remaining 3 positions on the array.

Once the solution had dried on the membrane the DNA was fixed by a 2 min incubation of the membrane on Whatman 3MM Chr paper that had been pre-wetted with 0.4 M NaOH/ 3 M NaCl. The membrane was then washed in 2x SSC for 5 min. The membrane was either used immediately in hybridisation experiments or dried and used at a later date.

2.11.4.1 DNA probes

In general purified PCR products were used as probes when a single gene detection was being performed. In other cases plasmid DNA or genomic DNA was used as the probe. In each case 2 µg of DNA was labelled with DIG-Chem Link from the Roche DIG-Chem Link Labelling and Detection Set as per the manufacturer's instructions. DNA probes were denatured by heating at 100°C for 5 min prior to addition to the hybridisation solution.

2.11.4.2 RNA probes

Total RNA was labelled with DIG using the Roche DIG-Chem Link Labelling and Detection Kit as per the manufacturers instructions. 1 µl of DIG-Chem link was used per µg RNA.

2.11.4.3 Prehybridisation and hybridisation

Membranes were pre-hybridised in DIG Easy Hyb (Roche) or Ultrahyb (Ambion) for 1 hr at 42°C or 50°C in a rotary hybridisation oven. Denatured probes were added directly to

the pre-hybridisation solution and the hybridisation reaction allowed to proceed overnight under the same conditions used for pre-hybridisation. Membranes were washed at low stringency in 2x SSC/0.1% (w/v) SDS at room temperature (2x 5 min) and then at high stringency 0.1x SSC/0.1% (w/v) SDS at 68°C (2x 15 min).

The bound DIG labelled probe was detected using the DIG-Chem Link Labelling and Detection Set. In brief, membranes were washed for 1-5 min in Washing buffer (Maleic Acid buffer (0.1 M maleic acid, 0.15 M NaCl, pH 7.5) supplemented with 0.3% (v/v) Tween 20) then blocked for 1 hr in Blocking solution (1 x Blocking Reagent (Roche) in Maleic Acid buffer). The membrane was then incubated with anti-DIG-AP conjugate diluted in Blocking solution (at between 1:10,000 and 1:100,000 depending on the application) for 30 min. The membrane was washed 2 to 3 times in Washing buffer (15 minutes each) and then equilibrated in Detection buffer (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5) for 5 min. The membrane was incubated with a small volume of alkaline phosphatase substrate (either CSPD or CDP-Star) for 5 min and then blotted on Whatman 3mm Chr paper, wrapped in plastic wrap and exposed to film.

2.12 Southern analysis

To transfer DNA from an agarose gel to a nylon membrane the gel was first soaked in 0.25 M HCl for 30 min to depurinate the DNA. After rinsing in water the gel was soaked in 0.4 M NaOH for 20 min to denature the DNA. The DNA was transferred to Hybond N^+ (Amersham Pharmacia Biotech, Buckinghamshire, UK) by capillary action as described by Sambrook *et al.*, 1989. DNA was cross-linked to the membrane by heating in a microwave oven for 2 min at 650 Watts. Southern blots were hybridised to DIG labelled DNA probes in a manner analogous to that described for DNA array hybridisation in Section 2.11.4.3.

2.13 Tissue culture techniques

2.13.1 Routine maintenance of cells

The mouse macrophage like cell line J774A.1 (ATTC No TIB-67) was grown at 37°C, 5% carbon dioxide in Dulbecco's modified Eagle medium (DMEM; TRACE Scientific, Australia) supplemented with 2 mM glutamine, 10% (v/v) foetal calf serum and 50 µM ß-mercaptoethanol. Macrophages were routinely grown in tissue culture flasks to 80-100% confluence. To passage the cells, the monolayer was washed twice in phosphate buffered

saline (PBS, 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HpO_4 and 1.5 mM KH_2PO_4), then incubated in a small volume of trypsin/EDTA (Gibco) for 5 min. The cells were detached from the flask by knocking it against a solid surface. To stop the activity of the trypsin FCS was added to at least 10% (v/v) . Cells were collected by gentle centrifugation (300 x *g*), resuspended in media and plated as required. When accurate cell numbers were required a haemocytometer in conjunction with trypan blue staining was used to determine the viable cell number. In brief an aliquot of cells was incubated with the supravital stain trypan blue (0.05% (w/v) final concentration) and then placed in the haemocytometer for cell counting.

2.13.2 Infection of J774 macrophages with *C. pseudotuberculosis*

An infection assay based on the gentamycin killing assay (Elsinghorst, 1994) was established. For the majority of experiments macrophages were plated in 6 well plates at $7x10^5$ cells/well 18 hr prior to addition of bacteria. For experiments in which fluorescence microscopy was the end point 22x 22 mm sterile glass coverslips were placed in the wells. Exponential phase bacteria (grown in BHI broth supplemented with 1% (v/v) Tween-80) were sonicated, diluted in pre-warmed DMEM and added to the macrophage monolayer. Following incubation for 1 hr the monolayer was washed 3 times with 37°C phosphate buffered saline (PBS) and then incubated with DMEM containing 100 µg/ml gentamycin. At further timepoints analysis of the infection could be analysed in a number of ways as described in sections 2.13.3 to 2.13.7.

In some experiments phagocytosis of the bacteria was prevented by incubation of the monolayer at 4°C for 20 min prior to addition of the bacteria, cooling all solutions to 4°C and performing all incubations at 4°C until 3 hr post addition of gentamycin containing media at which time the cells were transferred to 37°C.

2.13.3 Fluorescence microscopy

For microscopy, infected cells, grown on glass coverslips, were washed in PBS, fixed for 5 min in 3.7% (v/v) formaldehyde in PBS and then permeabilised by incubation in 0.1% (v/v) Triton X-100 for 5 min. The macrophages were counterstained with 0.4 µg/ml propidium iodide (PI) solution and then mounted onto glass slides using Dako fluorescent mounting medium. Slides were examined by fluorescent microscopy using an Olympus

BH-2 microscope. Photographs were taken with an Olympus C-35AD-4 camera using Fujichrome MS 100/1000 film.

2.13.4 Recovery of bacteria from macrophages for viable cell counts

Infected macrophages were washed in PBS. 0.5 ml 1% (v/v) Triton X-100 (in PBS) was added to each well and incubated for 5 min. Cells were dislodged with a cell scraper and 4 ml of PBS added. The contents of the well were transferred to a 5 ml tube and sonicated to disrupt bacterial clumps. The sonicate was serially diluted and plated on BHI agar. After 2 days at 37°C colony counts were performed to determine CFU/recovered per well.

2.13.5 Extraction of bacterial RNA from infected macrophage monolayers

All solutions were cooled on ice prior to use. Macrophages were washed in PBS then incubated with 1% TritonX 100 (in PBS) for 5 min. A cell scraper was used to dislodge cells from the plastic and the suspension was then transferred to a centrifuge tube. Following centrifugation (1400 x *g*, 5 min, 4°C) the supernatant was removed and 1 ml RNAzol™B added to the pellet. RNA was then extracted in the usual manner as outlined in Section 2.10.2.

2.13.6 Preparation of macrophage sonicates for flow cytometry

Infected macrophages were washed in PBS. 0.5 ml PBS was added to each well and the cells dislodged with a cell scraper. The contents of the well were transferred to a 1.5 ml tube and sonicated to release the bacteria and disrupt bacterial clumps. Sonicated samples were diluted in PBS then analysed by flow cytometry.

2.13.7 Preparation of infected macrophages for flow cytometry and FACS

Infected macrophages were washed in PBS then removed from the plastic by trypsinisation. To remove clumped macrophages the cell suspension was filtered through a 70 µM filter. The filtrate was diluted in PBS and then used in sorting experiments.

2.14 Flow cytometry and fluorescence activated cell sorter (FACS) analysis

Flow cytometry and FACS analysis was performed using a Becton Dickerson FACSCalibur machine using standard excitation and emission filters (530/30nm) with an argon laser emitting at 488nm. Data was collected and analysed using Cellquest software. 0.9% (w/v) NaCl was used as sheath fluid.

2.14.1 Preparation of *C. pseudotuberculosis* cultures for flow cytometry

To prevent clumping, cultures of *C. pseudotuberculosis* were grown in BHI supplemented with 1% (v/v) Tween-80. In general, for flow cytometry of *C. pseudotuberculosis* strains, 5 ml overnight cultures were cooled on ice, sonicated (XL-series sonicator, Heat using a standard microprobe set on 4) to break up clumps of cells, diluted to an OD_{600} of 0.1 and grown at 37°C. Prior to analysis cultures were sonicated again and diluted up to 500-fold in PBS.

2.14.2 FACS sorting of *C. pseudotuberculosis* and infected macrophages

Given the propensity of *C. pseudotuberculosis* to aggregate, bacteria were gated on the basis of log-scale forward and side scatter in addition to log-scale fluorescence intensity (log FL-1). This resulted in primarily single bacteria being sorted. In order to maximise the sorting efficiency the flow rate was maintained at less than 500 events/sec. The sorted population was collected in 50 ml polypropylene tubes coated with FCS. To recover the sorted population the tubes were centrifuged (1400 x *g*, 5 min, 4°C), the supernatant disgarded and the pellets resuspended in a small volume of BHI prior to plating on BHI agar containing 50 µg/ml kanamycin.

The same general principles with minor modifications were used to sort infected macrophages. Forward and side scatter data were collected on a linear scale instead of a log scale. The sorted population was sonicated prior to plating on agar plates in order to lyse the macrophages and disrupt bacterial clumps from within the macrophages.

Table 2

Plasmid
pACYC
pKEN1
pEP2
pEP2 M
pKK23
pRM13
pSM1
PSM2
pSM3
pSM4
pSM5
pSM11
pSM12
pSM13
pSM15
pSM16
pSM17
pSM18
pSM20
pSM21
pSM22
pSM23
pSM27
pSM28
pSM29
pSM30
pSM31
pSM32
pSM33
pSM39
pSM40
pSM4
pSM42

table

Table 2.3: Plasmids used in this study

Plasmid	Relevant Characteristics	Reference
pACYC184	chloramphenicol^R, tetracycline ^{R}	Rose, 1988
pKEN1*gfp*uv	Amp^R, pKEN1 backbone, source of FACS optimised mutant with *gfp* gene substitutions at F64L and S65T (Cormack), *srp* pomoter	A. Ryecroft (unpublished)
pEP2	kan^R, E. coli- *C. pseudotuberculosis* shuttle vector	Zhang *et al.*, 1994
pEP2 MCS	kan^R, derived from pEP2, used as the basis for promoter probe vector construction	R. Moore, unpublished
pKK233-2	source of rrnBT1T2 transcription terminator	Amann and Brosius, 1985
pRM130	kan^R, *pld* promoter	R. Moore, unpublished
pSM1	pEP2 MCS derivative, contains new MCS	this study
PSM2	pSM1 derivative, contains RBS	this study
pSM3	pSM2 derivative, *gfp*$^+$	this study
pSM4	pSM3 derivative, *trp*A transcription terminator	this study
pSM5	pSM4 derivative, *rrn*B T1T2 transcription terminator	this study
pSM11	pSM12 derivative containing *gfp* under control of *pld* promoter	this study
pSM12	pSM5 derivative containing correct *gfp* sequence	this study
pSM13	pSM12 derivative *gfp* under control of *srp* promoter	this study
pSM15	pSM12- pKEN1*gfp*uv fusion	this study
pSM16	pSM11 derivative containing RBS from pKEN1*gfp*uv	this study
pSM17	pSM13 derivative containing RBS from pKEN1*gfp*uv	this study
pSM18	pSM16 derivative without *pld* promoter	this study
pSM20	pSM18 derivative, promoterless, fusions allowed	this study
pSM21	pSM20 containing *gfp* under control of *srp* promoter	this study
pSM22	pSM20 containing *gfp* under control of *srp* promoter	this study
pSM23	pSM20 containing *gfp* under control of *dnaK* promoter	this study
pSM27	pSM20 containing *gfp* under control of *pld* promoter	this study
pSM28	pSM20 containing *gfp* under control of *pld* promoter	this study
pSM29	pSM20 deriviative in which *gfp* has been replaced with a destablised *gfp*	this study
pSM30	pSM29 containing destabilised *gfp* under control of *srp* promoter	this study
pSM31	pSM29 containing destabilised *gfp* under control of *pld* promoter	this study
pSM32	pSM27 derivative, *pld* promoter cutback	this study
pSM33	pSM27 derivative, *pld* promoter cutback	this study
pSM39	pSM46 derivative, *pld* promoter cutback	this study
pSM40	pSM46 derivative, *pld* promoter cutback	this study
pSM41	pSM46 derivative, *pld* promoter cutback	this study
pSM42	pSM46 derivative, *pld* promoter cutback	this study

table continues on next page

Plasmid	Relevant Characteristics	Reference
pSM43	pSM20 derivative without RBS and start codon of *gfp*	this study
pSM46	pSM43 derivative containing the *pld* promoter	this study
pSM48	pUC18 derivative containing kan^R	this study
pSM50	pSM48 derivative containing 5' *met* sequence	this study
pSM51	pSM50 derivative containing 3' *met* sequence, knockout cassette	this study
pSM52	pSM48 derivative containing 5' *sodC* sequence	this study
pSM53	pSM52 derivative containing 3' *sodC* sequence, knockout cassette	this study
pSM54	pSM48 derivative containing 5' *fagC* sequence	this study
pSM55	pSM54 derivative containing 3' *fagC* sequence, knockout cassette	this study
pTB111	kan^R, derived from pEP2 derivative containing *pld* under control of its own promoter	Tachedjian *et al.*, 1995
pUC18	Amp^R	Norrander *et al.*, 1983

Table 2.4: Primers used in this study

Table 2.4: Primers used in this study

Primer #	Sequence (5'→3')	Features	Gene and expected product
3	ccgaccggtcatgagtaaaggagaagaac	*AgeI*	forward, *gfp*
4	ccggggcccggcgccactagtttatttgtatagtttatc	*Spe*I, *Nar*I, *Apa*I	reverse, *gfp*. Use in conjunction with #3 to amplify *gfp* gene.
4b	ccggggcccggcgccactagtttatttgtatagttcatc	*Spe*I, *Nar*I, *Apa*I	reverse, *gfp*. Use in conjunction with #3 to amplify *gfp* gene.
5	ggtatcgacttggggactaccaactc$_5$g		*dnaK*, forward
6	gctgtaggttcgttgatgatacgc		*dnaK*, reverse. Use in conjunction with #5 to amplify an internal region of the *C. pseudotuberculosis dnaK* gene.
9	ggcgccgaactcagaagtgaaacg	*Nar*I	*rrn*B T1 T2 region, forward
10	ggcgccagggttattgtctcatgagc	*Nar*I	*rrn*B T1 T2 region, reverse. Use in conjunction with #9 to amplify the rrnB T1 T2 region of pKK233-2.
18	ccatgctccgcatcccttgcat		forward, *pld* promoter. Use in conjunction with #75 and #76 to amplify the *pld* promoter
18b	gatgcatgcccatgctccgcatcccttgcat	*Sph*I	forward, *pld* promoter. Use in conjunction with #108 to amplify the *pld* promoter
47	cgatatcggatcctctagatttaagaaggagatatactatgagtaaagg	*Eco*RV, *BamH*I, RBS	forward, for conversion of pSM18 into a fusion vector. Use in conjunction with #4b to amplify *gfp* gene of pSM18
48	ggatgcatgcctggcggaaatagaaga	*Sph*I	forward, *dnaK* promoter
49	gcggatccgggtgtttggttgtatttttaacg	*BamH*I	reverse, *dnaK* promoter. Use in conjunction with #48 to amplify the *C. pseudotuberculosis dnaK* promoter.
55	cgaagatcagcggtga		reverse *pld*,use in conjunction with #77 to generate a probe for Northern hybridisation

table continues on next page

Table 2.4

75	gcggatccactttctcctcatttttac	*BamH*I	reverse *pld* (from 4th codon), for amplification of *pld* promoter
76	gcggatccgtttcgataaatagagcaac	*BamH*I	reverse, *pld* promoter
77	gatgcatgcgctatggataccctaaaag	*Sph*I	forward, *pld* promoter. Use in conjunction with primers #75 and 108
95	ctctcactagtattaagctactaaagcgtagttttcgtcgtttgctgca ggcctttgtatagttcatccatgcca	*Spe*I site, LVA "instability domain"	reverse, *gfp*. Use in conjunction with #47 to generate a destabilised *gfp* gene
98	catatgtaatgcctgattg		forward, *pld* promoter, use in conjunction with primer #75
98b	gatgcatgccatatgtaatgcctgattg	*Sph*I	forward, *pld* promoter, use in conjunction with primer #108
100	gcaataaccctgattt		forward, *pld* promoter, use in conjunction with primer #75
100b	gatgcatgcgcaataaccctgattt	*Sph*I	forward, *pld* promoter, use in conjunction with primer #108
108	gcggatcctttctccctcatttttatc	*BamH*I	reverse *pld* (from 4th codon), for amplification of *pld* promoter
109	gatgcatgcgtccttaagttcaaaaacct	*Sph*I	forward, *pld* promoter, use in conjunction with primer #108
110b	gatgcatgccgtatctgacgcatttgaatacat	*Sph*I	forward, *pld* promoter, use in conjunction with primer #108
227	cgcggatccgctagcgagctctagatcttgcag	*Bam*HI, *Nhe*I	forward, kan resistance gene
228	gcgctgcaggctctgccagtgttacaac	*Pst*I	reverse, kan resistance gene. Use in conjunction with #227
229	cgcgaattccctacacaggtgtggatt	*Eco*RI	forward, *met*
230	gcgggatccactactgaataggagt	*Bam*HI	reverse, *met*. Use in conjunction with #229 to amplify 5' region of *met*
231	cgcctgcagactagttaacggaagacgagttgga	*Pst*I, *Spe*I	forward, *met*

table continues on next page

Table 2.4

232	caegcatgcgtccgcatcgctcaaaagc	*Sph*I	reverse *rnlB*. Use in conjunction with #231 to amplify 3' region of *met* and 5' region of *rnlB*
254	gcggaattctcgcagtcctgagcgcttgt	*Eco*RI	forward, *sodC*.
248	gcgggatcccccggtcttgagagtctcctc	*Bam*HI	reverse, *sodC*. Use in conjunction with #254 to amplify *sodC*
249	gcgctgcagcgatgccggcgcacgcttlg	*Pst*I	forward, *sodC*
250	gcggcatgccgacaacggacggcgagat	*Sph*I	reverse, *accD*. use in conjunction with #249 to amplify 3' end of sodC and 5' region of *accD*.
257	gcggaattcgagtcctcctcattgcgacca	*Eco*RI	forward, *fagB*,
258	gcgggatcccccatacgcgctgacgctgt	*Bam*HI	reverse, *fagC*. Use in conjunction with #257 to amplify 3' end of *fagB* and 5' end of *fagC*
259	gcgctgcagcttgcaatggttctagcccaaa	*Pst*I	forward *fagC*
260	gcggcatgcgtgtgctccataaatgcgtt	*Sph*I	reverse *pld*. Use in conjunction with #259 to amplify 3' end of *fagC* and 3' end of *pld*
261	actactgcagggatgctcgt		reverse *pld*. For screening *fagC* mutant strains. Use in conjunction with #257
276	gcggaattcgagctcggtacccggggatcctctagatttaagaag gagatatactatgactaggccacggttgcgttct	*Eco*RI, sequence from pSM22 (from *Eco*RI site to base prior to *gfp* start)	forward, *met*
277	cgcactagtaatcacaccttacggggcattt	*Sph*I	reverse *met*. Use in conjunction with #276 to amplify the *met* gene
278	cgcatcgatcacttcccgttaagtatcttc	*Cla*I	forward, *cat* gene from pACYC184.
279	cgcatcgattttgctttcgaatttctgccat	*Cla*I	reverse, *cat* gene from pACYC184. Use in conjunction with #278 to amplify *cat* and its promoter

Table 2.4

75	gcggatccactttctcccctcatttttac	*BamHI*	reverse *pld* (from 4th codon), for amplification of *pld* promoter

Table 2.5: Real time PCR primer pairs.

gene	forward primer	sequence (5'→3')	reverse primer	sequence (5'→3')	Clone[a]
pld	#101	gattgcccaccgcgttt	#102	tcgcaccgatcgcaact	L16587
16s rRNA	#103	cctgtataagaagcaccggctaa	#104	acgctcgcaccctacgtatt	X84255
dnaK	#111	cccgctgttcttggatgaga	#112	tcaagcagatcctgggtga	Q4-E5
maltose ABC transporter permease	#113	ttctggcgtgcgatttct	#114	tgcgcaacacgaggagagta	Q3-C1
aspartokinase	#115	agaacgtgctttacgacgat	#116	cctgggtgcgatttcatacc	Q4-E5
fagC	#117	cacaaacgcgtgaccgatt	#118	agaaccattgcaagccatacg	Q2-G12
cytochrome c1	#119	tgctaactacaacggaaagatcga	#120	gcgcaatttcaaacggaagag	Q4-C2
met	#121	gcaggccggcattcag	#122	gggttagaggcttctttgaggatt	Q12-E9
fas	#127	cggcggacgtagatctgtct	#128	cgaatgggcgggctactag	Q4-A3
rieske iron-sulfer protein	#129	gctagcccgtttgggtacag	#130	cgccttcgaccggaaaa	Q4-C2
hypothetical protein	#133	aactgagcgtcgcgcttt	#134	catacttgagccccaatttcct	Q7-B6
hypothetical metabolite BP	#135	gggcactcggctggatt	#136	tggcggaaccactaattacaaag	Q7-D6r
clpB	#137	gcatcatggtgtgcgtattca	#138	cgtgatatagcgatccgagaga	Q7-C2
dnaJ	#139	tgatgagttcaaggccatgatc	#140	cgcctccgttctgtccaa	Q12-C5
hspR	#141	cccgctcggacgtagaacta	#142	ccaggttaacgccttcttctg	Q12-C5
dtsR	#143	ttgttgatgttgccgcactta	#144	tgctttgaagtccgcgatct	C18
NO reductase	#145	cagcgttgggtgtggaaaa	#146	cgatgcgtccgtcttggta	C19
lpdA	#147	tgttcccgctggcttctg	#148	ctccaagcgcgtggtacat	Q11-F6

table continues on next page

homoserine kinase	#149	gtccgcagggtccttcct	#150	cacctggacacgttgaatcg	Q10-G2
					Q10-G2

lpdA	#147	tgttccgctggcttctg	#148	ctccaagcgcgtggtacat	Q11-F6

table continues on next page

Table 2.5

homoserine kinase	#149	gtccgcagggtccttcct	#150	cacctggacacgttgaatcg	Q10-G2
phosphoenolpyruvate sugar phosphotransferase	#151	aacgtacaaatgcgtcaggttc	#152	tcagagattccgcccagaag	Q10-G2
pfl	#153	tgagcactgggctcgttacc	#154	gcagcggccatgatcttc	Q10-B8
sodC	#155	tgaaggccgcgacaactac	#156	gcatcgccggtcttgaga	Q15-A3
acetyl CoA carboxylase	#157	atgcgcatcttccgttttt	#158	ccccacgaagccattactcc	Q15-A3
anion exchanger	# 164	tttaaggcccgaggcaaac	#165	tcatggtgaaagatccgagactt	Q10-B11f
hexose phosphate transport protein	#166	ttcgtattcccggctctca	#167	cgtccttgcccacaaacag	Q8-F11
translation initiation factor IF3	#168	gcagtctcgcccggaacta	#169	gccaaaatctgcgacatcatc	Q2-D10
ribosomal protein L11	#170	cctcaacgctaacgacatcga	#171	cccatggaacgggcagtac	Q2-G5
possible transcription factor	#172	ggaatgaccaaagagaggaaacc	#173	ccgtttggagaggcatttgt	Q2-B10
ABC-type cobalamin siderophore component	#174	tcgattgccaagggttatgc	#175	tgataggtggcacctgctgtag	Q1-D10
Rv2036c (stress family protein)	#188	ccgtgtccgatggtggtagt	#189	gccggctatgcaagcaa	Q16-A7r
hyp protein	#190	ctgggcagtcaatcactcgat	#191	caagcctgcaccggttttaa	Q13-B9f
aminopeptidase	#192	gggctaactccaggacagcat	#193	agtccctgaccggtatgtgaat	Q13-B9r
chrA	#200	ccacctggtacacatctactcaaaa	#201	ccgggctgcggtgat	Q15-G3f
amino acid carrier protein	#202	gcgcaaccgcttttgtg	#203	gccatccgcagttttggt	Q13-A4
cps	#267	caatctacggcgtcgatacca	#268	ttacggtgtcgctgaaaatatca	C13
dnaN	#269	agctgggttgcgcgtaac	#270	gaccttcatcatcggctgtga	C3

[a] Primer pairs were designed on the basis of the sequence found in the listed clone. Alternatively an accession number is listed for those genes whose sequences are in the NCBI databases.

Chapter 3

Identification of differentially regulated genes using DNA arrays

3.1 Introduction

The aim of the work described in this chapter was to use DNA array technology to identify up-regulated genes of *C. pseudotuberculosis* that would form candidates for further study. As discussed in Chapter 1 this aim is based on the premise the genes up-regulated in the *in vivo* environment or in models that mimic aspects of it may be intimately involved in the virulent phenotype.

Transcriptional profiling using array technology is a powerful methodology that can be used to simultaneously measure gene expression for a large number of genes. This type of approach is particularly useful when the genome sequence of an organism is known, as non-redundant DNA arrays may be constructed on the basis of one spot equals one gene. Given the lack of genome sequence for *C. pseudotuberculosis*, this type of approach was obviously not possible. In this study a *C. pseudotuberculosis* random clone DNA array was constructed that could be used as a screening tool to identify differentially regulated genes rather than to precisely monitor all changes in the total transcriptome. Macroarrays on nylon membranes were chosen as they could be established using standard laboratory procedures, without the requirement for expensive robotics equipment. In addition to whole genome transcriptional profiling DNA arrays may be used in other ways. DNA arrays have been used to study gene regulation in a small subset of genes involved in a particular function (Hu and Coates, 2001) or in conjunction with other screening technologies such as subtractive hybridisation (Li *et al.*, 2001) or RNA arbitrarily primed PCR (Trenkle *et al.*, 1998). Additionally arrays have be utilised to study genomic diversity (Behr *et al.*, 1999;Israel *et al.*, 2001;Dorrell *et al.*, 2001).

Two models were chosen for identifying differentially regulated genes of *C. pseudotuberculosis*. The first was a heat shock model. This was established in order to demonstrate the validity of the approach for identifying differentially regulated genes of *C. pseudotuberculosis*. A heat shock model was chosen because for many bacterial species the heat shock response is very well defined. The response is conserved across bacterial species and as such similar results would be expected for *C. pseudotuberculosis*. Using array technology the heat shock response has been measured in *E. coli* (Richmond *et al.*, 1999), *B. subtilis* (Helmann *et al.*, 2001) and *N. meningitis* (Guckenberger *et al.*, 2002). Although the type and duration of the heat shock has varied with each experimental system these studies

have shown that a subset of genes undergo changes in gene expression. This typically includes genes that encode proteins termed heat shock proteins, such as those of the *dnaK* and *groEL* operons. The array studies have also identified genes that more conventional scientific approaches have not previously shown to be part of the heat shock regulon.

To date there has been a single report of using transcriptional profiling to identify genes of bacterial pathogens that show changes in gene expression during infection of a host (Chuang *et al.*, 1993). This lack of studies is likely to be due to the complexities of isolating bacterial RNA from host tissues in a sufficient quantity for array probing. Additionally, methods that protect the RNA from degradation have only very recently been developed (Florell *et al.*, 2001). The use of tissue culture models of infection is a widely established practise in the field of microbiology, in which they are used to mimic the interaction between a pathogen and its host cell type. Both primary cell cultures and immortalised cell culture lines have been utilised. Given that *C. pseudotuberculosis* is an intracellular pathogen that primarily resides within macrophages, a macrophage-like cell line was chosen for the establishment of an infection model. Two macrophage-like cell lines that are commonly used to study the interaction between the host and the pathogen are RAW 264.7 and J774. J774 macrophages, which were utilised during this study, were derived from a sarcoma of a female BALB/c mouse and display typical macrophage characteristics (Ralph and Nakoinz, 1975). J774 have been used in a diverse range of host-pathogen studies including intracellular survival studies (Valentin-Weigand *et al.*, 1996;Kuehnel *et al.*, 2001), gene regulation studies (Freitag and Jacobs, 1999;Wilson *et al.*, 1997) and identification of differentially expressed genes by approaches such as DFI (Ramakrishnan *et al.*, 2000;Wilson *et al.*, 2001).

The majority of studies looking at bacterial gene expression inside eukaryotic cells have either been performed on single genes using sensitive techniques such as RT-PCR or have utilised reporter constructs. The advent of other screening technologies such as DFI has also allowed the identification of genes up-regulated following eukaryotic cell infection (see Chapter 4). There have been few reports of array technology being used to monitor changes in bacterial gene expression following infection of eukaryotic cells. A single study monitoring changes in *M. tuberculosis* gene expression following macrophage infection (Mangan *et al.*, 1999) has been reported, while a number of studies looking at the host response to a variety of bacterial

pathogens have also been reported (Cohen *et al.*, 2000;Eckmann *et al.*, 2000;Ichikawa *et al.*, 2000;Detweiler *et al.*, 2001).

In this chapter the establishment and optimisation of a transcriptional profiling system for identifying differentially regulated genes of *C. pseudotuberculosis* is described. In the first instance a DNA array of *C. pseudotuberculosis* library clones was constructed. To determine whether this could be successfully used to identify regulated genes the array was first applied to the identification of thermoregulated genes in a heat shock model. Having demonstrated the efficacy of the system the approach was then used to identify genes that were regulated in a density dependent manner and finally those that show regulation upon infection of a macrophage monolayer.

3.2 Construction of a *C. pseudotuberculosis* DNA array

To generate a *C. pseudotuberculosis* DNA array it was first necessary to construct a library of genomic fragments. Following construction of the library, it was grided into 96 well plates and DNA extracted from individual clones. The DNA was then arrayed onto nylon membranes to generate the completed array.

3.2.1 Construction of a *C. pseudotuberculosis* random library

In the majority of studies requiring a library of bacterial genomic fragments, restriction endonucleases are used to fragment the DNA so that it can be cloned into a suitable vector. The use of restriction endonucleases has some short-comings, as it limits the range of fragments that can be generated due to the sequence specificity of fragment generation. Secondly, owing to the variable frequency of a given restriction site within a genome some regions of DNA will not be cloned as they will either be too large to be cloned or will be selected against for being too small. To avoid these problems that are inherent when fragmenting DNA with restriction nucleases an alternative approach was utilised. Sonication, which fragments the DNA by cavetation, was chosen as it fragments DNA in an essentially random manner. Thus, overlapping DNA fragments should be generated, resulting in a more representative library. Many of the DNA fragments generated by sonication are likely to have 5' or 3' overhanging ends. It was necessary to remove these overhanging ends prior to

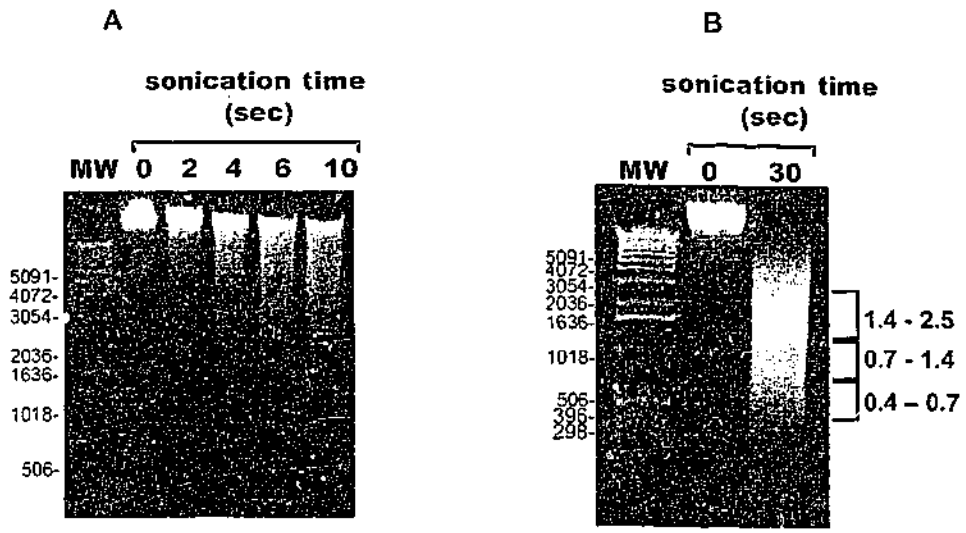

Figure 3.1: Fragmentation of C231 DNA by sonication.

C231 genomic DNA was sonicated using an XL-series sonicator (Heat Systems) using a standard microprobe set at 1.5. DNA was sonicated for 2, 4, 6, 10 (A) and 30 sec (B). Aliquots were taken at each timepoint and analysed by agarose gel electrophoresis. DNA in the ranges 0.4 to 0.7, 0.7 to 1.4 and 1.4 to 2.5 kb were size selected from the agarose. 1 kb DNA ladder is run in the first lane of each gel and band sizes are indicated to the left of this.

cloning. This was achieved by treating the DNA with mung bean nuclease (MBN), which specifically degrades single stranded DNA.

The degree of DNA fragmentation increases as either the sonication time or power increases. In order to minimise the number of different genes on any given plasmid, the aim was to shotgun clone 1 kb DNA fragments. In order to determine the required sonication conditions for generation of appropriately sized fragments, *C. pseudotuberculosis* C231 genomic DNA was sonicated for 2, 4, 6, 8, 10 and 30 sec using an XL-series sonicator (Heat Systems) with a standard microprobe set on 1.5. Samples were taken at each timepoint and analysed for fragmentation on a 1% agarose gel (Figure 3.1). A decrease in average DNA fragment size was observed that correlated with sonication time. After 30 sec a smear in the range of approximately 10 kb to 300 bp was observed. Given that the aim was to clone 1 kb fragments this sample was deemed suitable for library construction. As a control for the efficacy of the MBN reaction *Sau*3A1 fragments, which have a 5' overhang, were generated and as a control for ligation efficiency *Alu*I (blunt) fragments were generated.

Optimisation of conditions for MBN treatment was performed as it was observed that when the manufacturer's recommended instructions were followed, the ligation efficiency was very low. This indicated that the repair process had not been optimal. Approximately 3 µg of sonicated or *Sau*3AI fragmented DNA was treated with 30 U MBN at 30°C for 20 min (manufacturer's recommendation). Additionally 3 µg of sonicated DNA was treated in the same manner except at 37°C. Following MBN treatment the repaired DNA fragments were analysed by agarose gel electrophoresis. Sonicated DNA fragments in the ranges of 0.4 to 0.7, 0.7 to 1.4 and 1.4 to 2.5 kb were size selected from the agarose and DNA recovered from the gel. Additionally *Alu*I and *Sau*3AI fragmented DNA was also size selected, with DNA in the ranges 0.7 to 1.1 and 0.2 to 0.7 kb being selected. Upon reanalysis of the fragments by agarose gel electrophoresis, the effectiveness of the size selection could be observed (Figure 3.2). Fragments were ligated into commercially prepared pUC18 (Amersham Pharmacia Biotech, Buckinghamshire, UK) that had been digested with *Sma*I and then treated with bacterial alkaline phosphatase. Ligations were performed in 1x Rapid Ligation Buffer (RLB, Promega) which contains 5% (w/v) polyethylene glycol to increase ligation efficiency. Following incubation at 15°C for 16 hr, 3 µl of each 10 µl ligation was transformed into electrocompetent *E. coli* Top10 cells. The transformation mixture was plated onto LB agar plates containing X-gal and IPTG to allow for blue-white selection. Ligation of *Alu*I fragmented DNA into pUC18 was most efficient (Table 3.1). Contrary to the manufacturer's instructions MBN treatment at 37°C was found to be more effective than treatment at the recommended 30°C, resulting in 25-fold more colonies for sonication-derived DNA fragments.

Analysis of the inserts of 10 white colonies generated following ligation of repaired random fragments into pUC18 indicated that they all contained an insert and that the average insert size was 1.3 kb (Figure 3.3). Additionally each insert was of a different size indicating that there was randomness in the library. The remainder of the ligation was transformed into commercially prepared Epicurean Coli XL2-Blue MRF ultracompetent cells (Stratagene) with a calculated transformation efficiency of 5×10^9 cfu/µg pUC18 DNA, resulting in a library of approximately 4000 clones. Although this was not a large library it was considered sufficient as only 1536 colonies were required for construction of the array. The library was titled the pUC18-Q library.

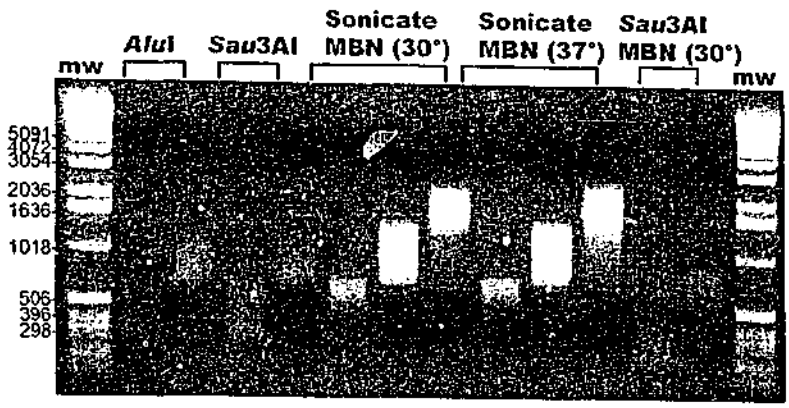

Figure 3.2: Size selection of sonicated or restriction endonuclease digested C231 DNA.

C231 DNA was fragmented by sonication or by incubation with *AluI* or *Sau*3A1. Some DNA samples were then treated with MBN at 30°C or 37°C. Treated DNA was analysed by agarose gel electrophoresis and DNA in various size ranges excised from the gel. Purified, size selected fragments were then reanalysed by agarose gel electrophoresis. 1 kb DNA ladder was run in the first and last lanes. Band sizes are indicated to the left of the gel.

Table 3.1: Ligation of genomic fragments into pUC18/*Sma*I treated with BAP

Ligation	Insert[a]	White colonies[b]	Blue colonies[b]
1	*AluI*, 0.7-1.1 kb	390	50
2	sonicate + MBN (30°C), 0.7-1.4 kb	10	30
3	sonicate + MBN (37°C), 0.7-1.4 kb	250	90
4	*Sau*3A1 + MBN (30°C), 0.7-.1 1kb	0	20

[a] A similar amount of insert DNA was used in each ligation reaction

[b] Colony number from transformation of 3 µl of 10 µl ligation into a 40 µl aliquot of electrocompetent Top10 F' *E. coli*.

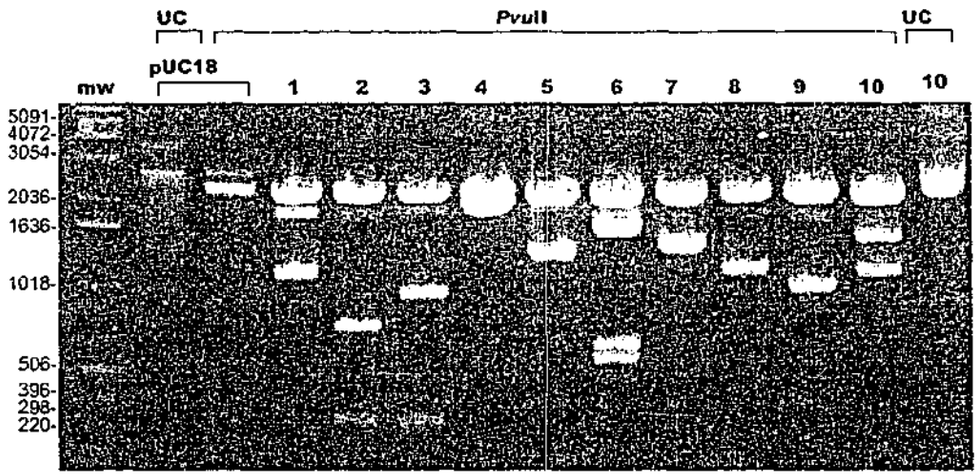

Figure 3.3: Analysis of plasmid inserts from the pUC18-Q library.

The library was generated by ligation of sonicated, MBN treated, size selected C231 DNA fragments into the *Sma*I site of pUC18. DNA was extracted from 10 randomly picked white colonies (1-10) then digested with *Pvu*II, which cuts out the insert DNA in addition to 322 bp of vector sequence. Uncut (UC) pUC18 and 10 were run in the 2nd and 14th lanes respectively. 1 kb DNA ladder is run in the first lane and band sizes are indicated to the left of this.

3.2.2 Griding of the pUC18-Q library and extraction of DNA from individual clones

1536 white colonies were picked from the pUC18-Q library plates into 16x 96 well plates containing Beaut broth. Following a 24 hr incubation to allow for bacterial growth, cultures in these plates were used to inoculate deep-well 96 well plates containing 1.2 ml Beaut broth, and then stored at -80°C as master plates. The deep-well plates were incubated for 24 hr with shaking at 300 rpm. DNA was then extracted using an alkaline lysis extraction protocol that has been adapted to the 96 well plate format (Section 2.6.3). DNA from the 96 well plate cultures was stored in 96 well plates. Plates were labelled Q1 through Q16. Individual DNA samples were given the title of the plate that they were from and the grid reference within that plate, for example, Q2-F6 or Q9-A8.

To ensure that the recovery of plasmid DNA was sufficient and that the majority of plasmids contained inserts, all 96-well plate samples were analysed by restriction nuclease digestion

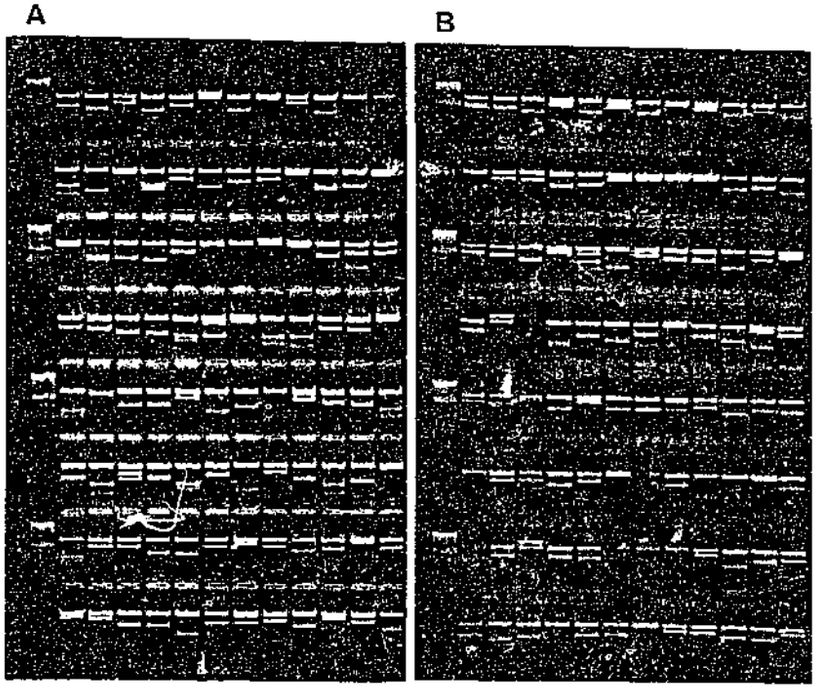

Figure 3.4: Analysis of plasmid DNA from clones of the pUC18-Q library.

The pUC18-Q library was gridded into 16x 96 well plates and DNA extracted from each clone. DNA was digested with *Pvu*II, which cuts out the insert and an additional 322 bp of vector sequence, and then analysed by agarose gel electrophoresis. Two representative plates are shown, (A) plate Q6 and (B) plate Q15. 1 kb DNA ladder is run in the first lane of the 1st, 3rd, 5th and 7th row of each gel.

with *Pvu*II. Two representative plates are shown in Figure 3.4. The majority of samples contained DNA. In the cases where the yield of DNA was low this could be correlated to poor growth of the clone at either of the two 96-well plate growth stages. For the clones that did grow, inserts were detected in the vast majority. This indicated that the blue/white screening process was efficient at identifying those clones that contained a plasmid insert. Due to the small scale of the gel it was not possible to obtain an accurate measure of insert size.

3.2.3 Library arraying

The 1536 DNA samples were consolidated into 4x 384 well plates using a 96 pin replicator. DNA in the 384 well plates was then replicated onto nylon membrane using a 384 pin replicator. Each sample was replicated in duplicate such that the entire library was represented

in 2 grids of 7x11cm containing 1536 spots each. The presence of DNA on the membranes was demonstrated by hybridising the array with DIG labelled pUC18 (Figure 3.5). Positive spots were observable for the majority of clones, indicating the presence of DNA on the membrane. When a spot was absent this could generally be attributed to the lack of DNA generated for that clone. Additionally variations in spot intensity could be attributed to the differences in the DNA concentration of the sample being arrayed. Given that DNA spot intensity between arrays rather than on the same arrays was being compared this was not considered to be an issue. In general the arraying of a specific DNA sample was found to be consistent.

3.3 Probing of the array with DIG labelled *C. pseudotuberculosis* cDNA

In the majority of bacterial transcriptional profiling experiments the target consists of labelled cDNA. In order to determine whether *C. pseudotuberculosis* gene expression could be detected using the array system, cDNA was generated then labelled with DIG. RNA was extracted from an exponentially growing culture of *C. pseudotuberculosis*. The RNA was DNaseI treated, then 1 µg of RNA was reverse transcribed using random hexamers as the primer. A subsequent RNaseH reaction was performed to degrade the RNA template. The cDNA was directly labelled with DIG using the Roche DIG Chem-Link Labelling and Detection Set, which chemically couples DIG to adenosine and guanosine bases via a cis-platinum linkage (Hoevel *et al.*, 1999). The labelled target was hybridised to the array overnight in DIG Easy Hyb (Roche). Following stringency washes, the bound probe was detected using an anti DIG alkaline phosphatase (AP) conjugate and subsequent incubation with the AP substrate, CSPD. Using such an approach 18 positives were detected (Figure 3.6). For 10 of the corresponding clones the sequence of the insert DNA was determined. The majority of clones coded for ribosomal RNA, with 2 corresponding to 16S rRNA and a further 5 to 23S rRNA. The remaining 3 clones had inserts coding for urease components, an ABC transporter ATP-binding protein and proteins of unknown function respectively. That the majority of positive clones coded for rRNA reflects that the vast majority of bacterial RNA is of ribosomal origin and hence the majority of the cDNA in the target is complementary to RNA genes.

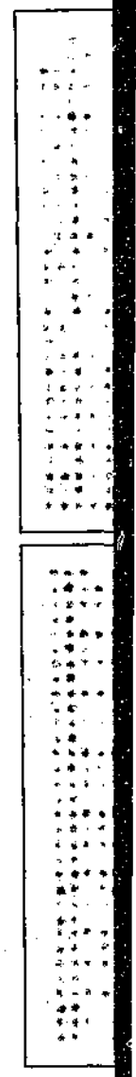

Figure
The C.
DNA ta
and the

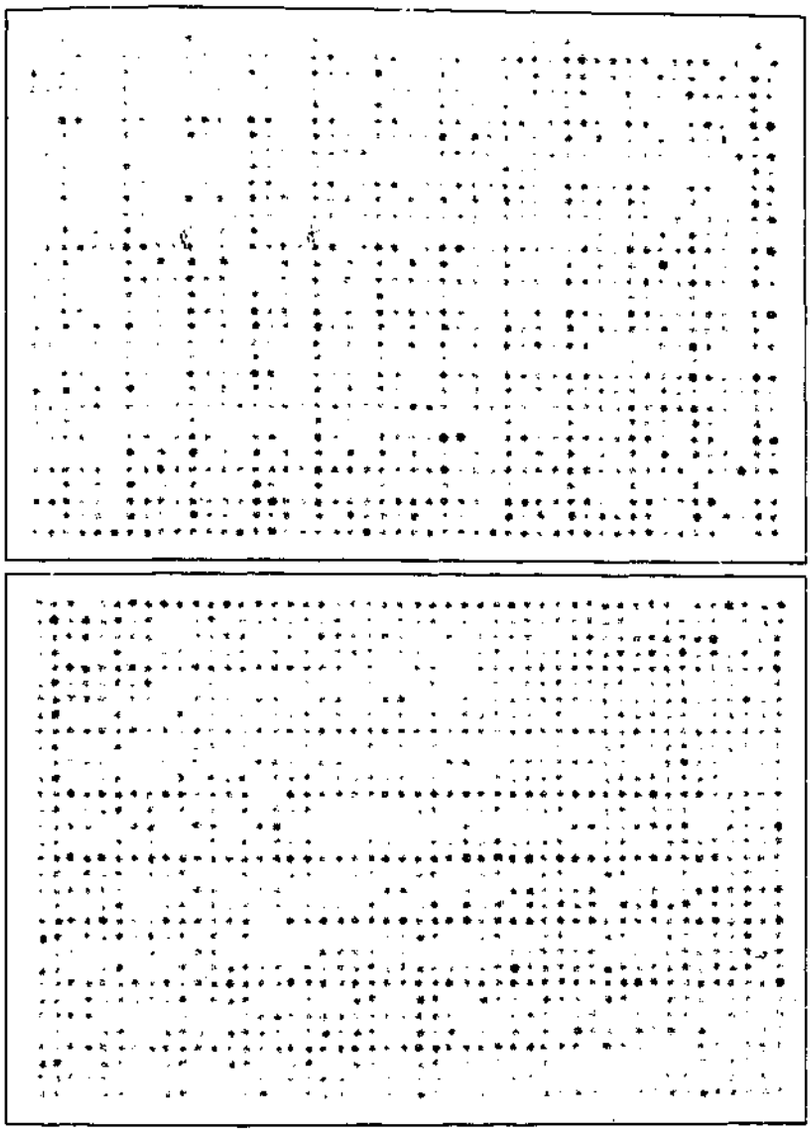

Figure 3.5: pUC18 positive clones of the *C. pseudotuberculosis* pUC18-Q DNA array.

The *C. pseudotuberculosis* pUC18-Q DNA array was hybridised overnight with a DIG labelled pUC18 DNA target. Following stringent washes, the bound DNA was detected using an anti DIG AP conjugate and the AP substrate CSPD.

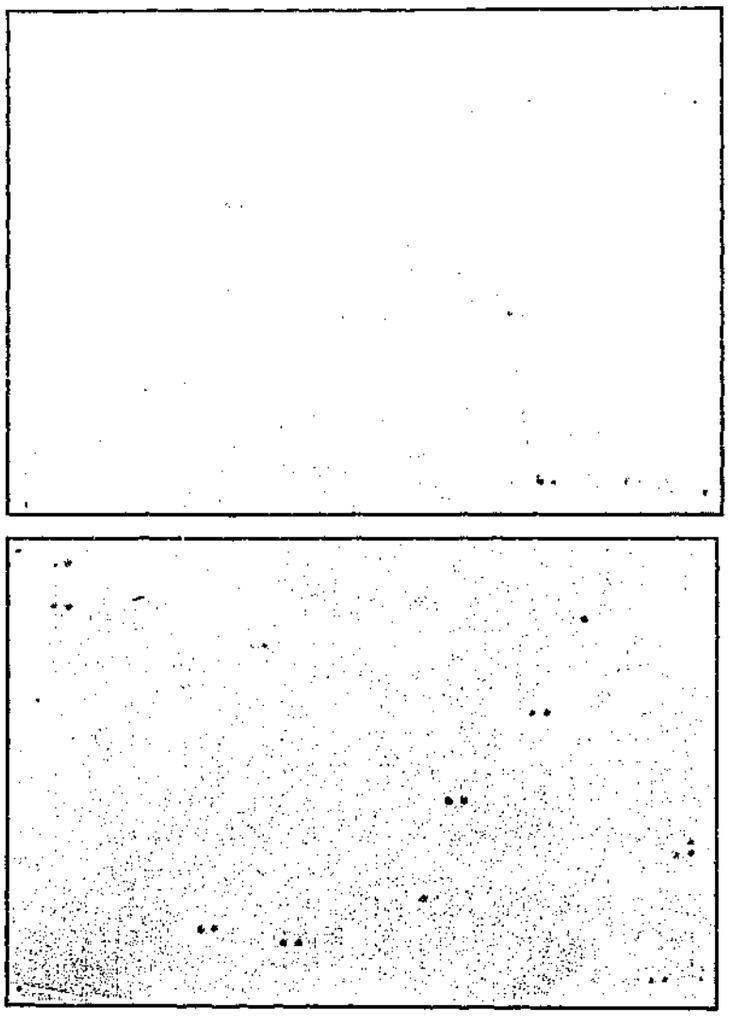

Figure 3.6: pUC18-Q DNA array probed with DIG labelled *C. pseudotuberculosis* cDNA.

C. pseudotuberculosis RNA was extracted from an exponentially growing culture. 1 µg of RNA was reverse transcribed, then labelled with DIG and used to probe a pUC18-Q array. Bound target was detected using an anti DIG AP conjugate and the AP substrate CSPD.

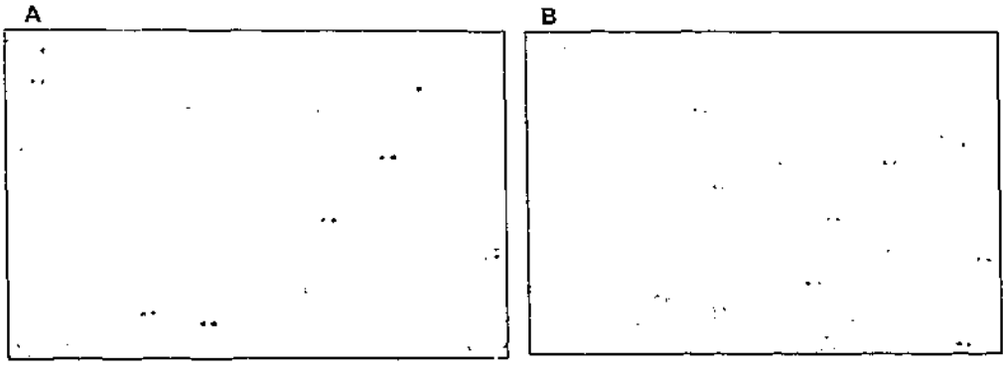

Figure 3.7: Comparison of RNA and cDNA as an array target.

C. pseudotuberculosis RNA was extracted from an exponentially growing culture. (A) 1 µg of RNA was reverse transcribed, then labelled with DIG and used to probe a pUC18-Q array. (B) 3.5 µg of RNA was DIG labelled and used to probe a pUC18-Q array. Bound target was detected using an anti DIG AP conjugate and the AP substrate CSPD. For each array only the grid corresponding to plates Q9 to Q16 is shown.

3.4 Optimisation of probing conditions

Given that so few positive spots were detected in the first array experiment it was necessary to optimise the array and probing conditions. A number of variables were tested, including choice of substrate for the target (DNA versus RNA), and the hybridisation and detection systems utilised.

3.4.1 RNA versus DNA as a target

Either DNA or RNA may be used as a substrate in the DIG Chem-Link system. To determine whether directly labelled RNA could be used as an array target its efficacy was compared to a DNA target. 3.5 µg of C231 RNA was labelled with DIG and used in an analogous manner to the cDNA probe in section 3.3. Analysis of the arrays indicated that most of the same clones were detected in each instance (Figure 3.7). It was therefore apparent that if appropriate RNase-free conditions were utilised it was possible to use directly labelled RNA as the target in array probing. Given that this precluded the need to perform an RT reaction, RNA was used as the target in all subsequent experiments.

3.4.2 Choice of hybridisation and detection systems

The amount of labelled target that binds to membrane bound DNA can be influenced by the hybridisation conditions, in particular the hybridisation solution utilised. Two commercially available hybridisation solutions were tested. The first was DIG Easy Hyb (Roche) which was used in the previous experiments. DIG Easy Hyb is a non-formamide based hybridisation solution that has been optimised for DIG based applications. The second was Ultrahyb™ Hybridisation buffer (Ambion). The manufacturer claims that this hybridisation solution maximises sensitivity without increasing background such that with an overnight hybridisation signal strength may be increased 20 to 50-fold over that achieved with traditional hybridisation buffers. The array experiment described in section 3.4.1 was repeated exactly except that Ultrahyb™ Hybridisation buffer was used in the place of DIG Easy Hyb as the hybridisation buffer. Although the same spots were detected regardless of the hybridisation solution utilised, the speed with which the positives became detectable on film was greater when using Ultrahyb™ Hybridisation buffer, such that only a 4 hr exposure was required compared to a 16 hr exposure. However, exposure of the Ultrahyb™ Hybridisation buffer blot for a longer time period did not increase the number of detectable positives, rather the background increased slightly.

As an additional way to increase the system sensitivity a second more sensitive alkaline phosphatase substrate was tested. The probing conditions were exactly as per the previous experiment. Washing and DIG detection conditions were also identical except that CDP-Star was used as the AP substrate. CDP-Star generates a luminescent signal that is approximately 10-fold more sensitive than CSPD. Upon exposure of the developed array to film, it was possible to detect strong positives within 4 min, additionally clones that had not previously been detected now came up as positives. Exposure of the array to film for a longer period of time (4 hr) led to an increase in background that made membrane analysis difficult.

To allow longer exposure periods it was necessary to optimise the DIG detection system. The two variables that were optimised were the concentrations of the anti DIG AP conjugate and the CDP star solution used for development. Antibody dilutions ranging from 1:20,000 to 1:100,000 and CDP-Star dilutions ranging from 1:100 to 1:1000 were tested (Figure 3.8). An optimal combination of antibody at 1:50,000 and CDP star at 1:500 was ultimately chosen as it

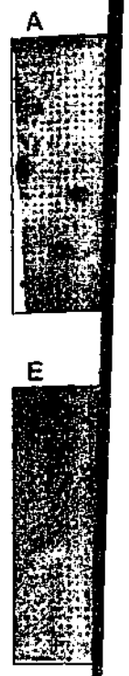

Figure 3.

Four pUC
labelled
H). Bou
1:35,000
at dilutio

allowed
backgr

3.4.3

In the
To dete
library
compa
Conco
detecte
arrays

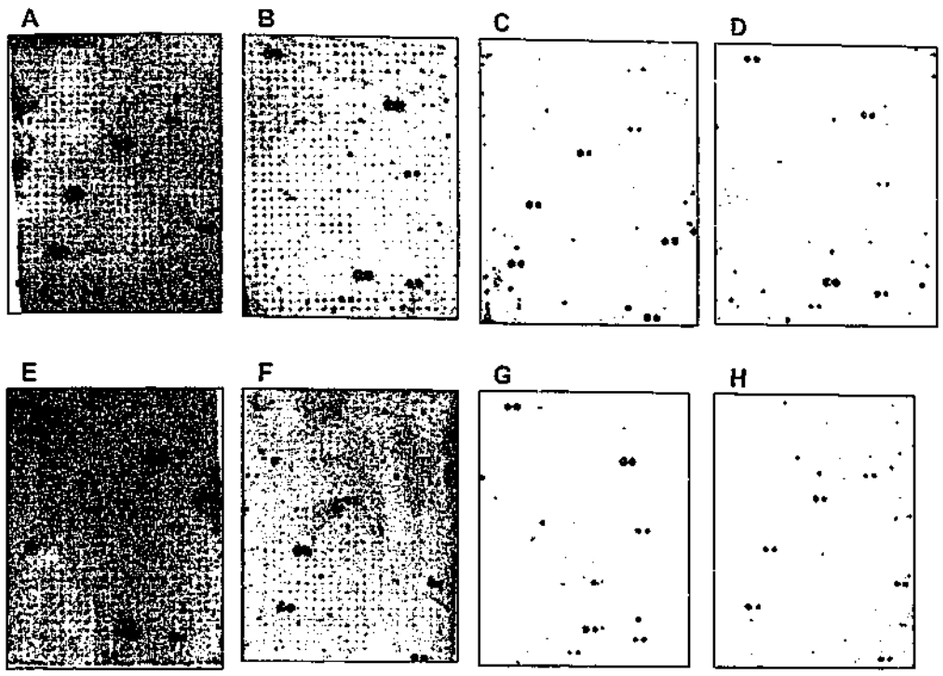

Figure 3.8: Optimisation of the DIG detection system.

Four pUC18-Q DNA arrays (corresponding to plates Q9 to Q16 only) were probed with 3.5 μg DIG labelled *C. pseudotuberculosis* RNA. Following stringency washes the arrays were cut in half (A to H). Bound target was detected using an anti DIG AP conjugate at dilutions of (A) 1:20,000, (B) 1:35,000, (C, E-H) 1:50,000 and (D) 1:100,000. AP was detected using the AP substrate CDP-Star at dilutions of (A-D and G) 1:500, (E) 1:100, (F) 1:300 and (H) 1:1000.

allowed for long exposures to maximise positive clone detection while minimising the level of background.

3.4.3 Amount of RNA in the probe

In the majority of reported array experiments a relatively large amount of target is labelled. To determine the effect of increasing the amount of target on the number of clones in the library that were detected, an array was probed with 20 μg of DIG labelled RNA and compared to an array in which approximately 6-fold less target was used (Figure 3.9). Concomitant with the increase in target amount was a large increase in the number of positives detected. At this stage the system was considered to be maximally optimised such that the arrays could be used for transcriptional profiling.

A

B

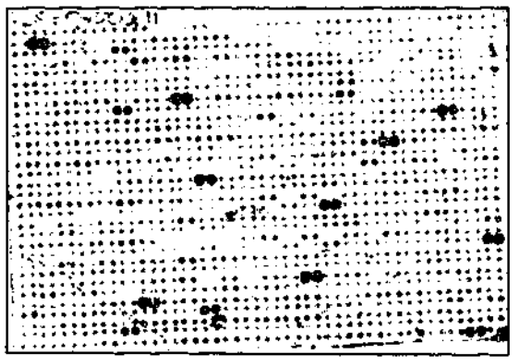

Figure 3.9: Effect of the amount of labelled RNA in the target on the number of array clones detected.

C. pseudotuberculosis RNA was extracted from an exponentially growing culture. 3.5 µg (A) or 20 µg (B) of RNA was DIG labelled and used to probe a pUC18-Q array. Bound target was detected using an anti DIG AP conjugate and the AP substrate CDP-Star. For each array only the grid corresponding to plates Q9 to Q16 is shown.

3.5 Identification of genes regulated by heat shock

3.5.1 Optimisation of conditions for heat shock

To determine an appropriate length of heat shock in which changes in expression of thermoregulated genes would occur, the effect of heat shock on expression of *dnaK* was determined. The heat shock gene, *dnaK*, was chosen as the gene sequence was available and the effect of shock on its expression had been measured in a wide variety of bacterial species.

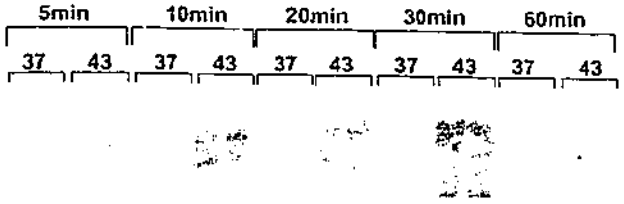

Figure 3.10: Regulation of *dnaK* expression by heat shock.

RNA was extracted from *C. pseudotuberculosis* control (37°C) and heat shocked samples (37°C to 43°C) at 5, 10, 20, 30 and 60 min post initiation of the shock. 15 µg of total RNA was analysed by Northern analysis using a DIG labelled *dnaK* PCR product.

A temperature increase of 6°C (from 37°C to 43°C) was chosen as it has previously been shown that induction of the *C. pseudotuberculosis dnaK* promoter occurs with this level of shock (Simmons, 1997). Aliquots of an exponentially growing C231 culture (37°C) were taken and then incubated at 37°C for a further 10 min. Half of the aliquots were maintained at 37°C while the other half were transferred to 43°C. At 5, 10, 20, 30 and 60 min, duplicate aliquots were removed from each temperature and RNA extracted. 15 µg RNA was run under denaturing conditions on a 1.2% formaldehyde agarose gel and then transferred to Hybond N+ membrane in 20x SSC. The northern blot was then hybridised with a DIG labelled *dnaK* PCR product (primers #5 and #6). Following stringency washes, the bound DIG was detected using an anti DIG-AP conjugate and the AP substrate CSPD and the membrane was then exposed to film (Figure 3.10).

dnaK mRNA was not observed in non heat shocked samples. At 43°C *dnaK* expression was first observed at 5 min post heat shock, the level increasing until it reached a plateau at 30 min. After one hour of heat shock the level of *dnaK* mRNA had decreased below that seen at earlier timepoints. Given that maximal expression of dnaK was observed at 30 min, this timepoint was deemed to be suitable for the identification of other thermoregulated genes.

3.5.2 Preparation of RNA for array probing

An overnight culture of C231 was diluted to an OD_{600} of 0.1 then grown to an OD_{600} of 0.8. Multiple aliquots were made and then incubated at 37°C for a further 10 min. At this time half

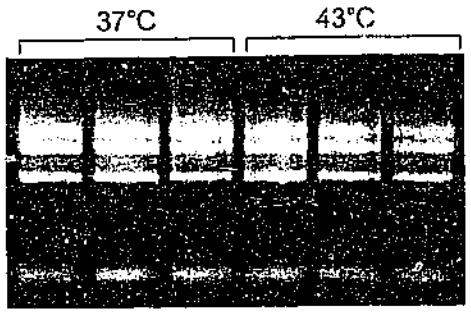

Figure 3.11: Analysis of RNA prepared for the heat shock array experiment.

RNA was extracted from triplicate cultures of C231 that had been incubated continuously at 37°C or that had been incubated at 37°C, then heat shocked at 43°C for 30 min. 2 µg of each RNA sample was analysed on a non-denaturing agarose gel.

of the cultures were transferred to 43°C. After a 30 min heat shock RNA was extracted from the control and heat-treated samples. The quality of the RNA and accuracy of the RNA quantitation was assessed on a non-denaturing agarose gel (Figure 3.11). The presence of discrete ribosomal bands indicated that the RNA was relatively intact, while the evenness of the intensity between lanes indicated that there were equivalent amounts of RNA in each tract.

3.5.3 Probing of the DNA arrays

20 µg of a 37°C sample and a 43°C sample were DIG labelled and hybridised to replica arrays at 42°C overnight. Following stringency washes the bound target was detected with an anti DIG antibody with an AP conjugate and subsequent incubation with the AP substrate CDP-Star. The blots were exposed to film (Figures 3.12 and 3.13).

3.5.4 Array analysis

The level of background on the blots made analysis of the blots using imaging software difficult. Instead differences in spot pattern were identified by visual inspection. Using this approach 17 clones were determined to be differentially regulated. The inserts of these clones were sequenced. To assign putative gene function the sequences were analysed using a BLAST-X search for sequence similarity to proteins in the NCBI protein databases. These data are presented in Table 3.2. A number of clones appeared to contain concatamers as indicated by a change in sequence similarity mid-sequence. A number of heat shock genes

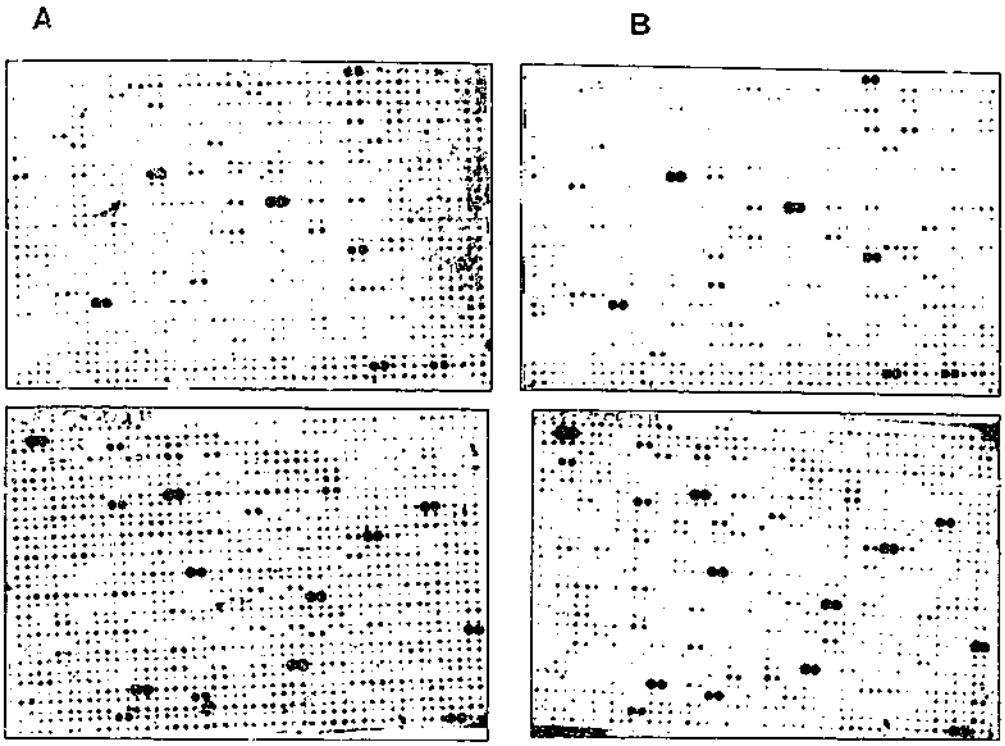

Figure 3.12: pUC18-Q DNA arrays probed with RNA derived from C. *pseudotuberculosis* incubated at 37°C or 43°C.

Cells were grown continuously at 37°C, or at 37°C, then heat shocked for 30 min at 43°C. RNA was extracted from the cells, labelled with DIG and then used to probe replica pUC18-Q DNA arrays. Bound probe was detected using an anti DIG AP conjugate and the AP substrate CDP Star. (A) 37°C and (B) 43°C.

were identified, including several from the *dnaK* operon and *clpB*. A number of other genes that are not considered heat shock genes were also identified, for example *pld* and a gene encoding a fatty acid synthase which we have termed *fas*. Several genes were represented on more than one clone, for example, *dnaK*, *dnaJ*, *pld* and *fas*. However in each instance the insert sequence was not identical indicating that these clones were not siblings.

3.5.5 Quantitation of gene regulation

As the method of selecting regulated clones was only semi-quantitative it was necessary to confirm regulation using a second approach. Additionally, a number of clones contained more than one gene so it was necessary to determine whether each gene within the plasmid insert

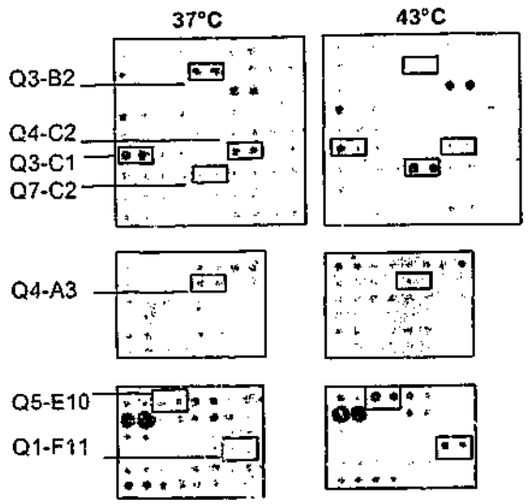

Figure 3.13: Close-up analysis of pUC18-Q DNA arrays probed with RNA derived from *C. pseudotuberculosis* incubated at 37°C or 43°C.

Magnified regions of the arrays presented in Figure 3.12 are shown. Clones which were identified as being regulated are boxed and the clone identification shown.

was regulated. RNA dot blots were initially utilised, however this approach was limited in its sensitivity and quantitative ability. The advent of real time PCR allowed this powerful technology to be utilised in conjunction with reverse transcription to measure relative changes in gene expression.

Real time PCR primers were designed for most of the genes of interest (Table 2.5). RNA samples were prepared from independent bacterial cultures in the same manner as for the array experiment. 1 µg of DNase I treated RNA was reverse transcribed using MuLV reverse transcriptase to generate cDNA which was used as a template for real time PCR. In addition to measurement of the mRNA of interest, 16S rRNA levels were also determined as a normalisation control. For each RT reaction a no RT control was performed in which the reactions were identical except for the absence of RT. As expected all of the heat shock genes were strongly induced following heat shock (Table 3.3). Another gene which coded for a putative maltose/maltodextrin ABC transport system permease component demonstrated a

Table 3.2: Sequence analysis of clones from the pUC18-Q library that were regulated by heat shock.

Clone	regulation[a]	insert size	Species and % identity[b]	gene homology[b]

Table 3.2: Sequence analysis of clones from the pUC18-Q library that were regulated by heat shock.

Clone	regulation[a]	insert size	Species and % identity[b]	gene homology[b]
Q1-F11	up	0.8	*Mycobacterium paratuberculosis*, 85%	*dnaK*
Q2-G12	down	1.9	*Yersinia enterocolitica*, 55%	*fagC* (iron permease component)
			C. pseudotuberculosis, 100%	*pld*
			Streptomyces coelicolor, 48%	exodeoxyribonuclease
Q3-C1	down	1.2	*Streptcoccus pyogenes*, 47%	maltose/ maltodextrin ABC transport system permease
Q3-B2	down	2.2	*Staphlococcus aureus*, 34%	peptide chain release factor 3 (RF-3)
			Xanthomonas campesiris, 49%	proline imminopeptidase
			C. pseudotuberculosis, 100%	*pld*
			E. coli, 44%	*fagC* (iron permease component)
Q4-A3	down	1.2	*Corynebacterium ammoniagenes*, 71%	*fas* (fatty acid synthase)
Q4-C2	down	1.0	*Corynebacterium. glutamicum*, 81%	cytochrome c1
			C. glutamicum, 72%	rieske iron-sulfur protein
Q4-E9	up	1.2	*Corynebacterium flavescens*, 78%,	aspartokinase
			M. paratuberculosis, 79%	*dnaK*
Q5-E10	up	0.9	*M. paratuberculosis*, 84%	*dnaK*
Q7-C2	up	1.3	*C.glutamicum*, 76%,	*clpB*
Q7-B6	down	2.2	*Mycobacterium tuberculosis*, 41%,	*serB2* (phosphoserine phosphatase)
			M. tuberculosis, 41%	hypthetical 42.3kDa protein Rv0433
Q7-D6	up	2.2	*Bacillus. subtilis*, 50%,	hypothetical metabolite transport protein
Q11-G1	down	0.9	*C. ammoniagenes*, 68%,	*fas* (fatty acid synthase)
Q11-G7	up	1.1	*M. paratuberculosis*, 75%,	*dnaK*

table continues on next page

Clone	regulation[a]	insert size	Species and % identity[b]	gene homology[b]
Q12-C5	up	1.5	*M. tuberculosis*, 51%,	*dnaJ*
			M. tuberculosis, 57%	*hspR*
Q12-E9	up	1.1	*S. coelicolor*, 32%	metallopeptidase
			Streptococcus pneumoniae, 63%	dTDP-glucose 4,6-dehydratase
Q13-G5	down	0.6	*S. coelicolor*, 43%,	probable substrate binding protein
Q16-G1	up	2.0	*S. coelicolor*, 43%,	*dnaJ*
			M. tuberculosis ,	GTP-binding protein LEPA (probable lepA)

[a] regulation refers to whether clones were identified as up or down-regulated compared to 37°C following heat shock

[b] Sequences were scanned against the NCBI protein databases for protein homologies in all 6 reading frames using a translating BLAST program (BLASTX). Proteins that showed the highest identity and the species in which this occurred are listed.

Table 3.

Clone
Q7-C2
Q12-C5,
Q12-C5
Q1-F11,
Q5-E10,
Q3-C1
Q4-E9
Q12-E9
Q7-B6
Q4-C2
Q4-C2
Q7-D6
Q2-G12,
Q4-A3, C
Q2-G12,

[a] fold indu

used to p

greater t

However

the two

genes tha

metaboli

and *pld*.

gene exp

3.6 Ar

The obse

regulatio

that *pld*

expressio

ordinatel

Table 3.3: Induction of genes following heat shock.

Clone	gene	fold induction[a]
Q7-C2	*clpB*	45.73
Q12-C5, Q12-G1	*dnaJ*	22.16
Q12-C5	*hspR*	17.51
Q1-F11, Q4-E9, Q5-E10, Q11-G7	*dnaK*	12.92
Q3-C1	maltose/maltodextrin ABC transporter permease	2.36
Q4-E9	aspartokinase	1.73
Q12-E9	metallopeptidase	0.90
Q7-B6	hypothetical protein	0.68
Q4-C2	rieske iron-sulfur protein	0.61
Q4-C2	cytochrome c1	0.53
Q7-D6	hypothetical metabolite transport protein	0.44
Q2-G12, Q3-B2	*fagC* (iron permease component)	0.11
Q4-A3, Q11-G1	*fas* (fatty acid synthase)	0.06
Q2-G12, Q3-B2	*pld*	0.04

[a] fold induction was determined on RNA samples prepared in an analogous manner to those used to probe the DNA array and was determined by RT-real time PCR.

greater than two-fold induction. This clone was originally identified as being down regulated. However, reanalysis of the array showed that this may have been a false result as only one of the two replicates for this clone appeared to be down regulated at 43°C (Figure 3.13). Four genes that demonstrated a greater than two fold repression were identified. These coded for a metabolite transport protein, an iron permease component (*fagC*), a fatty acid synthase (*fas*) and *pld*. Additionally several clones contained genes that showed less than a 2-fold change in gene expression, suggesting that they were false positives.

3.6 Array analysis of genes regulated in a cell density dependent manner

The observation that *pld* is down regulated by heat shock led to further studies of its regulation. While performing these studies (which are described in Chapter 6) it was observed that *pld* expression is also regulated in a cell density dependent manner such that its expression increases with increasing cell density. Given that virulence genes are often co-ordinately regulated, it is possible that other genes important for virulence may also be

regulated in the same manner. Additionally, in a number of gram-positive pathogens, some virulence determinants have been shown to be regulated by quorum sensing, in which their gene expression is directly related to the density of the bacterial population. For these reasons attempts were made to identify genes whose expression pattern correlated with *pld* expression such that they were expressed in a cell density dependent manner.

3.6.1 Preparation of RNA for array probing

To identify density dependent gene expression RNA was extracted from cultures of three different densities. An overnight culture of C231 was subcultured in duplicate to an OD_{600} of 0.1 and then grown at 37°C. At 30 min, 3 hr and 6 hr, aliquots of the cultures were taken and RNA extracted from the bacteria. At these timepoints the OD_{600} of the cultures were 0.16, 0.47 and 1.5 respectively. As for the heat shock experiment the integrity and quantitation of the RNA were assessed by agarose gel electrophoresis. Additionally the density dependent expression of *pld* in these samples was confirmed by Northern analysis (Figure 6.7).

3.6.2 Probing of the DNA arrays

20 µg of the 0.5, 3 and 6 hr RNA samples were DIG labelled and hybridised to replica arrays at 50°C overnight. Following stringency washes, the bound target was detected with an anti DIG antibody with an AP conjugate and subsequent incubation with the AP substrate CDP-Star. The blots were exposed to film (Figures 3.14 and 3.15).

3.6.3 Clone analysis

The density DNA arrays were analysed in a manner analogous to the arrays for the heat shock experiment. The inserts of 22 clones were sequenced and analysed for homologies to proteins in the NCBI protein databases (Table 3.4). As expected spots corresponding to clones Q3-B2 and Q2-G12, which contain a portion of the *pld* gene, increased in intensity at the later timepoints. A number of other clones that were identified as containing heat-regulated genes were also identified in this study. This included clones containing sequences for fatty acid synthase, rieske iron sulphur protein and *fagC*. A number of genes whose expression may be growth-related were identified. These included a number of ribosomal proteins, translation initiation factor 3 and fatty acid synthase. For three of the genes, density dependent

Figure 3.14: pUC18-Q DNA arrays probed with RNA derived from *C. pseudotuberculosis* cultures of different densities.

Cells were grown continuously at 37°C from a starting OD_{600} of 0.1. At 0.5, 3 and 6 hr RNA was extracted from the culture, labelled with DIG and then used to probe replica pUC18-Q DNA arrays. Bound target was detected using an anti DIG AP conjugate and the AP substrate CDP-Star. (A) 0.5 hr, (B) 3 hr and (C) 6 hr.

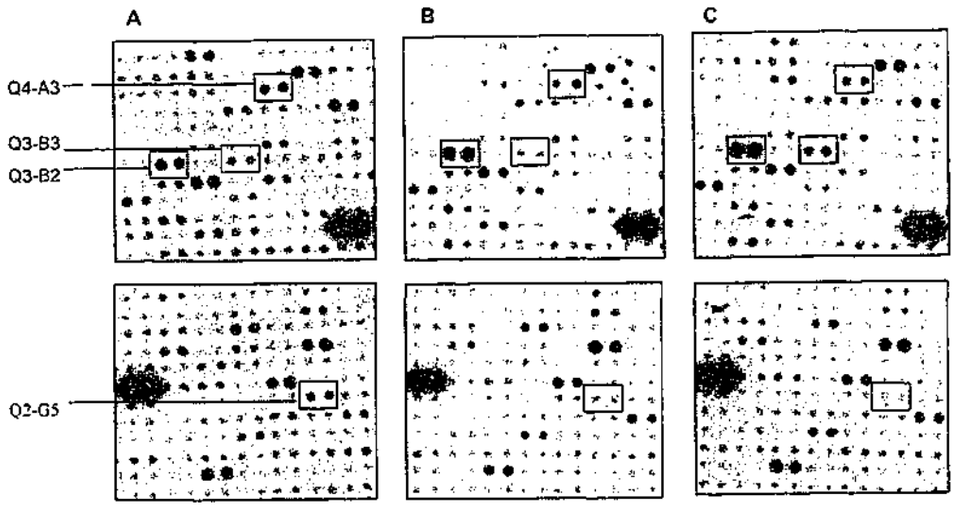

Figure 3.15: Close-up analysis of pUC18 Q DNA arrays probed with RNA derived from *C. pseudotuberculosis* cultures of different densities.

Magnified regions of the arrays presented in Figure 3.14 are shown. Clones which were identified as being regulated are boxed and the clone identification shown. A, B and C correspond to 0.5, 3 and 6 hr culture derived probes.

expression was confirmed by RT-real time PCR. *fagC* and *pld* were up-regulated approximately 2.5 and 5.6 fold respectively in high density cultures compared to low density cultures whereas *fas* was down-regulated approximately 2.6 fold at high density compared to low density.

3.7 Identification of *C. pseudotuberculosis* genes induced in a macrophage infection model

C. pseudotuberculosis is primarily an intracellular pathogen that resides within the macrophage. Within the macrophage *C. pseudotuberculosis* is likely to be exposed to altered conditions including changed pH, reduced nutrient availability and an increased concentration of reactive oxygen and nitrogen intermediates. In order to survive and replicate within this environment it would be expected that significant changes in bacterial gene expression would occur in comparison to bacteria cultured under nutrient rich *in vitro* conditions. To identify

Table 3.4: Sequence analysis of clones from the pUC18-Q library that were regulated in a density dependent manner.

Table 3.4: Sequence analysis of clones from the pUC18-Q library that were regulated in a density dependent manner.

clone	regulation[a]	species and % identity[b]	gene homology[b]
Q1-C7	up	*Propionibacterium freundenreichii*, 80%	methyl malonyl coA carboxyl transferase
Q2-A2	down	*Mycobacterium leprae*, 82%	DNA-directed RNA polymerase beta chain
Q2-D10	down	*M. tuberculosis* , 79%	translation intitiation factor IF-3,
		M. tuberculosis , 63%	50S ribosomal protein L35
Q2-G5	down	*C. glutamicum*, 70%	50S ribosomal protein L11
		C. glutamicum, 70%	50S ribosomal protein L1
Q2-G12	up	*Y. enterocolitica*, 55%	*fepC* (iron permease component)
		C. pseudotuberculosis, 100%	*pld*
		Streptomyces coelicolor, 48%	exodeoxyribonuclease
Q3-B2	up	*S. aureus*, 34%	peptide chain release factor
		Xanthomonas campestris, 49%	proline imminopeptidase
		C. pseudotuberculosis, 100%	*pld*
		E. coli, 44%	*fepC* (iron permease component)
Q3-B3	up	*Neisseria meningitis*, 47%	*nor* (nitric oxide reductase)
Q4-A3	down	*C. ammoniagenes*, 71%	*fas* (fatty acid synthase)
Q4-E5	up	*C. flavescens*, 78%,	aspartokinase
		M. paratuberculosis, 79%	*dnaK*
Q4-G4	up	*C. glutamicum*, 81%	rieske iron-sulfur protein
		C. glutamicum, 86%	cytochrome B subunit
Q5-C9	up	*Y. enterocolitica*, 53%	*fepC* (iron permease component)
		E. coli, 54%	*fepG* (iron permease component)
Q5-D1	up	*S. coelicolor*, 53%	putative reductase flavoprotein subunit

table continues on next page

clone	regulation[a]	species and % identity[b]	gene homology[b]
Q6-A5	up	*Lactococcus lactis*, 29%	maltose ABC transporter permease protein
Q6-D3	up	*M. tuberculosis*, 63%	*pepD* (aminopeptidase)
Q6-E1	down	*C. glutamicum*, 70%	preprotein translocase secY subunit
Q6-H7	down	*M. tuberculosis*, 57%	hypothetical ABC transporter
Q9-C5		*Clostridium pasteurianum*, 50%	*pfl* (formate acetyl transferase)
Q10-A12	down	*M. leprae*, 32%	hypothetical secreted protein
		M. leprae, 62%	phosphoserine aimnotransferase
Q10-B11	up	*Cavia porcellus*, 24%	anion exchange protein
Q10-C7	down	*M leprae*, 55%	Ribonuclease III
		M leprae, 53%	formamidopyrimidine-DNA glycosylase
		M. tuberculosis, 36%	hypothetical protein Rv2926c
Q10-B8		*Zymomonas mobilis*, 60%	*pfl* (formate acetyl transferase)
Q11-B11	up	*S. coelicolor*, 67%	phosphoglycerate mutase

[a] regulation refers to whether clones were identified as up or down-regulated in a high density culture compared to a low density culture.

[b] Sequences were scanned against the NCBI protein databases for protein homologies in all 6 reading frames suing a translating BLAST program (BLASTX). Proteins that showed the highest homology and the species in which this occurred are listed.

macrop
Second
monola
derived
3.7.1
A tissu
commo
In ord
and the
various
between
gfp ex
were i
membr
phago
18 hr
The m
(MOI
were
releas
were
soluti
bacte
genta
detcr

macrophage regulated genes it was first necessary to establish a suitable infection model. Secondly, a protocol for the extraction of bacterial RNA from an infected tissue culture monolayer needed to be developed. Finally DNA arrays were probed with macrophage derived bacterial RNA.

3.7.1 Establishment of a macrophage infection model

A tissue culture cell system using the gentamycin survival assay (Elsinghorst, 1994) is a commonly used infection model. In this assay the antibiotic gentamycin, which is unable to cross the eukaryotic plasma membrane, is added to monolayers of eukaryotic cells that have been experimentally infected with bacteria. Due to its specific location gentamycin kills extracellular but not intracellular bacteria. Therefore, researchers can be sure that they are studying bacteria that have been internalised by the eukaryotic cells.

In order to establish this assay J774 macrophages were infected with *C. pseudotuberculosis* and the number of viable bacteria remaining within the macrophages was determined at various times post-infection by performing bacterial counts. Additionally the interaction between the bacteria and the macrophages was monitored by fluorescence microscopy using a *gfp* expressing strain of *C. pseudotuberculosis*. In order to confirm that the bacteria detected were intracellular, rather than for example extracellular but associated with the plasma membrane in a way that protects them from antibiotic, an assay was performed in which phagocytosis of the bacteria was prevented by incubation of the monolayer at 4°C.

18 hr prior to infection macrophages were plated in 6 well plates at 5.5 x 10^6 cells per well. The macrophages were infected with *C. pseudotuberculosis* at a multiplicity of infection (MOI) of 4 and MOI 16 for 1 hr. After this time the cells were washed thoroughly in PBS and gentamycin-containing media was added to the cells. For half of the cells phagocytosis of the bacteria was prevented by preincubating the macrophages at 4°C for 20 min, then using cooled solutions and incubating the cells at 4°C until 3 hr post infection. At this timepoint the cells were returned to 37°C. At various timepoints the macrophage monolayers were lysed to release bacteria from the macrophages and then sonicated to disrupt bacterial clumps. Lysates were serially diluted then plated on BHI agar. After 48 hr at 37°C the number of colonies was determined (Figure 3.16). Incubation at 4°C greatly reduced the number of bacteria

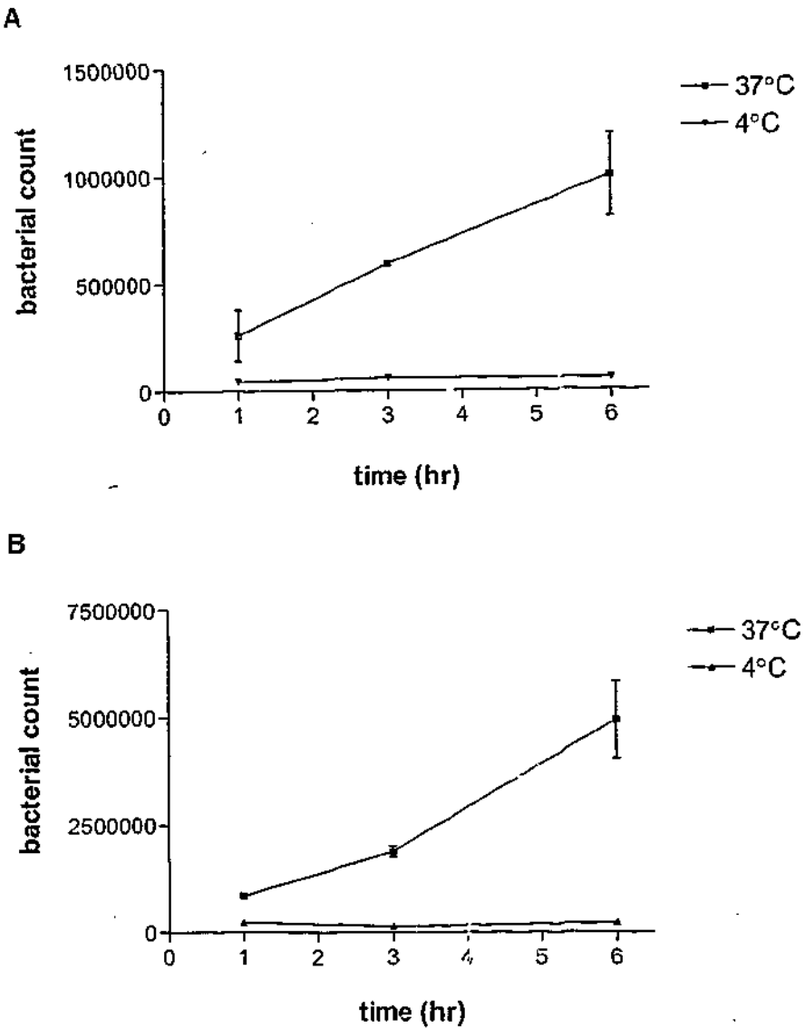

Figure 3.16: Infection of J774 macrophages with *C. pseudotuberculosis*.

J774 macrophages were infected with *C. pseudotuberculosis* at MOI 4 (A) and MOI 16 (B) for 1 hr. After this time the cells were washed and gentamycin containing media added to the cells. For some cells phagocytosis of the bacteria was prevented by pre-incubating the macrophages at 4°C for 20 min, then using cooled solutions and incubating the cells at 4°C until 3 hr post infection. At various time points the macrophage monolayers were lysed and sonicated to release then disrupt bacterial clumps. Lysates were serially diluted then plated on BHI agar. After 48 hr at 37°C the number of colonies was determined.

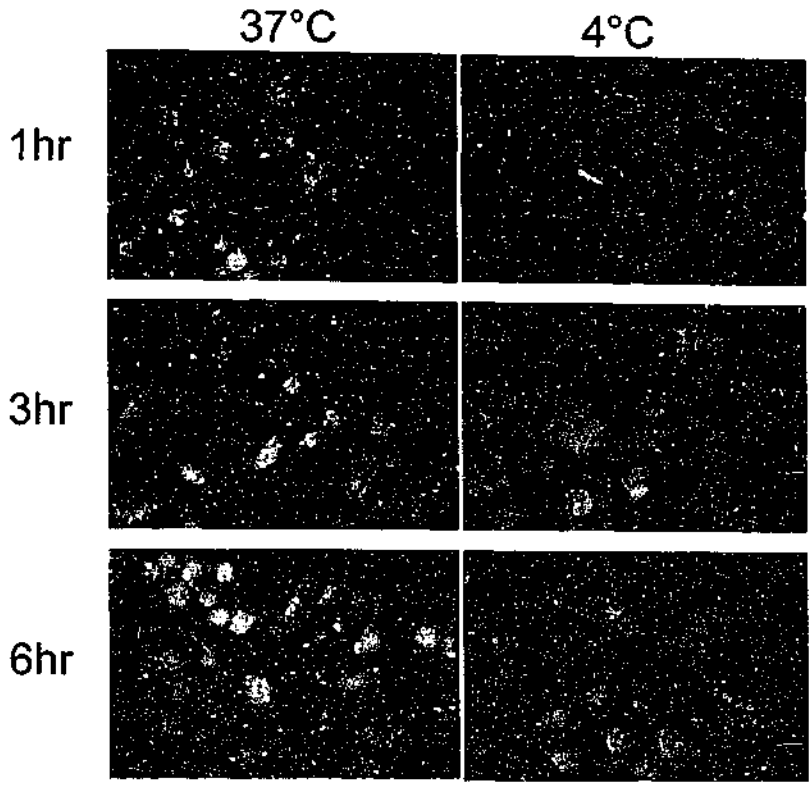

Figure 3.17: Infection of J774 macrophages with *gfp* expressing *C. pseudotuberculosis*.

J774 macrophages were infected at MOI 5 with C231(pSM22) at either 4°C or 37°C. 4°C samples were maintained at this temperature until 3 hr post infection at which time they were transferred to 37°C. At 1, 4 and 6 hours post infection cells were fixed in paraformaldehyde, counterstained with propidium iodide and viewed using an Olympus BH-2 fluorescence microscope. Images were taken with an Olympus C-35AD-4 camera.

phagocytosed by the macrophages. This is apparent at both MOI with approximately 6 and 3.5-fold fewer bacteria being taken up at MOI 4 and MOI 16 respectively. For samples incubated at 37°C the bacteria could be seen to be replicating within the macrophages as determined by increased bacterial cell counts at later timepoints. The rate of bacterial replication appeared to be slightly greater in cells infected at MOI 16 than MOI 4.

To monitor the experiment visually a similar approach was taken except that the macrophages were plated on glass coverslips that had been placed on the bottom of the wells of 6 well plates. A *gfp* expressing strain of *C. pseudotuberculosis* termed C231(pSM22) was utilised (Section 4.9.3). The infection experiment was performed exactly as for the first experiment using a MOI of 5. At various timepoints the coverslips were removed from the 6 well plates, fixed, permeabilised then counterstained with propidium iodide. Mounted coverslips were viewed by fluorescence microscopy (Figure 3.17). After infection for 1 hr at 37°C a proportion of macrophages were infected with one or more bacteria. At later timepoints the number of clumps of bacteria tended to remain constant however the number of bacteria in each clump tended to increase which is indicative of bacterial replication. At 4°C the number of bacteria phagocytosed by the macrophages was less and the rate of replication of those bacteria that were phagocytosed was less as indicated by reduced numbers of bacteria in the bacterial clumps observed at the later timepoints.

The macrophage infection model was used extensively throughout experiments described in this thesis. Although the basic principle of the assay always remained the same modifications were made depending on what type of experiment was being performed. In particular the MOI used and the experimental timepoints were altered depending of the experimental aims.

3.7.2 Preparation of bacterial RNA for array probing

20 µg of bacterial RNA is required for probing an array. In order to generate sufficient bacterial RNA from an infected monolayer it was necessary to perform large scale infections using a high MOI. For infection experiments J774 cells were plated in 100 mm dishes at 1.2 $x10^7$ cells/dish, 18 hr prior to infection. Exponential phase bacterial cultures were sonicated to break up clumps and then added to the monolayer at a MOI of 60. Following an invasion period of 1 hr the bacterial suspension was removed from the cells, the cells washed 3 times in

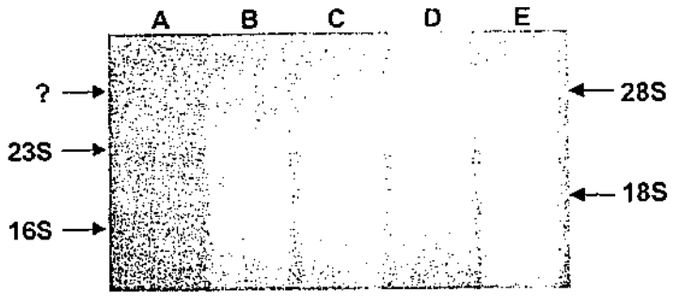

Figure 3.18: Analysis *C. pseudotuberculosis* RNA extracted from infected J774 macrophages by agarose gel electrophoresis.

Macrophages were infected at MOI 60 with *C. pseudotuberculosis*. At 6 hr post infection the monolayer was lysed, collected by centrifugation and bacterial RNA extracted (B and C). RNA was extracted in the same manner from uninfected macrophages (A) and from bacteria grown for the same period of time in DMEM (D). J774 macrophage RNA was also extracted and analysed (E). For each sample 2 µg of RNA was analysed on a non-denaturing agarose gel. The size of the bacterial and eukaryotic ribosomal bands is indicated to the left and right of the gel respectively. The identity of the high molecular weight band (?) present in lanes B to D is not known. It may correspond to a conglomeration of rRNA molecules or a pre rRNA species. This band is also seen in RNA samples analysed in Figure 3.11.

phosphate buffered saline (PBS) and then DMEM containing 100 µg/ml gentamycin was added. After a further 5 hours the monolayer was washed in ice-cold PBS and then lysed to release the bacteria by the addition of 3 ml of ice-cold 1% (v/v) Triton X-100 in PBS. Following a 5 min incubation the monolayer was scraped to dislodge attached cells, transferred to a 10 ml tube and pelleted by centrifugation. The supernatant was removed and RNA extracted from the pellet. To generate the *in vitro* control sample, the same number of bacteria were inoculated into 8 ml DMEM in a 100 mm dish at the start of the experiment. After 6 hr the bacteria were collected by centrifugation and RNA extracted. By performing a cell count on the monolayer lysate it was estimated that at the 6 hr timepoint there were approximately 75 bacteria per macrophage. To ensure that the extraction procedure isolated predominantly bacterial rather than macrophage RNA an identical extraction procedure was performed on uninfected J774 cells. This yielded an essentially undetectable amount of RNA. In addition when the RNA extracted from infected monolayers was run alongside J774 RNA on an agarose gel it could be clearly seen that the ribosomal bands in the two samples were of

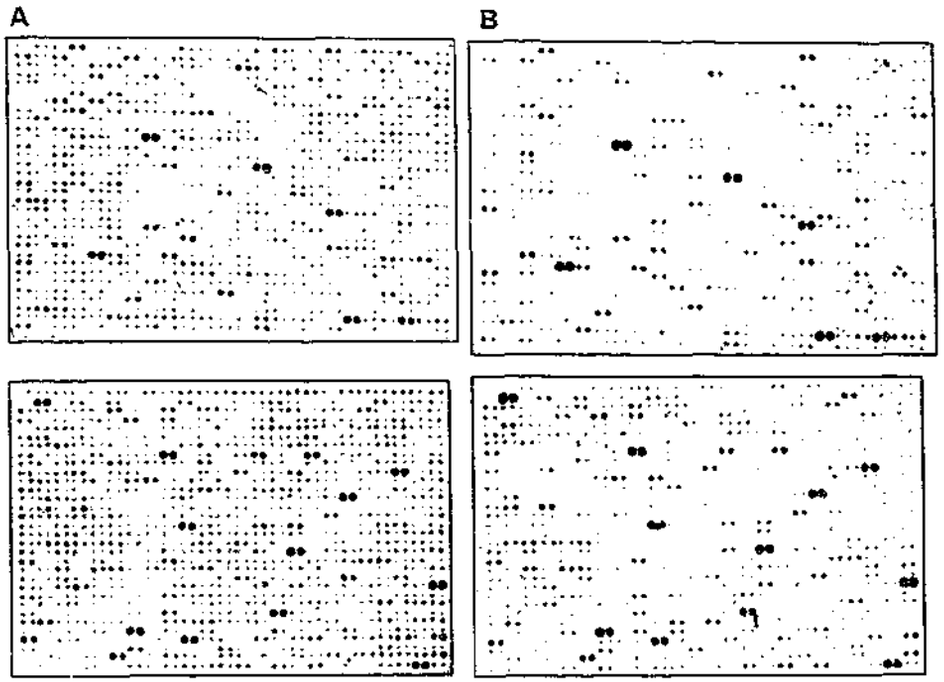

Figure 3.19: pUC18-Q DNA arrays probed with bacterial RNA derived from J774 macrophages infected with *C. pseudotuberculosis*.

J774 macrophages were infected with *C. pseudotuberculosis* at an MOI 60 and an equivalent amount of bacteria was also inoculated into DMEM. After 6 hr bacterial RNA was extracted from the infected monolayer and the DMEM culture. RNA was labelled with DIG and then used to probe replica pUC18-Q DNA arrays. Bound target was detected using an anti DIG AP conjugate and the AP substrate CDP-Star. (A) macrophage infected and (B) DMEM control.

different sizes reflecting the differences in prokaryotic and eukaryotic ribosomal RNA (Figure 3.18). Additionally this analysis showed that it had been possible to extract relatively good quality bacterial RNA from the infected macrophage monolayer.

3.7.3 Probing of the DNA arrays

Replica arrays were probed overnight at 50°C with 20 µg of DIG labelled RNA derived either from infected macrophages or from *C. pseudotuberculosis* C231 grown in DMEM. Following stringency washes the bound target was detected with an anti DIG antibody with an AP conjugate and subsequent incubation with the AP substrate CDP-Star (Figures 3.19 and 3.20).

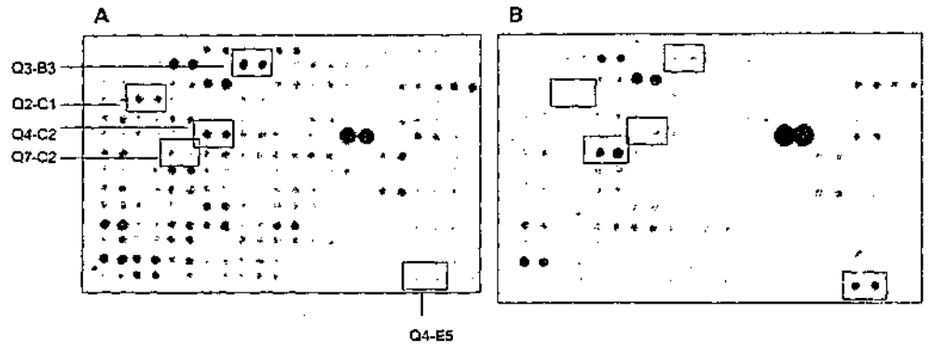

Figure 3.20: Close-up analysis of pUC18 Q DNA arrays probed with bacterial RNA derived from J774 macrophages infected with *C. pseudotuberculosis*.

A magnified region of the arrays presented in Figure 3.19 is shown. Clones which were identified as being regulated are boxed and the clone identification shown. A and B correspond to DMEM control and macrophage infected respectively.

3.7.4 Clone analysis

The inserts of 25 clones that may have been differentially regulated were sequenced and putative gene function assigned (Table 3.5). All of the clones that had been identified during the heat shock experiment as containing heat shock genes were again identified as up-regulated clones. However, none of the other genes that had been shown to be thermoregulated were identified as being macrophage regulated. A number of clones containing genes involved in metabolic functions, iron homeostasis and protection against microbicidal agents were identified. Additionally a number of hypothetical genes for which no gene function has been assigned were also identified.

3.7.5 Quantitation of gene regulation

Real time PCR primers were designed for each of the genes of interest (Table 2.5). Reverse transcription and real time PCR were performed in an analogous manner as for the heat shock study. The majority of genes listed in Table 3.5 were analysed. As indicated by their expression on the macrophage infected RNA, the heat shock genes showed the greatest induction (Table 3.6). A number of other genes were also up-regulated, but to a much smaller degree. These included a putative metallopeptidase, an acetyl coA carboxylase subunit, a 50S ribosomal protein and a member of the universal stress protein family. A number of other

Table 3.5: Sequence analysis of clones from the pUC18-Q library that were regulated following macrophage infection.

Table 3.5: Sequence analysis of clones from the pUC18-Q library that were regulated following macrophage infection.

Clone	regulation[a]	insert size	Species and % identity[b]	gene homology[b]
Q1-D10	down	1.1	*C. glutamicum*, 56%	ABC-type cobalamin/ Fe^{3+} siderophore permease component
			Thermologa maritima, 43%	hypothetical iron (III) ABC transporter
Q1-F11	up	0.8	*M. paratuberculosis*, 85%	*dnaK*
Q2-B10	up	0.9		no significant homologies
Q2-C1	down	1.5	*Haemophilus influenzae*, 68%	*pfl* (formate acetyl transferase)
Q2-D10	up	1.0	*M. tuberculosis* , 79%	translation initiation factor IF-3
			M. tuberculosis , 63%	50S ribosomal protein L35
Q2-G5	up	1.2	*C. glutamicum*, 70%	50S ribosomal protein L11
			C. glutamicum, 70%	50S ribosomal protein L1
Q3-B3	down	1.1	*N. meningitis*, 47%	*nor* (nitric oxide reductase)
Q4-C2	down	1.0	*C. glutamicum*, 81%	cytochrome c1
			C. glutamicum, 72%	rieske iron-sulphur protein
Q4-E5	up	1.9	*C. flavescens*, 78%,	aspartokinase
			M. paratuberculosis, 79%	*dnaK*
Q4-G4	down	0.9	*C. glutamicum*, 81%	rieske iron-sulphur protein
			C. glutamicum, 86%	cytochrome B subunit
Q5-E10	up	0.9	*M. paratuberculosis*, 84%	*dnaK*
Q7-C2	up	1.3	*C. glutamicum*, 76%,	*clpB*
Q8-F11	down	3.0	*Saccharopolyspora erythraea*, 37%	putative lysozyme
			Staphylococcus aureus, 49%	hexose phosphate transport protein
Q9-C5	down	0.8	*Clostridium pasteurianum*, 50%	*pfl* (formate acetyl transferase)
Q10-B8	down	1.2	*Zymomonas mobilis*, 60%	*pfl* (formate acetyl transferase)

table continues on next page

Clone	regulation[a]	insert size	Species and % identity[b]	gene homology[b]
Q10-B11	down	ND	*Cavia porcellus*, 24%	anion exchange protein
Q10-G2	up	ND	*C. glutamicum*, 76%	homoserine kinase
			C. glutamicum, 75%	phosphoenol pyruvate sugar phosphatase
Q11-F6	down	1.5	*M. tuberculosis*, 64%	*lpdA* (dihydrolipoamide dehydrogenase)
Q11-G7	up	1.2	*M. paratuberculosis*, 75%,	*dnaK*
Q12-C5	up	ND	*M. tuberculosis*, 51%,	*dnaJ*
			M. tuberculosis, 57%	*hspR*
Q12-E9	up	1.1	*Streptomyces coelicolor*, 32%	metallopeptidase
			Streptococcus pneumoniae, 63%	dTDP-glucose 4,6-dehydratase
Q13-B9	down	2.0	*C. glutamicum*, 94%	Hypothetical protein (metallo-beta-lactamase superfamily)
			S. coelicolor, 48%	aminopeptidase
Q15-A3	up	1.3	*M. tuberculosis*, 62%	*sodC* (Cu,Zn superoxide dismutase)
			C. glutamicum, 54%	*accD* (acetyl CoA carboxylase)
Q15-G3	up	4.0	*C. diphtheriae*, 100%	*chrA* (heme response regulator)
			B. subtilis, 27%	Conserved hypothetical protein
Q16-A7	up	2.0	*Clamydophila pneumoniae*, 31%	Conserved hypothetical protein
			M. tuberculosis, 67%	hypothetical protein Rv2026c (universal stress protein family)

[a] regulation refers to whether clones were identified as up or down-regulated following macrophage infection (compared to DMEM grown control).

[b] Sequences were scanned against the NCBI protein databases for protein homologies in all 6 reading frames using a translating BLAST program. (BLASTX). Proteins that showed the highest homology and the species in which this occurred are listed.

ND (not determined).

Table 3

Clone	[a] fold in array a
Q12-C5	
Q1-F11	
Q5-E10	
Q7-C2	
Q12-C5	
Q15-A3	
Q2-G5	
Q12-E9	
Q10-B1	
Q16-A7	
Q2-B10	
Q15-G3	
Q11-F6	
Q15-A3	
Q2-D13	
Q4-C2,	
Q10-G2	
Q13-B9	
Q3-B2,	
Q13-B9	
Q4-C2	
Q10-G2	
Q3-B3	
Q8-F11	
Q2-C1,	
Q10-B8	
Q1-D10	

Table 3.6: Induction of genes following macrophage infection.

Clone	gene	fold induction[a]
Q12-C5, Q12-G1	*dnaJ*	117
Q1-F11, Q4-E9, Q5-E10, Q11-G7	*dnaK*	93
Q7-C2	*clpB*	87
Q12-C5	*hspR*	36
Q12-E9	metallopeptidase	4.6
Q2-G5	50S ribosomal protein L11	4.22
Q15-A3	accD (acetyl CoA carboxylase)	3.39
Q10-B11	Low homology to anion exchanger (long ORF)	3.23
Q16-A7	conserved hypothetical protein Rv2026c (universal stress protein family)	3.0
Q2-B10	hypothetical protein	2.5
Q15-G3	*chrA* (heme response regulator)	2.34
Q11-F6	*lpdA* (dihydrolipoamide dehydrogenase)	1.91
Q15-A3	*sodC* (Cu,Zn Superoxide dimutase)	1.77
Q2-D10	translation initiation factor IF-3	1.5
Q4-C2, Q4-G4	rieske iron-sulfur protein	0.96
Q10-G2	homoserine kinase	0.95
Q13-B9	aminopeptidase	0.87
Q3-B2, Q2-G12	*pld*	0.84
Q13-B9	hypothetical protein	0.67
Q4-C2	cytochrome c1	0.57
Q10-G2	phosphoenol pyruvate sugar phosphatase	0.4
Q3-B3	*nor* (nitric oxide reductase)	0.39
Q8-F11	hexose phosphate transport protein	0.24
Q2-C1, Q9-C5, Q10-B8	*pfl* (formate acetyl transferase)	0.23
Q1-D10	ABC-type cobalamin/ Fe^{3+} siderophore permease component	0.14

[a] fold induction was determined on the same RNA samples as those used to probe the DNA array and was determined by RT-real time PCR.

(BLASTX). Proteins that showed the highest homology and the species in which the sequence is found.
ND (not determined).

genes were down-regulated including genes encoding an iron permease homologue, formate acetyl transferase and a hexose phosphate transport system.

3.8 Discussion

Transcriptional profiling on a macro scale was suitable for the identification of regulated genes of *C. pseudotuberculosis*. In particular genes that were thermoregulated, expressed in a cell density dependent manner or regulated during macrophage infection were identified. Although the DNA array utilised in this study was not representative of the entire genome, a sufficient number of regulated genes were identified to provide some information regarding gene regulation in *C. pseudotuberculosis* and to provide candidate genes for further study (Chapters 5 and 6).

The DNA array contained 3072 spots, corresponding to 1536 clones. DNA was however not successfully extracted from all clones, thus the size of the array was reduced by approximately 8%. The size of the *C. pseudotuberculosis* genome is unknown, however is likely to be similar to other corynebacterial species. The chromosome sizes of *C. glutamicum* and *B. lactofermentum* are each approximately 3.1 Mb (Correia *et al.*, 1994;Bathe *et al.*, 1996) while the *C. diphtheriae* genome is approximately 2.5 Mb (ftp://ftp.sanger.ac.uk/pub/cdip). Assuming that the genome size of *C. pseudotuberculosis* is also approximately 3.1 Mb, and that the average insert size is 1.3 kb, then the library represents an approximately 0.6-fold coverage of the genome. Many of the clones sequenced contained plasmid inserts that appeared to contain more than one DNA fragment suggesting that concatamers had formed during the ligation process. The formation of concatamers may have been enhanced by performing the ligation in the presence of polyethylene glycol 6000 (PEG), which is present to increase the ligation efficiency (Hayashi *et al.*, 1986). The need for an arrayed random library to have several times coverage of the genome for it to be representative was demonstrated by the redundancy of the generated array. This was observed as a number of genes being represented on two or more clones, for example *dnaK* and *pld*. Similarly other heat shock encoding genes, such as *groEL* and *grpE,* that would be expected to be regulated were not identified, presumably because there was no corresponding clone in the small arrayed library.

An initial problem with the developed array system was the lack of system sensitivity. This was reflected in the relatively few positives detected when using small amounts of RNA in the target. This is a problem inherent with all array experiments performed on bacterial genomes and stems from the inability to easily enrich for the mRNA population. The usual manner in which this is overcome is to use a large amount of target. Alternatively when the genome sequence is known a reverse transcriptase reaction primed by sequence specific primers may be utilised. Removal of rRNA from total RNA using magnetic beads tagged with antisense rRNA sequences has been reported (Plum and Clark-Curtiss, 1994), but in general this type of approach has not been adopted by the scientific community for transcriptional profiling experiments.

In addition to increasing the amount of target used during hybridisation steps other modifications to the system were made to increase the system sensitivity. In particular the choice of hybridisation solution and AP substrate was found to be important for attaining maximal sensitivity. The only labelling system utilised was direct labelling of the target with DIG. Radioactively labelled target hybridised to nylon macroarrays has been shown to be at least as sensitive as fluorescently labelled cDNA hybridised to glass microarrays (Granjeaud *et al.*, 1999;Bertucci *et al.*, 1999). It is therefore likely that the use of a radioactive label in this system would have increased system sensitivity. Given the occupational health and safety risks associated with working with and storing radioactive materials this approach was not attempted. Also DIG labelled RNA or DNA can be stored and used over a long period unlike radioactive probes which are subject to degradation and loss of specific activity in a relatively short time period. An advantage of the DIG system utilised in this study was that the RNA could be directly labelled avoiding the need to perform a reverse transcriptase reaction.

Analysis of gene expression during heat shock identified 5 genes that were up-regulated and 4 genes that were down-regulated. All but one of the genes identified as being up-regulated belonged to the heat shock gene family. *dnaK*, *dnaJ* and *hspR* form part of the dnaK operon. In other closely related bacterial species the gene *grpE* is also found in this operon such that the gene order is *dnaK*, *grpE*, *dnaJ* and *hspR*. Sequencing across the *dnaK-dnaJ* region did indeed show that *grpE* was located between these genes in *C. pseudotuberculosis* (data not shown), thus indicating that the *dnaK* operon has the same gene order as in other gram-positive organisms such as *Mycobacterium* and *Streptomyces* (Cole *et al.*, 1998;Servant and

Mazodier, 2001). The first three genes of the operon code for proteins with chaperone function, while the last, *hspR* is a regulatory protein.

Two types of regulation of the *dnaK* operon have been described. In gram-negative species such as *E. coli* gene expression is transcriptionally activated by the alternate sigma factors, σ^{32} or σ^{E} (Wosten, 1998). The other type of regulation described involves transcriptional repression. Two types of system have been described. The first is the CIRCE/HrcA system of repression in which the repressor protein HrcA (heat regulation of CIRCE) binds to inverted repeats (termed Controlling Inverted Repeat of Chaperone Expression) found in the promoter regions of heat shock genes (Narberhaus, 1999). This type of regulatory system has been observed in a diverse range of eubacteria, including some gram-positive organisms (Narberhaus, 1999). The CIRCE element has been observed upstream of a number of different heat shock genes including *dnaK*, *groESL* and *clpB*. *hrcA* is generally the first gene of the *dnaK* operon.

The second type of transcriptional repression involves the negative regulator HspR, which acts with DnaK as a co-repressor of the *dnaK* operon. Control of *dnaK* operon expression by HspR has been described in *Streptomyces albus* G (Grandvalet *et al.*, 1997), *Streptomyces coelicolor* (Bucca *et al.*, 1997) and *M. tuberculosis* (Stewart *et al.*, 2001). In each case the gene order is as has been observed in this study for *C. pseudotuberculosis*. Two or more inverted repeats have been identified in the *dnaK* promoter region of these species that have been termed HAIR (for HspR associated inverted repeat). A model has been proposed in which DnaK and HspR bind together and act as the functional repressor at the HAIR sites. Repression is released following heat shock as DnaK now binds preferentially to denatured proteins, thereby releasing the repression of transcription and thus leading to an increase in the amount of detectable *dnaK* mRNA.

Sequence upstream of the *dnaK* gene of *C. pseudotuberculosis* has previously been identified (Simmons, 1997). Analysis of this sequence reveals the presence of two regions with homology to the consensus HAIR motif (Figure 3.21). It therefore seems likely that a similar mode of regulation of expression of the *dnaK* operon occurs in *C. pseudotuberculosis* as in *Mycobacteria* and *Streptomyces*. Additionally a region with high homology to the consensus *E. coli* σ^{70} promoter region has been identified (i.e. 4/6 and 5/6 at the –35 and –10 regions with

```
  1  ctggcggaaa tagaagattt taaaaaaatt gagtctagtg ggaacaactt
 51  tgcgtgagat gtcgttatgg atagtgtcag gttgagtgac acgcgctcaa
101  gatatctggt aagttcttca cttgtagttg gacggtcccg taaaattgag
151  tcgacgtcag tcaactttag gtaaaacgtt aaaaatacaa ccaaacaccc
201  aggaggacat cac atg gga cgc cca
                  M   G   R   P
```

Figure 3.21: The HAIR motifs of the *C. pseudotuberculosis dnaK* promoter region.

The *dnaK* promoter was cloned by Simmons (1997). Regions that show homology to the *E. coli* σ^{70} consensus –35 (TTGACA) and -10 (TATAAT) regions are boxed. A putative ribosome binding site is underlined. Amino acid sequence corresponding to the first four codons is indicated by the single letter amino acid code. Sequences that show homology to the consensus HAIR motif (5'-CTTGAGT-N7-ACTCAAG-3') (Grandvalet *et al.*, 1999) are bolded.

a spacing of 18bp) (Simmons, 1997). This further supports that expression of the *dnaK* operon is controlled by the normal σ factors and that repression is then applied to prevent inappropriate gene expression.

The other heat shock gene identified was *clpB*. ClpB belongs to the family of proteins called the Clp ATPases. As with the *dnaK* operon proteins, these proteins show a high degree of conservation across prokaryotic and eukaryotic species and are induced at high temperature and under a variety of other stresses such as oxidative stress and high salt concentration. In a number of *in vitro* studies ClpB has been shown to interact with DnaK, GrpE and DnaJ to suppress and reverse protein aggregation (Zolkiewski, 1999;Mogk *et al.*, 1999;Motohashi *et al.*, 1999). In *Streptomyces* and *M. tuberculosis, clpB* belongs to the *hspR* regulon as indicated by the presence of HAIR sequences in its promoter region. The sequence contained in clone Q7-C2 corresponds to an internal portion of the *clpB* gene. Therefore to determine whether this was also the case for *C. pseudotuberculosis* it would be necessary to first clone and then sequence the *clpB* promoter region.

The majority of other thermoregulated genes were down-regulated by heat shock. The three most down regulated genes were *pld, fas* and a third encoding an iron permease component.

The observation that the major virulence determinant of *C. pseudotuberculosis* was down regulated by heat shock was a novel finding. However, given that expression of virulence determinants is often tightly regulated, it is perhaps not surprising. Given that Pld is the major virulence determinant, further studies characterising its expression were performed. These studies are described in Chapter 6 and as such the functional significance of thermoregulation of *pld* expression is discussed there.

Immediately downstream of the *pld* gene but on the complementary strand is a gene coding for a protein with homology to an iron permease component found in *E. coli* (Ozenberger *et al.*, 1987) and *Yersina enterocolitica* (Schubert *et al.*, 1999) termed *fepC*. In these bacteria FepC along with the integral membrane proteins FepG and FepD comprise an enterobactin transport permease. In *E. coli fepC* along with *fepG* and *fepD* are likely to comprise an operon and as such, regulatory elements, including iron responsive ones, have been found upstream of *fepD*. The gene order in the operon is DGC. Immediately upstream of the *C. pseudotuberculosis* *fepC* homologue was located a second gene that showed homology to the *E. coli fepG*, indicating that the same gene order may exist in *C. pseudotuberculosis* as in *E. coli*. Wh writing this thesis, *C. pseudotuberculosis* sequence upstream of that generated during th study became available in GenBank. Four putative genes downstream of *pld* have been identified and termed *fagA*, *fagB*, *fagC* and *fagD*. The first three are on the strand complementary to *pld*, while *fagD* is located on the same strand. The order of the first three genes appears to be the same as in *E. coli* such that *fagA*, *fagB* and *fagC* correspond to *fepD*, *fepG* and *fepC* respectively. Based on sequence homology it is likely that *fagD* is a homologue of a gene found in *E. coli* termed *fepB* and is likely to code for a periplasmic binding protein that acts as the high affinity binding protein for the system. Also identified within the *C. pseudotuberculosis* sequence was a DtxR consensus sequence that lay between the divergently transcribed *fagA* and *fagD* genes. DtxR is an iron regulated transcription factor that represses gene expression in the presence of iron. DtxR was first identified in *C. diphtheriae* as a regulator of diphtheria toxin production, however homologues have subsequently been identified in a number of gram-positive bacteria. The presence of this site within the regulatory regions of these genes is unsurprising as many genes involved in iron homeostasis are metabolite regulated.

Corynebacterium possess a complex cell wall structure, and as such needs to be able to synthesise fatty acids for its maintenance. Most bacteria have only one fatty acid synthase (FAS), however some gram-positive bacteria have at least two. Two types of fatty acid synthase have been defined. The first is type I FAS (FASI), which is a multidomain enzyme that encodes all the activities required for fatty acid biosynthesis on a single large polypeptide. Type II FAS (FASII), the other type of FAS, consists of a series of independent enzymes. In some cases FASII uses FASI products as a substrate, such as in the production of mycolic acids by mycobacteria (Kremer *et al.*, 2000).

The FAS identified in this study shows homology to FASI. Studies of the closely related *C. ammoniagenes* have indicated that this species contains two FASI genes, which have been termed *fasA* and *fasB* (Stuible *et al.*, 1996). The *C. pseudotuberculosis* FAS identified shows greater homology to FasB than FasA protein (71% identity compared to 54% identity). In *C. ammoniagenes* FasA is predominant. Additionally strains in which *fasB* has been inactivated are phenotypically wild type whereas *fasA* mutants require oleic acid for growth (Stuible *et al.*, 1996). Thus the role of FasB in fatty acid synthesis is not entirely clear, however it may synthesise a subset of fatty acids that are not essential for survival under *in vitro* conditions. Whether *C. pseudotuberculosis* contains more than one FAS remains to be determined. To determine if, as in *C. ammoniagenes*, the *C. pseudotuberculosis* FAS plays a minor role in fatty acid biosynthesis it may be necessary to generate mutant strains and test for attenuation. Additionally purified protein could be generated and its enzymatic activity tested *in vitro*.

The identified thermoregulated genes belonged to a number of functional groups. The role of the heat shock proteins during thermal stress is well understood and primarily involves the prevention or reversal of heat-induced protein denaturation and aggregation that may be lethal to the cell (Lund, 2001). Modulation of membrane composition in response to temperature is a common theme for all organisms, allowing the maintenance of membrane fluidity (Magnuson *et al.*, 1993) and biochemical activity within the bilayer (Kremer *et al.*, 2000). It is therefore likely that the observed change in gene expression of *C. pseudotuberculosis fas* occurs to mediate this type of response. Additionally the expression of genes involved in fatty acid biosynthesis has been observed to decrease in response to reduced cell growth rate (Oh-Hashi *et al.*, 1986;Magnuson *et al.*, 1993), presumably as a mechanism of ensuring that an appropriate amount of fatty acid is produced. Given that the growth rate of *C.*

pseudotuberculosis is reduced at 43°C compared to 37°C, the reduced expression of *fas* could also be growth rate related.

Expression of proteins involved in iron acquisition is often tightly regulated such that they are only expressed in low iron concentration environments. Typically, this involves negative regulation by the iron-regulatory proteins Fur or DtxR (Braun and Killmann, 1999). Regulation by other mechanisms is much less common; however there have been some reports of thermal regulation of iron transport and homeostasis genes. These include a four gene operon of Group A *Streptococcus* which codes for the components of a ferrichrome transport system that is up-regulated at 40°C and 29°C compared to 37°C (Smoot *et al.*, 2001) and several genes involved in iron transport and storage in *Yersinia pestis* (Konkel and Tilly, 2000). Both of these pathogens are exposed to a range of temperatures during their infection cycle. It has been suggested that at the more extreme temperatures the availability of iron is less, hence repression of expression is removed. It is unlikely that *C. pseudotuberculosis* is exposed to long periods of thermal stress during its infection cycle. Firstly the route of infection is typically from the contaminated aerosols of one animal to the skin of a second (Ellis *et al.*, 1987;Paton *et al.*, 1995) and secondly although experimental infections are associated with a transient increase in temperature, this is not maintained for more than a few days (Pepin *et al.*, 1991b). Given this, the biological implications of down-regulation of an iron siderophore system during heat-shock are not clear.

A number of genes were also identified that were expressed in a culture density dependent manner. Regulation was confirmed by RT-real time PCR for *pld*, *fagC* and *fas*. Each of these genes was also thermoregulated, indicating that they are regulated by multiple mechanisms. A number of other genes were also identified, however, regulation was not confirmed at the RNA level by an independent method. A number of genes whose expression would be expected to be related to the growth rate of the cells were identified. In addition to *fas*, this included genes found on the two clones Q2-D10 and Q2-G5 which code for 50S ribosomal proteins L1, L11 and L35, in addition to translation initiation factor-3 (IF-3). A transcriptional profiling experiment of *E. coli* grown in rich medium or minimal medium has shown that of the 55 genes that encode ribosomal proteins, 42 were expressed more highly in the rich media (Tao *et al.*, 1999). Given that the bacteria undergo more rapid growth in rich media this is likely to be a growth rate effect. That these genes were identified as being down regulated at

high cell density is indicative that the rate of growth of the *C. pseudotuberculosis* culture has decreased at 6 hr in comparison to the earlier timepoints.

A macrophage infection model was successfully established, allowing the interaction between *C. pseudotuberculosis* and host cells to be studied. Additionally this study was successful in extracting bacterial RNA from the infected monolayer that was largely free of host RNA. In order to obtain a sufficient amount of RNA for array probing it was necessary to infect at a high MOI and use a relatively late timepoint post infection. The consequence of this is that genes that may be transiently expressed, for example, during phagocytosis of the bacteria would not be identified. The approach was however successful in identifying a number of macrophage induced genes. Genes belonging to the heat shock family showed the greatest degree of regulation. A number of other genes that were up-regulated were also identified, however the degree of regulation was small compared to that observed for the heat shock genes. A number of down-regulated genes were also identified.

The expression of heat shock genes of bacterial pathogens has been observed to be up-regulated following host cell infection with other bacterial species, including following infection of the human epithelial cell line, McCoy A with *Staphylococcus aureus (Qoronfleh et al., 1998)* and J774 macrophages with *Salmonella* (Buchmeier and Heffron, 1990) or *Yersinia enterocolitica* (Yamamoto *et al.*, 1994). Thus the observation that heat shock genes of *C. pseudotuberculosis* were strongly up-regulated is concordant with these studies. That a similar response occurs in the animal host is suggested by the observation that heat shock proteins are immunodominant antigens for a number of infectious organisms including *M. leprae, M. tuberculosis* (Young *et al.*, 1988), *Legionella pneumophilia* (Hoffman *et al.*, 1989), *S. aureus* (Qoronfleh *et al.*, 1993), *S. typhi (Tang et al., 1997)* and *Leptospira* (Guerreiro *et al.*, 2001). Although heat shock genes are up-regulated during infection, their precise role in virulence is at this stage not well defined. Presumably expression of heat shock proteins by intracellular pathogens provides protection against stresses such as changes in pH and osmolarity and protein damaging agents in a manner analogous to that observed in *in vitro* models. Additional roles for some of the heat shock proteins have also been defined in studies using mutant strains. For example, DnaK has been implicated to play a role in phagocytosis of *L. monocytogenes* by macrophages (Hanawa *et al.*, 1999).

A key microbicidal mechanism of activated phagocytic cells is the production of reactive oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI). Activation of phagocytic cells leads to a respiratory burst in which superoxide is produced and the up-regulation of inducible nitric oxide synthase (iNOS)synthesis occurs, leading to increased production of nitric oxide. These molecules and their products may have a deleterious effect on the pathogen by causing DNA, protein and lipid damage. In order to counter this host response, bacteria possess enzyme systems for removal of the free radical generating compounds. In particular superoxide dismutase and nitric oxide reductase are able to reduce the load of ROI and RNI.

Clone Q3-B3, which correlated to a down-regulated spot on the array (following macrophage infection) was found to contain sequence that showed homology to a nitric oxide reductase (NOR) of *N. meningitidis*. The apparent down regulation was confirmed using RT-PCR which indicated a 2-3 fold reduction in expression had occurred. Given the role of NOR in protecting bacteria from phagocyte produced RNI, this was a surprising observation. Although down regulated, some *nor* expression was still apparent in intracellular *C. pseudotuberculosis*.

Macrophages from different species differ considerably in their ability to synthesise NO. Both mouse and bovine macrophages synthesise increased amounts of iNOS following activation (Adler *et al.*, 1996;MacMicking *et al.*, 1997). Conversely, the level of detectable NO in activated ovine or caprine macrophages does not increase following cell activation (Bogdan *et al.*, 1997;Adler *et al.*, 1996). If a similar situation exists *in vivo* then the lack of up-regulation of *nor* expression may be a reflection of a non requirement for large amounts of this enzyme in the native host. An alternative explanation may be that the macrophages were not sufficiently activated during the course of the infection experiment to up-regulate iNOS synthesis. In order to ensure that the macrophages are sufficiently activated treatment with interferon-γ prior to infection could be performed.

Clone Q15-A3 contained sequence with homology to 2 proteins. At the 5' end of the insert was the 3' end of a gene with homology to *sodC* which codes for a Cu,Zn superoxide dismutase (Cu,ZnSOD). Superoxide dismutases catalyse the dismutation of superoxide radicals to oxygen and hydrogen peroxide. Three classes of SOD have been defined based on

the metal ion present at the active site which can be (i) iron (Fe), (ii) manganese (Mn) and (iii) copper and zinc. Fe and Mn SOD have an intracellular localisation where they participate in the removal of superoxide generated during aerobic metabolism. Conversely Cu,ZnSODs usually possess a signal peptide for protein export, with gram-positive bacteria possessing an additional lipid attachment site. This results in a periplasmic localisation for Cu,ZnSOD of gram-negative bacteria and an external, membrane anchored location for Cu,ZnSOD of gram-positive bacteria (D'Orazio *et al.*, 2001). Given that superoxide is unable to cross the cytoplasm it has been proposed that the role of Cu,ZnSOD is to protect bacteria from extracellular superoxide, in particular that produced during the respiratory burst by phagocytic cells.

sodC genes have been identified in a number of pathogenic and non-pathogenic, gram-negative and gram-positive bacteria, including *Actinobacillus pleuropneumoniae* (Langford *et al.*, 1996), *S. typhimurium* (Farrant *et al.*, 1997), *Neisseria meningitidis* (Wilks *et al.*, 1998), *M. tuberculosis* (Cole *et al.*, 1998) and *Haemophilus ducreyi* (San-Mateo *et al.*, 1999). For a number of these species, the role of Cu,ZnSOD has been investigated both *in vitro* and *in vivo*. A number of *sodC* mutants have been generated, all of which when tested, have shown increased sensitivity to exogenous superoxide in culture (Farrant *et al.*, 1997;Wilks *et al.*, 1998;Sheehan *et al.*, 2000;Dussurget *et al.*, 2001;Piddington *et al.*, 2001). However, varying effects of Cu,ZnSOD on virulence, ranging from having a significant effect to no effect, have been observed in both cell culture and animal models. Not only is there species variation, but with at least two species conflicting data has been generated (Tatum *et al.*, 1992;Latimer *et al.*, 1992;Dussurget *et al.*, 2001;Piddington *et al.*, 2001). The differences observed between two studies in *M. tuberculosis* may be indicative of the cell culture models utilised and the degree of the respiratory burst produced by each of those models (Piddington *et al.*, 2001). Interestingly *sodC* mRNA was shown by RT -PCR to be up-regulated in *M. tuberculosis* following macrophage infection, but the degree of up-regulation was not clear (D'Orazio *et al.*, 2001).

Following the *sodC* gene on clone Q15-A3 was an intergenic region, followed by a gene coding for a putative acetyl coA carboxylase (ACC) subunit. The gene identified showed high homology to the β chain of putative ACC carboxyltransferases in *C. glutamicum* and *M. tuberculosis,* which is generally coded by *accD*. In bacteria ACC is part of the family of

carboxylases that use biotin as a cofactor and bicarbonate as the source of the carboxyl group and is generally comprised of four subunits that are encoded by the genes *accB*, *accC*, *accA* and *accD*. These code for a biotin carboxyl carrier protein (BCCP), biotin carboxylase and the α and β subunits of carboxyltransferase respectively. In a two- step reaction, BCCP is first carboxylated at a covalently bound biotin residue in a reaction catalysed by biotin carboxylase. The activated CO_2 is then transferred to an acceptor molecule (acetyl coA) in a reaction catalysed by carboxyltransferase to generate malonyl coA. The location of *accD* in relation to other *acc* genes is quite variable between bacterial species. For example in *E. coli accB* and *accC* are co-located while *accA* and *accD* are found in separate parts of the chromosome (Li and Cronan, 1992), while in *Lactobacillus planarum* the genes are co-transcribed (Kiatpapan *et al.*, 2001). The gene order within the region of *accD* does not appear to be conserved between *C. pseudotuberculosis*, *C. glutamicum* or *M. tuberculosis*.

The α and β subunits of the carboxyltransferase subunits of this family of carboxylases tend to be highly conserved. Therefore, although it is likely that the identified *accD* gene is part of ACC, it is possible that it is in fact a subunit of a carboxyltransferase from another carboxylase. That is, it could have different substrate specificity, for example propionyl coA or methylmalonyl coA instead of acetyl coA. Additionally, some carboxylases have multiple substrates and as such are able to carboxylate different substrate molecules. To assign a definitive function to the protein it may be necessary to purify this protein and others of the ACC complex and determine what activity is present.

ACC is an important component of the type II FAS system and catalyses the first reaction of this fatty acid biosynthesis pathway. The importance of this system to bacterial survival is indicated by the observation that attempts to generate chromosomal knockouts of ACC components in *C. glutamicum* (Jager *et al.*, 1996) and *M. bovis* (Norman *et al.*, 1995) were not entirely successful. This suggests that ACC is essential for bacterial viability. *accD* was observed to be up-regulated following macrophage infection. As discussed earlier, genes involved in fatty acid biosynthesis are often regulated such that appropriate membrane fluidity and composition are maintained and the required amount of fatty acids are produced in relation to the growth rate of the bacteria. Up-regulation of *accD* in this study could be in response to an altered growth rate within the intracellular macrophage environment.

Iron is an essential element for all bacteria because of its role as an essential cofactor for many enzymes such as those in electron transport chains. Although iron is abundant in the environment, the concentration of available soluble iron is extremely low, thus bacteria have evolved a number of different mechanisms for scavenging iron. Additionally for pathogenic bacteria, mammalian hosts possess proteins that bind iron, keeping it in a soluble form that is sequestered from the bacteria. Mammalian iron binding proteins include transferrin, lactoferrin and ferritin, which are iron carriers in the blood, secretory fluids and cells respectively.

Some bacteria have systems for transporting iron or bound iron directly into the cytoplasm, usually by way of active transport via specific ABC transporters. The ability to take up Fe^{2+} and Fe^{3+} in their free form is generally restricted to bacteria in certain niches in which iron is more soluble, for example, low pH environments. When direct uptake of iron is not possible bacteria may synthesise and export siderophores for iron scavenging. Siderophores are low-molecular weight iron ligands that are able to compete with host proteins for binding of available Fe^{3+}. Upon binding of Fe^{3+} the siderophore binds to the bacterial cell surface (to a membrane bound lipoprotein) and is then actively transported across the membrane via an ABC transporter. Once in the cytoplasm, the iron is rapidly released from the siderophore. The FagC iron permease component identified as being thermoregulated belongs to such a transport system.

Several genes involved in iron acquisition and metabolism by *C. pseudotuberculosis* were identified as being macrophage regulated. These included a second putative iron permease homologue and a heme response regulator *chrA*. Although *fagC* and *fagB* were not found to be up-regulated following macrophage infection, this in part reflects the chosen control. Analysis of *fagC* expression in low and high density cultures indicates that it is up-regulated at high density (at least in BHI). Assuming that the same density dependent expression occurs when *C. pseudotuberculosis* is cultured in DMEM, then up-regulation following macrophage infection would have been observed if a low density DMEM culture had been used as the *in vitro* control. The issue of selection of the most appropriate *in vitro* controls for experiments such as this is discussed in Chapter 7.

Clone Q15-G3 contained sequence that showed complete homology at the amino acid level with the 3' end of a gene encoding a transcriptional regulator from *C. diphtheriae* termed ChrA. In *C. diphtheriae* ChrA comprises part of a two-component signal transduction system that activates gene expression in response to heme or haemoglobin (Schmitt, 1999). Upstream of the *C. diphtheriae chrA* gene is a second gene that has been termed *chrS*. *chrS* has the properties of a sensor histidine kinase. Upon binding of extracellular heme or haemoglobin to ChrS, the two-component signal transduction system becomes activated leading to activation of the ChrA. Via this system the promoter of a gene encoding heme oxygenase (*hmuO*), the enzyme that catalyses the degradation of heme to release iron, becomes activated. To complement this system *C. diphtheriae* also has a heme transport system to deliver the substrates for HmuO to the cytoplasm (Drazek *et al.*, 2000).

The third clone containing putative iron-related genes was Q1-D10. Located on this clone was a gene encoding a protein that showed 56% identity to a predicated ABC-type cobalamin/Fe^{3+} siderophore system permease component in *C. glutamicum*. The protein also showed homology to hemin transport proteins and *fecCD* family proteins of a variety of bacterial species. Thus there is little doubt that the putative protein is part of an ABC transporter system involved in the acquisition of iron, however, the identity of the molecule that is transported is not clear. Immediately following the identified gene was the start of a second gene that showed homology to the ATP-binding component of various ABC transporters. This indicates that the insert sequence found in Q1-D10 is probably part of an operon comprising the elements of an ABC transporter. The first gene of this clone was shown to be strongly down-regulated following macrophage infection. This was a surprising observation given that the other iron homeostasis genes identified were either up-regulated or did not show a change in expression, suggesting that their expression was essential for iron acquisition in the macrophage intracellular environment. It may be that iron source transported by this permease is in excess in the intracellular environment compared to in DMEM, hence the requirement for that molecule is reduced and expression of the relevant genes also decreases.

This study has shown that *C. pseudotuberculosis* possesses several different mechanisms to obtain iron. No doubt these have evolved to allow the maximum amount of iron to be scavenged from the environment, thereby maximising bacterial viability.

Apart from the heat shock genes the most up-regulated gene following macrophage infection was that encoding a putative metallopeptidase. Clone Q12-E9 contained a gene encoding a protein that showed homology to a putative zinc metallopeptidase of *S. coelicolor*. This class of protease require zinc at the catalytic site for proteolytic activity and as such the protein identified contains a zinc-binding motif. Proteases play an important role in normal cellular processes, but for some pathogens they may act as virulence determinants. Many metalloproteases are secreted and as such their targets are host proteins. Alternatively some metalloproteases cleave bacterially derived proteins, resulting in the activation of that protein. This for example occurs in *V. cholerae* where the A subunit of cholera toxin is cleaved by a metalloprotease leading to its activation (Booth *et al.*, 1984). Metalloproteases have been shown to have diverse effects on the host ranging from tissue damage, cytotoxicity, enhancement of vascular permeability and facilitation of invasion (Miyoshi and Shinoda, 2000). That the identified metallopeptidase was up-regulated during macrophage infection suggested that its substrate may be of host origin. Clone Q12-E9 contained the 3' end of the metallopeptidase gene. It was therefore not possible to determine whether this was a secreted protein. Given the degree of up-regulation of this protein during macrophage infection, it was chosen as a candidate for further study. This work is described in chapter 5.

A number of genes involved in metabolic processes were identified. Of particular interest were three clones containing a gene with homology to formate acetyl transferase that were identified as being down-regulated. This enzyme catalyses the non-oxidative dissimilation of pyruvate to formate and acetyl coA under anaerobic conditions. Pyruvate is metabolised by a different pathway under aerobic conditions, such that pyruvate is oxidatively decarboxylated by the pyruvate dehydrogenase complex. A component of this complex, lipoamide dehydrogenase component (E3) was identified as being up-regulated following macrophage infection. These observations suggest that at the timepoint investigated, *C. pseudotuberculosis* grown in DMEM are under more anaerobic conditions than those in the macrophage. This is perhaps surprising, but may reflect that the *in vitro* control cultures were not shaken, were grown in a high CO_2 environment and grew faster than the bacteria resident within the macrophages, thus potentially consuming available oxygen more quickly.

The results presented in this chapter have demonstrated that even when used on a small scale, macroarray technology can be utilised successfully to identify regulated genes. Additionally,

through this study the amount of sequence available for *C. pseudotuberculosis* has increased substantially. Thus, although not every clone identified was found to be regulated, information gained through sequencing has given insights into the functionality of *C. pseudotuberculosis*.

From this study four genes were chosen for further analysis. The role of metallopeptidase, Cu,ZnSOD and FagC in macrophage infection are investigated in Chapter 5 via the generation of mutant strains. Although the effect of *pld* on virulence of *C. pseudotuberculosis* has been comprehensively demonstrated, nothing is known about how the gene is regulated and how this may impact on its function *in vivo*. In order to address some of these issues an investigation of *pld* regulation was also undertaken. This work is described in Chapter 6.

Chapter 4

Identification of differentially regulated genes using Differential Fluorescence Induction

4.1 Introduction

A well developed premise in the study of bacterial pathogens is that genes that are up regulated during infection of the host are likely to play an important role in disease pathogenesis. From this premise have arisen a number of techniques that specifically identify *in vivo* regulated genes. Some of these techniques such as signature tagged mutagenesis are only applicable to truly *in vivo* studies and as such require the use of an animal model of infection. Others such as DD-PCR and DFI may additionally be used in other experimental systems, for example models that mimic aspects of infection such as heat shock or changes in pH.

In part, the reason little is known about the regulation of virulence determinants in *C. pseudotuberculosis* is that few suitable systems for studying gene regulation have been developed for this species. *lacZ* has been used in reporter constructs (Zhang *et al.*, 1994) but is limited in its use, as the bacteria must undergo severe mechanical disruption to release the β-galactosidase activity. This obviously precludes obtaining quantitative measurements of gene expression in real time. Given that Gfp has been shown to be a useful reporter in a variety of bacterial species including mycobacteria, it was likely that it could be adapted to monitoring gene expression in *C. pseudotuberculosis*.

DFI was chosen as a complementary technique to the transcriptional analysis approach described in Chapter 3. In comparison to the array technology, DFI has the potential to screen large libraries relatively easily. DFI may also be more suitable for identifying macrophage induced genes as the difficulties in extracting large amounts of RNA for array probing can be avoided. In addition the tools generated during this study, such as the promoter probe vectors with *gfp* as a reporter, could be and indeed were used in other aspects of this study.

This results chapter details the design, construction and optimisation of a promoter probe vector containing *gfp* as a reporter. A library, generated by insertion of *C. pseudotuberculosis* genomic fragments into the vector was used in conjunction with DFI to identify genes up regulated during heat shock and macrophage infection. Although it was not possible to identify thermoregulated genes a number of macrophage induced genes were identified, thus providing targets for further study.

4.2 Design and construction of a promoter probe vector

The key features required in a promoter probe vector for DFI are (i) a *gfp* gene as a reporter and (ii) the presence of appropriate upstream sequence required for protein translation and for the cloning of genomic fragments that are to be screened for promoter activity. Additional features of the promoter probe vector designed for this study are transcriptional terminators both up and downstream of the *gfp* gene. The purpose of the upstream transcriptional terminator is to ensure *gfp* expression is not driven by promoters in the plasmid other than those inserted upstream of the *gfp* gene. The purpose of the transcriptional terminator downstream of the *gfp* gene is to ensure termination of the *gfp* transcript. The construction of the promoter probe vector is described in the following section and is represented schematically in Figure 4.1.

The parent plasmid used for the construction of the promoter probe vector was pEP2MCS, which is an *E. coli* - *C. pseudotuberculosis* shuttle vector. It was therefore possible to perform vector and library construction in *E. coli*. pEP2MCS is kanamycin resistant and contains a *repA* gene, which is required for replication in both hosts (Zhang *et al.*, 1994). This plasmid contains a limited number of restriction endonuclease sites for cloning, including *Dra*III, *Bgl*II, *Xba*I, *Sac*I, MunI, *Sal*I, *Pvu*II and *Age*I. To increase the number of useful sites a polylinker containing a number of enzyme sites was introduced into pEP2MCS. Oligonucleotides PL#1U and PL#1L were annealed to generate a double stranded DNA fragment that contained *Aat*II, *Pst*I, *Sph*I, *Eco*RV, *Bam*HI, *Eco*RI and *Age*I sites and an overhanging sequence of ggcc at the 3' end of the polylinker. The annealed polylinker was ligated into pEP2MCS linearised with *Age*I and *Pvu*II to generate pSM1. An additional *Apa*I site was generated by this cloning step. The presence of the polylinker was demonstrated by showing that the appropriate restriction sites had been introduced into the vector by restriction endonuclease digestion.

A ribosome binding site and stop codon sequences in each of the three frames were placed into pSM1 by insertion of a second polylinker. A ribosome binding site based on that for the *C. pseudotuberculosis repA* gene was utilised. The sequence immediately upstream of the start of the *repA* gene is CTGGGAGAAATCACGGTC. In order to allow this sequence to be

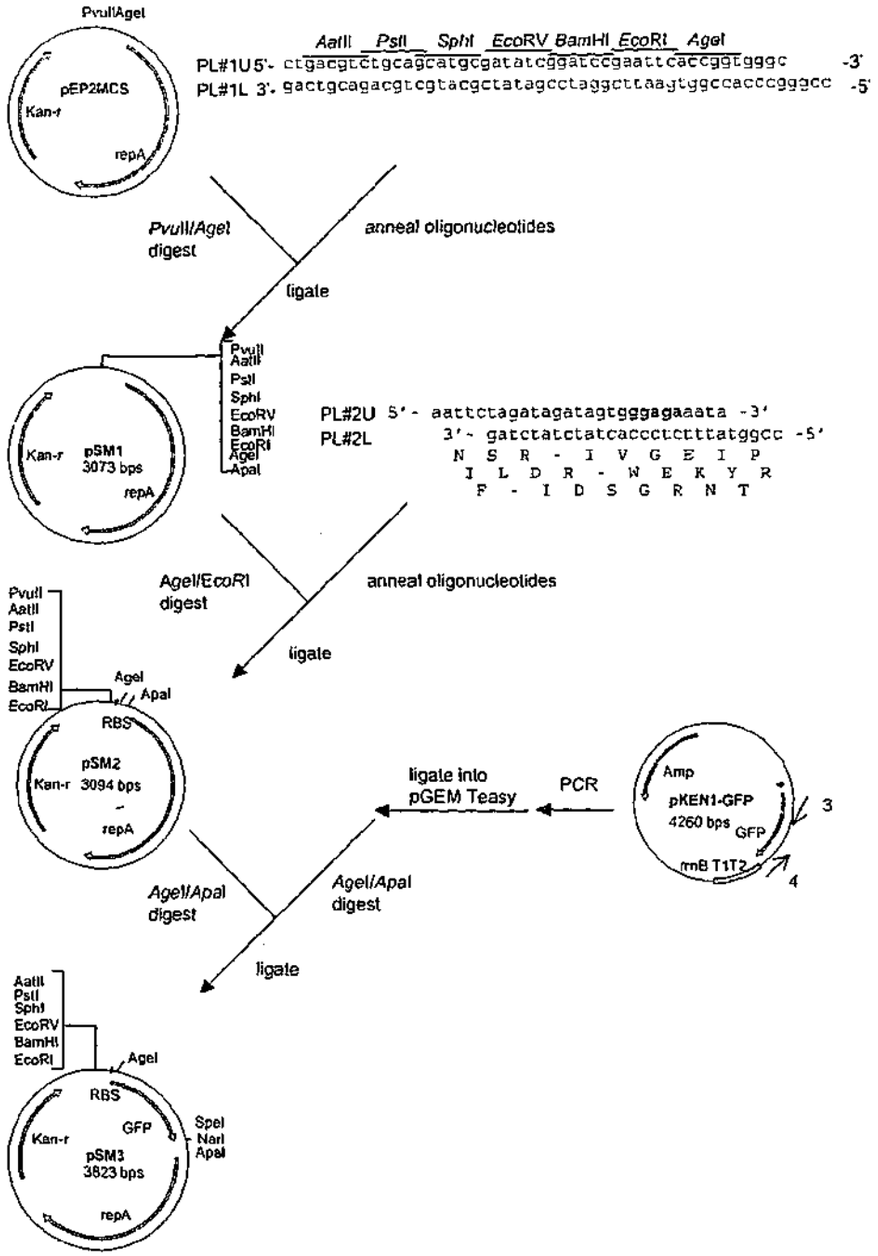

Figure 4.1: Figure continues on next page, refer to legend on page 119.

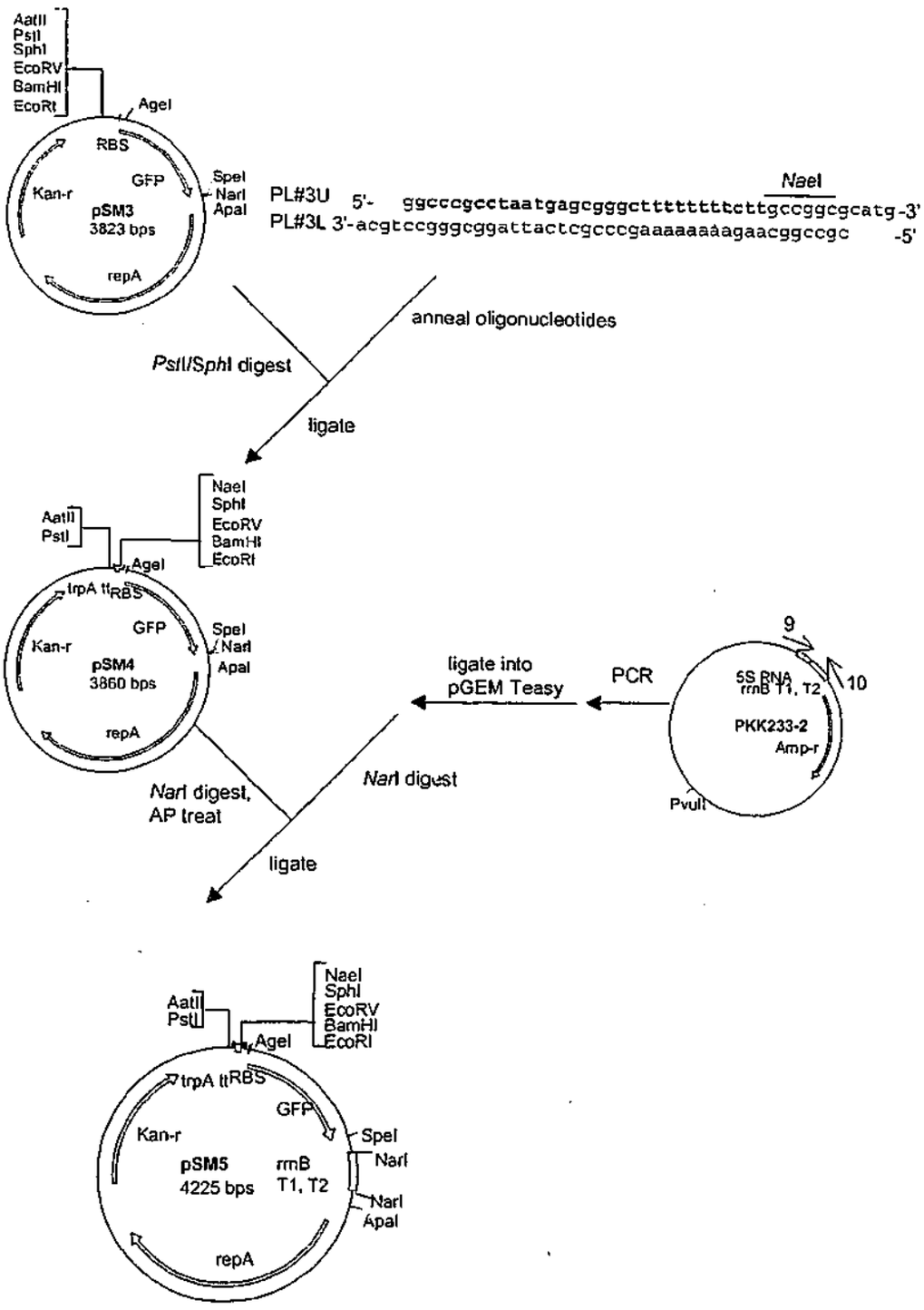

Figure 4.1: Construction of the promoter probe vector pSM5.

This schematic diagram outlines the steps leading to the construction of pSM5. The details relating to the construction of these plasmids can be found in Section 4.2.

cloned into the *Age*I site of pSM1 and maintain the *Age*I site for subsequent manipulations, the order of the C and A bases (at positions 12 and 13) was switched. Upstream of this 18 bp sequence was sequence coding for stop codons in each frame. These were included to prevent translational fusions between proteins coded for by inserted genomic DNA fragments and Gfp. For cloning purposes an overhanging *Eco*RI site was placed at the 5' end and an *Age*I site at the 3' end. Polylinker #2U and Polylinker #2L were annealed then ligated into pSM1 linearised with *Age*I and *Eco*RI to generate pSM2.

The *gfp* gene was sourced from pKEN1gfpuv1. This vector contains a FACS optimised mutant *gfp* gene that contains a mutated region between base pairs 190 and 193 in which the wildtype sequence of TTCT is replaced with CTGA giving rise to the amino acid substitutions at positions 64 and 65 in which the wildtype amino acids phenylalanine and serine are replaced with leucine and threonine respectively (Cormack *et al.*, 1996). The *gfp* gene was PCR amplified from pKEN1gfpuv1 using primers #3 and #4. Both primers contain additional sequence at their 5' ends for this and subsequent cloning steps. Primer #3 contains an *Age*I site while primer #4 contains *Spe*I, *Nar*I and *Apa*I sites at its 5' end. The PCR product was initially subcloned into pGEM®-T easy. This commercially prepared vector has been linearised with *Eco*RV and treated such that it has T overhangs to allow cloning of *Taq* DNA polymerase PCR products that contain A overhangs. The *gfp* gene was released from the pGEM®-T easy vector with *Age*I and *Apa*I and ligated into pSM2, that had been prepared in the same manner, to generate pSM3.

The rho independent *trpA* transcription terminator was placed upstream of the *gfp* gene. The polylinkers PL#3U and PL#3L were annealed to generate a double stranded DNA fragment that contained the transcriptional terminator, a *Nae*I site and, *Pst*I and *Sph*I compatible overhanging ends. The DNA was ligated into pSM3 linearised with *Pst*I and *Sph*I to generate pSM4. The presence of an additional *Nae*I site in the vector was used as a diagnostic indicator of successful insertion of the polylinker.

Transcription terminators were also inserted downstream of the *gfp* gene. A region containing the *E. coli rrnB* T1T2 transcription terminators was PCR amplified from the vector pKK233-2 with primers #9 and #10. Each primer contained a *Nar*I site at the 5' end. The PCR product

was first cloned into pGEM®-T easy, then released by digestion with *Nar*I. The fragment was ligated into pSM4, that had been linearised with *Nar*I and treated with Thermosensitive alkaline phosphatase, to generate pSM5. That the transcription terminator was appropriately orientated was determined by restriction endonuclease digestion with *Bsp*HI which cuts once within the *rrnB* T1T2 fragment and several times within the vector backbone.

The sequence of the *gfp* gene and surrounding sequence was confirmed by DNA sequencing. An error was found at the 3' end of the *gfp* gene that could be attributed to an error in primer #4. To fix this inaccuracy the *gfp* gene was reamplified from pKEN1gfpuv1 by PCR using primers #3 and #4b. The new *gfp* gene was inserted into pSM5 digested with *Age*I and *Spe*I to generate pSM12 in which the faulty *gfp* gene was replaced with one with the correct sequence.

4.3 Insertion of known promoters into pSM12

To determine the efficacy of pSM12 as a promoter probe vector two known promoters were inserted in the multiple cloning site upstream of the *gfp* gene. These were the *C. pseudotuberculosis* phospholipase D (*pld*) promoter and an artificial promoter, the *srp* promoter which is found in pKEN1gfpUV1. The *C. pseudotuberculosis pld* promoter was extracted from pRM130 which contains the *pld* promoter upstream of a chloramphenicol resistance gene in a pEP2 vector background (Figure 4.2). The *pld* promoter was removed as a *Bam*HI fragment and ligated into pSM12 digested with *Bam*HI and phosphatased by treatment with Thermostable alkaline phosphatase to generate pSM11. Restriction endonuclease digestion was performed to identify clones in which the promoter had been inserted in the correct orientation.

The artificial *srp* promoter (sterically repressed promoter) is responsible for *gfp* expression in pKEN1gfpuv1. The *srp* promoter consists of consensus -10 and -35 regions separated by 17 base pairs of sequence that is closely related to the *E. coli* lac O_1 operator (Figure 4.2) (Ezaz-Nikpay *et al.*, 1994). The *srp* promoter was chosen as it gives rise to highly fluorescent *E. coli* that carry this vector. The promoter region was excised from pKENgfp1uv1 by digestion with *Fsp*I and *Bam*H. The *Fsp*I site is located 5' of the *srp* promoter while the *Bam*HI site is located between the 3' end of the *srp* promoter and the start of the *gfp* gene. The *srp* promoter fragment was ligated into pSM12 digested with *Bam*HI and *Eco*RV to generate pSM13.

A

```
 1  aacgacggcc agtgaattgt aatacgactc actatagggc gaattgacat
                                    EcoRI
51  tgtgagcgga taacaatata atgtgtggaa ttcgagctcg gtacccgggg
    BamHI
101 atcctctaga tttaagaagg agatatacat atg agt aaa
                                      M   S   K
```

B

```
-410 gatccgtcga cgatgctccg catcccttgc atctgtaccg ggtctttcct
-360 gtggacgttt ttgccgttct tttttaggag attgccgcag acgttgcagt
-310 gattccggta tggatgtgaa gccataatca ataattagca gctttaaatc
-260 tgctatggat accctaaaag ttacaatttt ggacacattt tttgtccttt
-210 aagttcaaaa accttgaacg tatctgacgc atttgaatac atatcattca
-160 tatgtaatgc ctgattgata acgccaacct tttaaatggg gctcccgctc
-110 tcctctctag gggatggttt tctgtgtcaa ctgtcagctt ccgaaaaata
-60  tgagtataaa catatgcaat aacccttgat tttattgtta tttaagtttc
-10  ataatgggga Tatagcgcag ggtgttcacg ctggtaaagg gctatttttg
 41  ggcacctctc cgtgggaaat gtatccgcaa ttcattaaag aggagaaatt
 91  aactatgaga ggatcgcatc accatcacca tcacg
```

Figure 4.2: DNA sequences of the *srp* promoter region of pKEN1gfpuv1 and the *pld* promoter region of pRM130.

For the *srp* promoter (A) the –35 and –10 regions are bolded. Upstream of the *srp* promoter is a T7 promoter region (underlined). Sequence corresponding to the first three amino acids of the *gfp* gene is shown in the amino acid single letter code. Pertinent restriction enzyme sites are marked.

For the *pld* promoter (B) the –35 and –10 regions are bolded. Downstream of the –10 region of the *pld* promoter is the transcriptional start point (underlined uppercase **T**) that was previously identified using primer extension assays (Hodgson *et al.*, 1990). Within the *Bam*HI fragment the *pld* promoter sequence terminates at bp 68 which is 55 base pairs prior to the *pld* initiating methionine in *C. pseudotuberculosis*. The *Bam*HI fragment contains a further 55bp of sequence that is not homologous to the *pld* promoter (underlined).

4.4 Flow cytometry analysis of promoter containing constructs.

Analysis of colonies of *E. coli* Top10 F' carrying pSM11 or pSM13 under UV light indicated that these constructs gave rise to fluorescent bacteria. To determine whether the constructs gave rise to fluorescence in *C. pseudotuberculosis* the two plasmids were transformed into *C. pseudotuberculosis* and fluorescence quantitated by flow cytometry. Overnight cultures of C231(pSM11), C231(pSM12) and C231(pSM13) were diluted to an OD_{600} of 0.1 and then grown at 37°C for 2 hr. Cultures were sonicated to disrupt bacterial clumps, thereby ensuring a single cell suspension and then analysed for Gfp fluorescence by flow cytometry (Figure 4.3). Both the strains carrying intact promoter constructs were fluorescent with mean fluorescence intensities of 9 and 8 for C231(pSM11) and C231(pSM13) respectively. In comparison the promoterless construct C231(pSM12) was non-fluorescent with a mean fluorescence intensity of 1. An explanation of mean fluorescence intensity can be found in Appendix 1.

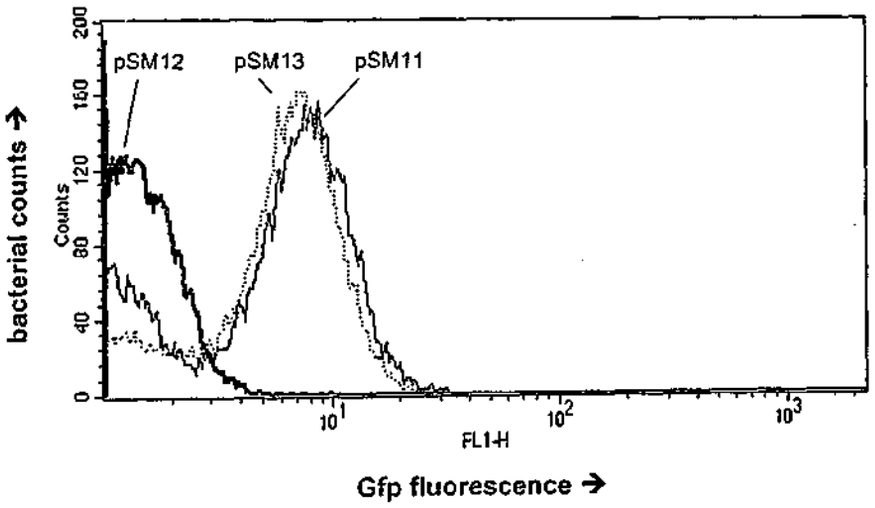

Figure 4.3: Flow cytometry analysis of *gfp* expression from *C. pseudotuberculosis* harbouring *pld* promoter (pSM11), *srp* promoter (pSM13) or promoterless (pSM12) constructs.

For each strain 30,000 events were counted. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria.

4.5 Construction of a random library in pSM12

To generate a library of *C. pseudotuberculosis* clones for DFI analysis a two step approach was taken. The first step was to construct a library in an *E. coli* host and to then extract plasmid DNA from the library. The second step was to transform this plasmid DNA into *C. pseudotuberculosis* to generate the final library. This approach was taken as the transformation efficiency in *C. pseudotuberculosis* is relatively low, suggesting that it could be difficult to make a large library by direct transformation of library ligations into *C. pseudotuberculosis*.

Random fragments generated by sonication were chosen for library construction for the same reasons as outlined in Section 3.2.1. These fragments were inserted upstream of the *gfp* gene of pSM12. pSM12 was digested with *Eco*RV then treated with thermostable alkaline phosphatase. Random genomic fragments in the size range of 0.7 to 1.4 that had been treated with MBN (Section 3.2.1) were ligated into the linearised vector overnight at 15°C, then a proportion of the ligation (2 µl) was transformed into electrocompetent Top10F' cells. To determine the percentage of clones containing inserts and the average insert size, plasmid DNA was extracted from 12 randomly picked colonies. Restriction endonuclease digests with *Sph*I and *Bam*HI which cut on either side of the *Eco*RV site were performed. Gel analysis of the digests indicated that 75% of the clones contained inserts and that the average insert size was approximately 1 kb (Figure 4.4).

The remainder of the library ligation was transformed into electrocompetent Top10 F' cells. The library was estimated to contain 62,000 clones. When the library plates were viewed under UV light a small percentage of the colonies could be seen to fluoresce. A number of fluorescent colonies were streaked out onto fresh plates for further analysis (Section 4.7), while the remaining colonies were collected from the agar plates. A proportion of the collected cell suspension was used to inoculate 500 ml of TB. The remainder was then stored at -80°C in the presence of 15% glycerol. This library was termed the *E. coli* pSM12-Q library. After overnight incubation at 37°C plasmid DNA was extracted from the 500 ml culture. The DNA was diluted 1 in 20 and then used to transform electrocompetent C231. To ensure that an equivalent number of clones contained an insert as when the library was prepared in *E. coli*, DNA was extracted from 12 independent colonies. Restriction

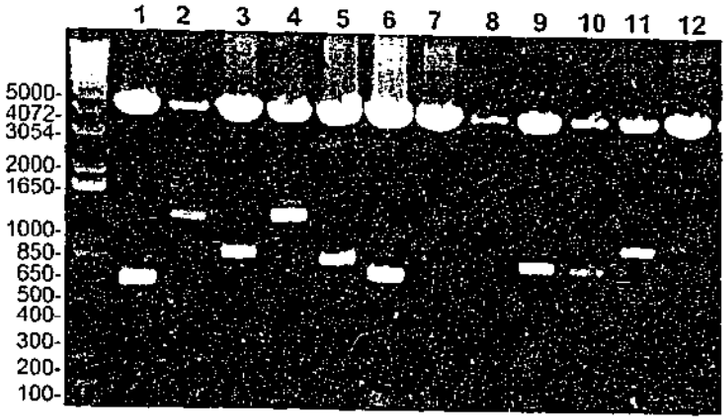

Figure 4.4: Analysis of inserts from the *E. coli* pSM12-Q library.

The pSM12-Q library was generated by insertion of sonicated fragments of C231 DNA into the *EcoRV* site of pSM12. The ligation was used to transform Top10 F' *E. coli* and DNA extracted from 12 individual colonies. DNA was digested with *Sph*I and *BamH*I which cut up and downstream of the *EcoRV* site respectively. 1 kb plus ladder is run in the first lane and band sizes (in bp) are indicated to the left of this. Clones 7, 8 and 12 did not contain inserts. The average insert size of the other clones is approximately 1 kb.

endonuclease digestion with *Sph*I and *Bam*HI however indicated that none of the 12 plasmids contained an insert.

In order to increase the number of clones in the *C. pseudotuberculosis* library that contained inserts the plasmid DNA extracted from the *E. coli* library was digested with *Eco*RV prior to transformation into C231. This step was performed to linearise plasmids without an insert. Upon transformation these clones will not be viable unless the DNA recircularises, thus removing them from the library. The disadvantage of such an approach is that it will select against plasmids that contain an insert with an *Eco*RV site. The digested DNA was transformed into electrocompetent C231. Analysis of the inserts of six plasmids indicated that four contained inserts (Figure 4.5). Analysis of the library plates under UV light indicated that a small proportion of the library contained plasmid inserts with promoter activity. Four colonies were streaked out for further analysis (see Section 4.7). To store the library, colonies were scraped off the agar plates into BHI, sonicated to disrupt bacterial clumps, centrifuged then resuspended in BHI and glycerol and stored at -80°C. The library was titled the C231-pSM12-Q library.

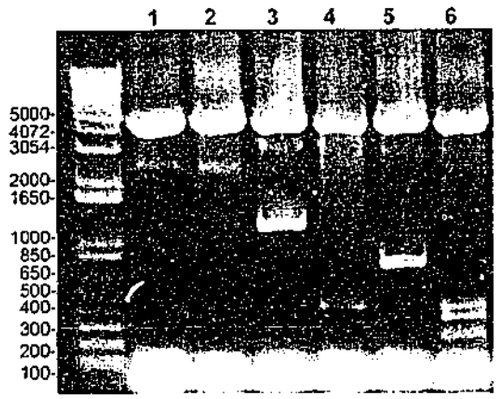

Figure 4.5. Analysis of inserts from the *C. pseudotuberculosis* pSM12-Q library.

The pSM12-Q library was generated by insertion of sonicated fragments of C231 DNA into the *Eco*RV site of pSM12. The ligation was used to transformTop10 F' *E. coli* and plasmid DNA extracted. The plasmid DNA was digested with *Eco*RV then used to transform into C231. DNA was extracted from 6 individual clones and then digested with *Sph*I and *Bam*HI which cut up and downstream of the *Eco*RV site respectively. 1 kb DNA ladder is run in the first lane and band sizes (in bp) are indicated to the left of this. Clones 1 and 2 do not contain inserts. The average insert size of the other clones is approximately 800 bp.

4.6 Flow cytometry of the C231-pSM12-Q library.

To accurately determine what percentage of the C231-pSM12-Q library contained bacteria harbouring plasmids with constitutive promoter activity flow cytometry was utilised. Overnight cultures of C231(pSM12) and the C231-pSM12-Q library were diluted 1 in 20 then incubated at 37°C for 1.5 hr. By flow cytometry analysis approximately 0.22% of the library clones were determined to contain plasmid inserts with promoter activity that gave rise to detectable fluorescence (Figure 4.6).

4.7 Analysis of Gfp expressing clones in *E. coli* and *C. pseudotuberculosis*

4.7.1 Analysis of clones that are fluorescent in *E. coli*

Eighteen individual fluorescent clones from the *E. coli* pSM12-Q library were further analysed. These clones were streaked out onto LB agar plates and titled gfp+1 to gfp+23.

forward scatter →

Figure 4.6: Flo

Overnight cultu
grown at 37°C
cytometry. For
scatter such th
contained with

Fluorescence
clones ranged
clones and tra
that only three
transformatio
fluorescing C
fluorescent ph

The sequence
determined.
with 100% ho
however this
within the ins
backbone wit

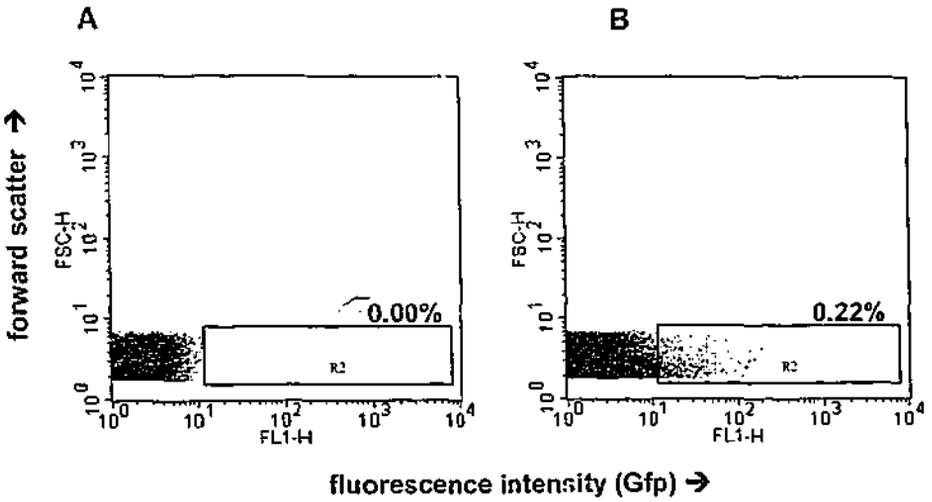

Figure 4.6: Flow cytometry analysis of the C231 pSM12-Q library

Overnight cultures of (A) C231 (pSM12) and the (B) C231 pSM12-Q library were diluted 1 in 20 then grown at 37°C for 1.5 hr. The percentage of Gfp positive clones in the library was determined by flow cytometry. For each culture 30,000 events were analysed. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria. The percentage of events that were contained within Region 2 (R2) is indicated.

Fluorescence intensity was assessed by viewing the bacteria under a UV light and for the 18 clones ranged from weak to very strong (Table 4.1). DNA was extracted from 15 of the clones and transformed into C231. Analysis by visual inspection under UV light indicated that only three remained fluorescent. To ensure that fluorescence had not been lost during the transformation process, plasmid DNA was extracted from three non-fluorescing and one fluorescing *C. pseudotuberculosis* strain. Upon retransformation back into *E. coli* the fluorescent phenotype was restored for all four clones.

The sequence of the inserts from the three clones that fluoresced in *C. pseudotuberculosis* was determined. gfp+5 contained a 43 bp insert. Within this sequence there was located a region with 100% homology to the -10 region of the *E. coli* σ^{70} promoter consensus (TATAAT), however this was located at the 5' end of the insert precluding the presence of a –35 region within the insert (Figure 4.7). 17 bp upstream of this region was a sequence in the vector backbone with 4/6 homology to the *E. coli* σ^{70} promoter consensus –35 region. This

Table 4.1: Gfp expressing clones from the *E. coli* pSM12-Q library

clone ID	fluorescence in *E. coli* [a]	transformed into C231	fluorescence in C231 [a]	retransformed into *E. coli* [b]	sequence information[c]
gfp+1	+++				
gfp+2	+++				
gfp+5	+++	yes	++		40 base pair insert, contains consensus –10 region
gfp+6	+++	yes	+++	yes (fluorescent)	leucyl tRNA synthase, middle of an enolase gene, consensus promoter sequence at 3' end
gfp+7	++	yes	++		334 bp insert, no significant homologies but a putative consensus promoter sequence at 3' end
gfp+8	++	yes	-		
gfp+10	+++	yes	-	yes (fluorescent)	
gfp+11	++	yes	-		
gfp+12	++	yes	-	yes (fluorescent)	
gfp+13	++	yes	-		
gfp+14	++	yes	-		
gfp+15	++	yes	-		
gfp+16	++	yes	-		
gfp+17	+	yes	-		
gfp+21	++	yes	-		
gfp+22	+++	yes	-	yes (fluorescent)	
gfp+23	++	yes	-		

[a] Flourescence of clones grown on agar plated was scored visually under UV light. Clones were designated as having weak fluorescence (+), medium fluorescence (++), strong fluorescence (+++) or being non-flourescent (-).

[b] DNA was extracted from fluorescent C231 clones then transformed into *E. coli*. Plasmid DNA was then extracted from *E. coli* and retransformed into C231.

[c] Sequences were scanned against the NCBI protein databases for protein homologies in all 6 reading frames using the translating BLAST program (BLASTX). Proteins that showed the highest homology are listed.

A

1 ctg

51 tag

101 g

B

bacterial counts

Figure 4.

(A) The

bolded a

that lies

homolog

(B) gfp+

polymer

carrying

the *PstI*/

Data wa

A

*Pst*I

1 ctgcaggccc gcctaatgag cgggcttttt tttcttgccg gcgcatgcga

*Sph*I

51 tagttgtata attcattga aagtttggaa acaaacaatg tatcggatcc

101 gaattctaga tagatagtgg gagaaatacc ggtatg

M

B

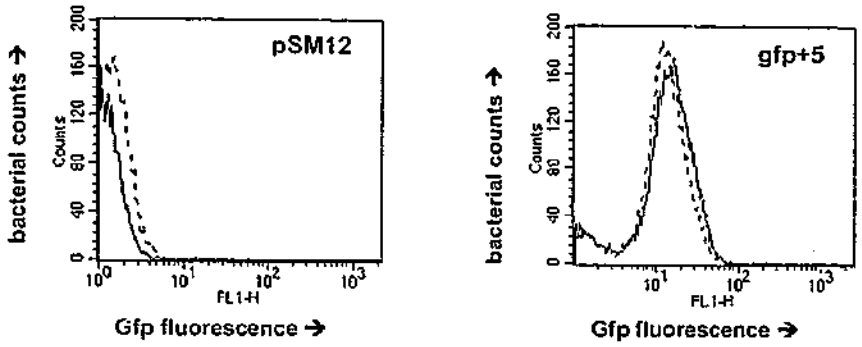

Figure 4.7: Sequence and fluorescence of gfp+5

(A) The sequence of the insert of gfp+5 and surrounding vector is shown. The insert of gfp+5 is bolded and a sequence with complete homology to the –10 region of *E. coli* σ^{70} promoter consensus that lies within this region is boxed. Within the upstream vector sequence is a region with 4/6 homology to the –35 region of *E. coli* σ^{70} which is also boxed.

(B) gfp+5 and pSM12 were digested with *Pst*I and *Sph*I, the ends blunted by treatment with T4 DNA polymerase and ligated to recircularise the vectors. The fluorescence of overnight cultures of C231 carrying the parent vectors (solid lines, pSM12 and gfp+5) was compared to the same vectors in which the *Pst*I/*Sph*I fragment had been removed (broken lines). For each clone 30,000 events were counted. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria.

observation lead to concern that these two sequences may be acting together to give rise to the promoter activity. In order to determine whether this was the case the sequence containing the region with homology to a –35 consensus was removed from the vector. As an additional control the same region was removed from the promoterless vector pSM20. pSM20 and gfp+5 were digested with *PstI* and *SphI*. The ends were filled in by treatment with T4 DNA polymerase and ligated to recircularise the plasmids. Sequencing was performed to confirm that the appropriate deletions had been made. Upon analysis of the constructs and their parents by flow cytometry it was observed that the fluorescence was equal regardless of the presence or absence of the putative –35 region. This indicates that full promoter activity was most likely contained within the 43 bp insert (Figure 4.7), although the possibility that a second pseudo –35 region was brought into an appropriate position by removal of the fragment cannot be completely ruled out.

Sequence analysis of gfp+6 indicated that it most likely contained an insert that was a concatamer of several genomic DNA fragments. At the 5' end of the insert was a gene with homology to leucyl tRNA synthetase, this was followed by an internal fragment of an enolase gene which in turn was followed by a region that did not show homology to any proteins but did contain putative promoter elements.

Sequence of gfp+7 revealed an insert of 334 bp that had no homology to proteins in the NCBI protein databases. There was however a consensus promoter sequence at the 3' end of the insert suggesting that a coding region would be found downstream of this region.

4.7.2 Analysis of clones that are fluorescent in *C. pseudotuberculosis*

Four fluorescent clones were streaked out when storing down the C231-pSM12-Q library and labelled gfp+25 to gfp+28. DNA was extracted from gfp+26, gfp+27 and gfp+28 and transformed into *E. coli*. In each case non-fluorescent colonies resulted. DNA was extracted and the inserts sequenced. Sequence analysis indicated that each of the three clones contained an identical insert of 838 bp. A BLASTX search was performed to identify protein homologues. The inserts contained sequence coding for the 3' end of an enolase gene (Figure 4.8). The sequence was most homologous to the enolase gene of *Streptomyces coelicolor* with 73% identity. The start of the insert corresponds to amino acid 152 of the enolase gene while

the enolase stop codon is located 10 bp from the 3' end of the insert. This suggests that the promoter activity resides within the enolase coding region. To confirm that promoter activity resided within the 3' end of the insert the 443 most 3' base pairs were removed from the plasmid by digestion with *Bam*HI. *Bam*HI cuts 396 bp into the insert and immediately downstream of the insert (Figure 4.8). The vector backbone was recircularised and transformed into C231. The removal of this portion of the insert ablated the fluorescence of the clone.

```
  1  ggtggcgcgc atgctgactc aggtgtcgat gttcaggagt ttatgatcgc tcccatcggt
       G  G  A   H  A  D   S  G  V  D   V  Q  E   F  M  I   A  P  I  G

 61  gcggagagct tctccgaggc gctgcgcatg ggtgctgagg tttatcactc gcttaaatct
       A  E  S   F  S  E   A  L  R  M   G  A  E   V  Y  H   S  L  K  S

121  gtgatcaagt ctaaggggct atctaccgga cttggtgatg aaggcggttt cgcgccttcc
       V  I  K   S  K  G   L  S  T  G   L  G  D   E  G  G   F  A  P  S

181  gtagagtcca ccaaggctgc tctggatctt atcgttgagg caatcgaaaa agccggcatg
       V  E  S   T  K  A   A  L  D  L   I  V  E   A  I  E   K  A  G  M

241  aaaccaggta cggatattgc cctggcattg gacgtggctt cctctgagtt cttcaaagat
       K  P  G   T  D  I   A  L  A  L   D  V  A   S  S  E   F  F  K  D

301  ggcaagtacc actttgaggg tggcgagcac accgctgaag agatggcaaa ggtttacgct
       G  K  Y   H  F  E   G  G  E  H   T  A  E   E  M  A   K  V  Y  A
                                        BamHI
361  gacctcatcg agcagtatcc gattgtgtcc atcgaggatc cgctgcaaga ggacgattgg
       D  L  I   E  Q  Y   P  I  V  S   I  E  D   P  L  Q   E  D  D  W

421  gaggggtaca ccaatctcac ggcagctatt ggagacaagg tccaaatcgt cggcgatgat
       E  G  Y   T  N  L   T  A  A  I   G  D  K   V  Q  I   V  G  D  D

481  ttctttgtaa ccaaccctgt ccgtctgcag gaaggcatcg ataagaaggc tgccaatgcg
       F  F  V   T  N  P   V  R  L  Q   E  G  I   D  K  K   A  A  N  A

541  cttttggtaa aggtcaacca gatcggcacc ttgacggaga ccttcgacgc cgtcgatctt
       L  L  V   K  V  N   Q  I  G  T   L  T  E   T  F  D   A  V  D  L

601  gctcaccgca atggctaccg cacaatgatg tctcaccgct ccggtgaaac cgaagatact
       A  H  R   N  G  Y   R  T  M  M   S  H  R   S  G  E   T  E  D  T

661  acgattgccg atctttcggt tgccctcgga tgtggtcaaa ttaagaccgg agcacccgct
       T  I  A   D  L  S   V  A  L  G   C  G  Q   I  K  T   G  A  P  A

721  cgctccgaac gtgtggcaaa atataatcag ctgctgcgaa ttgagcagga gctaggagac
       R  S  E   R  V  A   K  Y  N  Q   L  L  R   I  E  Q   E  L  G  D

781  gctgctatat atgcgggtcg ctctgcattc ccgcatttcc ggaattaaac ataggtta
       A  A  I   Y  A  G   R  S  A  F   P  H  F   R  N  -
```

Figure 4.8: Sequence of the insert of gfp+26.

The protein sequence shows 73% identify to the C-terminal of an enolase protein of *Streptomyces coelicolor.*

4.8 Construction and analysis of a pSM12-pKEN1gfpuv1 fusion vector

In *E. coli* the fluorescence of bacteria harbouring pSM13 was less than in those harbouring pKEN1gfpuv even though both contain the *gfp* gene under the control of the *srp* promoter. This suggested that the region downstream of the *srp* promoter, which differs in the two plasmids, is important for full gene expression. To determine whether this difference was maintained in *C. pseudotuberculosis*, a vector containing the pKEN1gfpuv1 sequence in a pEP2 background was generated by fusing the promoterless vector pSM12 and pKEN1gfpuv1. Both plasmids were digested with *PstI* then ligated together to generate pSM15. Analysis of Top10 F' (pSM13) and Top10 F' (pSM15) by UV light indicated that bacteria harbouring pSM15 were more fluorescent than those with pSM13. Upon transformation of the plasmids into C231 this observation was again made. Measurement of the fluorescence of *C. pseudotuberculosis* harbouring these plasmids indicated that C231(pSM15) was almost 10 fold more fluorescent than C231 (pSM13) (Figure 4.9).

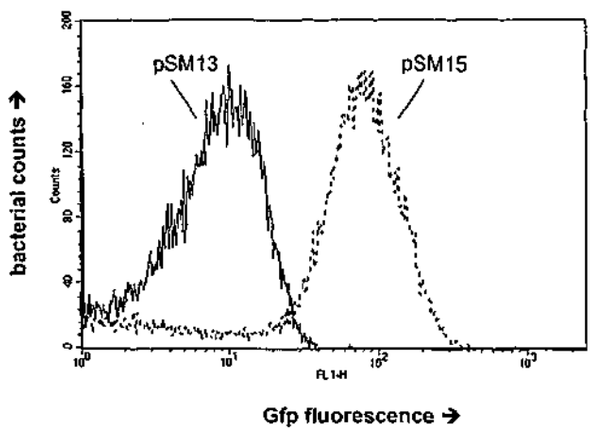

Figure 4.9: Effect of the region containing the RBS on *gfp* expression controlled by the *srp* promoter.

The fluorescence intensity of bacteria in overnight cultures of C231(pSM13) and C231(pSM15) was determined by flow cytometry. For each clone 30,000 events were counted. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria.

4.9 Modifications to the promoter probe vector and library construction process

Analysis of the pSM12-Q library indicated that it contained only 0.22% gfp positive clones. Compared to libraries prepared for DFI analysis of gene expression in *M. marinum* and *M. bovis BCG* which contained 6 and 9% of Gfp positive clones respectively, this percentage is very low. It is likely that this low percentage is related to the way in which the promoter probe vector and library were constructed. Relatively poor expression was observed from known promoters inserted upstream of the *gfp* gene in pSM12. This was particularly evident by comparing the fluorescence of pSM13 and pSM15 which both contain *gfp* under the control of the *srp* promoter. The principal difference between these two promoters is the sequence between the –10 region and the start of the *gfp* gene. This includes the region that contains the RBS and it therefore seems likely that this sequence may be important for maximal Gfp translation. The low percentage of Gfp positive clones may also be a result of passaging of the library through *E. coli*, leading to biases in the *C. pseudotuberculosis* library. This is evidenced by the observation that upon first transforming the library into *C. pseudotuberculosis* no clones with inserts were detected even though when in *E. coli* 75% of clones did in fact have inserts. Additionally empirical evidence suggests that very strong promoters may produce a level of Gfp that is toxic to *E. coli*, thereby removing some clones from the library.

To improve the quality of the library to be used for DFI, modifications were made to the promoter probe vector and to the way in which the library was generated in *C. pseudotuberculosis*. In particular the region immediately upstream of the *gfp* gene of pSM12 was altered and modifications were made so that translational fusions would also be recovered. Secondly a process was developed in which library ligations could be directly transformed into *C. pseudotuberculosis* such that there was no requirement for the intermediate library generation in *E. coli*.

4.9.1 Alteration of the region surrounding the RBS in the promoter probe vector (pSM12)

Inspection of the region between the *srp* promoter and the *gfp* gene in pSM13 and pKEN1gfpuv1 revealed several differences (Figure 4.10). Firstly there is additional sequence in pSM13 such that the distance between the –10 region and the methionine codon is 65 bp

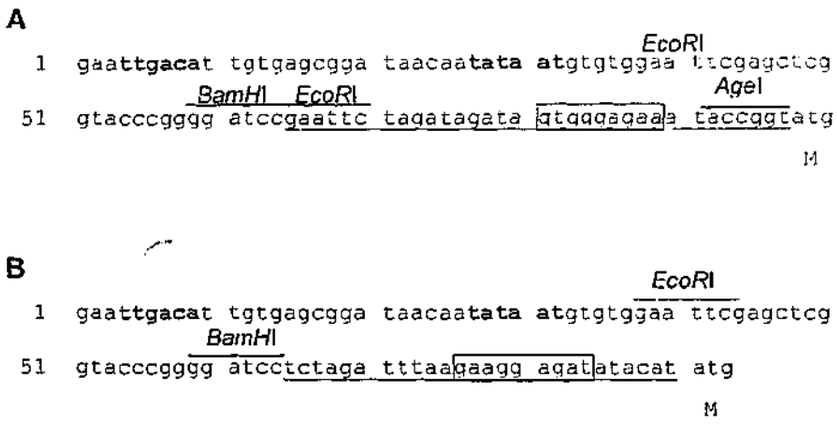

Figure 4.10: Comparison of the *srp* promoter region of pSM13 and pKEN1gfpuv1.

DNA sequence from pSM13 (A) and pKEN1gfpuv1 (B) are shown. The –35 and –10 regions are bolded. Predicted RBS's are boxed and the sequence that differs between the two plasmids is underlined. In each case the last three base pairs code for the initiating methionine.

compared to 58 bp. Secondly the sequence composition of the last 33 bp and 26 bp differs between the two plasmids. Thirdly the putative RBS is different in each construct consisting of gtgggagat in pSM13 and gaaggagat in pKEN1gfpuv. Although it was unclear whether the lesser levels of Gfp obtained from pSM13 are as a result of reduced transcription or translation it was thought that replacement of this region in pSM13 with that from pKEN1gfpuv1 would enhance the level of Gfp detected from the promoter probe vectors.

Both vectors were digested with *Eco*RI and *Hpa*I. *Eco*RI cuts both plasmids at the same point in the vector just downstream of the –10 region (see Figure 4.10). A single *Hpa*I site exists in both vectors and is located within the *gfp* gene. The promoter-*gfp* fragment of pKEN1gfpuv1 was ligated into the linearised vector backbone of pSM13 to generate the new vector pSM17.

Concurrently a modified vector containing the *pld* promoter was also constructed. pSM11 was digested with *Eco*RI and *Hpa*I to produce a linearised vector into which the *Eco*RI/*Hpa*I fragment from pKEN1gfpuv1 was ligated to generate pSM16. A new promoterless construct was then generated by digesting pSM16 with *Bam*HI to release the *pld* promoter fragment. The vector backbone was recirularlised by ligation to generate pSM18.

4.9.2 Flow cytometry analysis to determine the effect of the new RBS region on Gfp fluorescence

Cultures of C231(pSM11), C231(pSM13), C231(pSM16) and C231(pSM17) were grown overnight then subcultured to an OD_{600} of 0.1. The cultures were grown for a further two hr at 37°C, sonicated, and then analysed for Gfp fluorescence by flow cytometry (Figure 4.11).

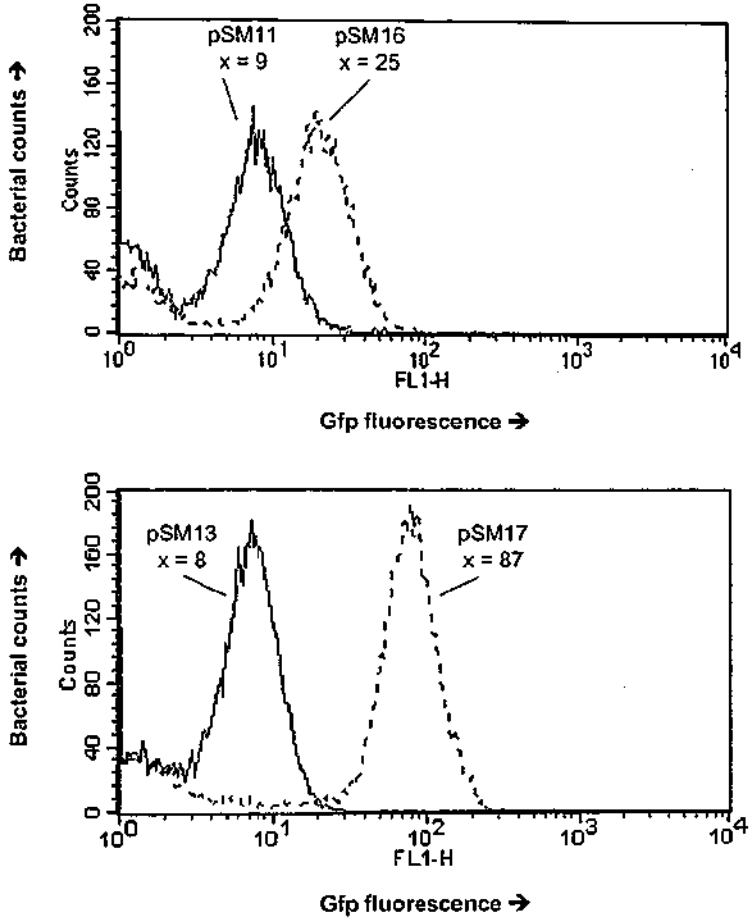

Figure 4.11: Effect of the region containing the RBS on *gfp* expression controlled by the *pld* or *srp* promoters.

Overnight cultures of the *pld* promoter containing strains C231(pSM11) and C231(pSM16) and the *srp* promoter containing strains C231(pSM13) and C231(pSM17) were diluted to an OD_{600} of 0.1 then incubated at 37°C for 2 hr. The fluorescence intensity of bacteria in each culture was determined by flow cytometry. For each clone 30,000 events were counted. Fluorescence intensity, x, is a measure of the mean fluorescence for each bacterial clone. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria

Inclusion of the new RBS region into pSM16 and pSM17 substantially increased the amount of fluorescence detected. C231(pSM16) was approximately 3-fold more fluorescent than C231(pSM11) whereas for the *srp* promoter constructs the fluorescence of C231(pSM17) was approximately 10-fold greater than for C231(pSM13).

4.9.3 Conversion of the promoter probe vector into a possible fusion vector.

In the original promoter probe vector, stop codons were placed immediately upstream of the RBS to prevent the formation of translational fusions. Upon reanalysis it was determined that there could be some advantages in allowing the formation of translational fusions between inserted genes and the *gfp* gene. Currently this was not possible due to the presence of sequence coding for stop codons immediately upstream of the gfp gene in pSM18 (Figure 4.12A). It is possible that translation may not be reinitiated following the stop codons It was therefore hoped that by allowing for translational fusions an increase in the number of Gfp positive clones would occur. To remove these stop codons a nucleotide between the stop codons and the methionine was removed. An A base located 3 bases from the ATG codon was chosen as this would destroy an *Nde*I restriction site, thus making diagnosis easy. A PCR primer was designed that spanned this region but was missing the A residue. The primer #47b is underlined in Figure 4.12B and was used in conjunction with primer #4b to PCR amplify the entire *gfp* gene. The PCR product was digested with *Spe*I and *Bam*HI and ligated into pSM18 that had been prepared in the same manner to generate pSM20. That the correct sequence was present was confirmed by diagnostic digests using *Nde*I and sequence analysis.

To confirm that this change had not altered Gfp fluorescence the *srp* promoter was released from pSM17 by digestion with *Sph*I and *Bam*HI and ligated into pSM20 prepared in the same manner to generate pSM22. The fluorescence of C231(pSM17) and C231(pSM22) was measured by flow cytometry (Figure 4.13). Both strains demonstrated essentially identical levels of fluorescence indicating that the single base mutation had no effect on *gfp* fluorescence driven by the *srp* promoter.

A

EcoRV *BamHI* *NdeI*

1 gcatgcgata tcggatcctc tagatttaag aaggagatat ac[a]tatgagt aaagg

A C D I G S S R F K K E I Y I - V
H A I S D P L D L R R R Y T Y E -
M R Y R I L - I - E G D I H **M S K**

B

EcoRV *BamHI*

1 gcatg cgata tcggatcctc tagatttaag aaggagatat actatgagt aaagg

A C D I G S S R F K K E I Y Y E -
H A I S D P L D L R R R Y T **M S K**
M R Y R I L - I - E G D I L - V

Figure 4.12: Conversion of pSM18 into a fusion vector.

The sequence immediately upstream of the *gfp* gene of pSM18 is shown (A). The three possible reading frames are shown using the amino acid single letter code. Sequence corresponding the first three amino acids of Gfp are bolded. Genomic DNA fragments are inserted into the *Eco*RV site. Between this site and the *gfp* gene are two stop codons which would prevent translational fusions occurring between inserted sequence and the Gfp. To allow for the possibility of translational fusions an A base (boxed) was removed from pSM18 (A) using a PCR based approach. The *gfp* gene and some upstream sequence of pSM18 was PCR amplified using primer #47b whose sequence is underlined (and differs from the sequence of pSM18 in that the boxed A is missing. Primer #4b was used as the reverse primer. The PCR product was inserted into pSM18 linearised by digestion with *Spe*I and *BamH*I to generate the new promoter probe vector pSM20 (B). The absence of the *NdeI* site in pSM20 was used for diagnostic digests to confirm that the new sequence had been inserted.

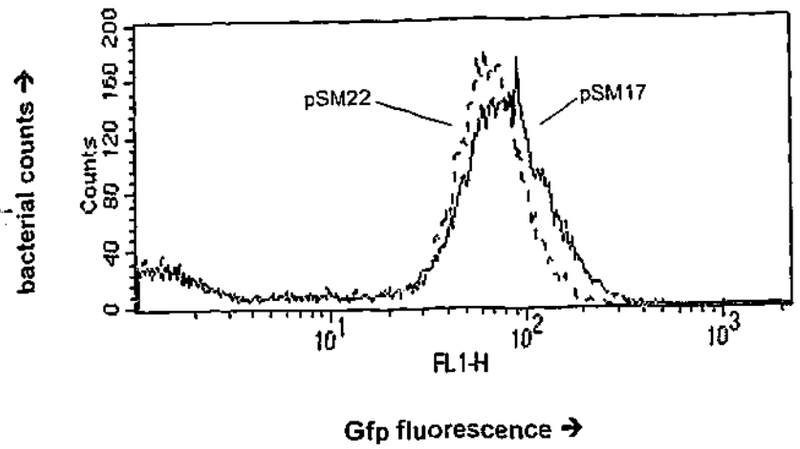

Figure 4.13: Effect of the fusion modification on Gfp fluorescence from the *srp* promoter

Overnight cultures of C231(pSM17) and C231(pSM22) were diluted to OD_{600} of 0.1 and then incubated at 37°C for 4.5 hr. The fluorescence intensity of bacteria in each culture was determined by flow cytometry. For each clone 30,000 events were counted. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria.

4.10 Library construction

The initial aim was to produce a random library by inserting sonicated fragments into pSM20. Technical difficulties were experienced when repairing the ends of sonicated DNA fragments with mung bean nuclease. The reasons for the difficulties were unclear, however it meant that it was not possible to prepare new random fragments in this manner. Therefore DNA fragments were instead generated by digestion with *Alu*I. Prior to transformation of library ligations into *C. pseudotuberculosis* the library was analysed for the percentage of clones with inserts and insert size in *E. coli*.

C231 genomic DNA was digested with *Alu*I and analysed by agarose gel electrophoresis (Figure 4.14A). DNA fragments in the size range of 0.7 to 1.2 Kb were excised from the gel, the DNA extracted and then reanalysed (Figure 4.14B). Fragments were ligated into pSM20 that had been linearised with *Eco*RV and subsequently treated with alkaline phosphatase. After an incubation at 15°C overnight a proportion of the ligation was transformed into Top10 F'. 2 µl of the ligation when transformed into a 40 µl aliquot of electrocompetent Top10 F' yielded 40,000 colonies. To analyse the percentage of clones that had inserts and the average insert size, 16 colonies were randomly chosen and plasmid DNA extracted. Plasmids were

digested with *Sph*I and *Bam*HI which cut up and downstream of the *Eco*RV site respectively. Analysis of the digests by agarose gel electrophoresis indicated that all the plasmids contained inserts (Figure 4.14C) and that the average insert size was approximately 950 bp.

Transformation of 1 µl of the same ligation into a 50 µl aliquot of electrocompetent C231 yielded 192 colonies. Additional empirical evidence indicated that increasing the volume of DNA transformed in an aliquot of electrocompetent cells above 1 µl substantially reduced the number of transformants. On the other hand an increase in DNA concentration within the 1 µl volume led to an increase in the number of transformants. To generate a suitably large library it was therefore necessary to perform large scale ligations (of the order of 200 µl). Following an overnight incubation at 15°C, the volume of the ligation was reduced by purification of the reactions using the Qiaquick nucleotide removal kit (Qiagen). Multiple 1 µl aliquots of the DNA were then individually transformed into 50 µl aliquots of electrocompetent C231. With such an approach it was possible to generate a library of 15,300 clones. Inspection of the plates under a UV light indicated that a small proportion of the library was fluorescent. Eight fluorescent colonies were streaked out for further analysis (Section 4.12).

To ensure that the library in *C. pseudotuberculosis* contained clones with inserts 8 colonies were randomly chosen, DNA extracted and then retransformed into *E. coli*. Plasmid DNA was isolated from *E. coli* cultures, then digested with *Pvu*II which indicated that 7 of the 8 clones contained inserts and that the average insert size was approximately 700 bp (Figure 4.15).

The library was stored as three sub-libraries each containing approximately 5000 clones. Colonies were scraped off plates in BHI, sonicated to disrupt bacterial clumps then stored at -80°C in the presence of 15% glycerol. The three libraries were called A1-3, A4-6 and A7-9.

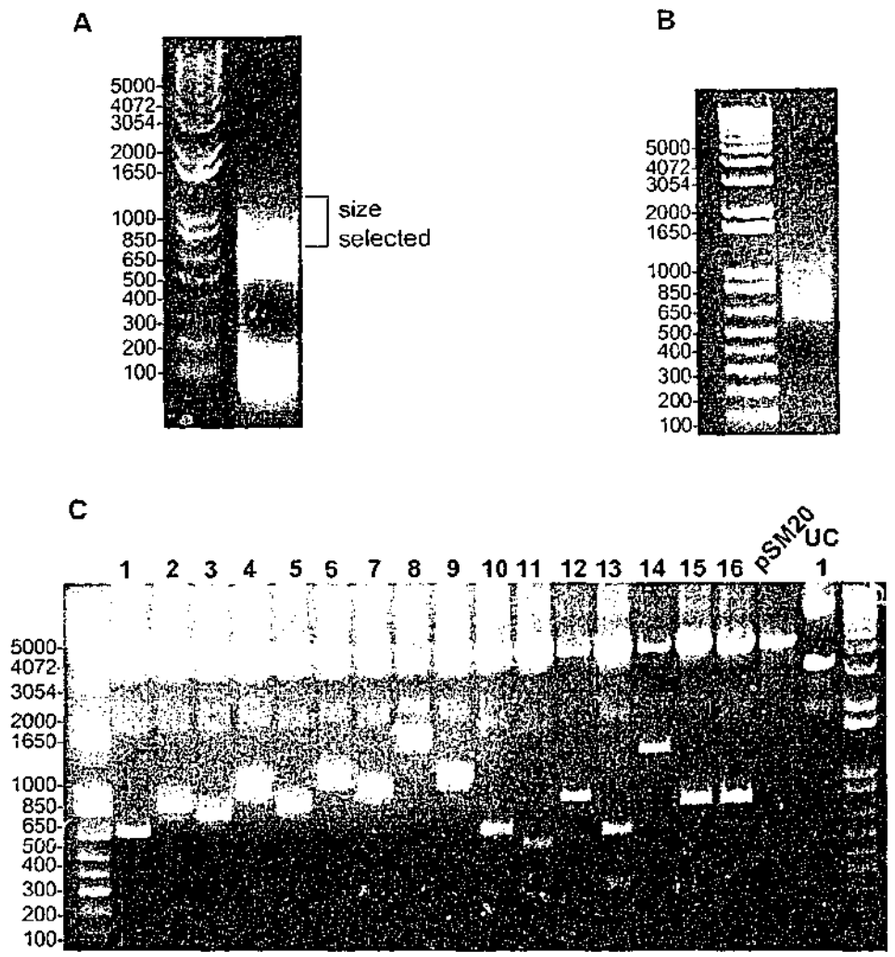

Figure 4.14: Construction of an *AluI* library in pSM20

C231 genomic DNA was completely digested with *AluI* and then analysed by agarose gel electrophoresis (A). DNA in the size range of 0.7 to 1.2 kb was extracted from the agarose gel and reanalysed again by agarose gel electrophoresis (B). The size selected fragments were ligated into pSM20 previously linearised with *Eco*RV and treated with alkaline phosphatase. Upon transformation of the library ligation into Top10 F' *E. coli* 16 colonies were randomly picked and plasmid DNA extracted. Plasmid DNA was digested with *Sph*I and *BamH*I which cut up and downstream of the *EcoRV* site respectively (C). 1 kb plus ladder is run in the first and last lanes and band sizes are indicated to the left of the gel.

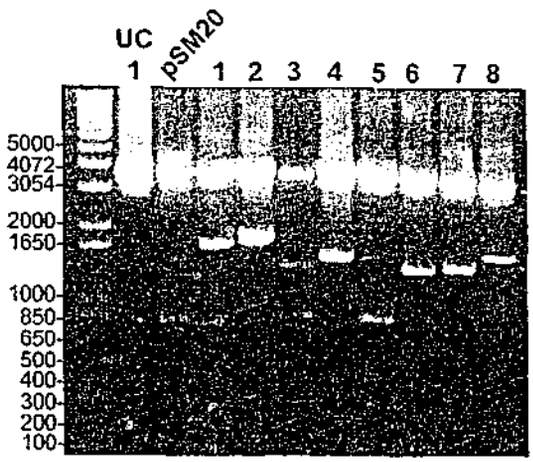

Figure 4.15: Analysis of inserts from pSM20-*Alu*I library in *C. pseudotuberculosis*.

The pSM20-*Alu*I library was generated by insertion of *Alu*I digested, size selected fragments of C231 DNA into the *EcoRV* site of pSM20. The ligation was transformed into C231 and DNA extracted from 8 individual colonies (1 to 8). DNA from the 8 clones and pSM20 was digested with *Pvu*II which cuts out the insert plus an additional 768 bp of vector sequence. Uncut DNA of clone 1 was run as an additional control (UC 1). 1 kb plus ladder is run in the first lane and band sizes are indicated to the left of this.

4.11 Flow cytometry analysis of the *Alu*I sub-libraries

Cultures of the three sub-libraries were generated in BHI from the glycerol stocks and grown overnight. The libraries were diluted 1 in 20 then grown for a further 1.5 hr. The libraries were analysed by flow cytometry (Figure 4.16). In comparison to C231(pSM20) which contained 0.01% fluorescent bacteria, sub-libraries A1-3, A4-6 and A7-9 contained 5.23, 2.77 and 2.92% fluorescent clones indicating that the library contained a proportion of clones with plasmid inserts that possessed promoter activity.

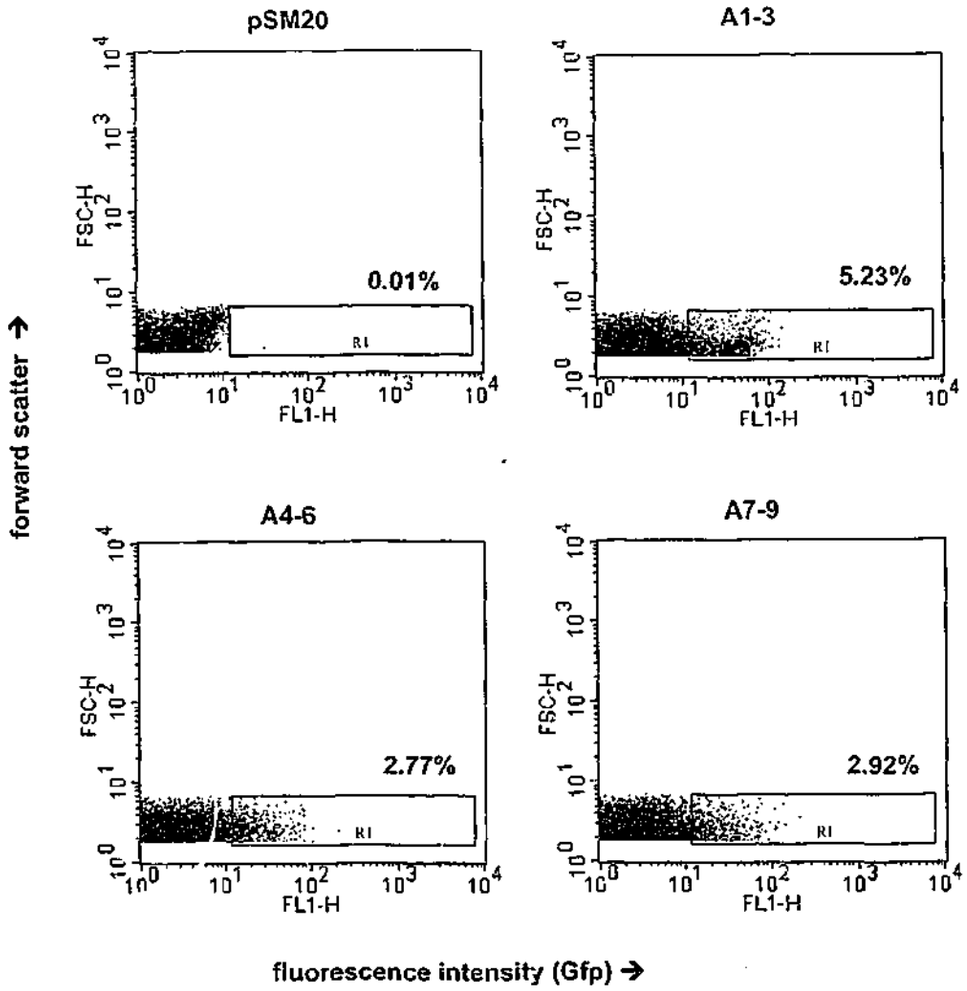

Figure 4.16: Flow cytometry analysis of *Alu*I libaries A1-3, A4-6 and A7-9.

Overnight cultures of the *C. pseudotuberculosis* sub-libraries A1-3, A4-6, A7-9 and C231(pSM20) were diluted to OD_{600} of 0.1 then grown at 37°C for 2 hr. The percentage of Gfp positive clones in each library was determined by flow cytometry. For each culture 30,000 events were analysed. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria. The percentage of events that were contained within Region 1 (R1) is indicated.

4.12 Analysis of *gfp* expressing clones from the *Alu*I library

Ten Gfp positive clones were selected from the *Alu*I library . Gfp positive clones in this series are titled gfp+A1, gfp+A2 onwards. gfp+A1 to gfp+A8 were selected by viewing the plated library under UV light and selecting fluorescent clones (Section 4.10). The remaining Gfp

positive clones were enriched by using a FACS to sort fluorescent clones in the A7-9 library. The fluorescence of gfp+A1 to gfp+A8 was measured in a qualitative manner using visual inspection under a UV light whereas the fluorescence of overnight cultures of the remaining clones was measured by flow cytometry. Sequence analysis indicated that two clones were identified twice. These were gfp+A1 and gfp+A5 which were identified by visual inspection of the library and gfp+A3 and gfp+A12, one each of which was picked using the alternate approaches. The fluorescence and sequence of each clone is summarised in Table 4.2.

In general insert sequences could be divided into two categories. The first included those such as gfp+A11 and gfp+A13 that contained the 5' portion of a gene at the 3' end of the insert such that the insert gene ran into the *gfp* gene. In these cases it is likely that the sequence upstream of the gene contains the promoter activity and as such the gene for the promoter is known. The second group contained those sequences in which the insert contained the 3' end of a gene followed by a region which did not show homologies to proteins in the NCBI protein databases. For a number of clones such as gfp+A2 and gfp+A8 the amount of sequence at the 3' end of the insert where promoter activity is likely to lie is around 70 to 90 bp. Additionally for a number of clones it was possible to identify consensus promoter sequence elements within this region. To determine whether these regions correspond to true promoters it would be necessary to identify downstream sequence and look for the presence of genes. One clone gfp+A1 contained sequence with no major homologies and as such is likely to contain only promoter sequence.

4.13 Identification of heat regulated genes by DFI

4.13.1 Optimisation of heat shock conditions

To determine the appropriate heat shock required to induce heat regulated genes and to be able to observe the induction by flow cytometry a heat shock model using a promoter probe vector containing the *dnaK* promoter was established. The *dnaK* promoter of *C. pseudotuberculosis* (Figure 4.17) was identified by C. Simmons in work that is described in his PhD thesis (Simmons, 1997). The promoter was PCR amplified using primers # 48 and 49. Primer #48 contains an *Sph*I site at its 5' end and primer #49 contains a *Bam*HI site at its 5' end. Both

Table 4.2: Sequence analysis of clones from the pSM20-*Alu*I library that constitutively express Gfp

Clone identification	fluorescence[ab]	insert size[c]	species homology[c]	gene homology and other features[c]
gfp+A1,gfp+A5	+++	692		no major homologies, possible –35 and –10 regions
gfp+A2	++	750	*M. tuberculosis*	3' end of hypothetical Rv3683 which ends 67bp prior to the end of the insert. Consensus –35 and –10 regions in this region
gfp+A3,gfp+A12	+++ 84		*M. tuberculosis*	complementary strand probable CoA ligase (faD14) which starts 75bp from 3' end of the insert.
gfp+A8	+++	580	*Neisseria meningitidis*	3' end of a nitric oxide reductase which ends 85bp from 5' end of the insert. Consensus –35 and –10 regions in this region but poor spacing
gfp+A9	53	ND	*C. glutamicum*	3' end of urease accessory protein that ends 210bp from the end of the insert. Consensus –35 and –10 regions in this region
gfp+A10	51	ND	*E. coli*	hypothetical 7.5kDa protein in the CSTA-AHPC intergenic region located in the middle of the insert
gfp+A11	26	ND	*C. glutamicum*	dihydripicolinate synthase starts 300bp upstream of the insert end and runs into the *gfp* gene.
gfp+A13	39	ND	*M. tuberculosis*	L14 rplN (50S ribosomal protein) starts 300bp upstream of the insert end and runs into the *gfp* gene.

[a] BHI was inoculated with individual colonies of each clone and incubated overnight at 37°C. The mean fluorescence of individual bacteria in each culture was determined by flow cytometry.

[b] Fluorescence of clones grown on agar was scored visually under UV light. Clones were designated as having weak fluorescence (+), medium fluorescence (++) or strong fluorescence (+++).

[c] Sequences were scanned against the NCBI protein databases for protein homologies in all 6 reading frames using the translating BLAST program (BLASTX). Proteins that showed the highest homology and the species in which this occurred are listed.

ND, not determined

the PCR product and pSM20 were digested with *Sph*I and *Bam*HI, then ligated together to generate pSM23.

A heat shock from 37°C to 43°C has been shown to be sufficient to induce expression of *dnaK* (Simmons, 1997 and Chapter 3). To determine whether the heat shock effect on *dnaK* could be observed using pSM23, the effect of heat shock on *gfp* expression from this plasmid was determined. To ensure that any effect observed was specific to the *dnaK* promoter, fluorescence from *C. pseudotuberculosis* harbouring pSM20 or pSM22 was also determined. Overnight cultures of C231(pSM20), C231(pSM22) and C231(pSM23) were diluted to an OD_{600} of 0.1 (multiple cultures were set up for each strain) and then incubated at 37°C for 2 hr. At this stage half of the cultures were transferred to 43°C (t =0). At various time-points cultures were sonicated and the fluorescence intensity of individual bacteria measured by flow cytometry (Figure 4.18). As expected no fluorescence was detected in bacteria harbouring pSM20, the promoterless vector. The level of fluorescence detected in the C231(pSM22) cultures increased approximately 15% at 43°C as compared to 37°C. In comparison, the effect of heat shock on expression from pSM23 was of a greater magnitude and increased over the timecourse of the experiment (up to 300%) although almost maximal levels were achieved by 60 min. From this timecourse a 90 min heat shock was deemed suitable for the detection and enrichment of other heat inducible genes in the library.

4.13.2 Library sorting

The A1-3 sub library was grown from an OD_{600} of 0.1 for 45 min. At this time the culture was split with half being further incubated at 37°C while the remaining aliquot was incubated at 43°C. After a 90 min heat shock, the fluorescence of the two cultures was measured by flow cytometry (Figure 4.19A). 6% of the bacteria in the 37°C culture were fluorescent while 8% of the clones in the heat shock culture were fluorescent. The fluorescent population of the heat shock culture was enriched for using a FACS. Two sorts were performed, the first aimed to collect all fluorescent bacteria while the second collected a subset of this population containing the most fluorescent bacteria. The sorted bacteria were recovered and plated onto BHI agar. After incubation at 37°C for 48 hr colonies were collected from the plates grown in

```
      48 →
  1   ctggcggaaa tagaagattt taaaaaaatt gagtctagtg ggaacaactt

 51   tgcgtgagat gtcgttatgg atagtgtcag gttgagtgac acgcgctcaa

101   gatatctggt aagttcttca cttgtagttg gacggtcccg taaaattgag
                                                  ← 49
151   tcgacgtcag tcaactttag gtaaaacgtt aaaaatacaa ccaaacaccc

201   aggaggacat cac atg gga cgc cca gta gcc atg
                  M   G   R   P   V   A   M
```

Figure 4.17: The *C. pseudotuberculosis dnaK* promoter region.

The dnaK promoter was cloned by Simmons (1997) and the sequence presented here was obtained from his PhD thesis. Regions that show homology to the *E. coli* σ^{70} consensus –35 (TTGACA) and -10 (TATAAT) regions are boxed. A putative ribosome binding site is bolded. Amino acid sequence corresponding to the first seven codons is indicated by the single letter amino acid code. The binding sites of primers used for PCR amplification of the *dnaK* promoter are underlined and the primer number shown. Primer #48 contains an *Sph*I site at its 5' end and primer #49 contains a *BamH*I site at its 5' end for cloning purposes.

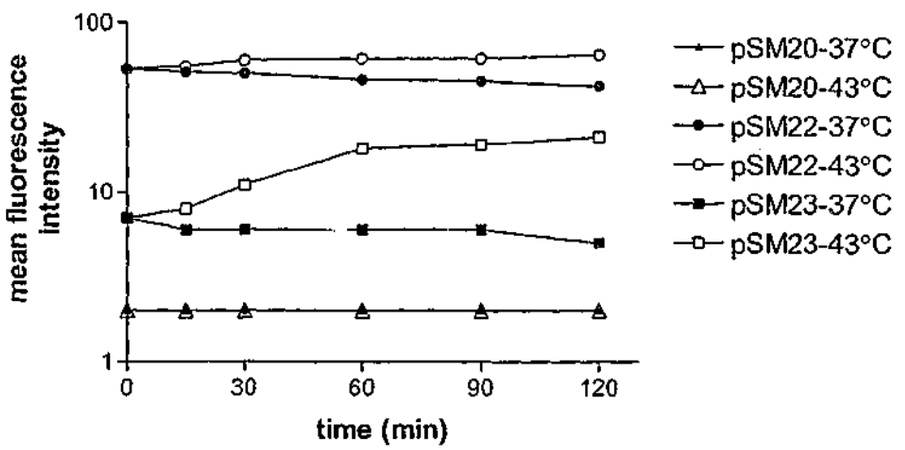

Figure 4.18: Expression of *dnaK* following heat shock of *C. pseudotuberculosis*.

Fluorescence of exponential phase cultures of *C. pseudotuberculosis* harbouring pSM20, pSM22 or pSM23 (promoterless, *srp* promoter and *dnaK* promoter respectively) grown at 37°C or 43°C (following a preincubation for two hours at 37°C) was measured by flow cytometry. For each sample 30,000 events were counted and the mean fluorescence of the population determined to give mean fluorescence intensity.

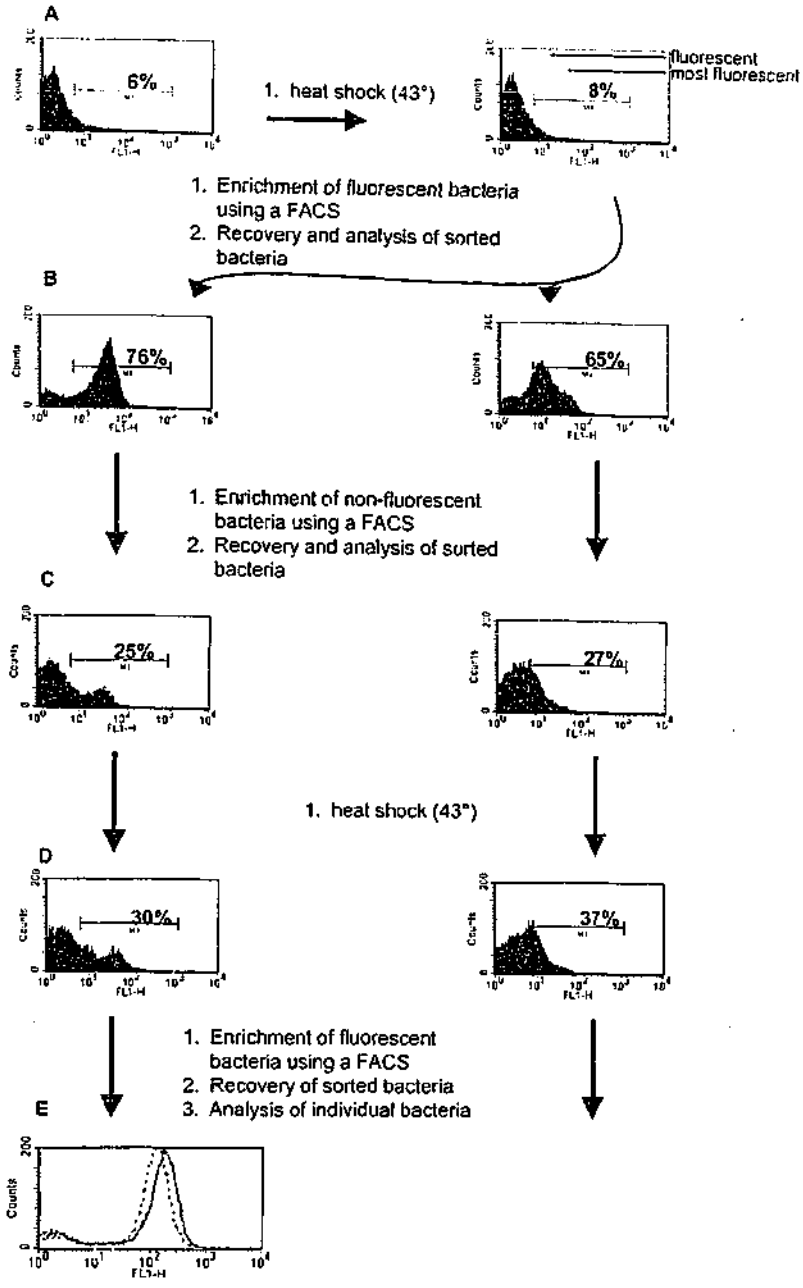

Figure 4.19: FACS sorting of C231 genomic sub-library A1-3.

Analysis of the A1-3 sub-library indicated that it contained 6% Gfp positive clones when incubated at 37°C and 8% when incubated at 43°C (A). Following a 90 min heat shock two populations of bacteria were collected, those that were fluorescent and a subpopulation of this group that were most fluorescent. Bacteria were recovered on BHI agar plates and then reanalysed by flow cytometry (B). The enriched bacterial population was grown at 37°C and a negative sort performed to enrich the non-fluorescent population. Bacteria were recovered on BHI agar plates and then reanalysed by flow cytometry (C). The recovered population was again heat shocked, analysed (D) and fluorescent clones enriched for. Sorted bacteria were recovered on BHI agar plates and then analysed individually in the heat shock model (E). FL1-H corresponds to Gfp fluorescence intensity while counts corresponds to cell number.

BHI at 37°C. The fluorescent population now contained 65% Gfp positive bacteria and the most fluorescent population contained 76% Gfp positive bacteria (Figure 4.19B). This corresponded to fold enrichments of 11 and 13 respectively. To remove clones containing promoters with constitutive activity these cultures were sorted using a FACS, this time enriching for the non-fluorescent population. The sorted bacteria were recovered and plated onto BHI agar. Analysis of the collected bacteria indicated that the percentage of Gfp positive clones of fluorescent and most fluorescent populations had been reduced to 25 and 27% respectively (Figure 4.19C). Upon heat shock of these cultures the percentage of fluorescent clones rose to 30 and 37% respectively (Figure 4.19D). As a final round of positive selection these cultures were incubated at 43°C for 90 min then the fluorescent population enriched for. The sorted bacteria were recovered on agar plates and analysed individually (Figure 4.19E).

4.13.3 Analysis of individual clones

To determine whether the clones obtained in the final library sort were indeed heat inducible, clones were analysed individually in the heat shock model. Thirty colonies were randomly picked, grown overnight in BHI then diluted 1 in 10 in duplicate. After a 45 min incubation at 37°C half of the cultures were transferred to 43°C. After a further 90 min the fluorescence of individual bacteria within each culture was determined by flow cytometry. Of the 30 clones analysed none showed fluorescence induction. For six of the clones this is represented graphically in Figure 4.20.

4.14 Identification of *C. pseudotuberculosis* genes induced in a macrophage infection model

The macrophage infection model that was described in Chapter 3 was also used for the following experiments.

4.14.1 *dnaK* promoter activity in the macrophage infection model

From data presented in Chapter 3 it is evident that expression of *C. pseudotuberculosis dnaK* is up-regulated during macrophage infection. In order to pick suitable time points for sorting of macrophage induced genes of *C. pseudotuberculosis* the C231(pSM23) strain was tested in the macrophage infection model and the level of promoter induction determined.

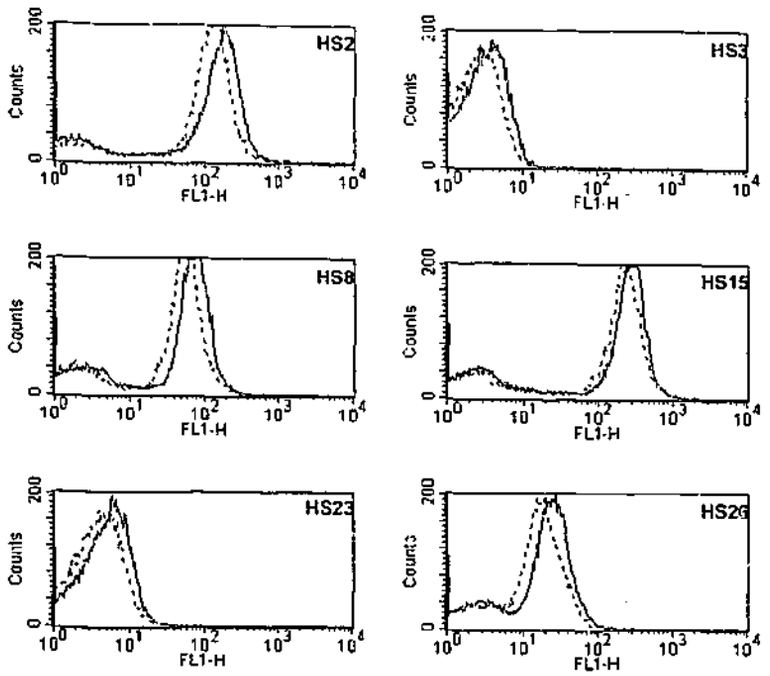

Figure 4.20: Fluorescence induction of individual clones from the *C. pseudotuberculosis* A1-3 sub-library following heat shock.

DFI enriched clones were either incubated at 37°C or 43°C for 90 min. The fluorescence of individual bacteria within each culture was determined by flow cytometry. For each clone the unbroken line corresponds to the culture grown at 37°C, while the broken line corresponds to the culture grown at 43°C. FL1-H corresponds to Gfp fluorescence intensity while counts corresponds to cell number.

J774 macrophages were plated at 7x 10^5 cells/well in six well plates 18 hr prior to infection. The macrophages were infected with exponentially growing C231(pSM23) and C231(pSM22) at MOI of 0.5 and 3. At 1 hr post infection the macrophages were washed and gentamycin containing media added to the cells. Cells infected at MOI 3 were analysed at 4 hr post infection and those infected at MOI 0.5 were analysed at 16 hr post infection. Sonicates of the macrophage monolayers were generated and the fluorescence of individual bacteria within the sonicate determined by flow cytometry. In addition the fluorescence of the bacteria added to the macrophages was also determined to allow for the calculation of fold induction following macrophage infection.

Following infection with C231(pSM22) no induction of fluorescence was observed at either timepoint. However, for C231(pSM23) fluorescence inductions of 1.5 and 2.6 were detected

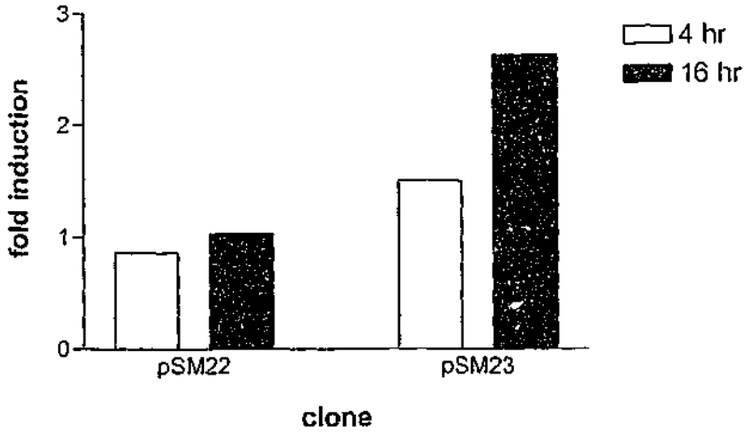

Figure 4.21: Induction of *C. pseudotuberculosis dnaK* expression following macrophage infection.

J774 macrophages were infected with exponentially growing C231(pSM22) and C231(pSM23). At 4 and 16 hr post infection the macrophage monolayers were lysed, then sonicated and the fluorescence of individual bacteria determined. The fluorescence of macrophage derived bacteria was divided by the fluorescence of the bacteria initially added to the cells to give fold induction.

at 4 and 16 hr respectively (Figure 4.21). The level of fluorescence detected at 16 hr in C231(pSM23) was directly comparable to that observed in the same strain grown *in vitro* following incubation at 43°C for 2 hr.

Thi. experiment showed that it was possible to detect regulated genes in the macrophage infection model. From this experiment it was decided to look for other macrophage regulated genes 16 hr post infection as at least for the *dnaK* promoter this appeared to maximise the differential gene expression detected.

4.14.2 Optimisation of macrophage sorting conditions

Given the technical difficulties with the heat shock experiments it was decided to sort infected macrophages rather than bacteria released from infected macrophages. This was likely to increase the sorting efficiency as, due to their size, macrophages are more easily sorted than bacteria. Additionally at 16 hr post infection macrophages may contain several bacteria which, assuming that they are derived from a single bacterium, would increase the relative

fluorescence of the sorted event. This has the potential to further separate the fluorescent from the non-fluorescent population leading to more efficient sorting.

As macrophages were being sorted rather then single cell suspensions it was necessary to infect the bacteria at a low multiplicity of infection. This was done so that on average a macrophage was infected by one or less bacteria. This was essential to avoid infecting macrophages with a mixed population of bacteria that contained both fluorescent and non-fluorescent clones.

Upon analysis of J774 macrophages by flow cytometry it was apparent that they were intrinsically fluorescent. To dampen this effect two approaches were taken, the first was to reduce the laser power used during flow cytometry, the second was to use tissue culture media that does not contain phenol red. Phenol red is used as a pH indicator in tissue culture media, but fluoresces upon excitation with the FACS laser. Phenol red, located on the macrophage cell surface or that has been internalised by the macrophages, would be expected to increase the intrinsic fluorescence of the cells.

To optimise sorting conditions macrophages were infected with mixed populations of fluorescent and non-fluorescent *C. pseudotuberculosis*. The fluorescent and non-fluorescent strains C231(pSM22) and C231(pSM20) were utilised. J774 cultures were infected at a low multiplicity of infection (1:1) with either C231(pSM20) or C231(pSM22) individually or as mixed populations for a period of 16 hr. Macrophages containing fluorescent bacteria could be readily distinguished from the remaining population (Figure 4.22). There was however a slight increase in fluorescence in bacteria infected with C231(pSM20) (Figure 4.22C) indicating that the intrinsic fluorescence of all infected macrophages is increased. A FACS was used to sort macrophages containing fluorescent bacteria from uninfected macrophages or macrophages infected with bacteria carrying promoterless constructs. Macrophages were sorted on the basis of fluorescence intensity and forward and side scatter . The sorted macrophages were collected by centrifugation and plated. After 48 hr at 37°C the total number of bacteria and the number of fluorescent bacteria was determined by viewing the plates under UV light. When infected with mixed populations of non-fluorescent and fluorescent bacteria (at 80:20, 90:10 and 95: 5 ratios), C231(pSM22) made up approximately 92 to 94% of the recovered sorted bacteria representing relative enrichments of 4.6, 9.4 and

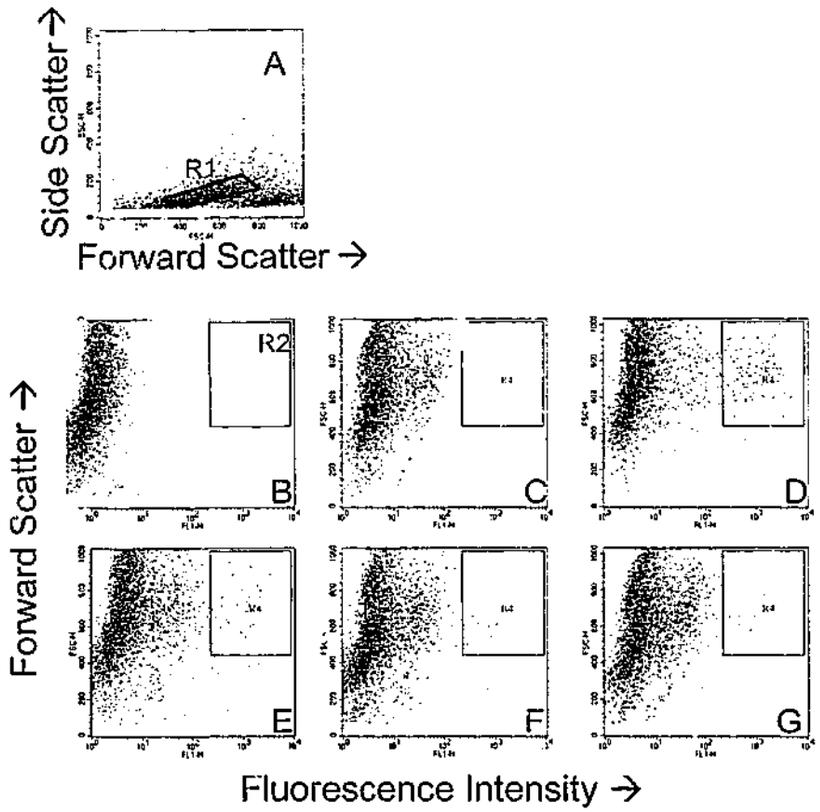

Figure 4.22: Flow cytometry of infected macrophages.

Macrophages were infected with C231(pSM 0) (A and C), C231(pSM22) (D) or 80:20, 90:20 and 95:5 ratios of the two strains (E, F and G respectively). Compared to uninfected macrophages (B) the fluorescent population is clearly visible (C to G).

18.6. This indicated that a reasonably good sorting efficiency could be achieved when infecting with a mixed population of bacteria. Additionally in a mixed culture of C231(pSM20) and C231(pSM22) in which the non-fluorescent bacteria were the minority species (90% fluorescent: 10% non fluorescent) it was possible to use a FACS to collect the non-fluorescent population. Upon plating of the sorted population the percentage of fluorescent bacteria had dropped from 90% to 7% and the non-fluorescent population had risen from 10% to 93%.

In part the reason why it was possible to so successfully sort C231(pSM22) from C231(pSM20) was because C231(pSM22) is a highly fluorescent strain. There is therefore a large difference in the fluorescence intensity of the two strains. Initial experiments of sorting the *Alu*I libraries indicated that it was not possible to achieve the same level of enrichment with the libraries as with the mixed populations of pSM20:pSM22. This was probably because the *Alu*I libraries contain only a very small proportion of highly fluorescent clones with a larger population of weak to medium fluorescing clones. When the difference between the fluorescence intensity of clones that are to be sorted and those that are to be abandoned becomes smaller, the sorting efficiency lowers. To deal with this problem an additional enrichment step was incorporated into the standard DFI protocol as outlined in Figure 1.1. This step comprised a second round of enrichment in the macrophage model such that bacteria collected after one macrophage infection were used to reinfect macrophages. The bacteria collected from this second infection were then grown under *in vitro* conditions and a negative sort performed.

4.14.3 Library sorting

The sub library A7-9, which contained 5% Gfp positive clones, was used to infect J774 macrophages at a MOI of 2. 16 hr post infection the macrophages were removed from the plastic surface by scraping into PBS. Analysis of the infected macrophages by flow cytometry indicated that 0.3% of the macrophages were strongly fluorescent. This is much lower than the percentage of Gfp positive *C. pseudotuberculosis* clones within the library as a low MOI is utilised. The consequence of this is that only a relatively small population of macrophages phagocytose a bacterium resulting in a reduction in the percentage of Gfp positive "events". The fluorescent population was enriched by FACS and plated onto BHI agar. After incubation at 37°C for 48 hr the bacterial colonies were collected from the plate and grown overnight in BHI. Flow cytometry analysis of this culture indicated that 53% of the population was fluorescent, corresponding to an 11 fold enrichment over the A7-9 sub library. A second round of macrophage infection was performed with this culture and after 16 hr analysis of the infected macrophages by flow cytometry indicated that 1.9% of the macrophages were strongly fluorescent. The fluorescent population was again enriched and plated out as before. Analysis of the fluorescence of this population indicated that 85% of bacteria were fluorescent corresponding to a further 1.6 fold enrichment. To remove clones containing promoters with

constitutive activity the collected bacteria were grown in DMEM for 16 hr. Flow cytometry analysis indicated that 58% of the bacteria were fluorescent. A negative sort was then performed in which the non-fluorescent population was enriched for. Plating out of the sorted population resulted in 111 colonies after incubation at 37°C for 48 hr. The sorting process is represented schematically in Figure 4.23.

4.15 Analysis of individual sorted clones

4.15.1 Sequence analysis

The final sort yielded 111 colonies. Twenty colonies were randomly chosen for further analysis and labelled C1 to C20. The fluorescence of each clone grown overnight in BHI was determined by flow cytometry and the fluorescence of each clone on agar plates assessed visually under UV light (Table 4.3). In addition DNA was extracted from each clone and used to retransform *E. coli* so that plasmid DNA could be prepared for sequencing. For a number of clones the expression of *gfp* in *E. coli* was high enough so as to be toxic to the cells. This was observed as an initial reduction in *E. coli* growth followed by a return to normal growth but a reduction in plasmid copy number. As it was not possible to extract sufficient plasmid DNA from these clones, the inserts of C3, C4 and C17 were PCR amplified using primers #20 and #21 then subcloned into pGEM®-T easy. These plasmids were then used as templates for sequencing. The fluorescence characteristics of each clone is summarised in Table 4.3. Sequence analysis indicates that 12 independent clones were represented within the 20 clones. One clone was represented twice, one thrice and a third five times (Table 4.4).

4.15.2 Measurement of fold induction

To determine what proportion of the clones were actually macrophage induced each clone was analysed individually one or more times in the macrophage infection model. J774 macrophages in 6 well plates were infected at a MOI of 2 with each clone. At 4 and 16 hr post infection the fluorescence intensity of individual bacteria extracted from the macrophage monolayer was determined by flow cytometry. For calculation of fold induction, fluorescence of (i) cultures of the same bacteria grown overnight in BHI and (ii) bacteria grown in DMEM under the same conditions and for the same period of time as the macrophage monolayers was measured. A third control was to include C231(pSM22) which contains the constitutively expressed *srp* promoter. Fold induction was determined by dividing the fluorescence of

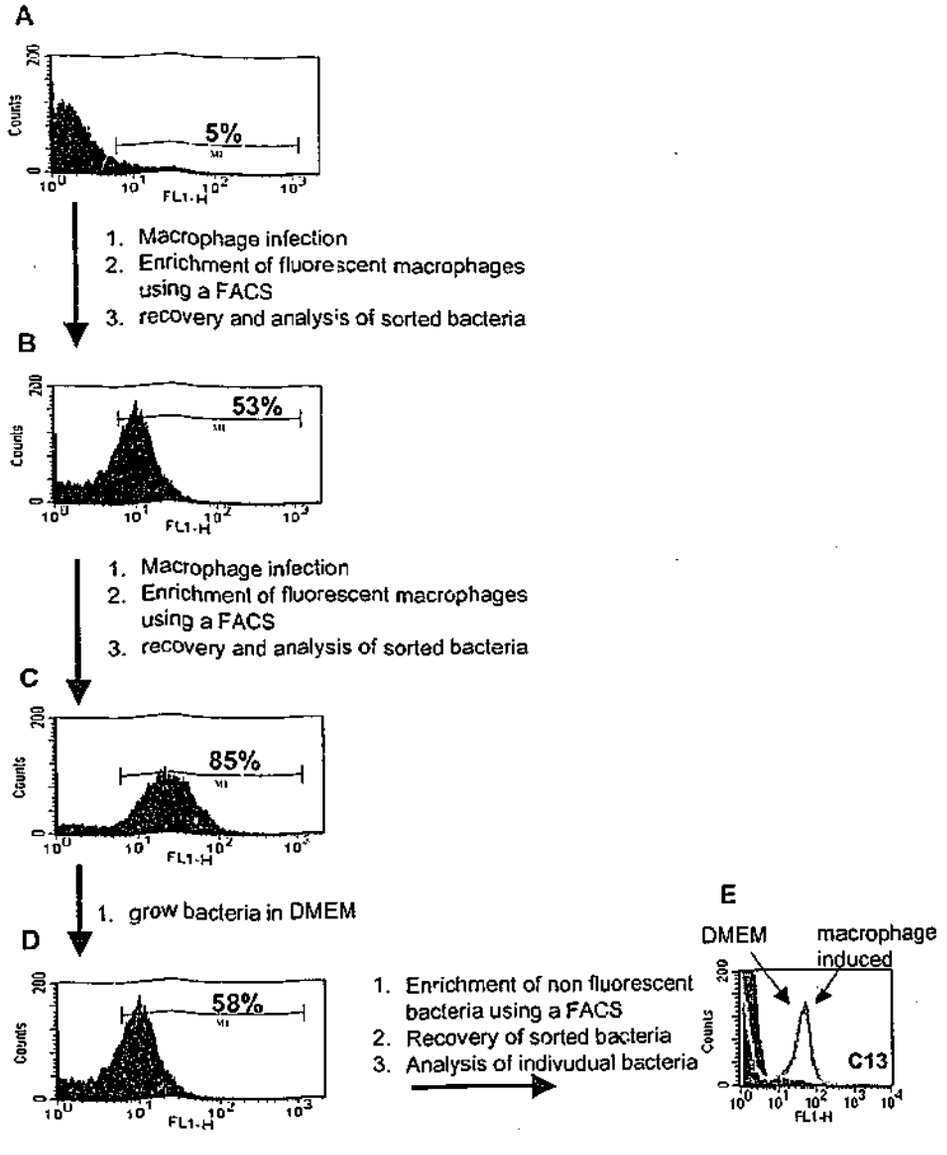

Figure 4.23: FACS sorting of C231 genomic sub-library A7-S.

Analysis of the A7-9 sub-library indicated that it contained 5% Gfp positive clones (A). J774 macrophages were infected with the library and 16 hr later fluorescent cells were enriched for using a FACS. Bacteria were recovered on BHI agar plates and then reanalysed by flow cytometry (B). The enriched bacterial population was used to reinfect macrophages and a second round of sorting performed. Collected bacteria were again analysed (C). Collected bacteria from the second round of sorting were grown in DMEM for 16 hr (D), then a FACS was used to enrich for the non-fluorescent population. Sorted bacteria were recovered on BHI agar plates and then analysed individually in the macrophage infection model (E). FL1-H corresponds to Gfp fluorescence intensity while counts corresponds to cell number.

macrophage derived bacteria by the fluorescence of DMEM derived bacteria (Figure 4.24A). Almost all clones showed fluorescence induction at both timepoints in macrophages compared to in DMEM. In general when different clones containing the same insert or the same clones were studied in separate experiments a reproducible fold induction was observed, although there were occasional differences. Although all the DFI clones showed induction so did the C231(pSM22) control strain which at 16 hr was induced by 2.5-3 fold in macrophages compared to when grown in DMEM. A number of clones were induced past this level, in particular C1, C13 and C18 . As an alternative analysis of the data, fold induction was expressed as relating to the fluorescence of the added bacteria. In general this meant taking the highest recorded fluorescence for a clone when grown overnight in BHI. Growth in such a manner generally produced a consistent level of fluorescence for a given clone. When the data was analysed this way C231(pSM22) was uninduced, perhaps indicating that this was a more suitable control (Figure 4.24B). Only C18 showed induction at both 4 and 16 hr post infection when analysed in this manner. A number of clones showed a greater than 1.5 fold induction at 16 hr. These were C3,C4, C12, C17, C6, C15 and C18. Of these C3, C6, C15 and C18 were chosen for further study. Given that C13 showed the greatest induction compared to bacteria grown in DMEM it too was chosen for further study.

4.15.3 Identification of downstream sequence

The DFI approach aims to identify regulated promoters. Once such promoters were identified it was necessary to determine the identity of the genes that these promoters transcribed. For C3, C6, C13 and C15 there was no indication of what gene was associated with the detected promoter activity. For C18 the 5' end of the gene was present in the insert however the function of the protein from this small region was not entirely clear. To allow these clones to by studied further it was necessary to identify downstream sequence. Two approaches were taken. The first utilised a PCR based approach. C231 genomic DNA was digested with *Sma*I or *Bam*HI and ligated into pUC18 digested with *Sma*I or *Bam*HI respectively. For each promoter a forward primer was designed that hybridised to DNA sequence approximately 200 bp 5' of the end of the insert. A PCR using the ligations as the template was performed using the forward primer and either the M13 forward or M13 reverse primers which hybridise to sequences in pUC18 as the reverse primer. The PCR products generated were analysed by agarose gel electrophoresis, which in general revealed the presence of one or more bands of

Table 4.3: Fluorescence characteristics of clones that may show fluorescence induction following macrophage infection

Clone ID	mean fluorescence [a]	C231 fluorescence on BHI agar plate [b]	Top10 fluorescence on LB agar plate [b]
C1	17	+	++
C2	32	++	++
C3	24	++	*
C4	23	++	*
C5	23	++	*
C6	12	-	++
C7	4	-	++
C8	11	-	++
C9	4	-	++
C10	23	++	*
C11	10	+	++
C12	24	++	*
C13	31	++	++
C14	25	++	+++
C15	11	-	+
C16	4	-	+
C17	24	++	*
C18	2	+++	+++
C19	17	-	++
C20	35	+++	+

[a] BHI was inoculated with *C. pseudotuberculosis* colonies of each clone and incubated overnight at 37°C. The mean fluorescence of individual bacteria in each culture was determined by flow cytometry.

[b] Fluorescence of clones grown on agar plated was scored visually under UV light. Clones were designated as having weak fluorescence (+), medium fluorescence (++), strong fluorescence (+++) or no detectable fluorescence (-).

*. upon transformation into *E. coli* apparent toxicity was observed by the presence of a large number of pin prick sized colonies plus a smaller number of normal sized colonies. When plasmid DNA was extracted from cultures derived from the larger colonies, a lower than usual DNA yield was obtained suggesting that there has been a reduction in copy number due to toxicity as a result of high levels of *gfp* expression.

Table 4.4: Sequence analysis of the plasmid inserts of FACS enriched clones from the A7-9 sub-library

Clone ID	insert size	species homology [a]	gene homology and other features [a]
C1, C11, C12	630	*M. tuberculosis*	3' end of a lysyl-tRNA synthase which ends 230bp into the insert, some putative –35 and –10 regions downstream of this.
C2	640	*M. tuberculosis*	3' end of a polyphosphate glucokinase, reasonably long ORF starts shortly after this.
C3, C4, C5, C10, C17	596		no significant homologies
C6	920	*M. tuberculosis*	conserved hypothetical protein Rv2226 at 5' end which ends approximately 400bp in, possible consensus sequences at 3' end
C7, C16		*Streptococcus pyrogenes*	very low homology to a putative biotin synthase which starts 170bp into the fragment and runs to the end
C8	600	*M. tuberculosis*	glycerol-3-phosphate dehydrogenase (NAD dependant). gene ends approximately 100 bp prior to the 3' end of the insert
C9	800		low homologies to non bacterial proteins
C13	591		no homology to bacterial proteins 400bp ORF at the 5' end of the gene
C15	900	*M. tuberculosis*	3' end of a conserved hypothetical protein at the 5' end of the insert that ends approximately 400bp into the insert
C18	492	*C. glutamicum*	*dtsR* gene which starts at the 3' end of the insert and runs into the *gfp* gene.
C19	770	*N. meningitidis*	nitric oxide reductase gene starts 300bp into the insert and runs into the *gfp* gene.
C20	700	*Pseudomonas stutzeri*	in the middle of the insert is a sequence that shows homology to part of a hypothetical 12.7 kda protein in the ptx operon 5'....

[a] Sequences were scanned against the NCBI protein databases for protein homologies in all 6 reading frames using against a translating BLAST program (BLASTX). Proteins that showed the highest homology and the species in which this occurred are listed.

Figure 4
sub-libr
J774 ma
hr. The
by flow c
of the sa
induction
(B) was
example

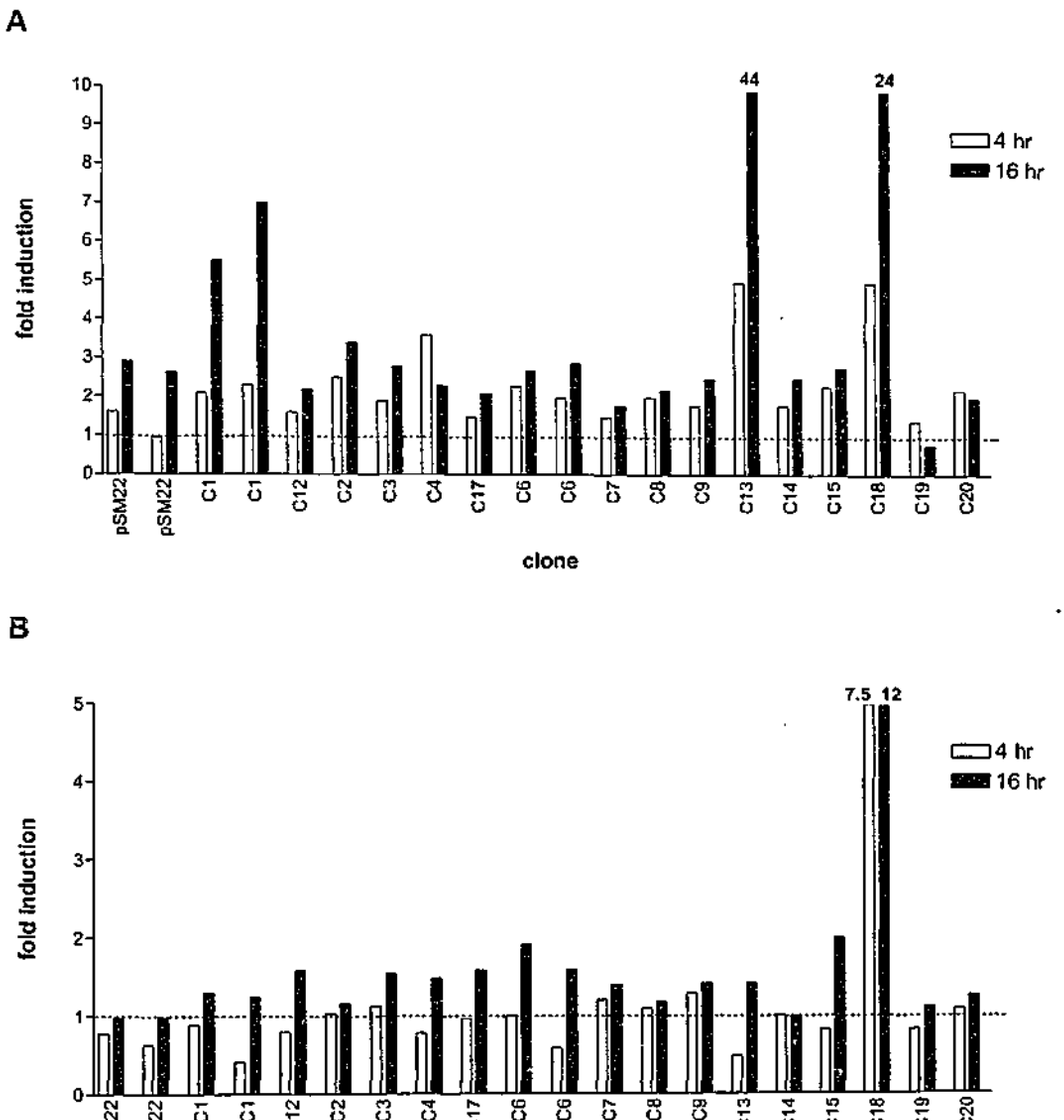

Figure 4.24: Fluorescence induction of individual clones from the *C. pseudotuberculosis* A7-9 sub-library following macrophage infection.

J774 macrophages were infected with DFI enriched clones C1 to C20 or with C231(pSM22) for 4 or 16 hr. The fluorescence of individual bacteria extracted from the macrophage monolayer was measured by flow cytometry and the mean fluorescence of each clone determined. Additionally the fluorescence of the same clones incubated for the same period in DMEM or overnight in BHI was determined. Fold induction following macrophage infection compared to DMEM grown bacteria (A) or BHI grown bacteria (B) was then calculated. Clones that contain the same plasmid insert are grouped together, for example C3, C4 and C17.

interest. Individual bands were extracted from the gel and cloned into pGEM®-T easy. Sequencing was then performed to identify clones that contained the downstream sequence. This approach was successful for a proportion of the clones identifying downstream sequence for C3, C13 and C18. Due to technical difficulties and limitations with this approach it was not possible to identify downstream sequence for C6 and C15.

A second approach was utilised to identify downstream sequence of C6 and C15. This involved using the insert of these clones as a probe on the DNA arrays generated for transcriptional analysis (Chapter 3). A positive spot was identified for C15 however sequence analysis of the corresponding clones indicated that it contained upstream sequence rather than downstream sequence. No positives were identified when using the C6 insert as a probe. Given the difficulties in identifying the contiguous genes for C6 and C15 these clones were not analysed further. It is anticipated that the screening of larger genomic libraries would allow the genes to be identified.

4.15.4 C3

This clone was enriched for more than any of the others. Although sequence analysis was only performed on C3, C4 and C17, it is likely that C5 and C10 contain the same insert as the insert size of these clones is the same and the plasmids demonstrate the same toxic effects in *E. coli*. As described in section 4.15, *gfp* expression from this promoter in *E. coli* is toxic to the cells resulting in a reduction in plasmid copy number. The insert was 596 bp and 56% AT rich. No homologies to bacterial proteins were observed following BLASTX analysis of the sequence. A further 350 bp of sequence was identified using the PCR approach (Figure 4.25). A BLASTX search indicated that a protein with homology to the β chain of DNA polymerase III starts 13 bp prior to the end of the original insert. The β chain of DNA polymerase III is encoded for by *dnaN*. The *dnaN* gene of *C. pseudotuberculosis* shows greatest homology to the gene in *M. smegmatis* with 53% identity over a 101 amino acid region (Figure 4.26). In many bacteria including mycobacteria the order of genes in this region is conserved such that the gene upstream of *dnaN* is *dnaA* which codes for the chromosomal replication initiator protein. In addition, the intergenic region between *dnaA* and *dnaN* is often the site of the chromosomal origin of replication (*oriC*) and as such the presence of this region on a plasmid may be sufficient to render a nonreplicative plasmid capable of autonomous replication. This

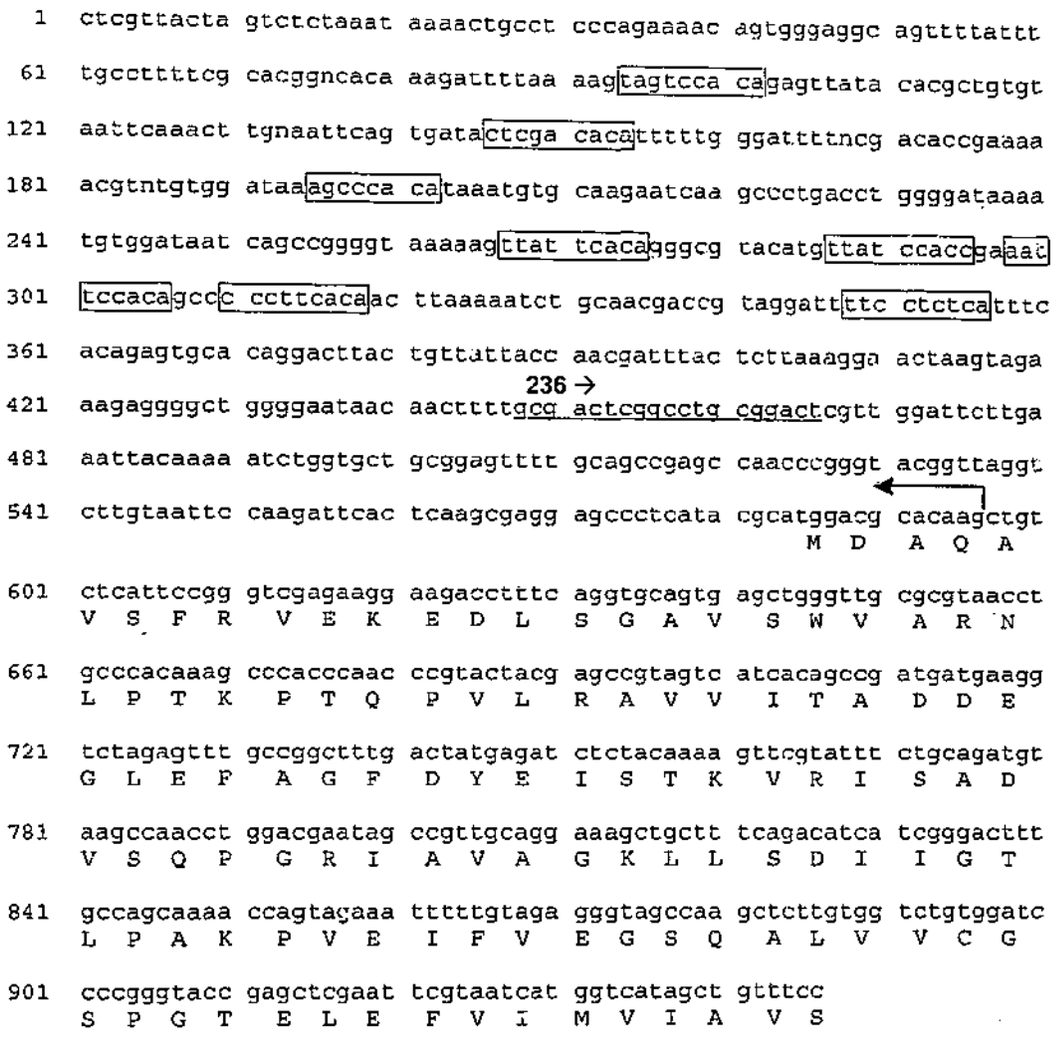

Figure 4.25. DNA sequence of the plasmid insert of C3 and further downstream sequence.

All sequence upstream of the arrow is from the original C3 clone. Sequence downstream of the arrow corresponds to sequence identified from the PCR generated clone. The sequence of primer #236, which was used to identify the downstream sequence, is underlined. Protein sequence with homology to the beta chain of DNA polymerase III (coded by *dnaN*) is shown using the amino acid single letter code. Shaded or boxed sequence indicate the positions of sequence with 1 to 3 mismatches to the *M. smegmatis* DnaA-box consensus (TT(G/C)TCCACA) are marked. Sequences that are boxed run in the forward direction while shaded sequences indicate that the consensus sequence lies on the complementary strand. A putative RBS is doubly underlined.

```
C.ptb  -----MDAQAVSFRVEKEDLSGAVSKVARNLPTKPTQPVLR-AVI ADDEGLEFAGFDYE 54
M.sm   MATTTAGLTDLKFRVVREDFADAVAKVARSLPTRPTIPVLAGVLLTGTDEGLTISGFDYE 60
              .   :.*** :**::.**:****.***:** ***  .::*. **** ::*****

C.ptb ISTKVRISADVSQPGRIAVAGKLLSDIG-TLPAKPVEIFVEGSQALVVCGSPGTELEFVI 113
M.sm  VSAEVKVSAEIASAGSVLVSGRLLSDITKALPAKPVEVSVEGTRVSLTCGSARFSLPTLA 120
      :*::*::**::::..* : *:*:***** :******: ***::. :.***.  .*  :

C.ptb MVIAVS---- 119
M.sm  VEDYPALPAL 130
      :    :
```

Figure 4.26. Clustal alignment of the predicted amino acid sequences of the N-terminal end of DNA polymerase III β chain from *C. pseudotuberculosis* (C.ptb) and *Mycobacterium smegmatis* (M.sm).

Residues that are identical in both species are indicated by an asterisk. Conserved amino acid substitutions are indicated by two dots while semi-conserved substitutions are indicated by a single dot. Alignment was performed using CLUSTAL W (1.82).

region is characterised by a number of features including an A+T rich cluster and DnaA boxes which have a consensus sequence of TT(G/C)TCCACA. Inspection of the C17 sequence revealed the presence of at least 11 putative DnaA boxes. These ranged from 8/9 to 6/9 identical to the consensus sequence.

4.15.5 C13

C13 contains an insert of 591 bp that is rich in A and T bases (64%). An 87 amino acid ORF is present at the 5' end of the insert, this however does not show homology to any proteins in the NCBI protein databases. Sequence downstream of the C13 insert was identified and analysed by BLASTX (Figure 4.27). A protein with 43% homology to a non-ribosomal peptide synthase of *Streptomyces avermitilis* was identified (Figure 4.28). This gene is predicted to start at a gtg codon positioned 102 bp after the end of the C13 insert. Homologies to a large number of other peptide synthases was also observed, although the C13 coding region often showed homology to the middle of these protein rather than the start. We have termed this gene *cps* for corynebacterial peptidase synthase.

```
   1  tgaaaatttc agaccactcg ccttcaatga atatcggaag attaaaaact ccccctgccg

  61  ttcatttagg cctgccccac cctataaaac agccagacga tatatcagta ctttggccat

 121  tctttcgcc gcctacccca gatcacaaac aggaaaagaa agcattccgt ccagccctta

 181  ttcctatttt caccgaccta ggcaaatagg tttatttac ttatataaga tttggaaatt

 241  gcctattaaa agaactaatt aagcaatgcc taaccgtata tttcccgatt aaaaagtatc

 301  atttttcaca tgcgttcaaa caactcgcaa tagtgactaa agggcacttc aaatggcttt

                                                          242→
 361  gcaatagttt ttaaatccct taaaatgtgc ttaatttcac accttgtatt cccgtaatgc

 421  atacgcgaaa agatagaaaa gcctcacaac cagcacggaa ataaaaaccc ctatttttcta

 481  tcgataaaat aataatattt cactgatttg catatgactt aaatatgtta ttaagaatat

 541  ttcatcaggt tgagtgtaaa ttactcttat cgctcaatcg gaaactatta gctttctaaa

 601  aagcacagac tatatctggg caatagttca ctgacagggg tctccgctaa caaactttgg

 661  aggcttttaa agcaattgaa aggaactttt tcggtgggaa aaccagcccc atccgacgac
                                        V  G   K  P  A   P  S  D  D

 721  cgtacgtatg caatctacgg cgtcgatacc atccccgaac ctcgcacgct cgttgatatt
        R  T  Y   A  I  Y   G  V  D  T   I  P  E   P  R  T   L  V  D  I

 781  ttcagcgaca ccgtaatcag ctaccccaac gccaccgctc ttattggcac aaatgaatct
        F  S  D   T  V  I   S  Y  P  N   A  T  A   L  I  G   T  N  E  S

 841  ctgacttaca gcgagttagc agagcgcgtc gagaagcaaa tagagcgctt ggcggctctc
        L  T  Y   S  E  L   A  E  R  V   E  K  Q   I  E  R   L  A  A  L

 901  ggggtcggcc gcggagcgcg catcggaatt cgggtaccct caggcaccac cgacctgtac
        G  V  G   R  G  A   R  I  G  I   R  V  P   S  G  T   T  D  L  Y

 961  atagcaatac ttgccacgat ctgtgctggg gcggcctacg ttcctgttga ctgggacgat
        I  A  I   L  A  T   I  C  A  G   A  A  Y   V  P  V   D  W  D  D

1021  ccagattctc gagctaacac agtatgggaa gaagccaatg tcactgctgt ctatggcgcc
        P  D  S   R  A  N   T  V  W  E   E  A  N   V  T  A   V  Y  G  A

1081  gaaactttcc ctcgacctc
        E  T  F   P  R  P
```

Figure 4.27: DNA sequence of the plasmid insert of C13 and further downstream sequence.

All sequence upstream of the arrow is from the original C13 clone. Sequence downstream of the arrow corresponds to sequence identified from the PCR generated clone. The sequence of primer #242, which was used to identify the downstream sequence, is underlined. Protein sequence with homology to a non-ribosomal peptide synthetase from *Streptomyces avermitilis* is shown using the single letter code. Amino acid sequence with homology to a consensus sequence of the A domain is bolded. A putative RBS is doubly underlined.

```
C.ptb  -----VG-KPAPSDDRTYAIYG------VDTIPEPRTLVDIFSDTVISYPNATALIGTNE 48
S.av   MAALQQGPALALSDDEVRAEFGDRARFSAGSPASPRTLVDILDASVRAYPDEPALDDGRR 60
              *   * ***.. * :*       ..: ..*******:. :* :***: .** . ..

C.ptb SLTYSELAERVEKQIERLAALGVGRGARIGIRVPSGTTDLYIAILATICAGAAYVPVDWD 108
S.av  CLTYRALAVEIEALRRRLAAAGVGLGDRVGVRVPSGTNDLYIAVLAVLAAGAAYVPVDAE 120
      .***  ** .:*   .**** *** * *:*:******.*****:**.:.********** :

C.ptb DPDSRANTVWEEANVTAVYGAETFPRP--- 135
S.av  DPDERAELVFGEAEVRAVVGAGHHLTVDGT 150
      ***.**: *: **:* ** **  .
```

Figure 4.28. Clustal alignment of the predicted amino acid sequences of the N-terminal end of a non ribosomal peptide synthase from *C. pseudotuberculosis* (C.ptb) and *Streptomyces avermitilis* (S.av).

Residues that are identical in both species are indicated by an asterisk. Conserved amino acid substitutions are indicated by two dots while semi-conserved substitutions are indicated by a single dot. Alignment was performed using CLUSTAL W (1.82).

4.15.6 C18

The C18 insert was 492 bp in length. The 3' end of the plasmid insert of C18 codes for a protein with homology to DtsR from *C. glutamicum* (Figure 4.29). This protein has been suggested to contain propionyl CoA carboxylase activity however no other homologies to propionyl CoA carboxylase were identified. Identification of additional downstream sequence identified further sequence with homology to the *dtsR* gene but also showed homology to the β subunit of propionyl CoA carboxylase of many bacterial species (Figure 4.30). We have therefore termed this gene *pccB*. The insert has an overall AT content of 50% however the 140 bp upstream of the *pccB* start codon is 60% AT rich and contains a number of A and T tracts of 4 or more bases.

Two clones, C13 and C18, were completely non-fluorescent in DMEM compared to in macrophages where they exhibited significant fluorescence. Although C18 was non-fluorescent in DMEM or BHI cultures prepared for flow cytometry, it displayed high fluorescence intensity when streaked out onto BHI agar (Table 4.3). This suggested that expression of *pccB* was either regulated in a contact dependent manner or alternatively its regulation could be related to the Tween-80 that is present in broth grown cultures used for flow cytometry but not in agar plates. To determine whether *pccB* expression was Tween-80

```
PNATALIGTNE 48
PDEPALDDGRR 60
*: .** . ..

GAAYVPVDWD 108
GAAYVPVDAE 120
******** :
```

```
 1  ctgctcgcgc aaccactcga cgttgagagg ggttcgcgag tcacgactac cgccgtcgct

61  ttctacgaga gcgctttgtg ctgcgtcctt gtccatcgtt tctcctgacc ctatgctcat

121 tcgcatcatt ttagccgctt gctttattac tcgtcataac tcgcctgtcc cctatttccg

181 gcacagactg ggtgcctgat ttccagtttg tgggatggga cgtccttata aaaagcaccg

241 tagaaagaca tgcggaaaaa atttttttgt acacaaaaca acccggctat ggtgggaaaa

301 cacacttttc ccggaaaaaa tttaacattt tgtttctggaa actgttgcgc tagcagtcta

                                    143→
361 tgctgccaaa catgacactt tcctcacctt ttgttgatgt tgccgcactt aaagacatga
                  M  T  L   S  S  F   F  V  D   V  A  A  L   K  D  M

421 caaccaccgc cggcaagatc gcggacttca aagcacgacg cgctgaagca ggccagccca
     T  T  T   A  G  K  I   A  D  F   K  A  R   R  A  E  A   G  Q  P

481 tgggcacaaa agctctggat aaagttaaac aagccggcag gctcacggct cgcgaacggc
     M  G  T   K  A  L  D   K  V  K   Q  A  G   R  L  T  A   R  E  R

541 tcgattacct tctagatgaa ggctctttca ttgaaacgga ccagctcgca cgccaccgca
     L  D  Y   L  L  D  E   G  S  F   I  E  T   D  Q  L  A   R  H  R

601 ccttcgactt tggtatgcac aaaaaacgtc ccgtcaccga cggaattgta accggctggg
     T  F  D   F  G  M  H   K  K  R   P  V  T   D  G  I  V   T  G  W

661 gcactat--a cggccgcgag gtatgcatat tctcccaaga cggaaccgtt ttcggcggtg
     G  T      D  G  R  E   V  C  I   F  S  Q   D  G  T  V   F  G  G

721 ccctcggcga ggtgtacggc gaaaaaatga tcaagattat gaagctcgcc gtggccaccg
     A  L  G   E  V  Y  G   E  K  M   I  K  I   M  K  L  A   V  A  T

781 gacggccccc tatcggactc tatgaaggtg caggtgcccc
     G  R  P   L  I  G  L   Y  E  G   A  G  A
```

Figure 4.29: DNA sequence of the plasmid insert of C18 and further downstream sequence.

All sequence upstream of the arrow is from the original C18 clone. Sequence downstream of the arrow corresponds to sequence identified from the PCR generated clone. The sequence of primer #143 which was used to identify the downstream sequence is underlined. Protein sequence with homology to the *dtsR* gene of *C. glutamicum* is shown using the amino acid single letter code.

```
C.ptb MTLSSPFVDVAALKDMTTTAGKIADFKARRAEAGQPMGTKALDKVKQAGRLTARERLDYL 60
C.glu MTISSPLIDVANLPDINTTAGKIADLKARRAEAHFPMGEKAVEKVHAAGRLTARERLDYL 60
      **:***::*** * *:.*******:******* *** **::**: ************

C.ptb LDEGSFIETDQLARHRTFDFGMHKKRPVTDGIVTGWGTIDGREVCIFSQDGTVFGGALGE 120
C.glu LDEGSFIETDQLARHRTTAFGLGAKRPATDGIVTGWGTIDGREVCIFSQDGTVFGGALGE 120
      ***************  **:  ***.*******************************

C.ptb VYGEKMIKIMKLAVATGRPLIGLYEGAGA 149
C.glu VYGEKMIKIMELAIDTGRPLIGLYEGAGA 149
      **********:**: **************
```

Figure 4.30. Clustal alignment of the predicted amino acid sequences of the N-terminal end of PCC beta chain from *C. pseudotuberculosis* (C.ptb) and *C. glutamicum* (C.glu).

Residues that are identical in both species are indicated by an asterisk. Conserved amino acid substitutions are indicated by two dots while semi-conserved substitutions are indicated by a single dot. Alignment was performed using CLUSTAL W (1.82).

regulated C18 was grown overnight in BHI in the presence of between 1 and 0.01% Tween-80. Analysis of culture fluorescence intensities indicated that *pccB* expression was regulated by Tween-80 in a dose dependent manner such that as the concentration of Tween-80 increased the level of Gfp fluorescence detected decreased (Figure 4.31).

Tween-80 has previously been used as a major source of carbon and energy in a number of bacterial species including mycobacteria. Mycobacteria hydrolyse Tween-80 to yield oleic acid and sorbitol or its polyoxyethylene esters. We were interested in determining whether the effect of Tween-80 on *pccB* expression was mediated via an oleic acid intermediate. Cultures of C18 were grown in the presence of 0.5 and 0.05mM oleic acid solution and flow cytometry used to measure Gfp fluorescence. No conclusive evidence regarding the role of oleic acid in *pccB* regulation could be obtained as its presence in the culture medium dramatically reduced bacterial cell growth suggesting that it was toxic to the cells.

4.15.7 Analysis of DFI clone gene expression by reverse transcription-PCR

To confirm that the changes in gene expression identified by flow cytometry were real, relative mRNA levels for the genes of interest was determined by reverse transcription followed by real time PCR. Given that a low MOI was used for the DFI experiments it was

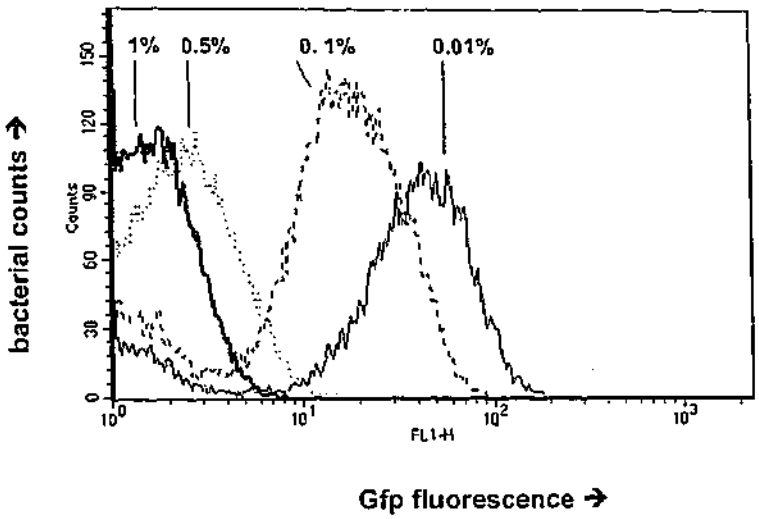

Figure 4.31: Effect of Tween-80 on *gfp* expression in C18.

Cultures of C18 were grown overnight in the presence of different concentrations of Tween-80 (from 0.01% to 1%). Fluorescence of each culture was determined by flow cytometry. For each clone 30,000 events were counted. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria.

not possible to extract sufficient RNA for analysis from identical conditions. Instead the *C. pseudotuberculosis* RNA extracted from macrophages infected with a high MOI for 6 hr (Section 3.7) was utilised (Figure 4.32B, D and F). The control for these samples was RNA from *C. pseudotuberculosis* grown in DMEM for the same time period. Additionally, to determine whether the *cps*, *dnaN* and *pccB* genes were regulated by changes in *in vitro* conditions, gene expression was also measured in BHI cultured bacteria at low and high density and following a 30 min heat shock in a high density culture (Figure 4.32A, C and E).

Analysis of *cps* showed that its expression was highly up-regulated in infected macrophages compared to in DMEM. Two separate macrophage derived samples showed 26 and 49 fold induction. Additionally expression of *cps* in BHI grown *C. pseudotuberculosis* was up-regulated by heat shock by approximately 7 fold. Expression of *dnaN* was essentially unchanged under all the treatments tested. Expression was always within a two fold induction and therefore any regulation under these conditions is negligible.

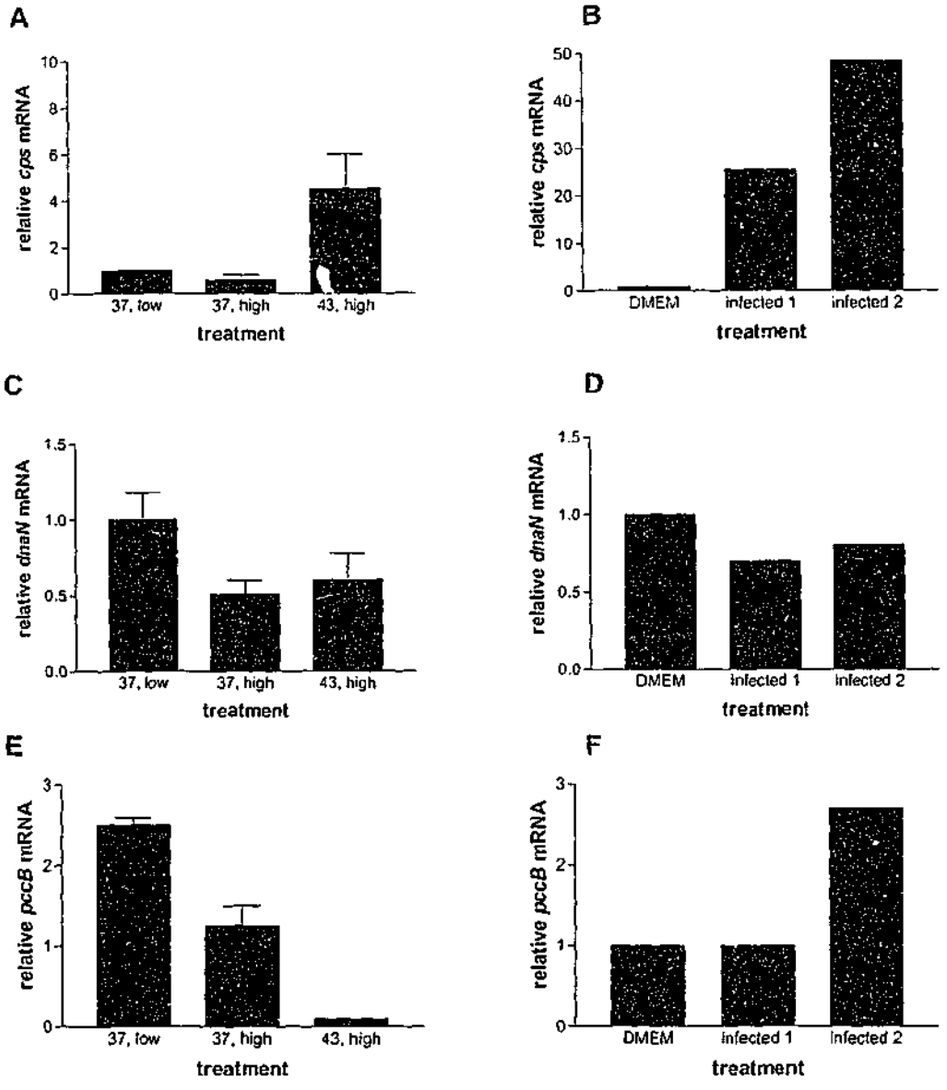

Figure 4.32. Effect of macrophage infection and changes in *in vitro* growth conditions on expression of *C. pseudotuberculosis cps, dnaN* and *pccB.*

(A, C, E) An overnight culture of *C. pseudotuberculosis* was diluted to an OD_{600} of 0.1 then incubated at 37°C. RNA was extracted from 30 min (37, low) and 6.5 hr (37, high) cultures. Additionally a 6.5 hr culture was prepared in which the last 30 min of the incubation was performed at 43°C (43, high). (B, D, F) Duplicate plates of J774 macrophages were infected with *C. pseudotuberculosis*. 6 hr post infection bacterial RNA was extracted from the cell monolayer. As a control DMEM was inoculated with bacteria for the same time period. The levels of *cps, dnaN* and *pccB* mRNA and 16S rRNA were determined by reverse transcription followed by real time PCR. Gene expression was normalised to 16S rRNA levels to take into account differences in starting amounts of RNA. For A, C and E results are represented as the mean of duplicate samples ± minimum and maximum values.

Given that the RNA samples analysed were from bacteria that had not been grown in the presence of Tween 80, *pccB* was detected in all samples. Of the two macrophage derived samples analysed one showed *pccB* up-regulation while the other showed no regulation compared to the DMEM derived control. *pccB* expression was shown to be regulated by changes in *in vitro* conditions. Expression was possibly density regulated such that expression decreased with cell density (approximately two fold) and was thermoregulated such that expression dropped approximately 20 fold following incubation at 43°C for 30 min.

4.16 Discussion

In this chapter the development and application of DFI to the identification of differentially regulated genes of *C. pseudotuberculosis* is described. Although heat regulated genes could not be identified, a number of genes that were regulated in the macrophage model of infection were identified thus demonstrating that the technology could ultimately be used in a successful manner.

This study has shown that *gfp* is a suitable reporter gene for *C. pseudotuberculosis*. *C. pseudotuberculosis*, because of its complex cell wall structure, must undergo severe mechanical disruption to release the cell contents. Therefore the use of traditional reporter genes such as *cat* or *lacZ* is somewhat problematic, both in terms of processing large sample numbers by methods such as sonication and in gaining reproducible data. Gfp, on the other hand, can be quantitated directly in living cells either by flow cytometry or fluorimetry without the need for sample preparation or other co-factors.

The initial vector, pSM12, was a relatively poor promoter probe vector. This was particularly apparent when comparing the fluorescence of pSM13 and pSM15 both of which contain the *srp* promoter upstream of the *gfp* gene in a pEP2 vector background. The sequence between the promoter and the *gfp* gene are different in the two constructs suggesting that this region is important for maximal expression from the *srp* promoter. Immediately upstream of the *gfp* gene in pSM13 is sequence almost identical to that found immediately upstream of the *C. pseudotuberculosis repA* gene. A minor modification was made to this region for cloning purposes by switching the order of the A and C residues which are located 12 and 13 bp upstream of the initiation codon. It is not possible to definitively conclude whether the poor

expression from the *srp* promoter in pSM13 was from reduced transcription, reduced translation or as a result of altered mRNA stability. However, the fact that the area 3' to the –10 region was important for increasing expression suggests that the effect is a post transcriptional one being either at the level of translation or RNA stability. Comparison of the sequence immediately upstream of the *gfp* gene in pSM13 and pKEN1gfpuv1 with the *E. coli* anti Shine-Dalgarno sequence (of the 16S rRNA) shows that there are more potential base-pairing interactions with the pKEN1gfpuv1 sequence than with the pSM13 sequence (Figure 4.33). This could potentially explain the differences in fluorescence intensity observed for the two constructs in *E. coli*. It was not possible to make a direct comparison with the same region of the 16S rRNA of *C. pseudotuberculosis*. Although the partial sequence of the 16S rRNA from a number of strains is available, none of the sequences are complete. This is because, in general, the 16S RNA gene has been PCR amplified using primers homologous to the end of the *E.* coli 16S rRNA (Riegel *et al.*, 1995), therefore the 5' and 3' sequences are often missing. Analysis of the anti Shine-Dalgarno sequences of the 16S rRNA of *M. tuberculosis* (Accession AB026698), *C. glutamicum* (Accession AF314192) and *C. diphtheriae* (ftp://ftp.sanger.ac.uk/pub/cdip) indicate that they are identical. Given the evolutionary relationship of *C. pseudotuberculosis* to these bacteria it is likely that the 3' end of the 16S rRNA of *C. pseudotuberculosis* is in turn identical to these sequences. Alignment of pSM13 and pKEN1gfpuv1 with this sequence (Figure 4.33) showed 7 possible base pair interactions for pKEN1gfpuv1 compared to 4 for pSM13, further supporting a role for translational efficiency in the level of Gfp detected in strains carrying the different plasmids.

The original library generated in *C. pseudotuberculosis* after construction in *E. coli* (pSM12-Q) contained only 0.22% Gfp positive clones. Compared to libraries generated for mycobacterial species that have been reported in the literature and the second library generated in this study this is a very low percentage of fluorescent clones. Libraries generated for DFI studies in *M. marinum* and *M bovis BCG* have yielded 6 and 9% Gfp positive clones respectively (Barker *et al.*, 1998;Triccas *et al.*, 1999). Interestingly due to low transformation efficiencies of these bacterial species, both libraries were generated in *E. coli* before transfer to the desired host. The low number of Gfp positive clones can partly be attributed to the vector sequence between the promoter and the *gfp* gene. Additionally the passaging of the library through *E. coli* was detrimental to the quality of the library generated. Empirical evidence from results described in this and other chapters of this thesis suggest that high levels of *gfp*

A

```
E. coli 16s rRNA 3'-AUUCCUCCA-5'
pKEN1gfpuv1      5'-GAAGGAGAUACAUAUG-3'
pSM13            5'-GUGGGAGAAUACCGGUAUG-3'
```

B

```
M. tuberculosis 16s rRNA 3'-UCUUUCCUCC
pKEN1gfpuv1              5'-AAGAAGGAGAUACAUAUG-3'
pSM13                    5'-UAGUGGGAGAAAUACCGGUAUG-3'
```

Figure 4.33: The putative Shine-Dalgarno interaction of pSM13 and pKEN1gfpuv1.

The 3' end of the *E. coli* (A) and *M. tuberculosis* 16s rRNA corresponding to the anti Shine-Dalgarno sequence is aligned with the region upstream of the *gfp* gene in pKEN1gfpuv1 and pSM13. For the two plasmid sequences the last three base pairs code for the initiating methionine. Bases that are complementary to the corresponding base in the anti Shine-Dalgarno sequence are shaded.

expression are toxic to the *E. coli* host. This was observed as an initial reduction in *E. coli* growth followed by a return to normal growth but a reduction in plasmid copy number and detectable Gfp fluorescence. It is therefore likely that clones that were strongly fluorescent in *E. coli* would have been selected against by passaging the library through *E. coli*. Although this could be seen as advantageous in removing clones that are highly fluorescent in both species it would also have removed clones that are highly fluorescent in *E. coli* but not as strongly fluorescent in *C. pseudotuberculosis*. C17 is probably an example of such a clone. Gfp is generally thought of as a non-toxic protein and as such there are few reports of it being toxic to prokaryotic cells (Lissemore *et al.*, 2000).

Analysis of the Gfp positive clones identified in both *E. coli* and *C. pseudotuberculosis* demonstrates that there are significant differences in what is recognised as a promoter in the different species. In particular a large number of promoter sequences were identified in the *E. coli* pSM12-Q library that did not demonstrate promoter activity in *C. pseudotuberculosis*.

The inadequacies with the initial promoter probe vector and library were overcome by making modifications to the vector and the manner in which the library was constructed. The substitution of the region containing the RBS of pSM12 with that from pKEN1gfpuv1 resulted in a significant increase in fluorescence when known promoters were inserted into the upstream multiple cloning site. Insertion of the *srp* promoter increased fluorescence by 10 fold while the *pld* promoter increased fluorescence 3 fold. It is unclear why the two promoters caused different increases in fluorescence intensity. The *pld* promoter is however highly regulated (discussed in Chapter 6) hence the differences may be related to this. The second modification to the promoter probe vector allowed the generation of fusion proteins between proteins coded for by inserted DNA sequence and Gfp. Of the clones with promoter activity identified during this study no translational fusions were observed. Both amino- and carboxy-terminal fusion proteins between Gfp and a second protein of interest are commonly used (Phillips, 2001) so it is unlikely that the lack of fusion proteins detected is a reflection of either a toxicity effect or an alteration in protein conformation leading to a loss of fluorescence. Given that only 4 clones were identified in which coding sequence was present at the 3' end of the insert, lack of fusions detected may simply reflect the relatively small number of this type of clone identified. Statistically one would expect that 1 in 3 such clones would contain a gene in frame with the *gfp* gene.

The second library generated was smaller than the first but had the advantage of being directly constructed in *C. pseudotuberculosis*. This prevented biases against plasmids that strongly express Gfp in *E. coli*. The final library contained approximately 15,300 clones with an average insert size of 700bp. The size of the *C. pseudotuberculosis* genome is unknown, however is likely to be similar to other corynebacterial species. The chromosome sizes of *C. glutamicum* and *B. lactofermentum* are each approximately 3.1 Mb (Correia *et al.*, 1994;Bathe *et al.*, 1996) while the *C. diphtheriae* genome is approximately 2.5 Mb (ftp://ftp.sanger.ac.uk/pub/cdip). Assuming that the genome size of *C. pseudotuberculosis* is approximately 3.1 Mb, the average insert size is 700 bp, DNA fragments are inserted into the promoter probe vector in either orientation equally and that 7 of every 8 plasmids have an insert, then it is likely that the library represents a 1.5 fold coverage of the genome. It is therefore likely that there are some sequences that are absent from the library. Given that this study did not aim to screen every promoter of *C. pseudotuberculosis* but rather aimed to prove the effectiveness of the methods and identify some regulated genes for further study, this was

not a major issue. It would be necessary to increase coverage by several fold to be confident that a representative library was being screened.

A number of clones that expressed *gfp* constitutively in the *Alu*I libraries were analysed individually. All of these clones expressed *gfp* at a reasonably high level. This primarily reflects the manner in which the clones were selected, as only relatively fluorescent colonies can be detected visually under UV light and sorting using a FACS was set up such that only the highly fluorescent clones were enriched for. Analysis of genes identified as being constitutively expressed indicated that a subset contained a gene at the 3' end of the insert reading into the *gfp* gene, upstream of which was the candidate promoter. A second set of clones contained the 3' end of a gene within the insert followed by a region with no homology to proteins in the NCBI protein databases. Presumably sequence with promoter activity is located within this region and as such for several clones sequences with homologies to *E. coli* σ^{70} consensus promoter sequences were identified close to the 3' end of the inserts. With this type of insert it would be necessary to identify downstream sequence and the genes that it contains to confirm that the detected promoter activity was indicative of a real promoter.

Changes in gene expression during the heat shock response in *C. pseudotuberculosis* and during macrophage infection could be monitored using the promoter probe vector with the *dnaK* promoter inserted upstream of the *gfp* gene. Expression from the *dnaK* promoter in a non-stressed state (37°C) was low and, as expected, the level of detectable fluorescence increased during heat shock. To ascertain how accurately the induction of fluorescence detected by flow cytometry reflected *dnaK* promoter activity, the pattern of fluorescence induction was compared to that observed in Chapter 3 when measuring *dnaK* mRNA levels in response to heat shock. The major differences observed between the two sets of data were in the rate of induction and the length of time during which the reporter protein or mRNA stayed elevated. *dnaK* mRNA levels peaked by 15 min, then returned to just above basal levels by 90 min, whereas the rate of fluorescence induction was slower and continued over the timecourse of the experiment. The slowness of the fluorescence response can be attributed to the time required for the synthesis and folding of the Gfp protein (Cormack *et al.*, 1996). This time means that there is a lag between promoter activation and the end point of fluorescence. The differences between the patterns of the protein and mRNA timecourses can be explained by the very different half lives of the Gfp protein and the *dnaK* mRNA. In a number of systems

Gfp has a half life of greater than 24 hours (Andersen *et al.*, 1998;Corish and Tyler-Smith, 1999;Corish and Tyler-Smith, 1999) and would therefore persist within the bacteria. The typical half-life of bacterial mRNA species is between 2 and 5 min. *dnaK* mRNA is likely to have a similarly short half life and would therefore be rapidly degraded leading to the reduction of *dnaK* mRNA observed at 1 hour. These data suggest that although changes in gene expression can be observed, the precise timecourse of regulation should be confirmed by more direct methodologies such as Northern analysis. In some respects the long half-life of Gfp protein may be an important advantage because it may allow transient up-regulation events to be more easily identified than if direct mRNA measurements are relied upon.

The application of DFI to identify heat regulated genes of *C. pseudotuberculosis* was unsuccessful. Thirty individual clones enriched for during the screen process were analysed however none were regulated. The inability to identify heat regulated genes probably stems from the difficulties in sorting the bacteria at a sufficiently high efficiency using a single round of sorting. It is likely that if further rounds of enrichment at 43°C had been performed, followed by rounds of selection for non-fluorescent bacteria at 37°C that thermoregulated genes would have been identified. The usefulness of multiple rounds of enrichment could be observed during the macrophage studies where this aided in sufficiently enriching the population such that regulated genes could be identified. Expression of *dnaK* increased less than three fold following a 90 min heat shock (as measured by fluorescence) which is an order of magnitude less than the change in *dna*K mRNA levels as measured by RT-PCR. Given that *dnaK* is highly induced it is likely that the fluorescence induction of other clones would be three fold or less. This relatively small difference probably contributed to the relatively poor sorting efficiency achieved when looking for heat regulated genes. When sorting macrophages this problem was partially alleviated by sorting macrophages instead of bacteria. Since each macrophage contained multiple bacteria the fluorescence of an individual event increased significantly.

That the application of DFI to the identification of macrophage induced genes was successful was demonstrated by the observation that two of the clones identified (C13 and C18) were highly fluorescent in macrophages but completely non-fluorescent in DMEM. In addition a third clone C1 showed approximately 6 fold induction in macrophages at 6 hr post infection.

Even assuming that all the other clones identified were false positives this is 15% success rate. The DFI studies in the literature have reported varying successfulness in the ability to efficiently enrich for the regulated population. Of studies performed in macrophages efficacy has ranged from approximately 2% in *M. tuberculosis*, up to 100% in *M. marinum*, 33 to 50 % in *S. typhimurium* and 25% in *L. monocytogenes* (Triccas *et al.*, 1999;Ramakrishnan *et al.*, 2000;Guckenberger *et al.*, 2002;Wilson *et al.*, 2001). This reported variability probably reflects the ease in which some bacterial species can be worked with, the number and type of enrichment steps performed and the type of FACS which was utilised. A major problem that was encountered during this study was that all clones that were studied individually were less fluorescent in DMEM compared to bacteria inside the macrophages. This was even apparent for the control *srp* promoter construct pSM23 which has generally been observed to be constitutively expressed under a variety of *in vitro* conditions during this study. Whether this is a form of regulation or is as a result of a generalised effect of the growth conditions is unclear. Growth of the bacteria in DMEM was under quite different conditions to the bacteria grown in BHI with shaking. In particular the cultures were not aerated and were incubated in an atmosphere containing 5% CO_2. Given that oxygen is an essential co-factor for Gfp fluorescence it could be possible that reduced oxygen levels resulted in a generalised reduction in the detectable fluorescence. The phenomenon of altered Gfp fluorescence in DMEM is investigated and discussed further in Chapter 6. This effect does however clearly have implications for DFI studies such as this, if truly regulated genes are to be enriched for rather than all fluorescent clones which all in turn happen to be less fluorescent under the *in vitro* conditions. As an alternative control, fold induction was determined compared to the fluorescence of clones grown in BHI at 37°C with aeration. In general this produced a consistent level of fluorescence for a given clone. When the data were analysed in this manner only C18 still showed a large fold induction following macrophage infection. The choice of the most appropriate controls for studies such as this is discussed further in Chapter 7.

The conclusion from this study is that DFI is suitable for identifying regulated clones when a clone is completely non-fluorescent under control conditions and then induced under experimental conditions. Conversely when clones show some fluorescence under control conditions it was difficult to identify truly regulated genes. It may be possible to solve this problem by using different gates during sorting such that bacteria in different fluorescence

ranges are collected into different groups. Using such an approach it may be possible to identify clones that are fluorescent under control conditions but more fluorescent under experimental conditions. One of the major difficulties in this study was performing negative sorts in which the non-fluorescent population was collected. This, in part, stems from the small size of the bacteria which makes them difficult to distinguish from background noise. In order to alleviate this problem approaches such as bacterial labelling or alternative selection procedures have been utilised. For example, in a study of *M. marinum*, bacteria were stained with antisera which could be detected by a different channel of the FACS. It was therefore possible to differentiate bacteria from noise (Ramakrishnan *et al.*, 2000). For DFI of a *M. tuberculosis* library a reporter vector was constructed which in addition to *gfp* contained a second gene called *sacB*. SacB confers sensitivity to sucrose (Triccas *et al.*, 1999). Thus instead of performing a negative sort by FACS, bacteria were plated onto sucrose containing media. Those clones that possessed constitutive promoter activity were not viable and hence were removed from the population.

Although the identification of *dnaN* as a macrophage induced gene of *C. pseudotuberculosis* appears to have been a false positive in that its expression was not macrophage induced, its identification in this study has pointed to the probable origin of replication (*oriC*) of *C. pseudotuberculosis*. In many eubacteria including the closely related mycobacteria, the *oriC* is found in the intergenic region between the *dnaA* and *dnaN* genes. The size of this intergenic region is variable, however it typically contains two key features; these are the presence of an A+T cluster and a number of DnaA boxes. DnaA boxes are 9 bp DnaA recognition sequences. Binding of DnaA protein to DnaA box sequences triggers a cascade of events that ultimately lead to initiation of DNA replication and synthesis of a new chromosome. A consensus sequence has been defined for the DnaA boxes of Gram positive bacteria (Rajagopalan *et al.*, 1995). Putative DnaA boxes were identified in the sequence upstream of the *C. pseudotuberculosis dnaN* gene. Eoght were found in the forward orientation and 3 were found in the reverse orientation. These matched the *M. smegmatis* consensus sequence at between 6 and 8 of 9 nucleotides. This degree of homology to the consensus sequence is seen in a variety of mycobacterial species (Qin *et al.*, 1997;Qin *et al.*, 1999;Madiraju *et al.*, 1999). Various numbers of DnaA boxes are found in the *dnaA-dnaN* intergenic region of different bacterial species and in the mycobacterium genus varies considerably ranging from 16 in *M. avium* (Madiraju *et al.*, 1999), 20 in *M tuberculosis* (Qin *et al.*, 1999) to 5 in *M. smegmatis*

(Qin *et al.*, 1997). No obvious A+T cluster was present in the *dnaN* promoter region, however this may simply be because it lies upstream of the C3 insert. In *M. tuberculosis* and *M. smegmatis* the A+T cluster is found at the 5' end of the *oriC*. To confirm that the plasmid insert of C3 does indeed contain the *oriC* for *C. pseudotuberculosis* it would be necessary to clone further upstream sequence and to then determine whether the region is sufficient to confer autonomous replication to a plasmid that was previously non-replicative in *C. pseudotuberculosis*.

Given that DNA replication must occur at the correct time in the cell cycle it would be expected that the expression of the required genes would be tightly regulated. *dnaN* expression has been shown to be regulated in both *E. coli* and *Synechococcus*, although in both of these bacterial species the *oriC* is not located immediately upstream of the *dnaN* gene. In the cyanobacterium *Synechococcus* sp strain PCC 7942 expression of *dnaN* is rhythmically expressed suggesting that it may be under circadian control (Liu and Tsinoremas, 1996). Given that in cyanobacteria cell division is light dependent this observation is perhaps unsurprising. In *E. coli dnaN* maps between *dnaA* and *recF*. In exponentially growing cells, *dnaN* and *recF* are expressed predominantly from the *dnaA* promoters. Upon entry of the cells into the stationary phase, *dnaN* promoters are activated leading to changes in expression of the *dnaA* operon genes, in particular *dnaN* and *recF* increases. Consequently the synthesis of the β subunit of DNA polymerase III and RecF protein increases as cell metabolism is slowing down.

Non ribosomal peptide synthetases are used by microorganisms to synthesise small polypeptides of less than about 50 amino acids by a non ribosomal pathway. A diverse range of cellular products are produced or modified by non ribosomal peptide synthases including some antibiotics (Stachelhaus and Marahiel, 1995), glycopeptidolipids (Billman-Jacobe *et al.*, 1999) and siderophores (Yu *et al.*, 1998). Non-ribosomal peptide synthetases are large multi-enzyme complexes. Each protein consists of a series of modules which function as "individual enzymes" and are responsible for the addition of the next amino acid to the nascent peptide chain. A module is defined as the unit required to complete one elongation cycle. The simplest type of module consists of three domains termed the A, T and C domains (Mootz and Marahiel, 1999). The A domain is involved in recognition, binding and activation by adenylation of the appropriate amino acid. Following activation the amino acid is

transferred to a thioester linkage on the cofactor 4'-phosphopantetheine in a reaction performed by the T domain. Finally the C domain catalyses the condensation of the activated amino acid with the one of the next module leading to chain elongation. A number of other types of domain also exist that may make further modification to the amino acids before chain elongation, for example epimerisation or methylation. The linear organization of the modules within the protein directly correlates with the sequence of the peptide while the number of modules determines the size of the amino acid.

Within each domain are conserved motifs that are usually associated with the specific enzyme reactions catalysed by the domains. Six motifs have been mapped to the A and T domain (Stachelhaus and Marahiel, 1995). Five of these are involved in known interactions such as ATP binding and 4-phosphopantetheine binding while the function of the sixth is unknown. Within the gene termed *cps* a region with homology to the first motif of the A domain was observed (Figure 4.27). 7 of the 11 amino acids within the consensus were conserved. Interestingly when comparing this region to the same motif in a peptide synthase of *M. smegmatis* (Billman-Jacobe *et al.*, 1999) a greater degree of homology was observed suggesting that a modified motif may exist in gram positive bacteria. As sequence corresponding to only the first 133 amino acids was identified no other motifs could be observed. It would be expected that these could be observed if further downstream sequence was identified. Additionally the sequence of the peptide or its function cannot be determined at this stage. As the general understanding of peptide synthase structure and chemistry increases it is likely that it will become possible to readily predict the sequence of the synthesised peptide from the sequence of the peptide synthetase. This type of approach has already been used to predict the structure of a siderophore of *Streptomyces coelicolor* (Challis and Ravel, 2000). The next steps in determining the role of *cps* during macrophage infection would be to identify the entire gene. This would give information regarding the length of the peptide synthesised and modifications that it may contain. Secondly mutant strains could be generated and used in studies to elucidate the function of Cps.

There are relatively few reports of regulated expression of peptide synthetases. The production of a number of non ribosomally synthesised siderophores has been shown to be iron regulated (Yu *et al.*, 1998;Wertheimer *et al.*, 1999;Devescovi *et al.*, 2001) suggesting that expression of the relevant peptide synthetases may be occurring. A role for non ribosomal

peptides in virulence is possible. Obviously the role of siderophores in iron scavenging is likely to be critical for some bacteria and these genes could then be considered to be virulence determinants. A role in virulence of a non ribosomally derived peptide has been demonstrated in strains of *Streptomyces* that cause scab disease in root and tuber crops. In the absence of the peptide synthetase gene *txtA* which synthesises thaxtomin A, wild-type growth was observed *in vitro*, however the mutant strains were avirulent being unable to infect potato tubers (Devescovi *et al.*, 2001).

The PccB protein whose expression was regulated by the promoter in C18 was determined to be most homologous (97%) to a protein of *Brevibactum lactofermentum* that has been termed DtsR (for detergent sensitivity rescuer). In *B. lactofermentum* the large scale secretion of L-glutamate is stimulated in the presence of some detergents, for example polyoxyethylenesorbitan monopalmitate (PESP). A strain of the bacteria was generated that was more sensitive to detergent than the wild type strain. Supplementation of the mutant with the *dtsr* gene restored the wild type phenotype (Kimura *et al.*, 1996). Additionally mutants with a disrupted *dtsR* gene required either oleate or oleate ester for growth (Kimura *et al.*, 1997). This requirement was quite specific as neither palmitate ester or stearate ester were able to support growth of the mutant. Analysis of other proteins with which C18 and the *dtsr* gene contain homology indicate that they both are likely to code for the β chain of propionyl CoA carboxylase (PCC). PCC is an enzyme that is composed of α and β subunits and catalyses the conversion of propionyl CoA to (S)- methylmalonyl-CoA ((S)-MCC). In subsequent enzymatic reactions (S)-MCC is first converted to (R)-methylmalonyl-CoA and then to succinyl-CoA. Succinyl-CoA can then enter the citric acid cycle. Alternatively MCC may act as a precursor for the synthesis of branched-chained fatty acids and polyketide.

The appearance of *C. pseudotuberculosis* colonies on agar plates was significantly altered by the presence of Tween-80 in the agar. Instead of being dull, rough and easily breakable, the colonies took on an appearance more like that observed of an *E. coli* colony, being smoother and shiny. This change in phenotype is probably indicative of changes in the composition of the cell wall surface. Similar changes in morphology have been reported for *M. paratuberculosis* following treatment with Tween-80 (Van-Boxtel *et al.*, 1990).

The effect of Tween-80 on the mycolic acid composition of three cutaneous corynebacterial species has previously been investigated (Chevalier *et al.*, 1988). Species investigated were *C. striatium* and two belonging to the pathogenic JK group corynebacteria. 0.3% Tween-80 was sufficient to increase the synthesis of mycolic acids in particular C_{36} mycolic acids and to increase the degree of unsaturation of the side chains. It has been postulated that the increase in the amount of $C_{36:2}$ corynemycolic acid that occurs following treatment with Tween-80 is as a result of a synthesis that involves the condensation of two molecules of oleic acid ($C_{18:1}$) (Chevalier *et al.*, 1988), this however has not been demonstrated.

It seems likely that PCC plays a direct role in fatty acid metabolism in *C. pseudotuberculosis* as evidenced by decreased expression of *pccB* in the presence of Tween-80, and the fact that PCC mutants in closely related species require oleate for growth. Within the cell it is likely that the Tween 80 is metabolised to yield oleic acid. Oleic acid has been postulated either to be able to act directly as a substrate for fatty acid biosynthesis or is alternatively broken down to yield acetyl CoA which can in turn be used for fatty acid biosynthesis or enter the citric acid cycle. It thus seems likely that *pccB* down-regulation in the presence of Tween-80 is in response to metabolite excess, in particular the excess of fatty acid. That is, that in the presence of Tween-80 there is a lesser requirement for de novo synthesis of fatty acids via a PCC pathway.

In addition to being regulated by Tween-80 the expression of *C. pseudotuberculosis pccB* was found to be regulated in response to changes in cell density and temperature. *pccB* is regulated in a similar manner to the fatty acid synthase (FAS) identified during the array studies described in chapter 3. This similarity of regulation gives further support to PCC playing a role in fatty acid metabolism. Many microbial organisms respond to shifts in environmental temperatures by altering the nature of the fatty acids that they produce which may explain the changes in *pccB* expression observed following heat shock. Additionally the effects on corynebacteriual membrane composition following treatment with Tween-80 observed by Chevalier are similar to those observed following changes in growth temperature in *Norcodia asteroides*, *Norcodia ruba* and *M. phlei*. This suggests that the mechanisms involved in altering membrane composition to the different stimuli are similar (Chevalier *et al.*, 1988).

The regulated expression of *pccB* highlights the need for very careful and extensive controls in DFI experiments. Although identified as a macrophage regulated gene, on further examination this was found not to be strictly the case. Rather the apparent regulation was an artefact of other differences in the assay procedure, in particular the supplementation of the control growth conditions with Tween-80.

The promoter probe vector generated for this study has potential applications outside of the area of DFI. Following insertion of known promoters upstream of the *gfp* gene it can be used to monitor changes in gene expression. pSM20 was used in such a way to monitor *pld* regulation (see Chapter 6). The ability to use *gfp* in fusion proteins also opens up the potential of performing protein localisation studies with relative ease in a manner that avoids the need to produce appropriate antibodies. Additionally we have shown that using a *gfp* expression construct, it is possible to monitor the interaction between *C. pseudotuberculosis* and the macrophage. This system will allow the monitoring of expression from known and unknown promoters within the macrophage. Such studies may indicate which genes are important *in vivo*. An extension of this type of approach would be to use *gfp* expressing bacteria to monitor the disease process *in vivo*. The Toxminus strain of *C. pseudotuberculosis* (Hodgson *et al.*, 1992) lacks a functional *pld* gene and has been suggested as a suitable veterinary vaccine vector for delivery of foreign antigens (Moore *et al.*, 2000;Moore *et al.*, 2001). To monitor how this strain interacts with the immune system a Gfp expressing version of this strain could be generated. Following infection, the *in vivo* fate of the bacteria could be monitored and the immune cells with which they interact established. In summary Gfp is a suitable tool not only as the reporter in DFI but also for downstream studies aimed at elucidating gene regulation and function and cellular interactions.

Chapter 5

Assessment of the role of regulated genes in macrophage infection

5.1 Introduction

The work described in this chapter is aimed at determining the functional significance of regulated genes of *C. pseudotuberculosis* during macrophage infection. In order to elucidate whether genes identified as being regulated do play an essential role in virulence mutant strains for a number of regulated genes were constructed. These mutant strains were then tested in the macrophage infection model to determine whether they were attenuated.

Three genes identified in Chapter 3 as being differentially regulated were chosen for further study. These genes were selected upon a variety of criteria. The first is a putative metallopeptidase that was selected on the basis that it showed up-regulation in the macrophage infection model. The second is the iron permease component *fagC*, which was chosen on the basis of a possible role in iron acquisition and the observation that its regulation was similar to that observed for the important virulence determinant *pld*. Given that virulence genes are often co-ordinately regulated this gene was deemed a good target. Finally *sodC*, a gene coding for a Cu,Zn superoxide dismutase was chosen. The degree of up-regulation observed was small, however, such genes have been implicated to play an important protective role against host derived oxygen radicals in other bacterial pathogens such as *M. tuberculosis* (Piddington *et al.*, 2001), *N. meningitidis* (Wilks *et al.*, 1998) and *Salmonella* (Farrant *et al.*, 1997).

To generate mutant strains of *C. pseudotuberculosis* a suicide plasmid containing the sequence of interest interrupted by an antibiotic resistance gene was utilised. This type of approach has been used to generate an *aroQ* mutant of *C. pseudotuberculosis* (Simmons *et al.*, 1997). Plasmids containing a ColEI origin of replication were used as previous studies have shown that plasmids with ColE1 replicons were either unable to replicate, or replicated very slowly in *C. pseudotuberculosis* (Simmons *et al.*, 1997). In general genes encoding kanamycin, erythromycin or hygromycin resistance have been used for gene disruption. Ampicillin has been avoided as it is rapidly degraded and cannot be effectively used to select against *C. pseudotuberculosis*. Upon transformation of the gene mutation cassette plasmids into *C. pseudotuberculosis*, transformants that are resistant to the antibiotic were selected for. Southern analysis was then performed to confirm that antibiotic resistance was a result of allelic exchange generated via a double crossover event.

In this chapter the characterisation of sequence in the array clones containing *met*, *fagC* and *sodC* sequence is described. To generate a knockout cassette general wisdom prescribes that the presence of at least 0.5 to 1.0 kb of sequence upon either side of the antibiotic resistance gene is necessary for allelic exchange to occur. It was therefore necessary to identify upstream sequence for both *sodC* and *met*. The knockout cassettes were constructed and the mutant strains generated. The growth of these strains was initially tested *in vitro* and then tested for attenuation in the macrophage infection model.

5.2 Metallopeptidase

5.2.1 Analysis of clone Q12-E9 insert sequence

Clone Q12-E9 contained a 1 kb insert that showed homology to two genes (Figure 5.1). At the 5' end of the insert was an ORF of 759 bp that showed homology to a putative zinc metallopeptidase of *C. glutamicum* and various streptomyces species. Using reverse transcription real time PCR this gene was shown to be up-regulated approximately 4 to 5-fold during macrophage infection (Chapter 3). That this putative protein was indeed a metallopeptidase was supported by the observation that it contained a sequence with complete homology to the zinc binding region signature found in neutral zinc metallopeptidases. This motif was identified by scanning the protein sequence against the PROSITE database of protein families and domains. Within the motif the two histidine residues and the glutamic acid residue act as the zinc ligands (Miyoshi and Shinoda, 2000). Zinc metallopeptidases may be further classified into distinct families. Further analysis of the metallopeptidase sequence identified a second region that showed homology to a motif found in zinc metallopeptidases that belong to the aminopeptidase family. This motif (LWLNEGX) is centred around a glutamic acid residue that is found 23 amino acids after the first histidine residue of the metallopeptidase motif. The sequence IWLNEGF was found in the appropriate position and differs at one amino acid from the consensus. The glutamic acid residue within this motif is an additional zinc ligand. Commencing approximately 20 bp downstream of the predicted end of the metallopeptidase gene, which we have termed *met*, was a second gene, *rmlB* that codes for a protein that is most homologous to the dTDP-glucose 4,6-dehydratase (RmlB) of *Streptococcus pneumoniae*. The insert of Q12-E9 contained approximately 230 bp, which when translated, shows amino acid sequence similarity to this protein. This enzyme catalyses

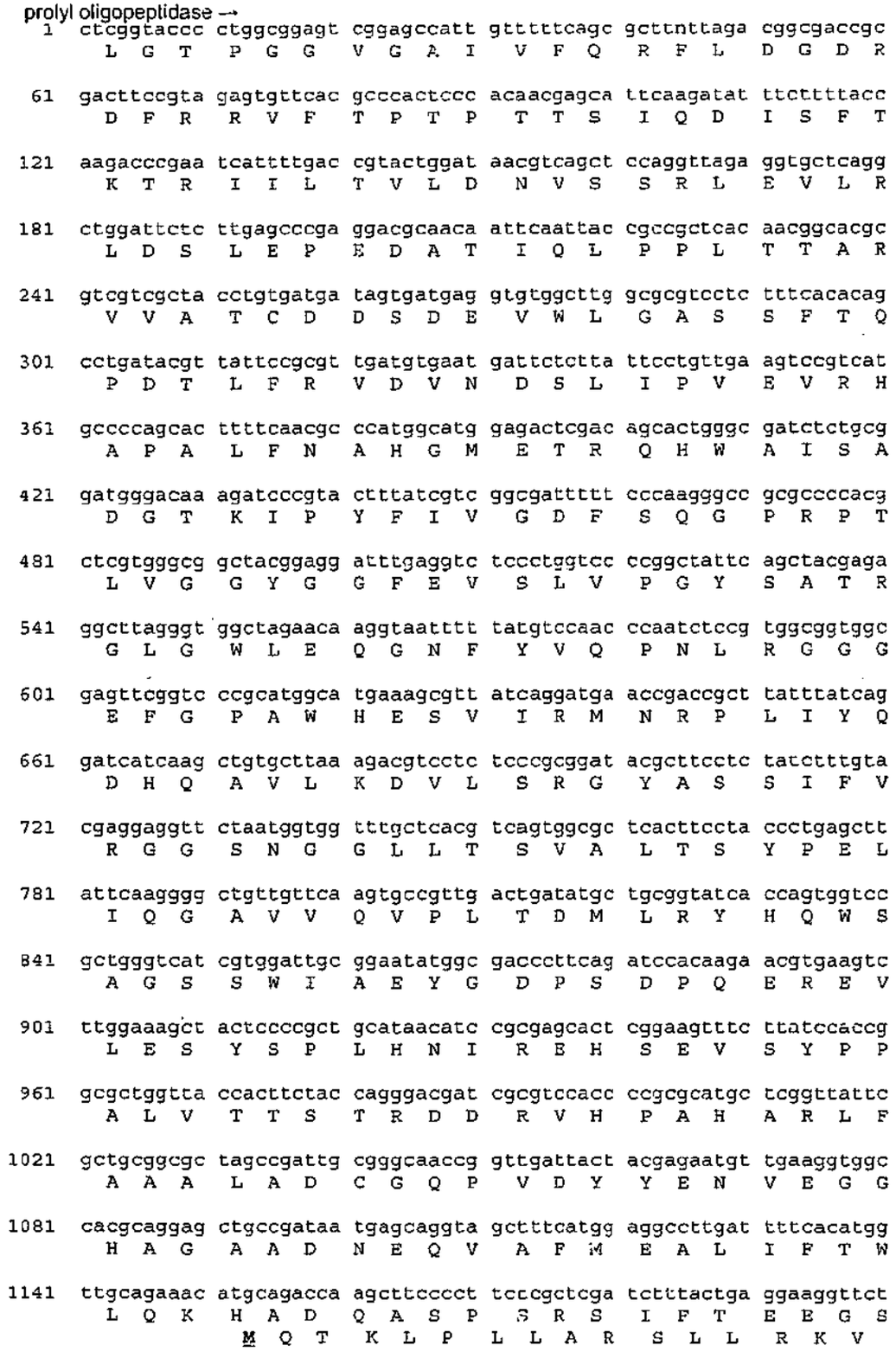

Figure 5.1: Please refer to legend on page 189.

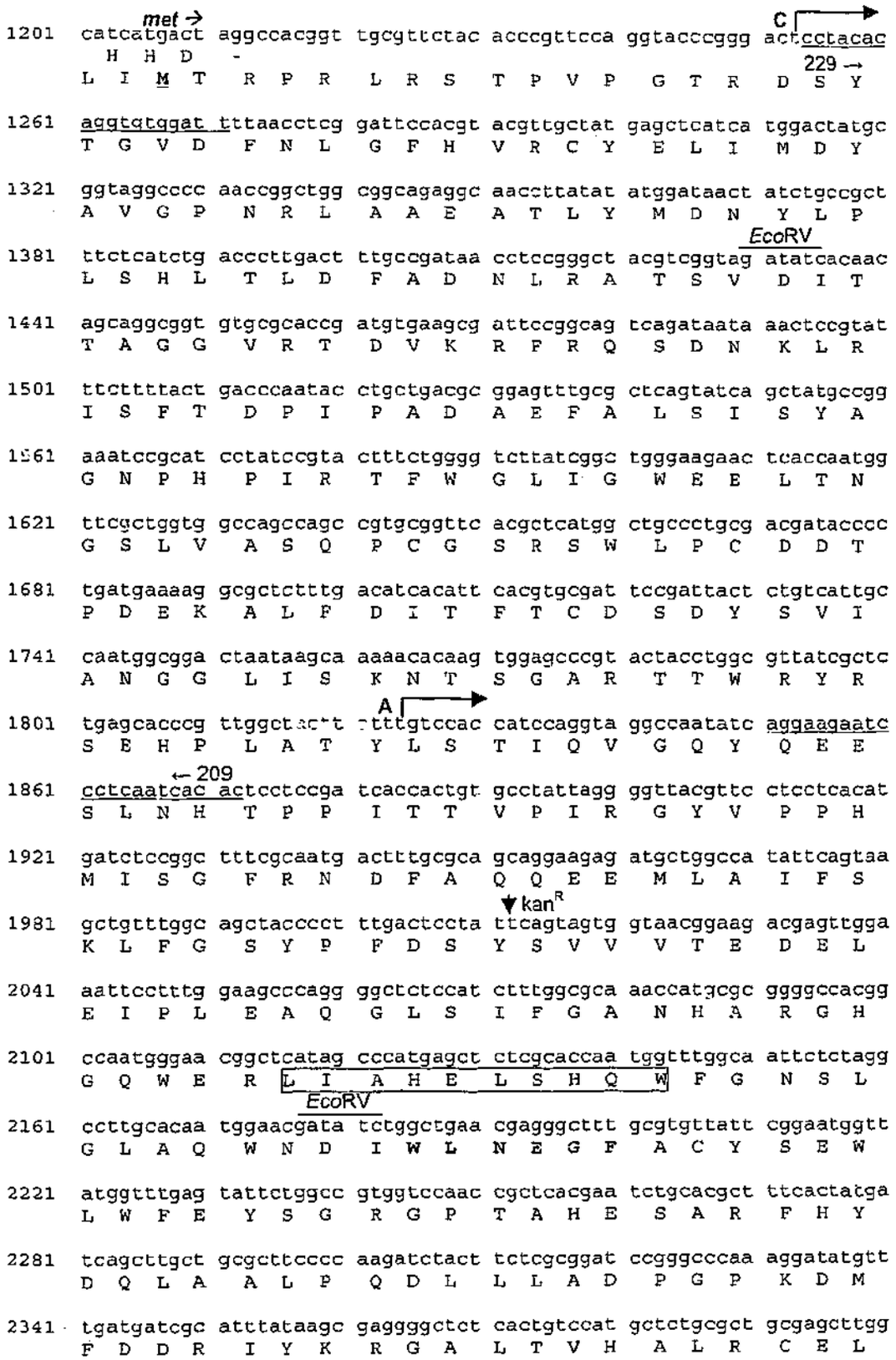

Figure 5.1: Please refer to legend on page 189.

```
2401  cgacgccccc ttcttccgcg ccgtccaacg ctatgtggcg gcaggccggc attcagtagt
        G  D  A  P    F  F  R    A  V  Q    R  Y  V  A    A  G  R    H  S  V

2461  agagccgcat gatctacgcc gagaaatcct caaagaagcc tctaacccac agcgttgtga
        V  E  P  H    D  L  R    R  E  I    L  K  E  A    S  N  P    Q  R  C

2521  agagctctgg gccagctggc tctatcgcac ggaactaccc gagttcccca acccataaaa
        E  E  L  W    A  S  W    L  Y  R    T  E  L  P    E  F  P    N  P  -
                                  → rmlB
2581  tgccccgtaa ggtgtgattc cgtgagaaag aatttgcttg ttaccggcgg cgccggcttt
                                  M  R  K    N  L  L    V  T  G    G  A  G  F

2641  atcggttcta attttgtccg ccgtactcta gaaacccgca gcgattatcg cgtgagggta
        I  G  S    N  F  V    R  R  T  L    E  T  R    S  D  Y    R  V  R  V

2701  ttagacaagc tgacgtatgc aggaaatagc gctaacctcg atggctgcga cgcggagctg
        L  D  K    L  T  Y    A  G  N  S    A  N  L    D  G  C    D  A  E  L
                                                                ← 232
2761  gtcgaaggcg atatttgcga tgctgcgctc gtcgataagc ttttgagcga tgcggacgtg
        V  E  G    D  I  C    D  A  A  L    V  D  K    L  L  S    D  A  D  V
      ←─┐ B
2821  atagtgcatt ttgctgcaga gtctcacaat gataactccc tggaagatcc ttctccgttt
        I  V  H    F  A  A    E  S  H  N    D  N  S    L  E  D    P  S  P  F

2881  attcatacca acgtagtggg cacgttcacg ctgctggaag ctgctcgacg ccaccgtgtg
        I  H  T    N  V  V    G  T  F  T    L  L  E    A  A  R    R  H  R  V

2941  catcttcatc acgtttctac ggatgaggtc ttcggcgatc tcgagcttga agaccctaac
        H  L  H    H  V  S    T  D  E  V    F  G  D    L  E  L    E  D  P  N

3001  cggttcacag agcgcacccc ttacaatcct tcctcccctt attcagcgtc caaggcgggt
        R  F  T    E  R  T    P  Y  N  P    S  S  P    Y  S  A    S  K  A  G

3061  tctgatcatt tggttcgagc atgggtgagg tcttttgatg tctcggctac tttgtctcat
        S  D  H    L  V  R    A  W  V  R    S  F  D    V  S  A    T  L  S  H

3121  tgctcgaaca attatggtcc ttaccagcac attgagaagt ttatcccccg gcagatcacc
        C  S  N    N  Y  G    P  Y  Q  H    I  E  K    F  I  P    R  Q  I  T

3181  aatatcttaa gcgggattcc acccaagctc tatggcacgg gtgaacaagt gcgagactgg
        N  I  L    S  G  I    P  P  K  L    Y  G  T    G  E  Q    V  R  D  W

3241  atccatgtag acgatcacaa cgatgctatc catgcgattc ttgatcgcgg tcgcataggt
        I  H  V    D  D  H    N  D  A  I    H  A  I    L  D  R    G  R  I  G

3301  gagtcctata tcatcggtgc cgacaatgac cacgtgaata ataaacaagt catcgagttg
        E  S  Y    I  I  G    A  D  N  D    H  V  N    N  K  Q    V  I  E  L

3361  atctgccatc tcatgggagc agattcttat gagcacgtcg ccgataggcc aggtcatgac
        I  C  H    L  M  G    A  D  S  Y    E  H  V    A  D  R    P  G  H  D

3421  atgcgttacg ctatggattc caccaaacta cgcacagagc tgggttggaa gcctcgattc
        M  R  Y    A  M  D    S  T  K  L    R  T  E    L  G  W    K  P  R  F

3481  acagacactg ccaccggcat gcgcgagggt ttgagctata ctattcagtg gtatcgcgac
        T  D  T    A  T  G    M  R  E  G    L  S  Y    T  I  Q    W  Y  R  D

3541  aacgagcact ggtggcgtcc tctcaaagac accgtggaaa acaactatgc caagcggggt
        N  E  H    W  W  R    P  L  K  D    T  V  E    N  N  Y    A  K  R  G
```

Figure 5.1: Please refer to legend on page 189.

```
                    → rmlC
3601  cagtgatgaa catagaaacc tgccctatcc agcgcatgat ggtcatccat ctcgacttgc
       Q  -
            M   N  I  E  T   C  P  I   Q  G  M   M  V  I  H   L  D  L

3661  atgaggactc tcgcggctgg ttcaaagaaa attggcaaga gtccaagctt gtcggcttag
       H  E  D   S  R  G  W   F  K  E   N  W  Q   E  S  K  L   V  G  L

3721  gattcaccgg tttccgacca gtgcaaaaca atatttcctt taatgctgag gccggagtaa
       G  F  T   G  F  R  P   V  Q  N   N  I  S   F  N  A  E   A  G  V

3781  cccgcggttt acacgcagaa ccatgggata aatacgtctc cgtagcaacc ggctcagtct
       T  R  G   L  H  A  E   P  W  D   K  Y  V   S  V  A  T   G  S  V

3841  ttggcgcctg gtgcgattta cgcgaagggt ccgcaacgtt tggcgccaca tttaccatca
       F  G  A   W  C  D  L   R  E  G   S  A  T   F  G  A  T   F  T  I

3901  ccattacccc tgagaccgct gtctttgttc cgcggggggt ggccaacgga ttccaggctc
       T  I  T   P  E  T  A   V  F  V   P  R  G   V  A  N  G   F  Q  A

3961  ttgaagctac cgcatacacc tatttagtca acgaccactg ggcaccagac gctcattatt
       L  E  A   T  A  Y  T   Y  L  V   N  D  H   W  A  P  D   A  H  Y

4021  cttttgtcaa tcttgctgat cctgcattaa acatttcttg gccgatcccc ctcgatcaag
       S  F  V   N  L  A  D   P  A  L   N  I  S   W  P  I  P   L  D  Q

                                                          → rmlD
4081  cgacccttcc caagaaagac acggcgcacc ctccgctagc atcagcggta cccgtaccac
       A  T  L   S  K  K  D   T  A  H   P  P  L   A  S  A  V   P  V  P

4141  ccaagaaagt actggttacc ggggcaaacg ggcagcttgg caggactcta aaaaaggtct
       P  K  K   V  L  V  T   G  A  N   G  Q  L   G  R  T  L   K  K  V

4201  tcccccatgc cgagttctgc tctcgcaccg acttagacat cactacagac attacagagg
       F  P  H   A  E  F  C   S  R  T   D  L  D   I  T  T  D   I  T  E

4261  cgcgtcgctg ggctgactac agcgtaatta tcaacgccgc ggcctacaca ggggtaccga
       A  R  R   W  A  D  Y   S  V  I   I  N  A   A  A  Y  T   G  V  P

4321  g
```

Figure 5.1: Sequence of clone Q12-E9 and further up and downstream sequence.

Sequence between arrows A and B is from the original clone Q12-E9. Sequence between arrows A and C corresponds to sequence identified from the PCR generated clone, while sequence upstream of arrow C and downstream of arrow B was generated by mutation recovery. Protein sequence with homology to a metalloprotease of *S. coelicolor* (coded by *met*), dTDP-glucose-4, 6-dehyrdratase (*rmlB*), dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase (*rmlC*) and dTDP-4-keto-L-rhamnose reductase (*rmlD*) of *C. glutamicum* and the 3' end of a prolyl oligopeptidase family protein is shown using the single letter amino acid code. A region which is homologous to the neutral zinc metallopeptidase, zinc binding region signature is boxed and a region that shows homology to the aminopeptidase family motif (LWLNEGX) is bolded. The glutamic acid residue that is located 23 amino acids after the first histidine residue of the metallopeptidase motif is conserved in this family and acts as the third zinc ligand. The two putative initiating methionine codons of *met* are bolded and underlined. Sequence complementary to primer #209 which was used to identify upstream sequence between arrows A and C is underlined. The insertion site of the kanamycin resistance gene in pSM51 is indicated. The location of primers #229 and 232 which were used to generate the PCR probe for Southern analysis (Figure 5.3) is indicted.

the dehydration of dTDP-D-glucose to an unstable intermediate dTDP-4-keto-6-deoxy-D-glucose (Tonetti *et al.*, 1998). This intermediate may act as a precursor for rhamnose biosynthesis.

5.2.2 Identification of upstream metallopeptidase sequence

Before a knockout cassette could be generated it was necessary to identify sequence upstream of that in Q12-E9. A PCR based approach analogous to that used to identify downstream sequence of DFI clones in Chapter 4 was utilised. Genomic DNA was digested with *Bam*HI and *Sma*I and shotgun cloned into pUC18. A reverse primer (# 209, Figure 5.1) was designed. A PCR reaction using primer # 209 and either M13 Forward or M13 Reverse, which hybridise to sequence of pUC18 on either side of the *Sma*I or *Bam*HI sites used for cloning, was performed. Analysis of the PCR reactions by agarose gel electrophoresis revealed the presence of a single band of approximately 0.7 kb. This product was purified from the agarose and then cloned into pGEM®-T easy. The insert was then sequenced.

Using this approach a further 570 base pairs of sequence was identified. Analysis of the new sequence indicated that it contained upstream *met* sequence coding from predicted amino acid 16 of the *C. glutamicum* protein. The total amount of sequence from this region now totalled 1580 bp, which was considered sufficient for generation of a mutant strain by allelic exchange.

5.2.3 Construction of a *met* knockout cassette

A three step cloning approach was utilised and is outlined in Figure 5.2. The kanamycin resistance gene and the two halves of the *met* gene were PCR amplified and then sequentially inserted into pUC18. The region containing the kanamycin resistance gene and its promoter was PCR amplified from pEP2MCS using primers #227 and 228. For cloning purposes Primer #227 contained *Bam*HI and *Nhe*I sites and primer #228 a *Pst*I site. Purified PCR product was digested with *Bam*HI and *Pst*I and ligated into pUC18 prepared in the same manner to generate pSM46. The upstream *met* gene fragment was PCR amplified using primers#229 and 230. This PCR product is expected to contain sequence coding for amino acids 16 to 210 of Met. Primer #229 and 230 contain *Eco*RI and *Bam*HI sites at their 5' ends respectively. The resulting 750 bp PCR product was purified, digested with *Eco*RI and *Bam*HI, and then ligated into pSM48 prepared in the same manner to generate pSM50. The second *met* fragment was PCR amplified using primers #231 and 232. This primer pair is

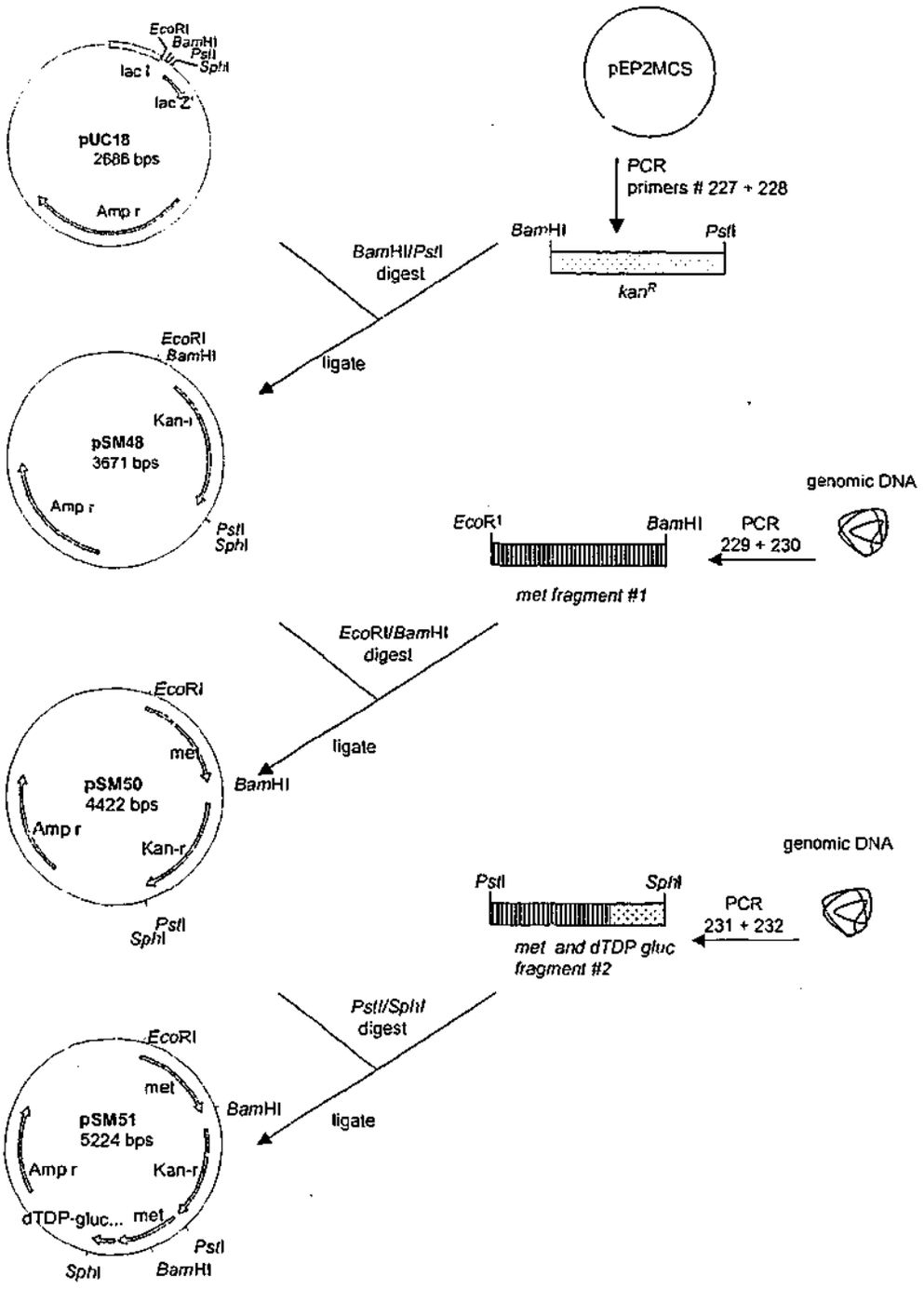

Figure 5.2: Construction of a metallopeptidase knockout cassette.

The diagram summarises the steps leading to the construction of pSM51. Details relating to the construction of this plasmid are presented in Section 5.2.3.

expected to amplify the remainder of the *met* gene in addition to approximately 210 bp of the *rrnlB* gene. *Pst*I and *Sph*I sites were located at the 5' ends of primers #231 and #232 respectively. A PCR product of 800 bp was generated, digested with *Pst*I and *Sph*I and ligated into pSM50 that had been prepared in the same manner to generate pSM51. Restriction endonuclease digestion and sequencing were used to confirm the presence of the three fragments in the new vector. The position of the kanamycin resistance gene within the cassette was such that it was inserted into the *met* gene just upstream of the zinc binding region signature (Figure 5.1). This site was chosen on the basis that the truncated protein generated should not be active as it is unable to bind the required metal cofactor.

5.2.4 Generation of *met* mutants by allelic exchange and confirmation by southern analysis

To generate a *met* mutant the metallopeptidase knockout cassette was introduced into C231 by electroporation. In order to determine the level of illegitimate recombination between plasmid DNA and the *C. pseudotuberculosis* chromosome control transformations were performed using pSM48. Additionally *C. pseudotuberculosis* was "transformed" with an equivalent volume of water to determine the level of spontaneous resistance to kanamycin.

50 µl aliquots of electrocompetent C231 were transformed with 5 µg (in 1 µl) of pSM48 or pSM51 or an equivalent volume of water. To maximise the time in which a crossover event between the plasmid and the *C. pseudotuberculosis* chromosome could occur, 1 ml of BHI was added to the transformed cells which were then incubated at 37°C with shaking for 20 hr. Bacteria were then plated onto BHI agar containing 20 µg/ml kanamycin and then incubated at 37°C for 48 hr. For each transformation of pSM51 between 2 to 3 colonies generally resulted. In general approximately twice as many colonies resulted following transformation with pSM51 as with the control plasmid pSM48. A similar number of colonies were observed following transformation with water as with pSM48.

Given that it was unlikely that an easily scored phenotypic change would be observed in the mutant strains it was necessary to screen for allelic exchange by southern analysis. Six transformants from pSM51 labelled met1 to met6 and 2 from pSM48 labelled con1 and con2 were selected for analysis. Chromosomal DNA from the 8 clones and the parent strain C231

was isolated. The DNA was then digested with *EcoR*V, which cuts twice within the *met* gene (Figures 5.1 and 5.3), and separated by agarose gel electrophoresis. The fragmented DNA was transferred to Hybond N+ and then analysed by Southern hybridisation. Replica Southern blots were probed with probes specific for the *met* and *rmlB* genes and pUC18 DNA. The *met/rmlB* probe was generated by PCR using primers # 229 and 232 to generate an approximately 1.6 kb DNA fragment (Figures 5.1 and 5.3) and then labelled with DIG. Additionally a pUC18 probe was generated by DIG labelling pUC18 plasmid.

Following hybridisation and a series of stringency washes the bound DIG labelled DNA was detected using an anti DIG AP conjugate and the AP substrate CDP-Star. In C231 a fragment of approximately 800 bp was detected with the *met/rmlB* probe which corresponded to the large *met* fragment generated by digestion with *Eco*RV. This band was also detected in met1, met4, met5, con1 and con2. This band was absent in met2, met3 and met6 which instead showed a band at approximately 1.8 kb, a shift of approximately 1 kb which corresponds to the size of the *Eco*RV fragment expected upon insertion of the kanamycin resistance gene into the metallopeptidase gene. Bands at 0.8, 1.8 and 3.4 kb were detected for met4, the first two bands correspond to the wildtype and mutant phenotypes respectively. The third band at 3.4 kb corresponded to the second fragment of pSM51 and was also detected with the pUC18 probe. The size of this band, in addition to the fact that it also hybridises pUC18 indicates that this clone has undergone a single rather than a double crossover event. Given that the probe used to detect the *met* gene was longer than the *Eco*RV fragment, it would be expected that the probe should hybridise to other fragments. The size of these fragments is not known, hence they may be large and as such would potentially be obscured by incompletely digested DNA at the top of the Southern. In summary the southern data indicate that met2, met3, met4 and met6 have undergone allelic exchange. The presence of an additional band for met4 which is also detected by the pUC18 probe indicates that this clone has undergone a single crossover event whereas the others have undergone true allelic exchange as a result of a double crossover event. met1, met6, con1 and con2 are presumably spontaneous kanamycin resistant mutants. The met6 clone was selected for further studies and renamed Cpmet.

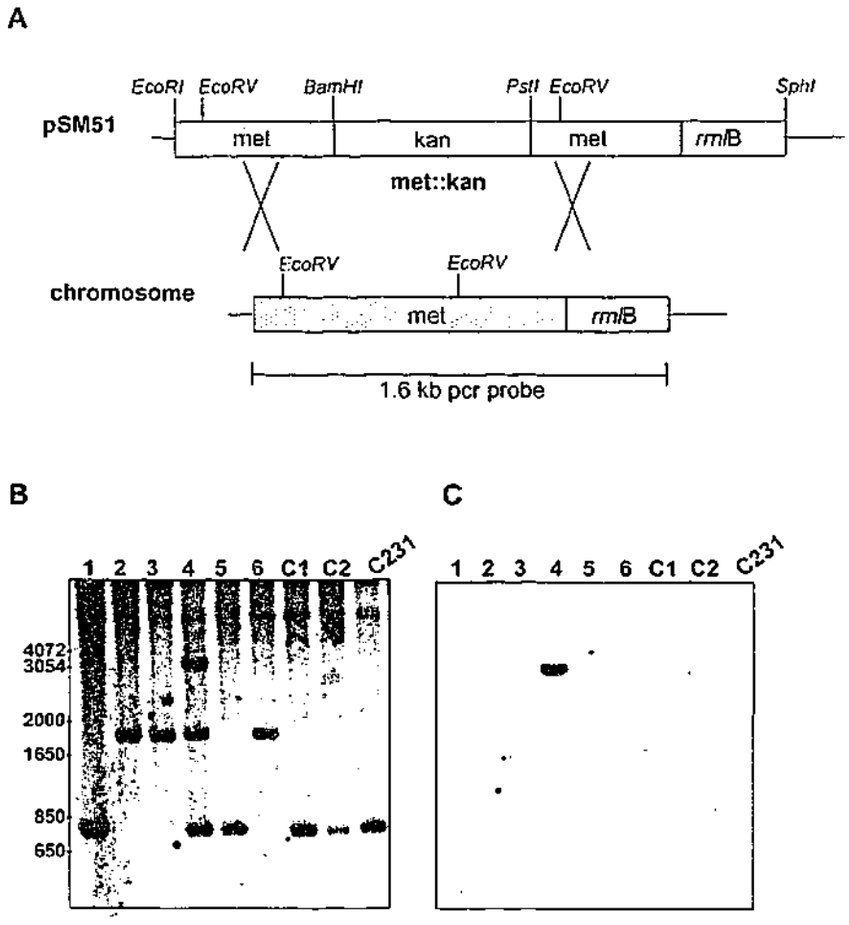

Figure 5.3: Verification of allelic exchange at the *met* locus.

The schematic diagram (A) represents the strategy used to generate a *met* mutant strain. The 1.6 kb PCR probe was generated using PCR primers #229 and 232. (B, C) Results of the Southern hybridisation of chromosomal DNA extracted from C231, met1 to met6 (1 to 6) and con1 (c1) and con2 (c2), digested with *Eco*RV and probed with the DIG labelled 1.6 kb probe (B) or DIG labelled pUC18 (C). Approximate DNA sizes are indicated to the left of B.

5.2.5 Identification of further upstream *met* sequence

The generation of the mutant strain allowed the remainder of the *met* gene to be identified from the genome by mutation recovery. Genomic DNA from CPmet was digested with a panel of blunt cutting restriction endonuclease enzymes, then shotgun cloned into pUC18. The ligation reactions were transformed into *E. coli* and colonies that were resistant to

kanamycin selected for on the basis that they should contain the kanamycin resistance gene and upstream sequence from the mutant strain. Three clones were obtained from a *Pvu*II digest. Sequencing indicated that they contained the appropriate fragment of DNA. Varying amounts of up and downstream sequence were obtained reflecting that partially digested DNA fragments had been ligated into pUC18.

Two methionine codons that may act as the initiating codon were identified in the 5' region of the *met* gene (Figure 5.1). Based on the observations that an upstream gene ended immediately after the second methionine codon, that a putative RBS could be identified upstream of the second methionine codon but not the first, and protein homologies, it is most likely that the second methionine codon corresponds to the start of the *met* gene. The predicted protein is 458 amino acids in length and is calculated to have a molecular weight of 51.7 kDa. The complete protein was most similar to homologues in *C. glutamicum* and various *Streptomyces* species (Figure 5.4).

Immediately upstream of the *met* gene is the 3' end of a gene coding for a probable peptidase that is likely to be a member of the prolyl oligopeptidase family. Additional sequence 3' of the metallopeptidase gene coded for the remainder of *rmlB*. This was followed by a protein that showed homology to RmlC and RmlD. RmlC codes for a protein with dTDP-4-keto-6-deoxyglucose-3,5-epimerase activity while *rmlD* contains dTDP-4-keto-L-rhamnose reductase activity. It seems likely that these two enzyme activities are found on the same protein in *C. pseudotuberculosis*, in a manner analogous to that in *C. glutamicum*.

The gene order in the vicinity of *met* was compared to that observed around the *met* homologue of *C. glutamicum* and *C. diphtheriae* (Figure 5.5). The annotated *C. glutamicum* genome has recently been released, hence annotated sequence was obtained from the NCBI Genbank database. The *C. diphtheriae* genome sequence is available but has not been annotated. A *met* homologue was identified by performing a BLASTN search. DNA in the vicinity of the homologue was then annotated on the basis of homologies to proteins in the NCBI protein databases.

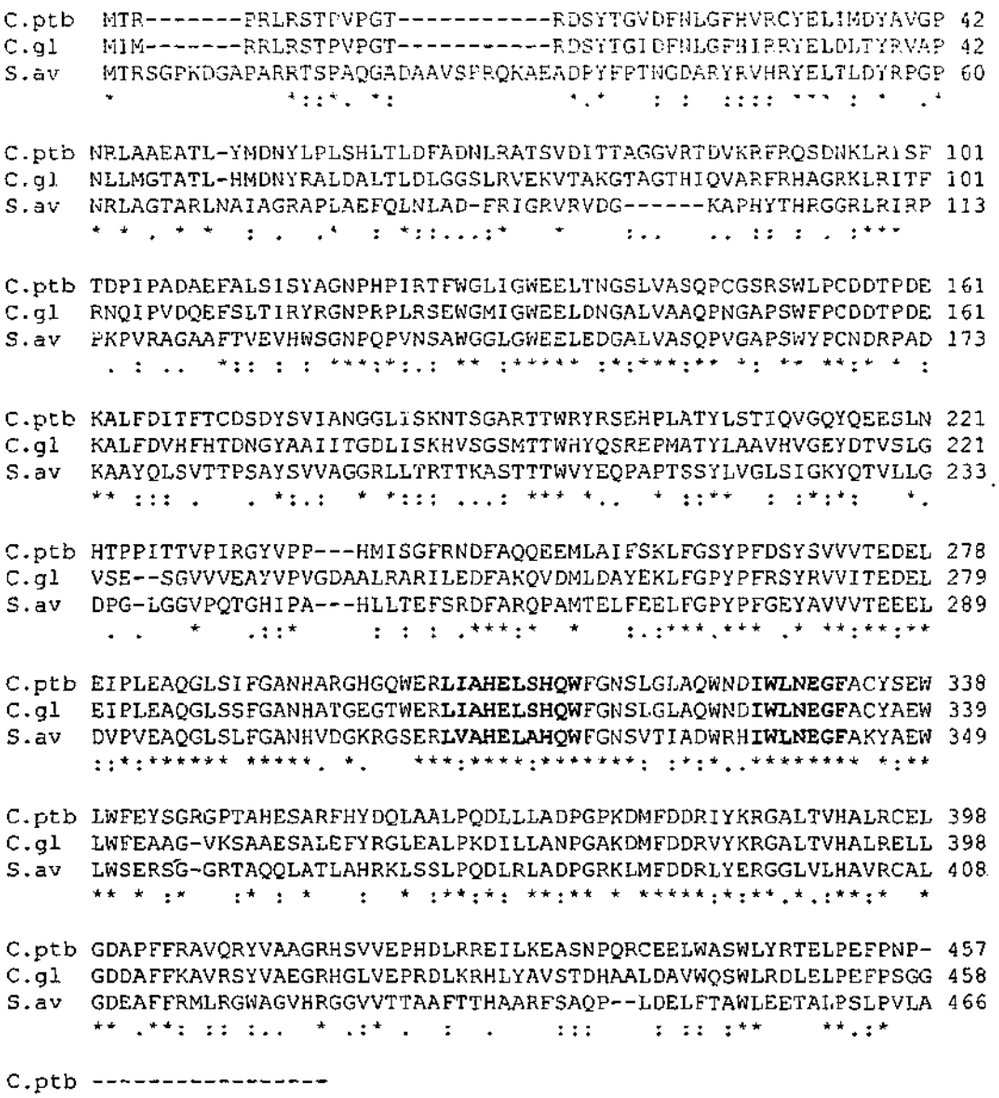

Figure 5.4: Clustal alignment of the predicted amino acid sequences of a metallopeptidase from *C. pseudotuberculosis* (C.ptb), *C. glutamicum* (C.gl) and Streptomyces *avermitilis* (S.av).

The two neutral zinc metellopeptidase zinc binding region signatures are shown in bold. Residues that are identical in all three species are indicated by an asterisk. Conserved amino acid substitutions are indicated by two dots while semi-conserved substitutions are indicated by a single dot. Alignment was performed using CLUSTAL W (1.82).

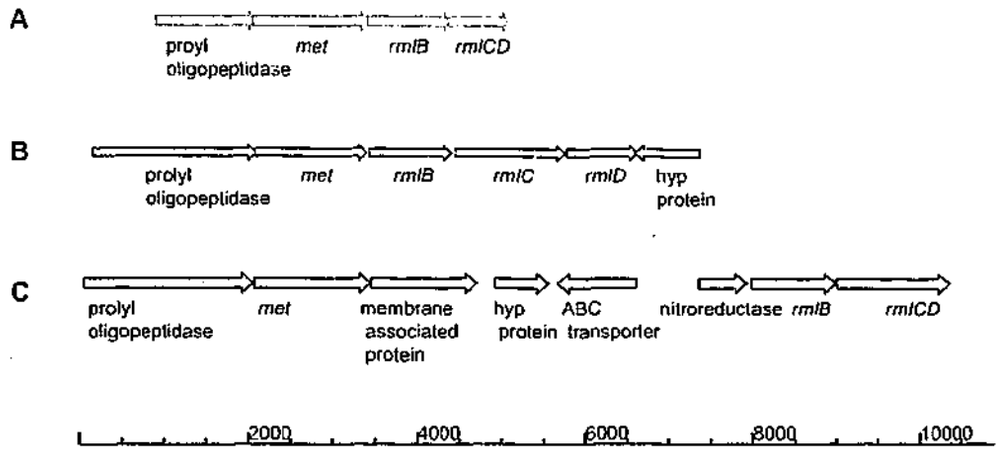

Figure 5.5: Organization of the *met* gene region in corynebacterial species.

The organization of genes surrounding the *met* gene of *C. pseudotuberculosis* (A), *C. diphtheriae* (B) and *C. glutamicum* (C) was determined. The *C. pseudotuberculosis* sequence was generated during this study and the *C. glutamicum* sequence was obtained from NCBI nucleotide database (Accession number NC003450). *C. diphtheriae* DNA sequence was produced by the Cdip Sequencing Group at the Sanger Institute and obtained from ftp://ftp.sanger.ac.uk/pub/cdip. This sequence was subsequently annotated based on homologies to proteins in the NCBI databases. Hyp protein represents a hypothetical protein while the ABC transporter of *C. glutamicum* is an ABC-type cobalamin/Fe^{3+}-siderophore transport system, periplasmic component.

Analysis of the *C. diphtheriae* sequence indicated that the gene order in this region was the same as that observed in *C. pseudotuberculosis*. The only difference was that *rmlC* and *rmlD* appear to code for separate proteins. In contrast only the *met* and prolyl oligopeptidase genes were co-located in the *C. glutamicum* genome. Genes coding from *rmlB* and *rmlCD* were however found a little further downstream. Like *C. pseudotuberculosis* the *rmlB* and *rmlD* proteins of *C. glutamicum* may comprise a single polypeptide.

Given that many proteases are secreted, in particular those that contribute to virulence, the Met sequence was analysed for the presence of a signal peptide using SignalP V1.1 (Nielsen *et al.*, 1997). This is a program that predicts secretory signal peptides. Using such an approach no signal peptide could be identified suggesting that Met is located within the cytoplasm.

Analysis of close homologues of *C. pseudotuberculosis* Met indicated that they did not contain a signal peptide either.

5.3 Copper, Zinc superoxide dismutase

5.3.1 Analysis of clone Q15-A3 insert sequence

Clone Q15-A3 contained a 1.2 kb insert that showed homology to two genes. At the 5' end of the insert was an ORF of 120 bp that showed homology to the 3' end of Cu,ZnSODs from a wide variety of prokaryotic and eukaryotic species (Figure 5.6). The protein was however most similar to Cu,ZnSOD homologues in mycobacterial species (Figure 5.9). While it is debatable as to whether this gene is up-regulated during macrophage infection (fold induction measured as 1.8) the purported role of this protein in virulence made it an interesting target for further study. A scan of the protein sequence against the PROSITE database identified a region that contained a Cu,ZnSOD signature (Figure 5.6). This motif (GDAGARFACGVI) is found near the C-terminal of the protein and contains a cysteine residue that is involved in a disulfide bond. The motif identified in the *C. pseudotuberculosis* Cu,ZnSOD matches the consensus motif at 11 of 12 amino acids (where the D residue is a G or N in the consensus). Two other motifs are usually found in gram-positive Cu,ZnSODs. The first contains amino acid residues that act as the copper ligand while the second is a prokaryotic membrane lipoprotein attachment site. These motifs would be expected to be located upstream of the sequence in Q15-A3 and hence were not present.

Commencing approximately 180 bp downstream of the predicted end of the *sodC* gene is a gene that shows homology to a putative carboxyltransferase subunit of acetyl CoA carboxylase that has been termed *accD*. The insert of Q15-A3 contained approximately 1 kb of sequence with homology to this gene.

5.3.2 Identification of upstream *sodC* sequence

Given that only 120 bp of the *sodC* gene was present on Q15-A3 it was necessary to identify upstream sequence before a Cu,ZnSOD knockout cassette could be generated. An analogous approach to that described in Section 5.2.2 for *met* was taken. Primer #156 was used as the reverse primer (Figure 5.6, Table 2.4). Two PCR generated fragments were identified with sizes of 200 and 520 bp. Both were analysed by sequencing and shown to contain the

1
61
121
181
241
301
361
421
481
541
601
661
721
781
841
901
961
1021
1081
1141
1201
1261
1321
1381
1441
1501

Figure

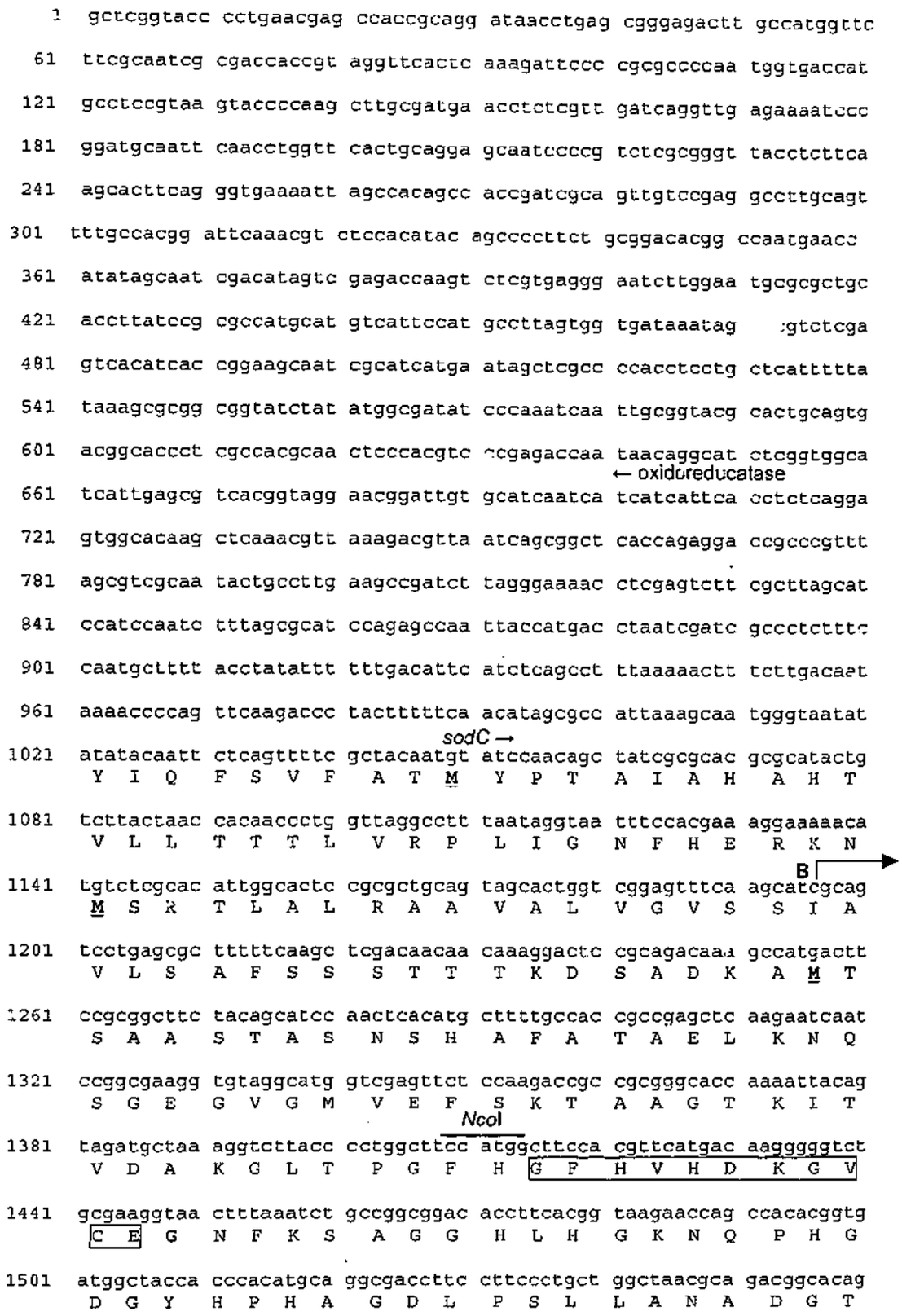

Figure 5.6: Please refer to legend on page 201.

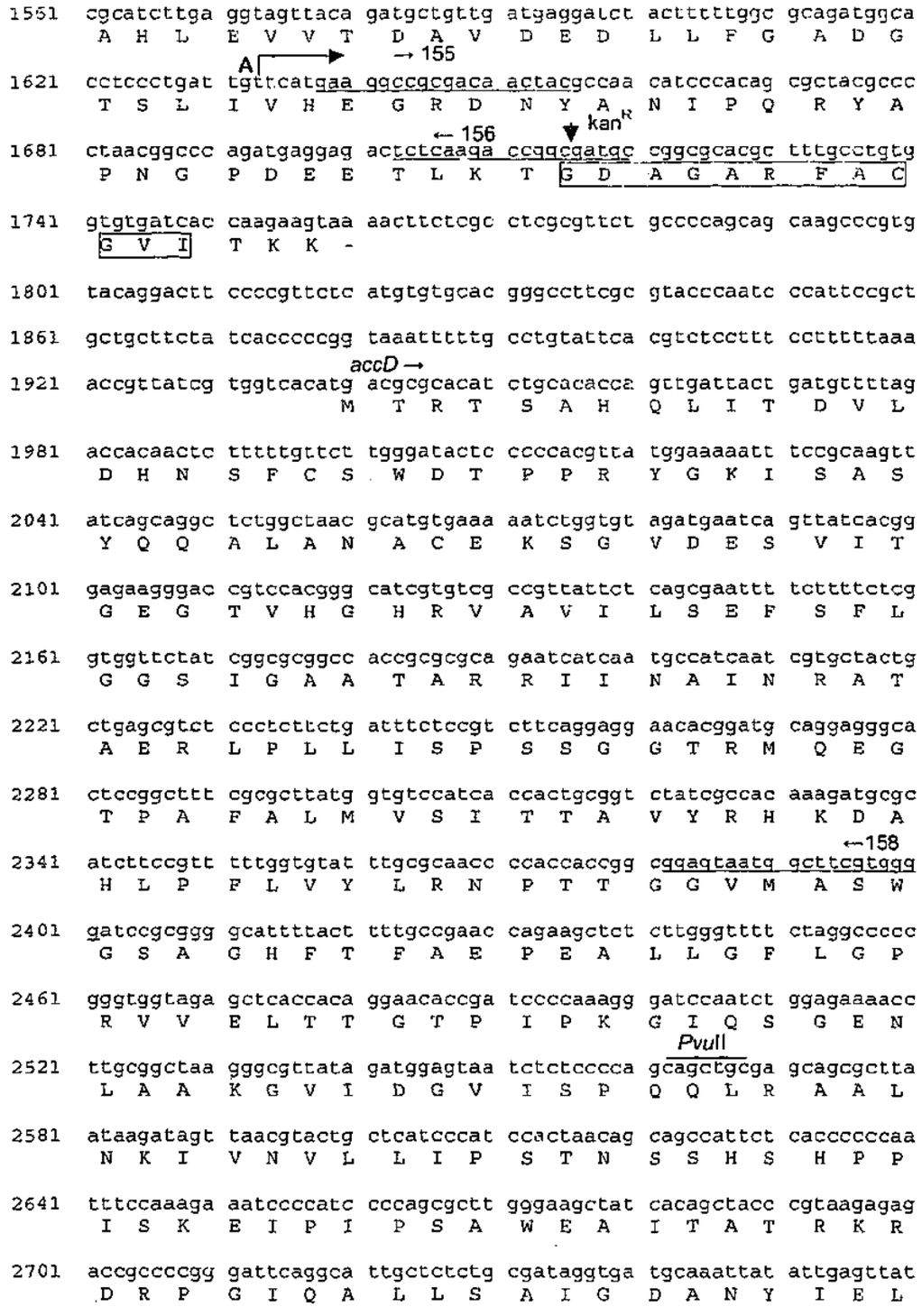

Figure 5.6: Please refer to legend on page 201.

```
2761  ccggaagcgg cgacggaagg atctcgccgt ccgttgtcgt cgcgttagcc cgaattgaag
       S  G  S   G  D  G  R   I  S  P   S  V  V   V  A  L  A   R  I  E

2821  ggcgtacggt ggtgattgtt gggcaagaca ggcactcgca gccccctttc gcagactctc
       G  R  T   V  V  I  V   G  Q  D   R  H  S   Q  P  P  F   A  D  S

2881  agctgggaac agaagcctta gatttgccc ggcggggcat t
       Q  L  G   T  E  A  L   R  F  A   R  R  G   I
```

Figure 5.6: Sequence of clone Q15-A3 and further up and downstream sequence.

All sequence downstream of arrow A is from the original clone Q15-A3. Sequence between arrows A and B corresponds to sequence identified from the PCR generated clones. Sequence upstream of the arrow B was identified by mutation recovery. Protein sequence with homology to a Cu,ZnSOD (coded by *sodC*) of mycobacteria and an acetyl coA carboxylase subunit of *C. glutamicum* (coded by *accD*) is shown using the single letter amino acid code. Upstream of *sodC*, on the complementary strand, is a gene coding for an oxidoreductase. The start of this gene is indicated. Two regions that show homology to the Cu,ZnSOD signature motifs are boxed. The first contains two histidine resides that bind the copper atom and the second contains a cysteine residue that is involved in a disulfide bond. Three putative initiating methionines of Cu,ZnSOD are in bold and underlined. Sequence complementary to primer #156 which was used to identify the upstream sequence is underlined. The insertion site of the kanamycin resistance gene in pSM53 is indicated. The location of primers #155 and 158 which were used to generate the PCR probe for Southern analysis (Figure 5.8) is indicated.

appropriate upstream sequence, leading to the identification of a further 430 bp of upstream sequence. This sequence coded for an additional 143 amino acids. An ATG codon was present near the 5' end of the sequence. It was not clear as to whether this was the initiating codon as the length of *sodC* genes varies considerably between bacterial species. Additionally sequence identity between species is less well conserved at the N-terminal end of the protein, making assignment based purely on homology difficult.

A scan of the new protein sequence against the PROSITE database identified the second Cu,ZnSOD signature motif. The region GFHVHDKGVCE (Figure 5.6) conforms to the defined consensus at all but one of the amino acids (where the second V in the sequence is a S, D or G in the consensus). This motif contains two histidine residues that bind the copper atom.

5.3.3 Construction of a *sodC* knockout cassette

As for the metallopeptidase construct a three step cloning process was utilised. pSM48 was used as the base plasmid and *sodC* sequence was inserted on either side of this. Given that only 560 bp of sequence 5' to the end of the *sodC* gene was available it was necessary to insert the kanamycin resistance gene quite close to the end of the *sodC* gene. A position was chosen that was immediately 5' of the motif involved in disulfide bonding (Section 5.3.1). Given that this motif is conserved in both prokaryotic and eukaryotic Cu,ZnSODs it was hoped that the absence of the disulfide bond would be sufficient to ablate superoxide dismutase activity even though the protein was only slightly truncated.

The 5' *sodC* fragment was PCR amplified from genomic DNA using primers #254 and #248. These contain *Eco*RI and *Bam*HI sites respectively at their 5' end for cloning. A PCR product of 520 bp was generated and cloned into pSM48 digested with *Bam*HI and *Eco*RI to generate pSM52. The 3' end of *sodC* and downstream sequence containing the *accD* gene was PCR amplified from *C. pseudotuberculosis* genomic DNA using primers # 249 and 250 which contain *Pst*I and *Sph*I sites respectively at their 5' ends for cloning. A fragment of 1.1 kb was generated and cloned into pSM52 linearised with *Sph*I and *Pst*I to generate the knockout cassette pSM53. The cloning steps required to generate pSM52 are summarised in Figure 5.7.

5.3.4 Generation of the *sodC* mutant strain and confirmation by Southern analysis

3x 50 µl aliquots of electrocompetent C231 were transformed with 3 µg of pSM53 and pSM48. After incubation at 37°C for 20 hr the transformation mixes were plated onto BHI agar containing 20 µg/ml kanamycin. On average two to three colonies per transformation were generated from pSM53 and one for pSM48. Seven colonies from the pSM53 transformations were chosen for further analysis and labelled sup1 to sup7. Chromosomal DNA was extracted and doubly digested with *Nco*I and *Pvu*II, which cut in the *sodC* and *accD* genes respectively (Figure 5.8), and then separated by agarose gel electrophoresis. The fragmented DNA was transferred to Hybond N+ and then analysed by Southern hybridisation. Replica Southern blots were hybridised with probes specific for *sodCIaccD* and pUC18 DNA. The *sodCIacc*D probe was generated from genomic DNA by PCR using DNA using primers #155 and #158, then labelled with DIG. This PCR product lies within the *Nco*I- *Pvu*II genomic fragment (Figure 5.8). The DNA incubated with *Pvu*II and *Nco*I was not

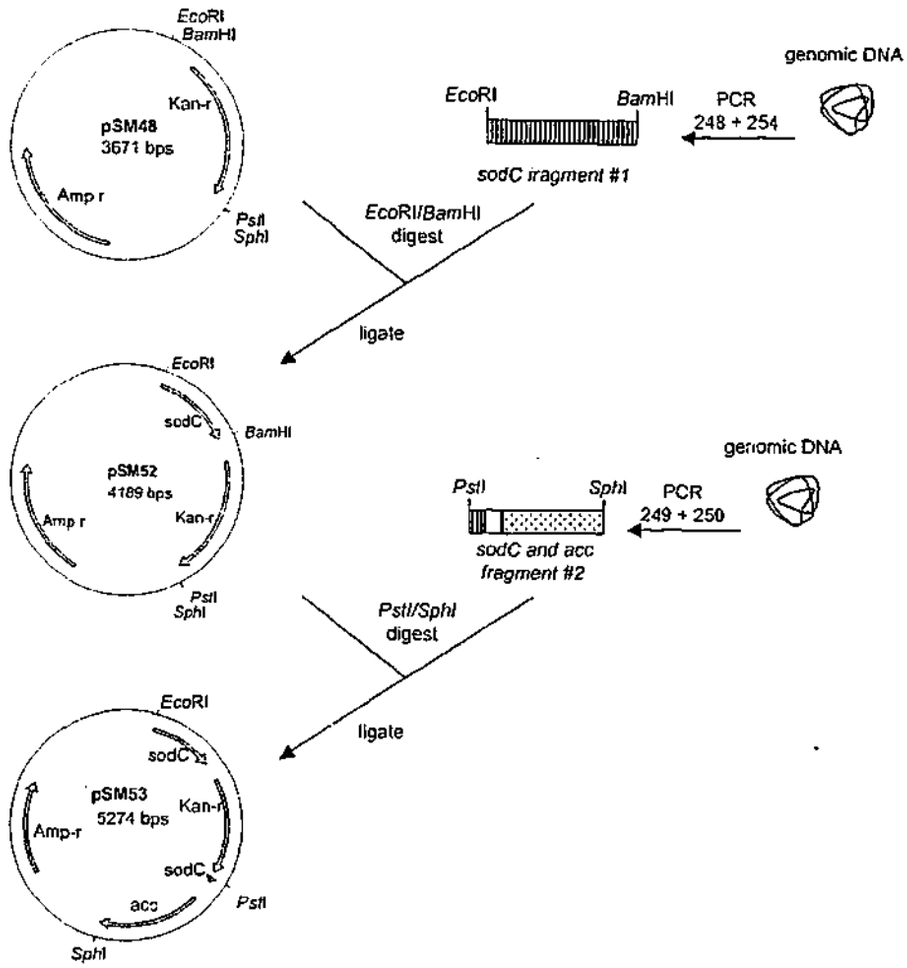

Figure 5.7: Construction of the *sodC* knockout cassette.

The diagram summarises the steps leading to the construction of pSM53. Details relating to the construction of this plasmid are presented in Section 5.3.3.

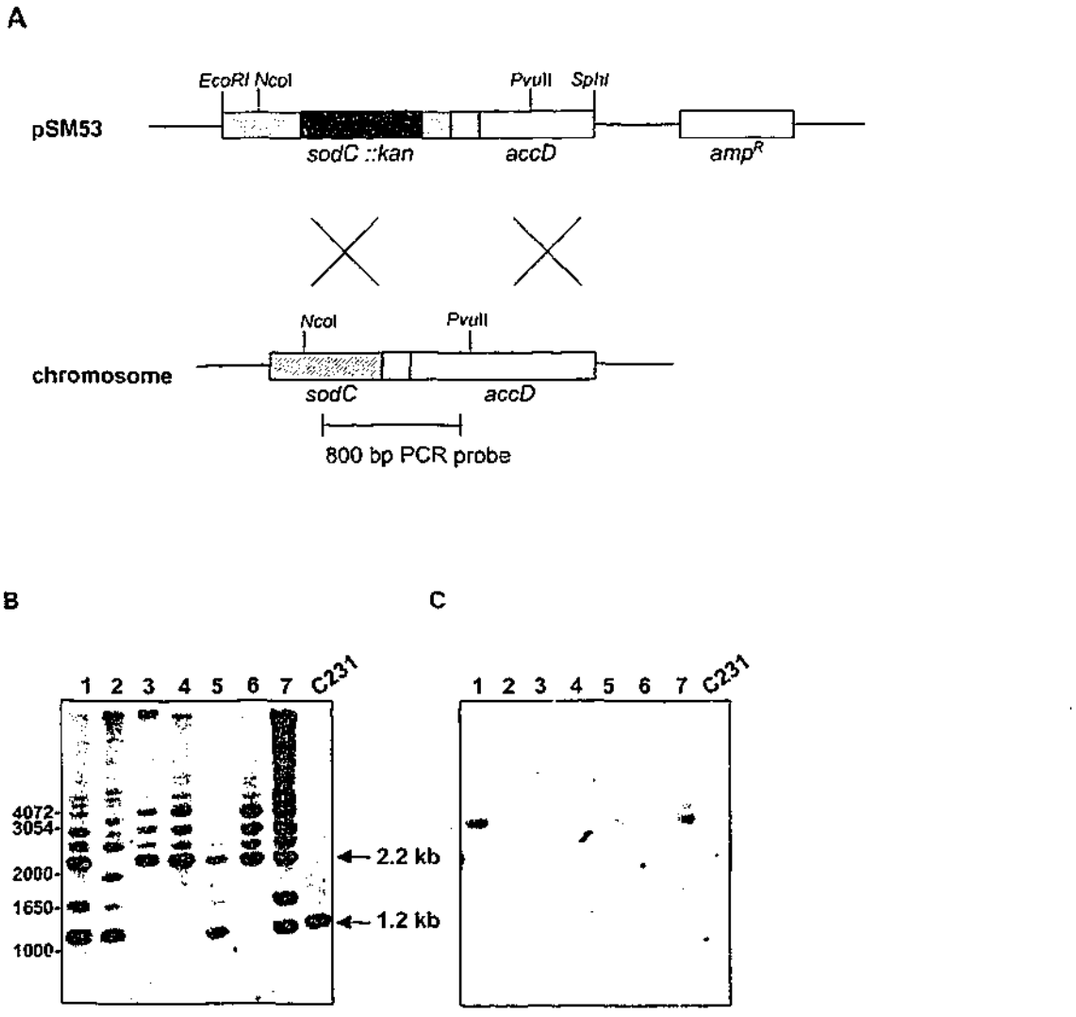

Figure 5.8: Verification of allelic exchange at the *sodC* locus.

The schematic diagram (A) represents the strategy used to generate a Cu,ZnSOD mutant strain. The 800 bp PCR probe was generated using PCR primers #155 and 158. (B, C) Results of the Southern hybridisation of chromosomal DNA extracted from C231 and sup1 to sup7 (1 to 7), digested with *NcoI* and *PvuII* and probed with the DIG labelled 800 bp probe (B) or DIG labelled pUC18 (C). Bands corresponding to the wildtype *NcoI*/*PvuII* fragment (1.2 kb) and kanamycin resistance gene interrupted fragment (2.2 kb) are indicated by arrows in (B). Approximate fragment sizes are indicated to the left of (B).

digested to completion, hence additional high molecular weight bands were present on the Southern. The results were however still interpretable. In C231 a fragment of approximately 1.2 kb was detected with the *sodC/accD* probe which corresponded to the wildtype size *Pvu*II/*Nco*I fragment. This band was also present in sup1, sup2, sup5 and sup7. This band was absent in sup3, sup4 and sup6 which instead showed a band at approximately 2.2 kb, a shift of approximately 1 kb which corresponds to the size of the *Pvu*II/*Nco*I fragment expected upon insertion of the kanamycin resistance gene into *sodC*. This second band was also present in sup1, sup5 and sup7. The presence of both bands in sup1, sup5 and sup7, in addition to the observation that these clones are positive for pUC18 indicates that they are as result of a single crossover event resulting in incorporation of the entire plasmid into the *C. pseudotuberculosis* genome. Given that sup3, sup4 and sup6 contain only the band at 1.2 kb and are not positive for pUC18 indicates that a double crossover event has occurred to generate the mutant strains. The strain sup3 was renamed CPsodC and used in further studies.

5.3.5 Identification and characterisation of upstream *sodC* sequence

Further upstream sequence was identified using mutation recovery in a manner analogous to that for the *met* gene. Genomic DNA was isolated from mutant strain CPsodC and digested with *Pvu*II prior to ligation into pUC18 linearised with *Sma*I. Following transformation of the library, selection for kanamycin resistant colonies was performed, resulting in approximately 30 colonies.

Using this approach a further 1.2 kb of upstream sequence was obtained. Upstream of *sodC* was a gene on the complementary strand that coded for an oxidoreductase. An intergenic region of approximately 450 bp was located between the two genes and presumably contains promoter sequence for both genes. Within the sequence three putative methionine initiation codons were identified for *sodC*. Based on homology to other *sodC* genes from other species it was not immediately apparent which of these was the most likely candidate for the start of the protein. Sequence at the N-terminal of Cu,ZnSODs is generally less well conserved than that found further downstream (Figure 5.9). Additionally the length of *sodC* is quite variable between bacterial species. For example the Cu,ZnSOD of *M. tuberculosis* and *M. leprae* is 240 amino acids, that of *M. paratuberculosis* is 226 amino acids and that of *S. typhimurium* is

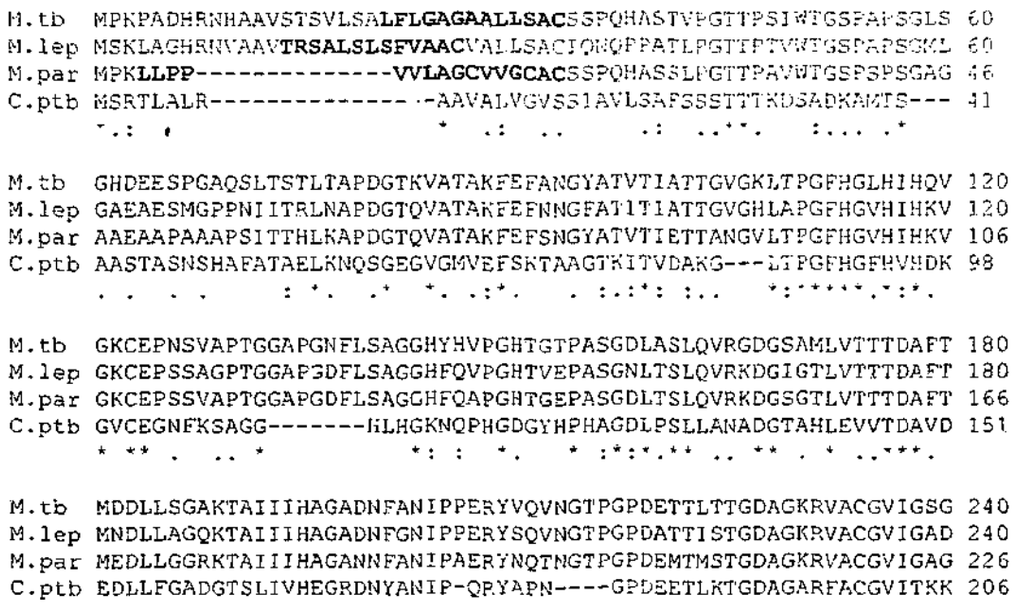

Figure 5.9: Clustal alignment of the amino acid sequences of Cu,ZnSOD from *C. pseudotuberculosis* (C.ptb), *M. tuberculosis* (M.tb), *M. leprae* (M.lep) and *M. paratuberculosis* (M.par).

The predicted prokaryotic membrane lipoprotein attachment site is shown in bold for each of the mycobacterial sequences. Residues that are identical in all species are indicated by an asterisk. Conserved amino acid substitutions are indicated by two dots while semi-conserved substitutions are indicated by a single dot. Alignment was performed using CLUSTAL W (1.82).

177 amino acids. The three putative methionine initiation codons would make the *C. pseudotuberculosis* protein 237, 206 or 168 amino acids in length. Based on information gained from other gram-positive Cu,ZnSOD proteins it would be expected that the *C. pseudotuberculosis* Cu,ZnSOD should contain a prokaryotic membrane lipoprotein lipid attachment site (Figure 5.9). This is a defined precursor signal peptide motif that is found near the N-terminal of lipoproteins. The signal peptide is recognised and cleaved by signal peptidase II (which is a specific lipoprotein signal peptidase). Cleavage occurs upstream of a cysteine residue and is concomitant with attachment of a glyceride-fatty acid lipid to the cysteine residue (Hayashi and Wu, 1990). Upon analysis of sequence in the region of the three methionine residues no prokaryotic membrane lipoprotein lipid attachment site could be detected, in particular there were no cysteine residues to act as the site of lipid attachment. To

```
C.ptb  MSR----TLALR-AAVALVGVSSIAVLSAFSSSTTTKDSADKAMTSAASTASN---SHAF 52
C.dip  MNTRAVFALSPRSKKLTMLGVG-FALASALAACGSGEQTENASMNASASETSSGEMGTPF 59
       *.     :*: *   :::**.::*: **:::. : ::: : :*.::** :*.   . .*

C.ptb ATAELKNQSGEGVGMVEFSKTAAGTKITVDAKGLTPGFHGFHVHDKGVCEGNFKSAGGHL 112
C.dip ATAKLVTQKGEDAGTVTFTTTQAGTEINVDAKGLEPGFHGFHVHDKGKCEGDFKSAGGHL 119
      ***:* .*.**..* * *:.* ***:*.****** ********** ***:********

C.ptb HGKNQPHG----DGYHPHAGDLPSLLANADGTAHLEVVTDAVDEDLLFGADGTSLIVHEG 168
C.dip SGPDHSHGSDHSGGYHPHAGDLPSLLAAKDGTAKLTVTTDALTPELLLGTDGASLIVHEG 179
       * ::.*'    .************** ****:* *.***:  :**:*:**:********

C.ptb RDNYANIPQRYAPNGPDEETLKTGDAGARFACGVITKK 206
C.dip RDNFANIPERYAPEGPDQDTLKTGDAGARFACGVIEAK 217
      ***:****:****:***::**************** *
```

Figure 5.10: Clustal alignment of the predicted amino acid sequences of Cu,ZnSOD from *C. pseudotuberculosis* (C.ptb) and *C. diphtheriae* (C.dip).

Alignment of the *C. pseudotuberculosis* sequence is based on the hypothesis that the second methionine residue (Section 5.3.5) is the initiating codon. The predicted prokaryotic membrane lipoprotein attachment site of the *C. diphtheriae* sequence is shown in bold. The *C. pseudotuberculosis* sequence is predicted to contain a Type I signal peptide. The cleavage site for this peptide is indicated by the arrow. Residues that are identical in both species are indicated by an asterisk. Conserved amino acid substitutions are indicated by two dots while semi-conserved substitutions are indicated by a single dot. Alignment was performed using CLUSTAL W (1.82).

determine whether any other signalling peptides were present the amino acid sequence was analysed using SignalP V1.1 (Nielsen *et al.*, 1997). Using such an approach a signal peptide was predicted with the cleavage site most likely to be positioned between amino acids 24 and 25 (LSA-FS) relative to the second methionine (Figure 5.10).

5.3.6 Comparison of *C. pseudotuberculosis* and *C. diphtheriae* Cu,ZnSODs

The *C. diphtheriae* genome has been sequenced but not annotated. Given that *C. diphtheriae* is a pathogenic bacteria it is likely to possess a *sodC* gene. *C. pseudotuberculosis* is more closely related to this organism than mycobacteria. We were therefore interested in determining the sequence of *C. diphtheriae* Cu,ZnSOD and whether it contained a prokaryotic membrane lipoprotein lipid attachment site. The *C. diphtheriae* genome sequence was searched for a Cu,ZnSOD homologue by performing a BLASTX search with the *C. pseudotuberculosis sodC* sequence. Highly homologous protein sequence was identified. The

A

```
 1  atgtctcgca cattggcact ccgcgctgca gtagcactgg tcggagtttc aagcatcgca
      M  S  R   T  L  A   L  R  A  A   V  A  L   V  G  V   S  S  I  A

61  gtcctgagcg ctttttcaag ctcgacaaca acaaaggact ccgcagacaa agccatgact
      V  L  S   A  F  S   S  S  T  T   T  K  D   S  A  D   K  A  M  T
```

B

```
 1  atgtctcgca cattggcact ccgcgctgca gtagcactgg tcggagtttc aagcatcgca
      M  S  R   T  L  A   L  R  A  A   V  A  L   V  G  V   S  S  I  A

61  gtcctgagcg ctttttcatg ctcgacaaca acaaaggact ccgcagacaa agccatgact
      V  L  S   A  F  S   C  S  T  T   T  K  D   S  A  D   K  A  M  T
```

Figure 5.11: Sequence analysis of the N-terminal of *C. pseudotuberculosis* Cu,ZnSOD.

DNA and corresponding protein sequence of the 5' end of the *C. pseudotuberculosis sodC* gene (A). Substitution of a single adenosine base (boxed, A) to a thymidine (boxed, B) would be sufficient to generate a prokaryotic membrane lipoprotein attachment site (bold, B).

corresponding DNA sequence was obtained and shown to code for a protein that showed homology to Cu,ZnSODs from a variety of bacterial species. Analysis of this sequence indicated the presence of a prokaryotic membrane lipoprotein lipid attachment site (Figure 5.10). 31 amino acids upstream of the cysteine residue of this sequence was a methionine residue that is likely to be the initiating methionine. In an attempt to further define the initiation codon of the *C. pseudotuberculosis* Cu,ZnSOD a CLUSTAL W alignment was performed against the *C. diphtheriae* sequence (Figure 5.10). Based on alignment against the *C. diphtheriae* sequence it is most likely that the second methionine is the start codon. This is further supported by the observation that a putative RBS is located upstream of this. This methionine is also in the appropriate position for cleavage of the putative signal peptide.

Further analysis of the *C. pseudotuberculosis* sequence in this region revealed that the sequence differed from the consensus pattern for a prokaryotic membrane lipid attachment site at one amino acid. In particular this was the absence of the cysteine residue within this motif. Interestingly the substitution of a single base within the gene sequence for this region would be sufficient to change a serine residue into a cysteine residue (Figure 5.11). Such a modification would generate a prokaryotic membrane lipid attachment site. To ensure that the

sequence obtained for this region was correct three independent genomic clones were sequenced in both directions. This confirmed that the sequence was correct as first annotated.

5.4 Iron permease component, *fagC*

5.4.1 Analysis of clones Q2-G12 and Q3-B2

Analysis of the sequence of the insert of clones Q2-G12 and Q3-B2 indicated that they contained a number of genes and in both cases it appeared that the insert was a concatamer of DNA fragments. However, both clones contained sequence coding for the *pld* gene and a protein with homology to the ATPase component of the Fep iron permeases of *E. coli* and *Y. enterocolitica*. This protein was most similar to the ferric enterobactin transport ATP-binding protein FepC of *E. coli* and *Yersinia enterocolitica*. Subsequent to the completion of laboratory work described in this thesis the sequence of this gene and others in the region was published (Billington *et al.*, 2002). In this publication the *C. pseudotuberculosis* homologue of *fepC* has been termed *fagC*. In order to maintain continuity this nomenclature has been adopted in this study. The *fagC* gene is on the complementary strand to the *pld* gene such that their transcripts converge. In Chapter 3 expression of *fagC* was shown to be up-regulated at higher cell densities and down regulated following heat shock. Although not further up-regulated during macrophage infection *fagC* was expressed by intracellular bacteria. Thus *fagC* demonstrated a similar expression pattern to the major virulence determinant *pld*.

A scan of the FagC protein sequence against the PROSITE database of protein families and domains identified two motifs that are typically found in this type of protein. The first is identified as an ATP/GTP binding motif and the second as an ABC transporter family signature (Figure 5.12).

A clone containing upstream sequence of *fagC* was identified when identifying genes of *C. pseudotuberculosis* that are expressed in a density dependent manner (Chapter 3.6). Clone Q5-C9 in addition to containing the 5' portion of *fagC* contained a gene coding for a protein homologous to FepG of *E. coli* and Y. *enterocolitica* that has subsequently been termed *fagB*. FepG is an integral membrane protein that forms part of the membrane permease. The organization of genes within this region is represented schematically in Figure 5.12. Two other genes that were identified by Billington et al (2002) include a gene termed *fagA* that is

A

```
  1  cccgtggggc tagtcactgt agttattggt ggcatttacc tgctatctct cattgcctca
       P  V  G   L  V  T   V  V  I  G   G  I  Y   L  L  S   L  I  A  S

 61  gcttcccgac gcgcctaagg attcctgtgc ataccctaa aactatctct acagatttac
       A  S  R   R  A  -
       S  F  P   T  R  L   R  I  P   M  H  T  P   K  T  I   S  T  D  L

121  ccactacagc tggcatttcc acggatcata cgtctgccct tcgcacagaa aacgtcaccc
       P  T  T   A  G  I   S  T  D   H  T  S  A   L  R  T   E  N  V  T

181  ttagctggga taagcatgtt gtaagcaccg accttagtgt tgacataccc acggggaagt
       L  S  W   D  K  H   V  V  S   T  D  L  S   V  D  I   P  T  G  K

241  tcacggctat catcggccct aatggttgtg gaaagtcaac gctgcttaaa agctgtgccc
       F  T  A   I  I  G   P  N  G   C  G  K  S   T  L  L   K  S  C  A

301  gtattcttgc cccagacact gggtcgattt atcttaatca gcacaatctc acccacctgc
       R  I  L   A  P  D   T  G  S   I  Y  L  N   Q  H  N   L  T  H  L

361  ataccaagga gatagccaaa cagcttgcgc ttctccccca aagcaccatc actccggccg
       H  T  K   E  I  A   K  Q  L   A  L  L  P   Q  S  T   I  T  P  A

421  atatcacggt agaagagctc gtccgccgcg gacgattccc tcatcacaat tggttacatc
       D  I  T   V  E  E   L  V  R   R  G  R  F   P  H  H   N  W  L  H

481  aatggaccgc cgaagatact cgtgccgtcg acgctgctct caatgccgcc agtgttaaag
       Q  W  T   A  E  D   T  R  A   V  D  A  A   L  N  A   A  S  V  K

541  aactcactca caaacgcgtg accgatttat ccggtggaca gcgtcagcgc gtatggcttg
       E  L  T   H  K  R   V  T  D   L  S  G  G   Q  R  Q   R  V  W  L

601  caatggttct agcccaaaat actcctacgg tacttctgga cgaacccacc accttcctag
       A  M  V   L  A  Q   N  T  P   T  V  L  L   D  E  P   T  T  F  L

661  acatcgccca ccagtaccaa cttcttgagt tagcacgatc tcttacgact cagctcaatc
       D  I  A   H  Q  Y   Q  L  L   E  L  A  R   S  L  T   T  Q  L  N

721  gcaccgttgt cgcggtactg cacgatctgc aacaggctgt ccgctacgcc gaccacctga
       R  T  V   V  A  V   L  H  D   L  Q  Q  A   V  R  Y   A  D  H  L

781  ttgttatgaa aaatggcagc atcgtagcta ccggttcccc acaagaagtg atcagcgcag
       I  V  M   K  N  G   S  I  V   A  T  G  S   P  Q  E   V  I  S  A

841  aactcatatc cgaagttttt aatatcacgg ttcgcaccat gcttatcgac gaccaactca
       E  L  I   S  E  V   F  N  I   T  V  R  T   M  L  I   D  D  Q  L

901  tagtcattcc aaccgaatta ccagtgccaa acgtcgctat gatgcagggc tagggcctct
       I  V  I   P  T  E   L  P  V   P  N  V  A   M  M  Q   G  -
```

B

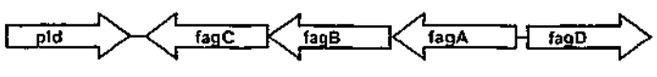

Figure 5.12: Sequence of *fagC*.

(A) DNA sequence of *fagC* was obtained from array clones Q2-G12, Q3-B2 and Q5-C9. Immediately upstream of *fagC* is the 3' end of *fagB* (underlined). A region that shows homology to a consensus motif found in ABC transporters is boxed and shaded while an ATP/GTP-binding site motif is boxed. (B). The gene order in the vicinity of *fagC* is shown (Billington *et al.*, 2002).

homologous to *fepD* of *E. coli* and *Y. enterocolitica*. Like *fagB*, *fagA* is an integral membrane protein and together with *fagC* they form the membrane permease. Based on the organization of genes *fagA*, *fagB* and *fagC* are likely to comprise an operon whose expression is driven by a promoter upstream of *fagA*. Upstream and divergently transcribed to *fagA* is a fourth gene termed *fagD*. Based on sequence homology *fagD* is likely to be a periplasmic iron-siderophore binding protein.

5.4.2 Construction of a *fagC* knockout cassette

To generate a *fagC* knockout cassette the same approach as used previously for the other constructs was utilised. The cassette was generated such that the kanamycin gene disrupted *fagC* approximately 510 bp from the initiating codon. This position was chosen as it lies immediately before the ABC transporter family signature, a region which is presumably important for protein function based on the conservation of sequence within this region. The 5' fragment which includes this *fagC* sequence and additional *fagB* sequence was PCR amplified from genomic C231 DNA using primers #257 and #258 which contain *Eco*R1 and *Bam*HI sites respectively at their 5' end for cloning purposes. A 790 bp product was generated and ligated into pSM48 digested with *Eco*R1 and *Bam*HI to generate pSM54. The 3' portion of *fagC* and the end of *pld* were PCR amplified using primers #259 and #260 which contain *Pst*I and *Sph*I sites at their 5' ends for cloning purposes. A PCR product of 960 bp was generated and ligated into pSM54 digested with *Pst*I and *Sph*I to generate the knockout cassette pSM55. The cloning steps required to generate pSM55 are summarised in Figure 5.13.

5.4.3 Generation of the *fagC* mutant strain and confirmation by PCR

3x 50 µl aliquots of electrocompetent C231 were transformed with 3 µg of pSM48 and pSM55. After incubation at 37°C in 1 ml BHI for 20 hr the transformation mixes were plated onto BHI agar containing 20 µg/ml kanamycin. On average 3 colonies per transformation were generated from pSM55 and 1 for pSM48. Eight colonies from the pSM55 transformations were chosen for further analysis and labelled fer1 to fer8. Chromosomal DNA was extracted and digested with *Acc*I which cuts once within the *fagC* gene and once in the *fagC pld* intergenic region (Figure 5.14). Unfortunately the digestion reaction did not go to completion, hence the blotted DNA was not suitable to be hybridised with a *fagC* probe.

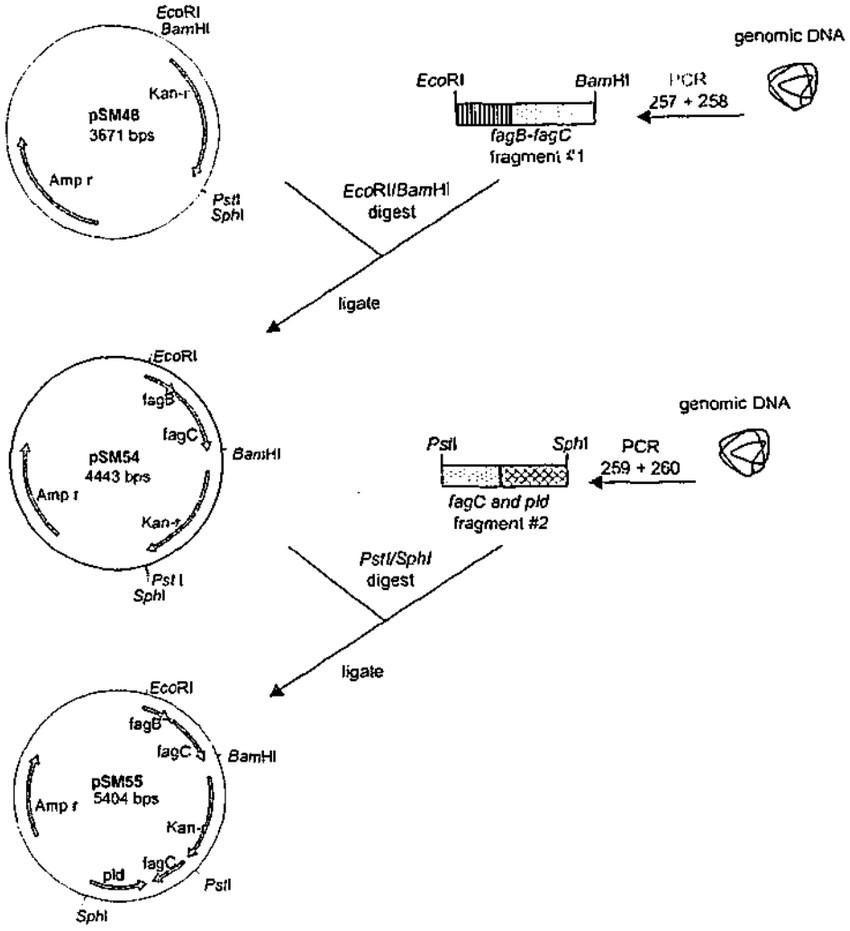

Figure 5.13: Construction of a *fagC* knockout cassette.

The diagram summarises the steps leading to the construction of pSM55. Details relating to the construction of this plasmid are presented in Section 5.4.2.

Given that useful information could still be obtained from probing the blotted DNA with pUC18 (to determine whether vector had been incorporated into the genome) this was still performed. Clones fer2 and fer3 were positive for pUC18. As an alternative approach to Southern analysis PCR was utilised to determine whether a double crossover event had occurred. In order to ascertain whether the *fagC* gene was disrupted by the presence of the kanamycin resistance gene all clones except fer2 were analysed by PCR using primers #257 and #261. Primer #257 is located in the *fepC* gene and was used as the forward primer to generate the 5' *fagC* fragment for the knockout cassette. Primer #261 hybridises immediately 3' to the *pld* sequence found in pSM55 (hence is not within the region found in pSM55). Using this approach two band sizes were observed (Figure 5.14). A band at approximately 1.8 kb was seen in C231, fer4, fer5, fer7 and fer8. The remaining three clones fer1, fer3 and fer6 contained a band at approximately 2.8 kb corresponding to the insertion of the kanamycin gene into this region. fer3 was positive for pUC18 and hence it is likely that this clone has undergone a single crossover event. That the 2.8 kb PCR band is observed for fer3 reflects that the reverse primer (#261) lies outside the sequence found in pSM55 and that the crossover event has occurred 3' to the kanamycin resistance gene. fer1 and fer6 were not positive for pUC18 and as such they represent allelic exchange as a result of a double crossover event. fer1 was renamed CPfagC and selected for future experiments.

5.5 *in vitro* growth of mutant strains in BHI

In order to determine whether the mutant strains demonstrated the same growth curves as C231 *in vitro*, overnight cultures of C231, CPmet, CPsodC and CPfagC were diluted to OD_{600} of 0.1 then incubated at 37°C and 300 rpm. At 2 hourly intervals the OD_{600} of each culture was measured (Figure 5.15). All strains demonstrated the same growth curve. It is likely that the starting density of CPsodC and CPfagC was slightly higher as these curves lie slightly above the other two.

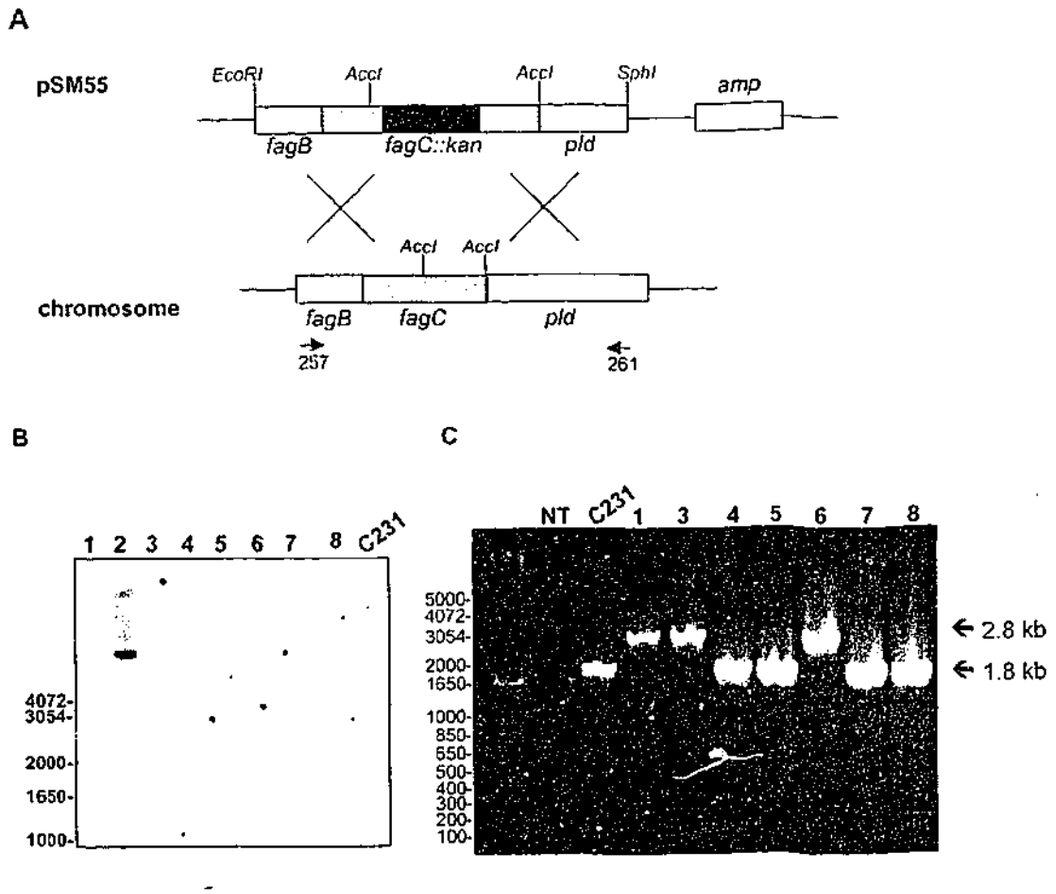

Figure 5.14: Verification of allelic exchange at the *fagC* locus.

The schematic diagram (A) represents the strategy used to generate a *fagC* mutant strain. (B) Results of the Southern hybridisation of chromosomal DNA extracted from C231 and fer1 to fer8 (1 to 8), digested with *AccI* and probed with DIG labelled pUC18. (C) PCR analysis was performed using primers #257 and #261. Chromosomal DNA from C231, fer1 (1) and fer3 to fer8 (3 to 8) was used as the template. A PCR amplicon of 1.8 kb was expected following amplification of the wildtype sequence, while a fragment of 2.8 kb was observed following insertion of the kanamycin resistance gene into *fagC*. A no template control was included (NT). 1 kb plus ladder is run in the first lane and band sizes (in bp) are indicated to the left of this.

5.6 Viability and replication of mutant strains in the macrophage infection model

In order to determine whether the mutant strains were attenuated in the intracellular environment macrophages were infected with the mutant strains and bacterial viability and rate of replication determined by recovering bacteria and determining the number of colony

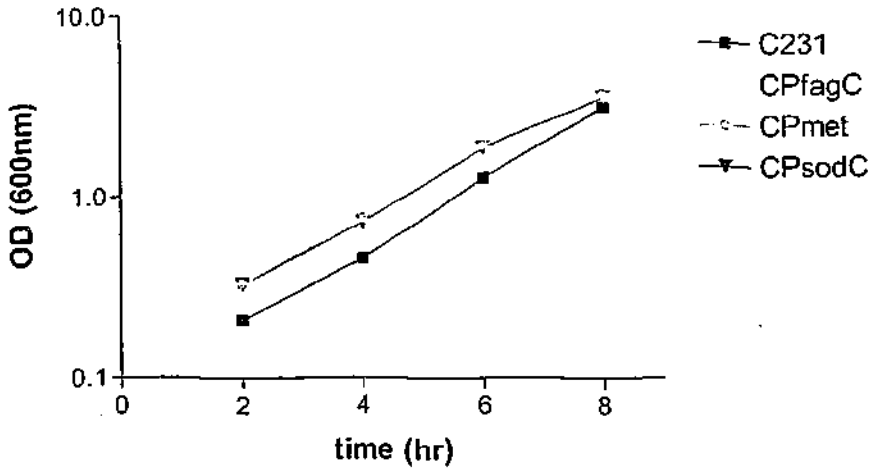

Figure 5.15: *in vitro* growth of *C. pseudotuberculosis* mutant strains.

Overnight cultures of C231, CPmet, CPsod and CPfag were diluted to OD_{600} of 0.1 then incubated at 37°C with shaking (300 rpm). At various timepoints the OD_{600} of each culture was measured.

forming units (CFU). Macrophages were plated at 7×10^5 cells/well in 6 well plates 18 hr prior to infection. Overnight cultures of C231, CPmet, CPsodC and CPfagC were diluted to OD_{600} of 0.1 and grown for 1 hr. Macrophages were incubated with bacteria at a MOI of 4 for 1 hr, then washed and fresh media containing 100 µg/ml gentamycin added to kill extracellular bacteria. At 1, 3, 6 and 9 hr post-infection the macrophages were lysed and then sonicated to release then disrupt bacterial clumps. After serial dilution lysates were plated onto BHI agar plates and incubated at 37°C for 48 hr. The results of two independent experiments are presented in Figure 5.16. CPfagC was not attenuated, demonstrating the same growth curve as C231. Both CPsodC and CPmet demonstrated altered growth curves from the wildtype strain. This effect was most pronounced for CPmet which at 9 hr showed between a 45 and 55% reduction in the number of CFU recovered. CPsodC growth was intermediate showing a 20 to 25% reduction in the number of viable bacteria at 9 hr post infection.

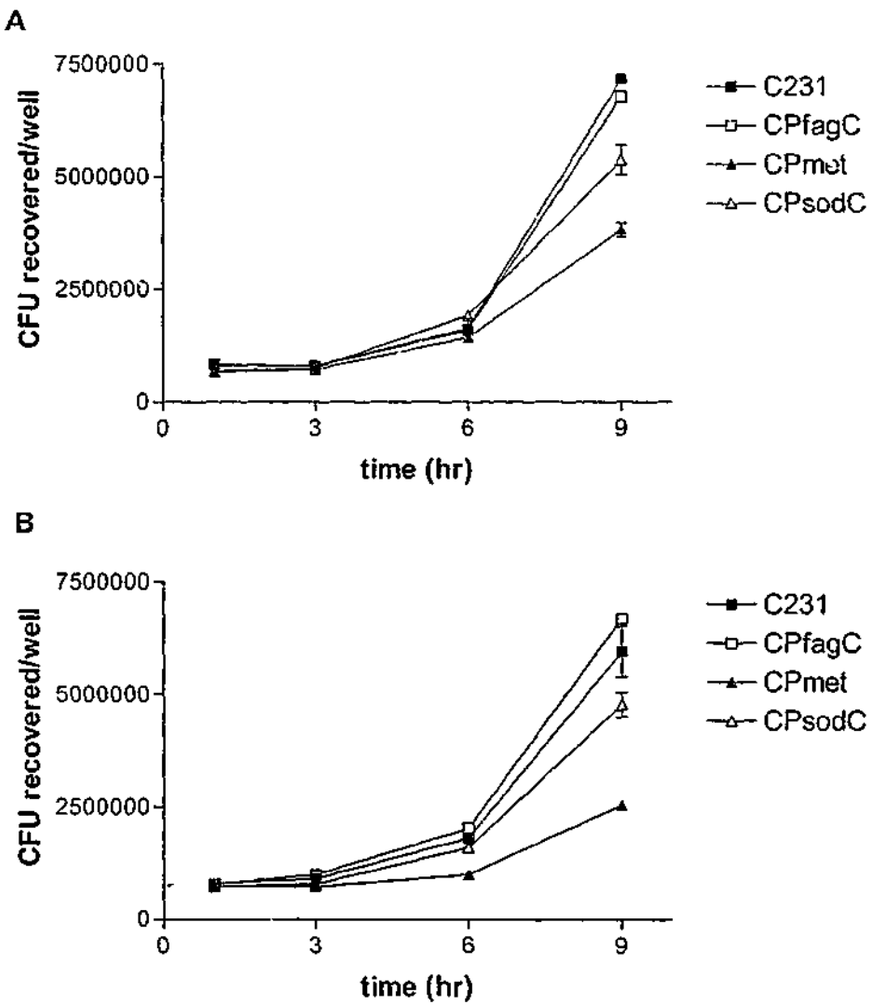

Figure 5.16: Attenuation of *C. pseudotuberculosis* mutant strains in the macrophage infection model.

18 hr prior to infection J774 macrophages were plated in 6 well plates at $7x10^5$ cells/well. Macrophages were infected at MOI 4 with C231 (■), CPfagC (□), CPmet (▲) and CPsodC(Δ) for 1 hr. After this time the cells were washed and gentamycin containing media added to the cells. At various timepoints the macrophage monolayers were lysed and sonicated to release then disrupt bacteria clumps. Lysates were serially diluted then plated on BHI agar. After 48 hr at 37°C the number of colonies derived from each well was determined. A and B represent replica experiments. Results are presented as mean ±S.E.M. of 3 replicates. A one-way ANOVA followed by a Dunnetts multiple comparison test was performed. In both replicates there were significantly fewer bacteria recovered following infection with CPmet ($p<0.01$) and CPsodC ($p<0.05$) than C231 at 9 hr post infection.

5.7 Complementation of CPmet

To determine whether the lack of a functional Met protein was responsible for the altered phenotype of CPmet is was necessary to complement the mutant strain by supplying the *met* gene in trans. Given that *met* appears to be part of an operon, for which not all upstream sequence has been obtained, it was not possible to construct a plasmid in which its expression was driven by its own promoter. Therefore the plasmid pSM22 was utilised. This plasmid was chosen as it contains the *srp* promoter which drives high expression of *gfp* in this plasmid. The *gfp* gene in the plasmid was replaced with the *met* gene.

In pSM22 the *srp* promoter is located on an *Sph*I/*Eco*RI fragment that is 46 bp upstream of the *gfp* gene. Given that there were no unique restriction sites between the *Eco*RI site and the start of the *gfp* gene, the *gfp* gene was removed from the vector as an *Eco*RI/*Spe*I fragment. In order to insert the *met* gene in exactly the same position as the *gfp* gene a forward PCR primer was generated that in addition to containing sequence complementary to the *met* gene contained additional 5' sequence that was the same as that found in pSM22 between the *Eco*RI site and the *gfp* gene. This primer was called primer #276 and was used in conjunction with primer #277 to PCR amplify the *met* gene. Primer #277 contains sequence complementary to the 3' end of *met* in addition to a *Spe*I site for cloning. A PCR product of 1450 bp was generated and then ligated into pSM22 digested with *Eco*RI and *Spe*I to generate pSM56. That the *met* gene had been correctly inserted was confirmed by sequencing across the *srp* promoter and the 5' end of the *met* gene. Given that CPmet is a kanamycin resistant strain it was necessary to insert an additional antibiotic resistance gene into pSM56. The chloramphenicol acetyltransferase gene and its promoter from pACYC184 was PCR amplified using primers #278 and 279. Both primers contain a *Cla*I site at their 5' end for cloning purposes. A PCR product of 1270 bp was generated and ligated into pSM56 that had been digested with *Cla*I and treated with shrimp alkaline phosphatase to generate pSM57 (Figure 5.17). Following transformation into *E. coli* transformants that were chloramphenicol resistant were selected.

Electrocompetent CPmet cells were prepared using the standard protocol outlined in Chapter 2.5.2. Unlike C231 cells that on centrifugation pellet relatively poorly and display a poorly

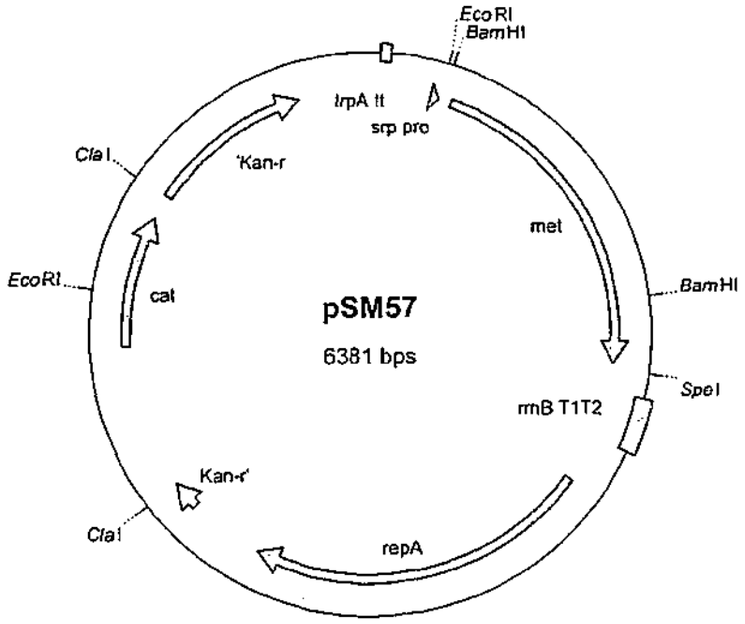

Figure 5.17: Plasmid map of pSM57.

Relevant genes and encoded products are: *met* encodes Met; *repA* encodes RepA; K*an-r* encodes kanamycin resistance gene and cat encodes chloramphenicol acetyltransferase. Expression of *met* is under the control of the *srp* promoter. *rrn*B T1T2 and *trp*A tt are transcription terminators.

compacted pellet, upon centrifugation, the CPmet cells pelleted relatively easily and displayed a pellet more like that usually seen with *E. coli* than *C. pseudotuberculosis*. That the cells were indeed *C. pseudotuberculosis* was confirmed by plating an aliquot of untransformed cells onto BHI agar that did not contain antibiotics. After 48 hr at 37°C typical *C. pseudotuberculosis* colonies of an appropriate size were visible on the agar. pSM57 was transformed into CPmet to generate the complemented strain CPmet+ pSM57.

5.8 Viability and replication of met mutants in the macrophage infection model

An infection experiment analogous to that described in Section 5.6 was performed. Macrophages were infected with C231, CPmet and CPmet+pSM57. The number of CFU/well was determined at 1, 3, 6 and 9 hr post infection (Figure 5.18). In comparison to the growth

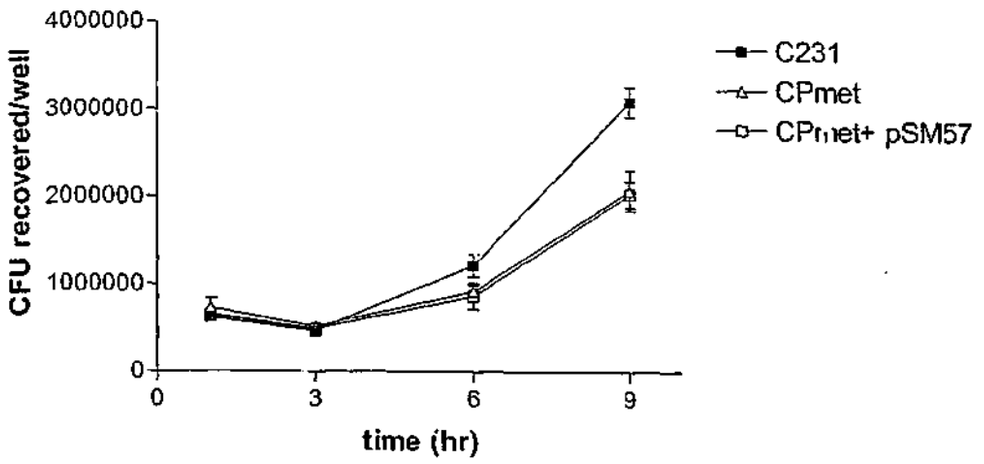

Figure 5.18: Complementation of CPmet with *met* supplied in trans.

18 hr prior to infection J774 macrophages were plated in 6 well plates at $7x10^5$ cells/well. Macrophages were infected at MOI 4 with C231 (■), CPmet (Δ) and CPmet+ pSM57 (□) for 1 hr. After this time the cells were washed and gentamycin containing media added to the cells. At various timepoints the macrophage monolayers were lysed and sonicated to release then disrupt bacteria clumps. Lysates were serially diluted then plated on BHI agar. After 48 hr at 37°C the number of colonies derived from each well was determined. Results are presented as mean ±S.E.M. of 3 replicates. At 1 and 3 hr there was no difference between recovered cell number. At 6 hr there were significantly fewer bacteria recovered following infection with CPmet+pSM57 ($p<0.05$) compared to C231. At 9 hr there were significantly fewer recovered bacteria following infection with either CPmet or CPmet+pSM57 ($p<0.01$) compared to infection with C231.

curve observed for C231, both CPmet and CPmet+pSM57 showed reduced growth as indicated by approximately 35% fewer CFU being isolated at 9 hr post infection. This indicated that the complementation of CPmet with the *met* gene in trans had not restored the wildtype phenotype.

5.9 Discussion

In this chapter the generation of three mutant strains of *C. pseudotuberculosis* was described. When tested in the macrophage infection model two of these mutants showed attenuation. In these mutants, genes coding for a metallopeptidase and a Cu,ZnSOD had been disrupted by

insertion of a kanamycin resistance gene into the coding region. In order to demonstrate that the CPmet phenotype was a result of inactivation of the *met* gene a complemented strain was generated by supplying the *met* gene in trans under the control of a constitutive promoter. This strain however demonstrated the same phenotype as CPmet indicating that inactivation of the *met* gene may not have been the cause of the observed phenotypic change.

In the macrophage infection experiments attenuation was observed as a reduced growth rate at the later timepoints, particularly in the samples assayed 9 hr post infection. This indicates that the mutated genes have no effect on the degree of phagocytosis of *C. pseudotuberculosis* by the macrophages. Additionally we observed that both the wild type and mutant strains exhibited very little growth in the first three hr. This may reflect that an initial adaptation to the intracellular environment occurs prior to replication or alternatively an overestimation of bacterial number at one hr post infection may have occurred as at this time extracellular bacteria had not been exposed to gentamycin.

The CPfagC mutant strain was not attenuated in this model. This indicates that the iron transport system of which FagC is a component, is not essential for obtaining iron within the intracellular environment of the macrophage. The DNA array experiments described in Chapter 3 identified components of other iron homeostasis systems including a heme regulator and a second iron permease component. The presence of these genes within the *C. pseudotuberculosis* genome indicates that this pathogen contains multiple systems for the scavenging of iron. It is therefore likely that the disruption of the FagABC permease was not sufficient to cause an iron deficiency. Rather it is most probable that the other systems were able to compensate such that an appropriate amount of iron for normal cellular function could be scavenged. Additionally given that the macrophages are cultured in an iron-rich tissue media, it may be that even within the macrophage that there is a reasonably abundant supply of iron relative to what would be found in an animal. The recent study identifying the *fag* genes of *C. pseudotuberculosis* (Billington *et al.*, 2002) included an analysis of a mutant strain in which the *fagB* gene (upstream of *fagC*) was insertionally inactivated. Using a goat strain of *C. pseudotuberculosis* they were able to show that a *fagB* mutant termed JGS283 had normal growth characteristics *in vitro* and was able to utilise iron from a variety of different sources including $FeCl_2$, $FeSO_4$, hemin, haemoglobin and transferrin. However, during animal infection JGS283 was significantly attenuated. Both the mutant and wildtype strains caused a

similar amount of abscessation at the site of infection. but only one of four goats infected with JGS283 developed abscesses in the lymph nodes and no bacteria could be isolated from goats infected with the mutant strain. The authors suggest that that the mutant has a decreased ability to survive once in the body, an observation that is consistent with an iron scavenging defect as the host can be considered to be a low iron environment. However the results need to be treated with caution as no complementation was demonstrated.

The mutant strain CPsodC was slightly attenuated at 9 hr post infection, such that the number of recovered bacteria was reduced by 20 to 25%. A complemented strain was not generated so this effect cannot be unequivocally assigned to the lack of a functional Cu,ZnSOD. Analysis of the region surrounding *sodC* on the *C. pseudotuberculosis* chromosome suggests however that it may be transcribed as a monocistronic unit. Upstream of *sodC* is a divergently transcribed gene that codes for an oxidoreductase. Downstream of *sodC* is a 180 bp intergenic region prior to the *accD* gene. It therefore seems unlikely that the observed effect would be polar in nature.

Analysis of the *C. pseudotuberculosis sodC* sequence failed to identify a prokaryotic membrane lipoprotein attachment site. To date the sequences of all other identified gram positive Cu,ZnSODs have shown the presence of this motif (D'Orazio *et al.*, 2001). In order to confirm that the sequence from this region was accurate a number of independent genomic clones were sequenced in both directions. In each case the same sequence was generated thus confirming that the motif was absent in the *C. pseudotuberculosis* Cu,ZnSOD. Whether this motif is absent in other strains of *C. pseudotuberculosis* remains to be determined. Given that a single base substitution in the *C. pseudotuberculosis sodC* sequence would be sufficient to convert a serine residue to a cysteine residue, thus creating a prokaryotic membrane lipoprotein attachment, suggests that this may be a relatively recent mutation.

The majority of gram negative Cu,ZnSODs have classical signal peptidase I cleavage sites which directs the protein to the periplasm, while as just detailed gram positive Cu,ZnSODs have sequences recognised by signal peptidase II that results in them becoming membrane attached lipoproteins. Using the SignalPV1.1 prediction server a putative signal peptide was identified at the N-terminal of *C. pseudotuberculosis* Cu,ZnSOD. This indicates that the protein is probably secreted from the cell and as such its substrates are likely to be

extracellular in nature. Thus like other Cu,ZnSODs the *C. pseudotuberculosis* protein is predicted to have a role in protecting the bacteria from host-generated superoxide.

Given the probable extracellular location of the *C. pseudotuberculosis* Cu,ZnSOD, it is likely that the attenuation seen for CPsodC is due to a decreased resistance to superoxide generated by the macrophages during a respiratory burst in response to the infection. All *sodC* mutants that have been generated and tested to date have shown reduced survival *in vivo* in the presence of extracellular superoxide (Farrant *et al.*, 1997;San-Mateo *et al.*, 1998;Wilks *et al.*, 1998;Sheehan *et al.*, 2000;Dussurget *et al.*, 2001;Piddington *et al.*, 2001). In order to further characterise CPsodC similar studies could be performed. The degree of attenuation observed for CPsodC is relatively small and is only seen at the last timepoint analysed. That attenuation is not seen earlier may reflect the amount of time required for the respiratory burst to become effective. In order to increase the magnitude of the respiratory burst that *C. pseudotuberculosis* is exposed to, the macrophages could be pre-treated with an agent that would activate them such as IFNγ. It would be expected that the *sodC* mutant strain would be further attenuated in such an environment and possibly at earlier timepoints post infection.

Analysis of studies with *sodC* mutants from a broad variety of pathogenic bacteria indicates that there is not always a correlation between sensitivity to superoxide *in vitro* and attenuation *in vivo* or in cell culture infection models (Sheehan *et al.*, 2000;Dussurget *et al.*, 2001). Additionally a mutant may be attenuated in a tissue culture model but not in an animal model (Farrant *et al.*, 1997;Piddington *et al.*, 2001) or attenuated in an animal model but not the natural host (Bong *et al.*, 2002). In order to determine whether the *C. pseudotuberculosis* Cu,ZnSOD is a virulence factor for this pathogen infection experiments using the native host should be performed with CPsodC and a complemented strain.

Many bacteria possess more than one type of SOD and as such it is likely that the enzymes have overlapping functions. This is supported by the observation that a number of *sodC* mutants are not attenuated *in vivo* (Sheehan *et al.*, 2000;Dussurget *et al.*, 2001). Even though the Fe and Mn SODs have an intracellular location they have in some cases been shown to enhance bacterial survival *in vivo*. For example, a *sodB* (coding for FeSOD) mutant was attenuated in a *Brucella pertussis* mouse respiratory infection model (Khelef *et al.*, 1996), while a *sodA* (coding for the MnSOD) mutant of *Y. enterocolitica* is attenuated in a mouse

model when delivered by an intravenous route (Roggenkamp *et al.*, 1997). It would be interesting to determine whether *C. pseudotuberculosis* contained *sodA* and *sodB* homologues and whether these had a synergistic effect with *sodC* in protecting the organism from free radicals produced during the respiratory burst.

The metallopeptidase identified in this study contained two motifs that identify it as being a zinc metallopeptidase. The first of these motifs is common to all zinc metallopeptidases while the second is found in zinc metallopeptidases that belong to the aminopeptidase family. Core residues within both these motifs act as the zinc ligand. Aminopeptidases are exopeptidases that catalyse the cleavage of the N-terminal amino acid from a peptide chain. To determine whether the target of Met was of intracellular or extracellular origin searches were performed to identify a signal peptide sequence at the N-terminal of the protein. No such sequence could be found, suggesting that the substrates of Met are located within the cytoplasm. Aminopeptidases may be involved in a number of different physiological functions including endogenous protein turnover, the maturation of proteins and the degradation of exogenous or abnormal proteins (Gonzales and Robert-Baudouy, 1996). During cell growth a dynamic equilibrium exists between the synthesis of new proteins and the catabolism of existing proteins. Within a cell the rate of protein turnover can vary significantly depending on the physiological status of the cells, for example cells in stationary phase degrade proteins at a several times greater rate than those in exponential phase (Gonzales and Robert-Baudouy, 1996). Modification to the rate of protein turnover may also occur in response to factors such as nutrient limitation. For example to produce new proteins during conditions of nutrient limitation it is necessary for normally stable proteins to be degraded. Additionaliy the degradation of proteins may occur to increase the amount of available carbon. Aminopeptidases may also allow a bacterium to use exogenous peptides as nutrients. This is particularly important for bacteria that are auxotrophic for some amino acids. *Lactococcus lactis,* for example, is dependent on proteases to appropriately degrade exogenously derived proteins to supply certain amino acids. This is demonstrated by the observation that certain peptidase mutants demonstrate a reduced growth rate in milk (Mierau *et al.*, 1996). This is presumably due to their inability to completely degrade proteins that are present within the milk. It is not known whether *C. pseudotuberculosis* is auxotrophic for any amino acids and therefore whether its proteases and peptidases play a similar role.

Like many proteins involved in nutrient scavenging the expression of some aminopeptidases has been shown to be metabolite regulated. In *L. lactis* expression of a number of peptidases is down-regulated in the presence of a peptide source (Guedon *et al.*, 2001), while a single aminopeptidase *pepP* was regulated by catabolite repression. *E. coli pepN* expression is increased under conditions of phosphate limitation, anaerobiosis and changes in carbon source from glucose to glycerol or succinate (Gharbi *et al.*, 1985). The up-regulation of *C. pseudotuberculosis met* following macrophage infection may reflect a change in the nutrient availability to the cells. Such an environmental change may lead to alterations in the rate of protein turnover or an increase in the requirement for protein scavenging from the environment.

Analysis of the sequence up and downstream of *met* suggests that it may be part of an operon. Immediately upstream was a second peptidase that is likely to be a prolyl oligopeptidase family member. Members of this family are serine peptidases, the substrates of which are peptides of less than 30 amino acids (Polgar, 2002). Downstream of *met* were two genes that code for enzymes of the rhamnose biosynthetic pathway (Figure 5.19). Rhamnose is a 6-deoxyhexose that occurs naturally in several complex carbohydrates. In gram-negative bacteria rhamnose is often a component of the lipopolysaccharide, whereas in gram-positive bacteria, such as mycobacteria, rhamnose is an important component of the cell wall. Immediately downstream of *met* is *rmlB* which codes for a protein with dTDP-D-glucose 4,6-dehydratase activity. This was followed by a second gene, *rmlCD,* which codes for a protein with dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase and dTDP-4-L-rhamnose reductase activity. The gene order in the vicinity of the *met* homologue in *C. glutamicum* and *C. diphtheriae* was compared to that observed in *C. pseudotuberculosis*. The arrangement of a prolyl oligopeptidase upstream of an aminopeptidase was conserved in all three species. The gene order downstream of the *met* gene was also conserved between *C. pseudotuberculosis* and *C. diphtheriae* such that the *rmlB*, *rmlC* and *rmlD* genes proceeded it. Given the close proximity of the genes in this region to one another and that they are transcribed in the same direction, it is quite likely that they comprise an operon. In each case the stop codon of one gene either overlaps or is immediately followed by the start codon of the next gene. The observation that the gene order resembles that found in *C. diphtheriae* is consistent with *C. pseudotuberculosis* being more closely related to this organism than C. *glutamicum* (Takahashi *et al.*, 1997). Analysis of the *C. glutamicum* genome indicated that the *rmlB* and

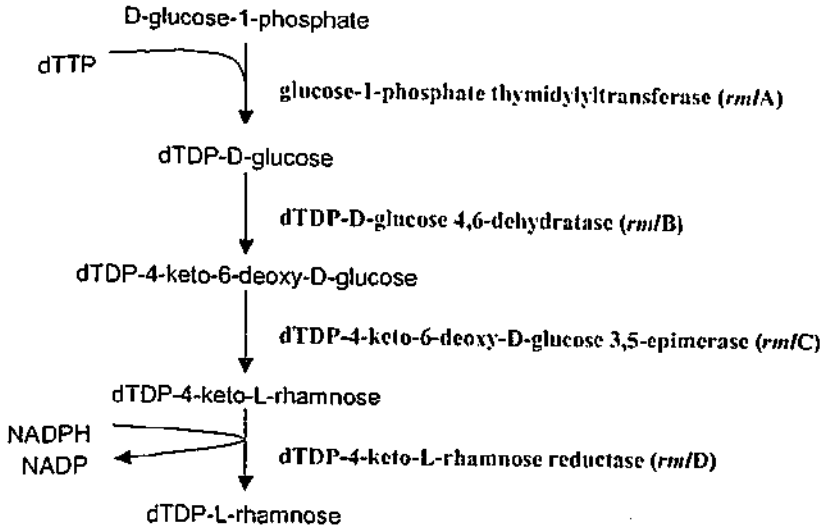

Figure 5.19: The biosynthetic pathway for the formation of dTDP-L-rhamnose.

The gene coding for each of the enzymes in the pathway is shown in brackets.

rmlCD genes were located within the vicinity of the *met* homologue but were unlikely to comprise an operon.

Although CPmet was significantly attenuated in the macrophage infection model the phenotype could not be restored by supplying the *met* gene in trans. This suggests that either *met* is not being expressed appropriately from pSM57 within the complemented strain, the mutation is polar such that downstream genes are not appropriately transcribed, the mutation is trans dominant or alternatively an unrelated spontaneous mutation has occurred. *met* expression from pSM57 was not measured, however, the gene was inserted into the same site as the *gfp* gene in pSM22. Given that *gfp* expression from this promoter was constitutive and strong it would be expected that *met* would be expressed in a similar manner. That a spontaneous mutation has occurred in the strain could be investigated by looking at the other *met* mutant strains that were generated in this study. If these all show the altered phenotype then it is unlikely that it is a result of an unrelated mutation.

CPmet displayed normal growth characteristics *in vitro* both in liquid cultures and on BHI agar plates. The only readily apparent phenotypic difference between CPmet and the wildtype strain was observed when preparing CPmet electrocompetent cells. In particular the pelleting properties of the *met* mutant were considerably different to that observed for the wildtype strain. Instead of a relatively unstable pellet being generated, a compact well-formed pellet was generated following centrifugation of a CPmet culture. This observation suggested that a phenotypic change at the cell surface might have occurred, thus modulating the surface properties of the bacteria and the way in which they can interact with each other.

Although the cell wall structure of corynebacteria has not been studied in any detail it is likely to be similar to that observed in other closely related bacteria such as mycobacteria and rhodococcii. These groups show similar cell wall structures that at a simplistic level consist of three layers (Sutcliffe, 2000;Ma *et al.*, 2001). The outermost layer consists of mycolic acids. These are esterified to arabinogalactan, the middle component. The innermost layer consists of peptidoglycan and is attached to the arabinogalactan via a linker disaccharide of α-L-rhamnosyl (1→3)α-D-N-acetyl glucosaminosyl-1-phosphate. Thus it is likely that rhamnose is also a minor but key component of the corynebacterial cell wall.

In general, bacteria contain only one rhamnose biosynthetic pathway. It is therefore very likely that the rhamnose biosynthetic genes identified during this study are responsible for the synthesis of the rhamnose moiety that is incorporated into the cell wall. When taken together, the data showing that the CPmet phenotype is not restored to wildtype upon complementation with *met* and that an observable phenotypic change in regard to cell-cell interactions has occurred, suggests that a polar effect has occurred such that the enzymes of the rhamnose biosynthetic pathway are not appropriately expressed. The insertion of the kanamycin resistance gene into the *met* sequence should not have resulted in a polar mutation as there is no transcriptional terminator 3' to the kanamycin gene. However, reports of the insertion of extraneous sequence such as this leading to alteration in the transcription of the remainder of an operon have been made (Gulig *et al.*, 1992;Baker *et al.*, 1998;Fischer *et al.*, 2001).

It has been suggested that strains of mycobacteria without a functional rhamnose biosynthetic pathway would not be viable due to an inability of the arabinogalactan and mycolic acid layers to attach to the inner peptidoglycan (McNeil *et al.*, 1990). This is supported by the

observation that in *M. tuberculosis* the enzyme rhamnosyl transferase is required for bacterial growth (Ma *et al.*, 2001). This enzyme is involved in the synthesis of the linker between arabinogalactan and peptidoglycan that contains rhamnose. Assuming a similar situation in *C. pseudotuberculosis* this would suggest that the expression of the *rml* genes has been reduced but not ablated. This could be determined by measuring the level of expression of these genes by Northern analysis in the wildtype and mutant strains. Northern analysis would also indicate as to whether a full-length transcript was being produced in the mutant strain. Assuming that the *met* gene supplied in trans is being expressed in the complemented strain then the simplest way to determine whether the lack of a complementable phenotype was due to a polar effect would be to complement CPmet+ pSM57 with the downstream rhamnose biosynthetic genes. Reversion to wildtype phenotype would be indicative of the *met* mutation having a polar effect on downstream gene expression, while non-reversion would be indicative of a trans dominant mutation. In order to avoid a polar effect on downstream gene expression several different approaches could be used to generate a *met* mutant. In particular either a gene replacement approach in which the *met* gene is entirely replaced by the selectable gene or the generation of a site specific mutation may be more suitable.

CPmet was attenuated in the macrophage infection model but demonstrated wildtype growth characteristics when grown in BHI media. It is not entirely clearly which gene or genes are responsible for the change in phenotype thus making it difficult to postulate on the reasons for the phenotypic change. If the change in phenotype is as a result of inactivation of the *met* gene then the changes in growth rate may reflect differences in nutrient availability within the two environments. In comparison to BHI the intracellular environment of the macrophage is likely to be relatively deficient in readily available nutrients. The reduced ability of a peptidase mutant to degrade peptides may lead to a slowed growth rate as a result of not being able to sufficiently quickly generate amino acids for protein synthesis or as a carbon source. Alternatively, a deficiency of rhamnose within the bacteria may have important implications for cell wall structure. Alterations in cell wall structure may influence factors such as: permeability, the ability of other proteins to integrate into the cell wall, cell motility and the functionality of protein secretion systems (Barry, 2001). That factors such as permeability may affect bacterial growth rate under certain conditions is supported by the observation that a *M. avium* variant with a rough phenotype is more virulent *in vivo* and multiplies more quickly than a smooth variant with which it shows similar growth characteristics *in vitro* (Kansal *et*

al., 1998). It has been suggested that the altered phenotype may be as a result of an alteration in cell wall structure such that the permeability of the smooth strain is greater than the rough strain (Kansal *et al.*, 1998). Alteration in cell wall integrity may influence bacterial susceptibility to toxic macromolecules and the ability to inactive reactive oxygen and nitrogen species (Daffe and Draper, 1998). Additionally the cell wall of wildtype bacteria is likely to be relatively resistant to degradation by host enzymes, a situation that may be altered in a rhamnose deficient mutant. The ability of *C. pseudotuberculosis* to clump has been suggested to contribute to virulence, a property that is thought to be attributable to lipid which occurs external to the cell wall (Batey, 1986b). It might be envisioned that in the presence of a disrupted cell wall there would be a disruption to this layer that may alter the clumping properties of *C. pseudotuberculosis*. Whether this would have implications for intracellular replication of *C. pseudotuberculosis* is not clear.

The aim of the work described in this chapter was to determine the functional significance of regulated genes of *C. pseudotuberculosis* during macrophage infection. While we were able to demonstrate that *sodC* and *met* mutants were attenuated in the macrophage infection model, it is not entirely clear which genes are responsible for the change in phenotype. The *fagC* mutant was not attenuated, however, it has recently been shown that disruption to the *fag* operon is sufficient to decrease the virulence of *C. pseudotuberculosis* in goats (Billington *et al.*, 2002). This indicates that there is not always a good correlation between what occurs in a model that mimics some aspects of an infection process and what occurs in the natural host. Given this, work aimed at determining the significance of *sodC*, *met* and genes downstream of *met* during infection of an animal should be performed.

Chapter 6

Regulation of phospholipase D expression

6.1 Introduction

In Chapter 3 the *C. pseudotuberculosis* phospholipase gene was identified as being thermoregulated, or more specifically its expression was down regulated following heat shock from 37°C to 43°C. The concept that expression of virulence determinants is regulated is not novel, however this was the first report of a regulated virulence gene of *C. pseudotuberculosis*. Given that Pld is undoubtedly essential for full virulence of *C. pseudotuberculosis* (Hodgson *et al.*, 1992;McNamara *et al.*, 1994;Hodgson *et al.*, 1994;Simmons *et al.*, 1998;Hodgson *et al.*, 1999), a study of its regulation was deemed worthwhile. Likely outcomes from such a study include an increased understanding of how virulence genes are regulated and further insights into *C. pseudotuberculosis* pathogenesis. In particular little is known about *pld* expression *in vivo*, therefore understanding how it is regulated by environmental factors may help to understand its *in vivo* mode of action.

Initial work was aimed at confirming and further characterising the heat shock response by northern analysis. To allow further regulation studies to be more easily performed the *pld* promoter was inserted upstream of the *gfp* gene in the reporter construct pSM20. Studies with this construct indicated that *pld* is also regulated in a cell density dependent manner. Previous studies have shown that under certain conditions Pld is toxic to erythrocytes and neutrophils (Hsu *et al.*, 1985;Yozwiak and Songer, 1993). To determine whether such findings are of biological significance, studies were performed to determine whether *C. pseudotuberculosis* residing in macrophages expressed *pld*. Having shown that *pld* was expressed by intracellular *C. pseudotuberculosis*, studies aimed at determining the effect of Pld on macrophage viability and function were performed.

The results presented in this chapter indicate that the expression of *pld* is highly regulated. Analysis of the *pld* promoter and upstream sequence was subsequently performed in order to elucidate the mechanisms behind the observed regulation.

6.2 Timecourse of regulation of *pld* expression by heat shock

Northern analysis was performed to determine the timecourse of *pld* mRNA down-regulation following heat shock. An overnight culture of C231 was diluted to an OD_{600} of 0.3 then grown until an OD_{600} of 0.65 was reached. 6 ml aliquots of this culture were prepared, half of

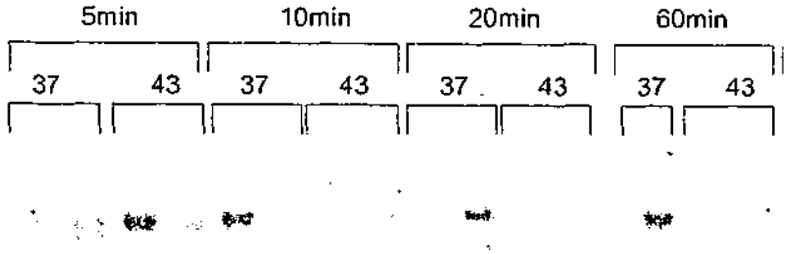

Figure 6.1: Timecourse of *pld* regulation following heat shock.

RNA was extracted from *C. pseudotuberculosis* control (37°C) and heat shocked samples (37°C to 43°C) at 5, 10, 20 and 60 min post initiation of the shock. 8 µg of total RNA was analysed by Northern analysis using a DIG labelled *pld* PCR product. One of the 37°C, 60 min replicates was lost during sample processing.

which were further incubated at 37°C while the remainder were incubated at 43°C. At 5, 10, 20 and 60 min post initiation of the shock two aliquots from each temperature were taken and RNA extracted. 8 µg RNA was run under denaturing conditions on a 1.2% (w/v) formaldehyde agarose gel and then transferred to Hybond N+ membrane in 20x SSC. The membrane was then probed with a DIG labelled *pld* PCR product (generated using primers #55 and #77). Following detection of bound probe with an anti DIG-alkaline phosphatase conjugate and the AP substrate CSPD, the membrane was exposed to film (Figure 6.1). *pld* mRNA was detected in all 37°C samples, however, following heat shock at 43°C, down-regulation of *pld* expression was apparent by 10 min as indicated by decreased band intensity. The level of *pld* mRNA continued to decrease past this time point to a minimal level detected at 20 and 60 min.

6.3 Construction of reporter construct containing the *pld* promoter upstream of the *gfp* gene

Often virulence genes are regulated by more than one environmental factor, for example temperature, osmolarity and iron availability. In order to develop a system whereby it was possible to readily assess *pld* expression under a variety of environmental shocks, reporter plasmids were constructed by inserting the *pld* promoter into the promoter probe vector

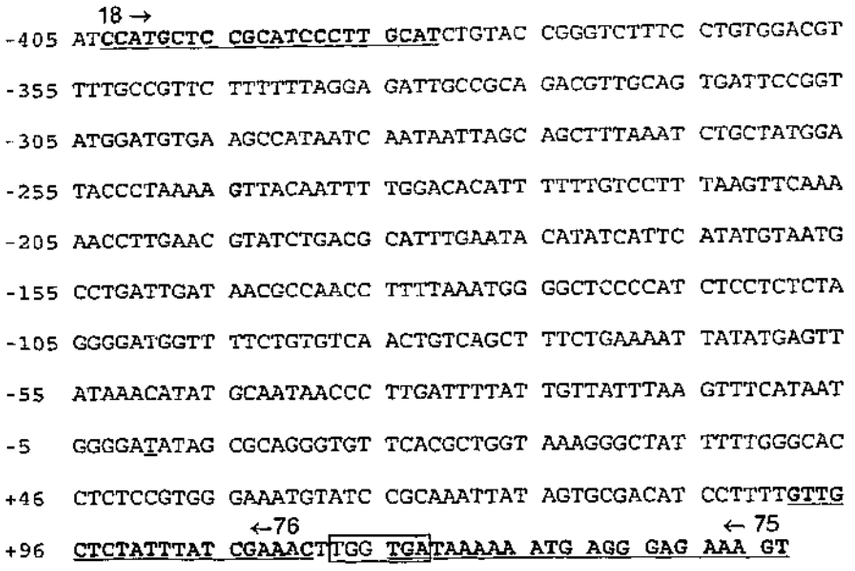

Figure 6.2: Sequence of the *pld* promoter region showing primers used for reporter construction.

The predicted –35 and –10 regions are doubly underlined. Assignment is based on the homology of these sequences to *E. coli* σ^{70} –35 and –10 promoter consensus sequences. Downstream is the transcriptional start point (underlined T) which was previously identified using primer extension assays (Hodgson *et al.*, 1990). Further downstream is a putative RBS (boxed TGGTGA) which lies 6 bases upstream of the start codon. Sequence for the first four amino acids is shown and the corresponding sequence is shown using the amino acid single letter code. Primers used to PCR amplify the *pld* promoter for cloning into pSM20 are bolded and underlined (18, 75 and 76). Primers 75 and 76 have additional sequence at their 5' end corresponding to a *Bam*HI site for cloning.

pSM20 (Chapter 4). Two constructs were generated which differ on the basis of whether the *pld* RBS and initiating codon were present (Figure 6.2). The *pld* promoter was PCR amplified from genomic C231 DNA using primer pair combinations #18 and #75, and #18 and #76. Primer #76 was used to generate a *pld* promoter fragment up to but not including the RBS while primer #75 gave rise to a PCR product that included the *pld* RBS and first 4 amino acids of the *pld* gene. A restriction site (*Bam*HI) was engineered into primers #75 and #76 to allow cloning into pSM20. PCR products 18-75 and 18-76 were digested with *Bam*HI and ligated into pSM20 cut with *Eco*RV and *Bam*HI to generate pSM27 and pSM28 respectively. pSM28

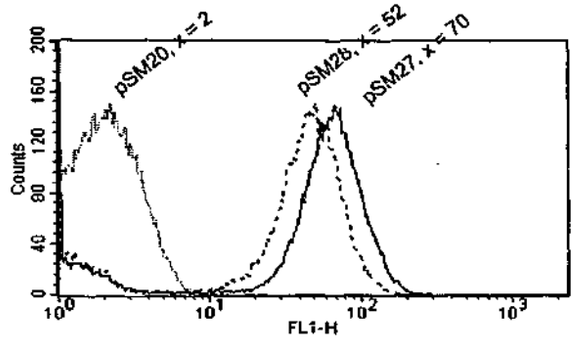

Figure 6.3: Flow cytometry analysis of *gfp* expression from *C. pseudotuberculosis* harbouring *pld* promoter constructs pSM27 and pSM28 or the promoterless construct pSM20.

For each clone 30,000 events were counted. Fluorescence intensity, x, is a measure of the mean fluorescence for each bacterial clone. Data was gated on forward and side scatter such that plotted data represents primarily single bacteria.

contains the *pld* promoter up until the RBS, hence the cells will use the RBS present downstream in the vector backbone. pSM27 is a translational fusion and would be expected to use the *pld* promoter RBS. In pSM27 the sequence coding for the first 4 amino acids of Pld is in frame with the *gfp* gene such that a fusion protein between the Pld amino acids, intermediate sequence and Gfp will be generated upon translation.

The constructs were transformed into C231, and Gfp fluorescence measured by flow cytometry. Overnight cultures of C231(pSM27) and C231(pSM28) were diluted to OD_{600} of 0.1, then grown for a further two hr prior to flow cytometry. Both strains gave rise to detectable fluorescence. C231(pSM28) was slightly less fluorescent (mean fluorescence, x= 52) compared to that observed from C231 containing the fusion construct (pSM27, mean fluorescence, x=72) (Figure 6.3). It is unclear why the expression from the fusion construct is greater, although it may reflect different translation efficiencies from the two RBSs.

6.4 Assessment of C231(pSM27) in a heat shock model

C231 (pSM27) was chosen for further studies as it was the more fluorescent construct and contained the entire region surrounding the *pld* promoter (i.e. including the RBS). To demonstrate that this vector was a suitable reporter for monitoring *pld* expression, attempts were made to replicate the observations made during the array and northern experiments regarding thermoregulation of *pld*. A similar experiment to that described for the northern experiment was performed but at a lower cell density. A lower cell density was chosen as analysis of *C. pseudotuberculosis* by flow cytometry appears to be most efficient for low density cultures as a result of less bacterial clumping. An overnight culture of C231(pSM27) was diluted to an OD_{600} of 0.1 and then grown at 37°C for one hr. Bacteria were then either incubated further at 37°C or transferred to 43°C. *pld* promoter activity was assessed by measuring Gfp fluorescence by flow cytometry at 0, 0.5, 1, 2 and 3.5 hr post initiation of the heat shock (Figure 6.4). Using this system a small decrease in *pld* promoter activity at 43°C was observed. However this was only apparent at the latter time points of 2 and 3.5 hr. Interestingly, in samples incubated at 37°C a decrease in Gfp fluorescence at these timepoints was also observed however the magnitude of the decrease was substantially larger than that observed at 43°C. This apparent density-related expression of *pld* was further investigated in Section 6.6. The apparent lack of effect of heat shock on *pld* promoter activity as measured by detectable fluorescence could possibly be explained by two factors. Firstly, by reduced cell growth at 43°C and secondly, the long half-life of the reporter protein. Gfp has a half-life of at least 24 hr (Andersen *et al.*, 1998), thus the way changes in *gfp* expression are monitored in this system are through its dissipation when cell growth is greater than the rate of expression, or its accumulation if the rate of *gfp* expression exceeds cell replication. At 43°C *C. pseudotuberculosis* demonstrates a reduced growth rate, therefore Gfp accumulates within the bacteria. In an attempt to solve the problem of working with a reporter gene with an exceptionally long half-life, a construct with a destabilised *gfp* gene was constructed.

6.5 Construction of a reporter construct containing a destabilised *gfp* gene

As indicated from the results obtained in Section 6.4 there are inherent problems with working with a reporter protein that has a long half-life. These stem from the fact that the long half-life of Gfp makes it extremely difficult to measure transient gene expression. To overcome this obstacle new mutant forms of Gfp which have reduced half-lives have been generated

relative Gfp fluorescence

Figure

Culture
incubat
time po
sample
to give

(Ande
Those
have b
and se
an add
rapid d
peptide
destruc
transcr
protein
switchi
(ANDE
recogn

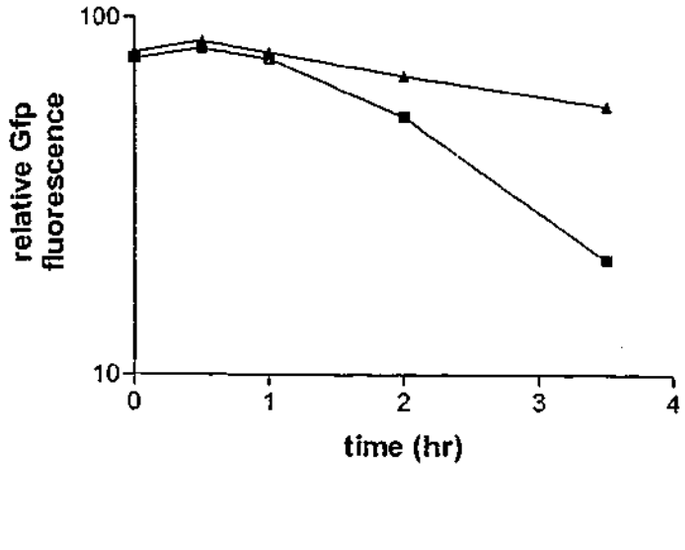

Figure 6.4: Effect of heat on *pld* promoter activity in *C. pseudotuberculosis*.

Cultures of C231(pSM27) were grown at 37°C from an OD_{600} of 0.1 for 1 hr. One culture was incubated further at 37°C (■) while the second was transferred to 43°C (▲) (t= 0). At various time points Gfp fluorescence of the cultures was measured by flow cytometry. For each sample 30,000 events were counted and the mean fluorescence of the population determined to give relative Gfp fluorescence.

(Andersen *et al.*, 1998;Li *et al.*, 1998;Deichsel *et al.*, 1999;Corish and Tyler-Smith, 1999). Those generated by Andersen *et al* (1999) may be useful in *C. pseudotuberculosis* as they have been generated for use in bacteria and demonstrate reduced half lives of between 40 min and several hr compared to over 24 hr for wildtype Gfp. The destabilised Gfp proteins have an additional 10 amino acids at their carboxyl terminal, which renders them susceptible to rapid degradation by carboxyl terminal-specific cellular proteases (Keiler *et al.*, 1996). The peptide tag is derived from a naturally occurring system in *E. coli*, which exists for the destruction of prematurely terminated polypeptides. In *E. coli ssrA* codes for an RNA transcript that possesses some tRNA like properties. This transcript is able to recognise proteins that are translated from incomplete or damaged mRNAs and by cotranslation switching from the nascent polypeptide chain to the *ssrA* transcript, the peptide tag (ANDENYALAA) is added to the carboxyl terminal of the polypeptide chain. The tag is recognised by cellular proteases, leading to rapid degradation of the protein. By altering the

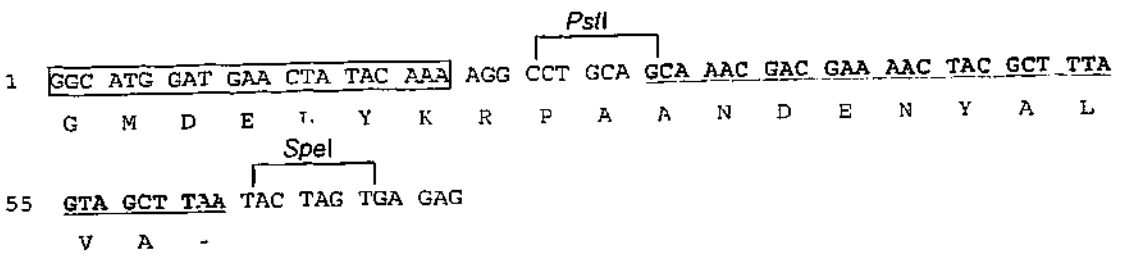

Figure 6.5: Sequence complementary to primer #95 used in the construction of a destabilised *gfp* gene.

The sequence corresponding to the 3' end of the *gfp* gene is boxed and the sequence coding for the 10 amino acid peptide tag is in bold and underlined. Corresponding amino acids are shown using the single letter code. A *Spe*I site used for cloning is marked.

last three amino acids of the peptide tag, protein stability can be further modulated. *ssrA* homologues have been identified in a number of gram negative and gram positive bacteria raising the possibility that corynebacteria have such a system.

In this study a destabilised Gfp with the tag ANDENYALVA was generated. In *E. coli* this tag has been shown to reduce the half- life of a FACS optimised Gfp mutant (gfpmut3) (Cormack *et al.*, 1996) to 40 min and in *Pseudomonas putida* to approximately 60 min (Andersen *et al.*, 1998). To generate a destabilised *gfp* gene a reverse PCR primer was designed in which sequence for the 3' end of the *gfp* gene, the additional destabilising amino acids, and an *Spe*I site for cloning were included (primer #95, Figure 6.5). The *gfp* gene was then PCR amplified from pSM20 using the new primer and primer #47 (Table 2.4, hybridises to sequence 5' of the *gfp* gene and contains *Eco*RV and *Bam*HI sites for cloning at its 5' end). The resulting PCR product was digested with *Bam*HI and *Spe*I and ligated into pSM20, pSM22 and pSM27, that had been prepared in the same manner, to generate pSM29, pSM30 and pSM31 respectively (promoterless, *srp* promoter and *pld* promoter respectively). DNA sequencing confirmed that the sequence for the tag had been added to the 3' end of the *gfp* gene. In *E. coli* Gfp fluorescence from pSM30 and pSM31 was less than that observed from the parental plasmids (pSM22 and pSM27) as determined by viewing bacterial colonies on an agar plate under UV light. The new plasmids were transformed into C231. Analysis of colonies on plates indicated that they did not fluoresce. This was confirmed by flow cytometry which showed that C231(pSM30) and C231(pSM31) were non-fluorescent. The

conclusion from this observation is that the peptide tag with the LVA modification is not suitable for *C. pseudotuberculosis*.

6.6 Density dependent regulation of *pld*

To determine whether *pld* expression is related to cell density the experiment performed in Section 6.4 was repeated at 37°C with the inclusion of a second strain of C231 containing pSM22 which expresses *gfp* under the control of the constitutive *srp* promoter. Overnight cultures were diluted to an $O.D._{600}$ of 0.1 and grown at 37°C for 8 hr. Fluorescence of the cultures was measured by flow cytometry every hour and optical density also determined. The two strains exhibited the same growth curve (Figure 6.6A) indicating that they are of the same density and that growth was not affected by the amount of Gfp in the cell. C231(pSM27) showed decreased fluorescence to an OD_{600} of 0.8 which was indicative of low expression from the *pld* promoter (Figure 6.6B). This was followed by a rapid increase in fluorescence indicating increased promoter activity. On the other hand, Gfp expression from C231(pSM22) was essentially constant until the culture density reached OD_{600} of 2. After this timepoint an increase in Gfp fluorescence was observed. This may correspond to a slowing in cell growth leading to Gfp accumulation. These results indicate that density dependent expression of *pld* is indeed occurring.

The density dependent expression of *pld* was further confirmed at the mRNA level by northern analysis. Cultures of C231 were subcultured to OD_{600} of 0.1 then grown at 37°C. At 30 min, 3 hr and 6 hr duplicate samples (corresponding to average OD_{600} of 0.16, 0.47 and 1.5 respectively) were harvested and RNA extracted. Northern analysis was performed as described in Section 6.2. The northern data confirmed that *pld* expression increases with cell density (Figure 6.7).

6.7 Is *pld* expression regulated by a quorum sensing mechanism?

It is now well established that bacterial pathogens can synchronise their behaviour based on population density. The mechanisms by which bacteria control gene expression in response to cell density have been given the term Quorum sensing. Gram-positive and gram-negative bacteria use different mechanisms for quorum sensing. Gram-negative bacteria use the well studied LuxI/LuxR signalling system (Bassler, 1999). Quorum sensing systems are not as

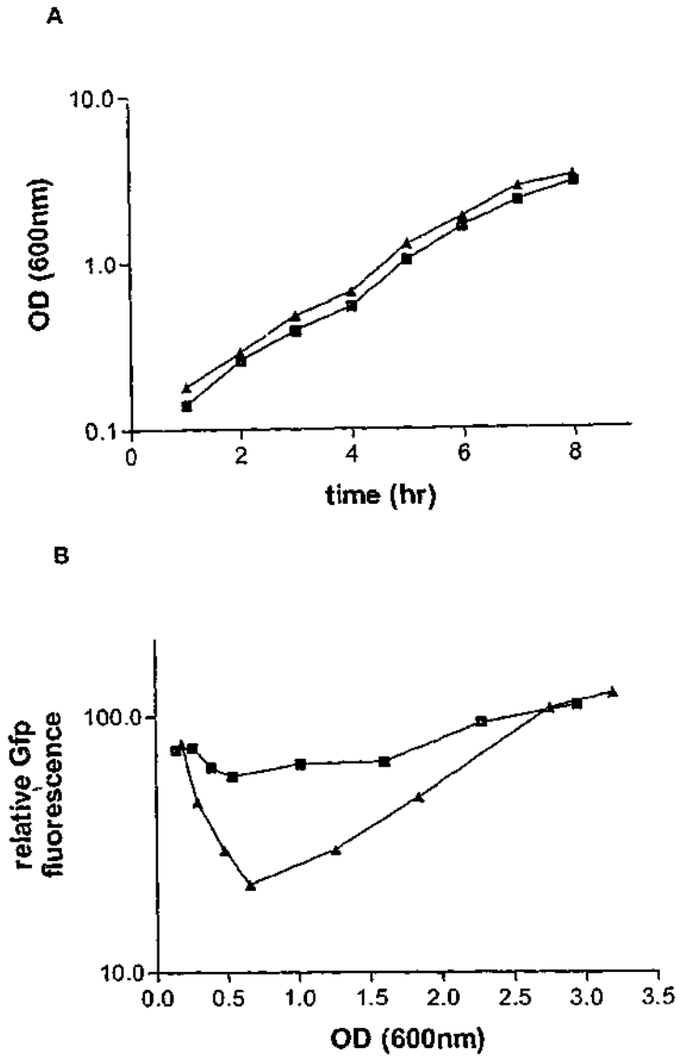

Figure 6.6: Effect of culture density on expression of *pld*.

Cultures of C231(pSM22) (■) and C231(pSM27) (▲) were diluted to an OD_{600} of 0.1 and then incubated at 37°C. At hourly intervals the growth of the culture was determined by measuring OD_{600} (A) and the level of Gfp fluorescence of each culture was determined by flow cytometry (B). For each sample 30,000 events were counted and the mean fluorescence of the population determined to give relative Gfp fluorescence.

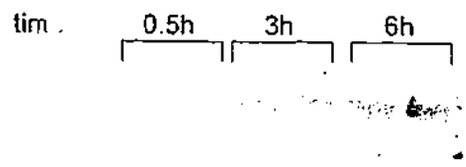

Figure 6.7: Density dependent expression of *C. pseudotuberculosis pld*.

C. pseudotuberculosis was grown from an OD_{600} of 0.1 for 30 min, 3 hr and 6 hr (corresponding to average OD_{600} of 0.16, 0.47 and 1.5 respectively) At each time point RNA was extracted from duplicate cultures. 8 µg of total RNA was analysed by Northern analysis using a DIG labelled *pld* PCR product

highly conserved in Gram-positive bacteria, however, they in general use variations of a two-component signalling mechanism. In general gram-positive bacteria produce a modified polypeptide (the autoinducer) that is actively transported from the cell by an ABC transporter, where it accumulates in the extracellular environment. When the concentration of the autoinducer reaches a sufficiently high concentration in the environment, it is detected by a two-component sensor kinase protein located in the membrane. The sensor kinase becomes activated and autophosphorylates a conserved histidine residue. The phosphoryl group is then transferred to cognate response regulators leading to their activation Activated response regulators in turn activate or repress transcription of specific target genes, thus leading to changes in the cell's phenotype. Given that the amount of autoinducer in the environment is proportional to the number of bacteria producing it, the system is usually triggered at a certain cell density.

Examples of quorum sensing in Gram-positive bacteria include the regulation of virulence gene expression in *Staphylococcus aureus* (Ji *et al.*, 1995), induction of competence in *Streptococcus pneumonia* (Havarstein *et al.*, 1995) and *B. subtilis* (Tran *et al.*, 2000), and bacteriocin production in *Lactobacillus* (Risoen *et al.*, 2000).

Given that Pld is a major virulence factor we were interested in determining whether the mechanism for the observed density dependent expression was quorum sensing. If *C. pseudotuberculosis* produces an autoinducer molecule it would be expected that it would be present in the conditioned media of a high density culture and as such should be able to induce *pld* expression in a low density culture. Thus to determine whether *C. pseudotuberculosis*

produces a quorum sensing molecule that regulates *pld* expression, conditioned media from a high density C231 culture was inoculated with bacteria at a low density. Flow cytometry was then used to determine whether components of the conditioned media could prevent the decrease in Gfp fluorescence usually observed in a low density culture. Conditioned media (CM) from C231(pSM20) cultures that had been grown from an OD_{600} of 0.1 for 5.5, 6 and 7 hr (OD_{600} = 1.8, 2.8, 2.8 respectively) was generated by pelleting bacteria by centrifugation and subsequent filtration of the supernatant through a 0.4 μm filter. When bacteria were inoculated into the CM a reduced growth rate was observed within 4 hr of inoculation. This was probably due to depletion of nutrients in the CM. To circumvent this problem 5 ml of conditioned media was supplemented with 1 ml of 4x BHI broth. An overnight culture of C231(pSM27) was inoculated into the 3 CMs to give an OD_{600} of 0.1. As a control, C231(pSM27) was inoculated into 5 ml BHI supplemented with 1 ml of 4x BHI broth. Cultures were incubated at 37°C and Gfp fluorescence measured by flow cytometry at hourly intervals for 5 hr. The bacteria incubated in the presence of CM demonstrated the same fluorescence profile as the control sample (Figure 6.8). It is possible that the concentration of an autoinducer molecule in the CM is maximal at a cell culture density other than the three tested. For this reason CM from a number of other different density cultures were also generated however none were able to induce *pld* expression at low culture density. It was therefore not possible to demonstrate that quorum sensing was the mechanism by which density dependent regulation of *pld* was occurring.

6.8 Does heat regulation of *pld* expression occur at all densities?

The studies performed to date show that *pld* is regulated by at least two mechanisms (thermoregulation and cell density dependent). The heat regulation experiments had been performed at medium to high cell densities thus raising the question as to whether thermoregulation also occurs at lower cell densities.

An overnight culture of C231 was diluted to an OD_{600} of 0.1 then grown at 37°C for 6.5 hr. RNA samples were prepared from 1, 3.5 and 6.5 hr cultures. For the last 30 min of each culture period heat shock samples were also prepared by incubation at 43°C. RNA was DNaseI treated, quantitated by spectroscopy and the integrity and quantitation confirmed by

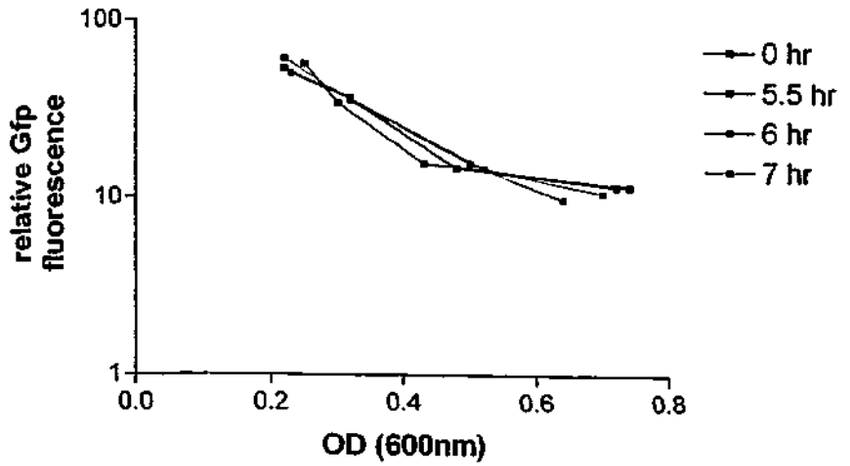

Figure 6.8: Effect of CM on density dependent expression of *pld*.

An overnight culture of C231(pSM27) was used to inoculate CM from 5.5, 6 and 7 hr *C. pseudotuberculosis* cultures (corresponding to OD_{600} of 1.8, 2.8, 2.8 respectively). Cultures were incubated at 37°C and at various timepoints the Gfp fluorescence of the cultures was measured by flow cytometry. For each sample 30,000 events were counted and the mean fluorescence of the population determined to give relative Gfp fluorescence.

agarose gel electrophoresis. 1 µg of total RNA was reverse transcribed to generate cDNA which was used as a template for real time PCR. In addition to measurement of *pld* mRNA, 16S rRNA levels were also determined as a normalisation control. As previously demonstrated by northern analysis expression of *pld* mRNA increased with increasing cell density (5.6-fold over the timecourse of the experiment, Figure 6.9). Thermoregulation of *pld* expression occurred following heat shock at all cell densities with fold reductions of 4.2, 10.2 and 17.8 at 1, 3.5 and 6.5 hr respectively. The level of *pld* mRNA detected was however essentially equal in all heat shocked samples, suggesting that heat shock reduced *pld* expression to a basal level. This study demonstrated that down-regulation by heat shock overrides up-regulation of expression of *pld* at higher cell densities.

1.1 Pld expression inside the macrophage

As Pld is required for virulence of *C. pseudotuberculosis*, it must be expressed *in vivo*. Although supramaximal amounts of Pld, when delivered systemically, causes lysis of ovine erythrocytes (Hsu *et al.*, 1985) it is not clear as to whether this is the main mechanism by

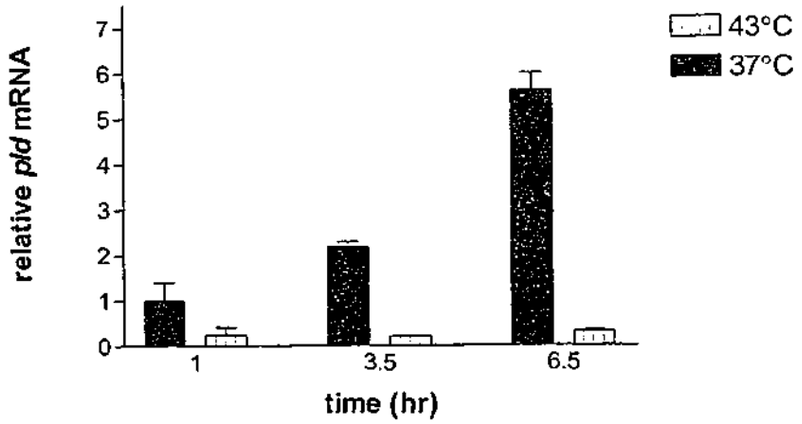

Figure 6.9: Effect of temperature and cell density on expression of *C. pseudotuberculosis pld*.

An overnight culture of C231 was diluted to an OD_{600} of 0.1 then incubated at 37°C. At 30 min, 3 hr and 6 hr aliquots of the culture were transferred to 43°C for 30 min. RNA was extracted from the heat shocked samples and those incubated at 37°C for the same period of time (i.e. 1, 3.5 and 6.5 hr). The levels of *pld* mRNA and 16S rRNA were determined by reverse transcription followed by real time PCR. *pld* expression was normalised to 16S rRNA levels to take into account differences in starting amounts of RNA. Results are represented as the mean of duplicate samples ± S.E.M.

which Pld increases virulence. Given that erythrocytes are not phagocytic this type of action implies an extracellular action by Pld. However following infection *C. pseudotuberculosis* is an intracellular pathogen primarily residing within the macrophage. For . is reason we were interested in determining whether *pld* is expressed inside the macrophage and, if so, is that expression regulated in an analogous manner to the *in vitro* observations. If *pld* is expressed inside the macrophage other studies aimed at elucidating its function could then be performed. Such studies may give insights into other mechanisms by which Pld may increase the virulence of *C. pseudotuberculosis*.

6.9.1 Expression of *pld* by intracellular *C. pseudotuberculosis*

The tissue culture infection model utilising the mouse macrophage-like cell line J774 is described in Chapter 3. Two independent infection experiments were performed. In the first,

macrophages were plated at 7 x 10^5 cells/well in 6 well plates 18 hr prior to infection. Overnight cultures of C231(pSM22) or C231 (pSM27) were diluted to OD_{600} of 0.1 and grown for 1 hr. Macrophages were incubated with bacteria at a MOI of 4 for 1 hr, then washed and fresh media containing 100 µg/ml gentamycin added to kill extracellular bacteria. At further time-points macrophage monolayers were washed, lysed, then sonicated to disrupt bacterial clumps. The sonicates were diluted in PBS and Gfp fluorescence of individual bacteria measured by flow cytometry (Figure 6.10A). Gfp fluorescence from the control C231(pSM22) strain was constant over the timecourse of the experiment. Fluorescence from C231(pSM27) was high throughout the timecourse of the experiment, however a decreased level of fluorescence at the 1.5 hr timepoint compared to the fluorescence of the added bacteria (t= 0 hr) was observed. Fluorescence of C231(pSM27) at the remaining timepoints gradually increased to the level initially observed in the added bacteria.

A second experiment was performed to confirm this data. An additional control in which the two strains were also incubated in DMEM alone was also included and additional time points (up to 12 hr) were analysed. Preparation of the macrophages and bacteria was the same as in the first experiment. For the *in vitro* controls the same number of bacteria were inoculated into 2 ml DMEM and incubated under the same conditions as the macrophages for the duration of the experiment (without washes and gentamycin containing media). The pattern of intracellular *pld* promoter activity in the second experiment closely followed that observed in the first experiment. The additional timepoints (8, 10, 12 hr) indicate that *gfp* expression from pSM27 plateaus, rather than continuing to increase (Figure 6.10B). The fluorescence detected from C231(pSM22) and C231(pSM27) incubated in DMEM decreased over the timecourse of the experiment. This effect was more pronounced for C231(pSM27) than C231(pSM22). Thus the fluorescence of these cultures did not follow the pattern of fluorescence observed in BHI with shaking.

To confirm that *pld* was highly expressed during infection, relative *pld* mRNA levels were measured by RT- real time PCR. The cDNA samples prepared from infected macrophages and control DMEM grown bacteria that were described and utilised in Chapter 3 were used for this experiment. Pld expression was found to be high in both DMEM and macrophage derived bacteria. There was no induction following macrophage infection. The design of the two

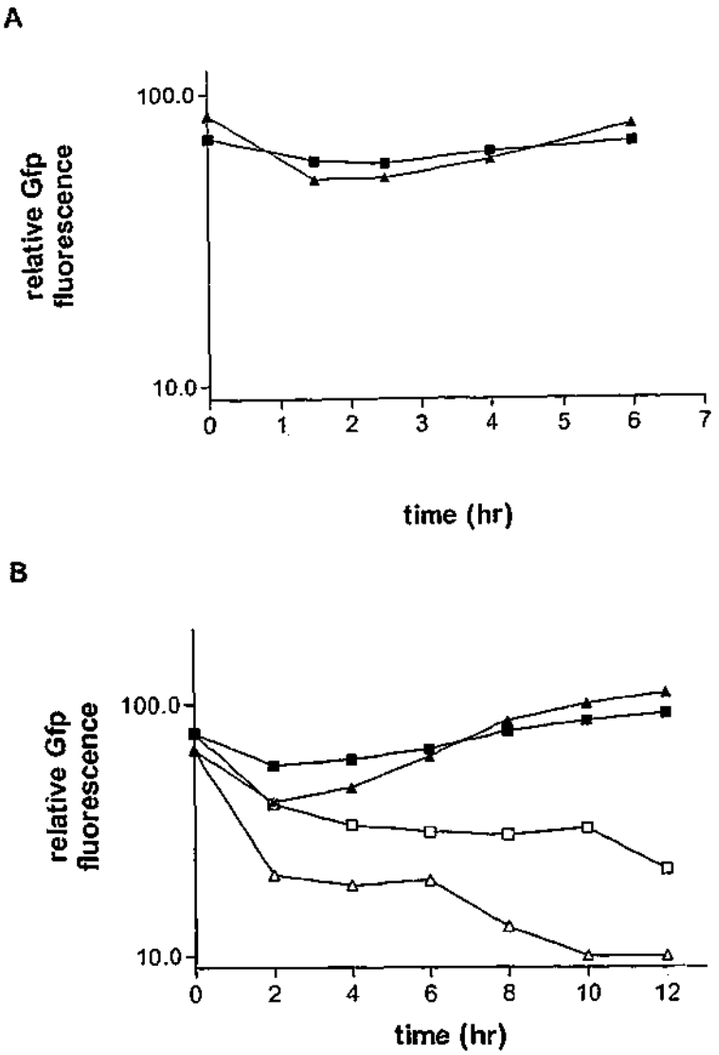

Figure 6.10: *pld* expression from *C. pseudotuberculosis* inside infected J774 macrophages.

(A and B) 18 hr prior to infection J774 macrophages were plated in 6 well plates at 7x 10^5 cells/well. Macrophages were infected with exponentially growing C231(pSM22) (■) or C231(pSM27) (▲) at MOI 4. Following an incubation period of 1 hr, the macrophages were washed and gentamycin containing media added. At various timepoints the macrophage monolayers were lysed, sonicated and the fluorescence of individual bacteria determined by flow cytometry. For each sample 50,000 events were measured and the mean fluorescence of the bacterial population determined to give relative Gfp fluorescence. (B) At the same time, equal numbers of C231(pSM22) (□) or C231(pSM27) (Δ)were inoculated into DMEM. At each timepoint a proportion of the culture was taken, sonicated and the fluorescence of individual bacteria determined by flow cytometry.

experiments is different however this result indicates that *pld* is expressed by intracellular bacteria. That *pld* was found to be highly expressed in DMEM is in contrast to the flow cytometry data.

6.9.2 Investigation of *pld* expression in DMEM

In Section 6.9 the expression of C231(pSM22) and C231(pSM27) in DMEM was observed to be different from that when the strains were cultured in BHI. For C231(pSM22) a gradual decrease in the level of fluorescence detected was observed which contrasted to the usually steady state level observed when the strain was grown in BHI. For C231(pSM27) the fluorescence of the bacteria decreased as is usually observed; this was not however followed by a subsequent increase in fluorescence as cell density increased.

In addition to being cultured in different growth medium the culture conditions also varied. BHI cultures are incubated with rapid shaking (300 rpm) to aerate the cultures, whereas the DMEM cultures are incubated in a stationary manner in an environment that contains 5% CO_2. In order to determine whether the observed *pld* promoter activity in the different media is related to the mechanism of culturing or the media used, bacteria were incubated in DMEM or BHI under the two different culture conditions. Overnight cultures of C231(pSM22) and C231(pSM27) were diluted into DMEM or BHI to give cultures with an OD_{600} of 0.1. For shaking cultures 30 ml glass McCartney bottles containing 10 ml of culture were utilised in a shaking incubator at 300 rpm. For CO_2 incubator cultures 60 mm tissue culture dishes containing 10 ml of culture were utilised. The fluorescence of cultures was determined by flow cytometry at 0, 1, 3 and 6 hr post-dilution (Figure 6.11). Expression of *gfp* from the *srp* promoter construct (pSM22) was high and essentially constant for shaking cultures regardless of the media type. Expression in the CO_2 incubator cultures was slightly lower in the BHI culture and much lower in the DMEM sample such that the fluorescence intensity of the C231(pSM22) DMEM culture decreased approximately two-fold over the timecourse of the experiment. In contrast to the *srp* promoter construct expression from the *pld* promoter appeared to be media related. As expected density dependent regulation of *pld* expression occurred in BHI shaking cultures. Additionally the BHI C231(pSM27) culture incubated in the CO_2 incubator showed reduced fluorescence followed by increased fluorescence. On the

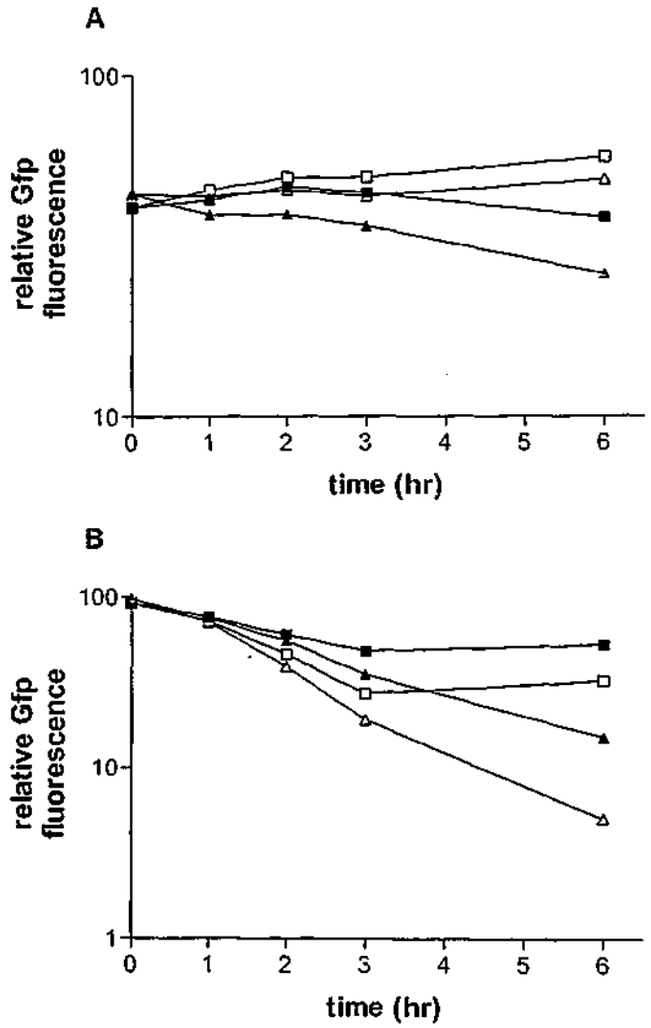

Figure 6.11: Effect of media and growth conditions on *pld* and *srp* promoter activity.

(A) C231(pSM22) and (B) C231(pSM27) were inoculated into BHI (■ and □) or DMEM (▲ and Δ) to give an OD_{600} of 0.1. Cultures were incubated in sealed McCartney bottles at 37°C with shaking at 300 rpm (□ and Δ) or in 60 mm tissue culture dishes in a tissue culture incubator at 37°C with 5% CO_2 (■ and ▲). At various timepoints Gfp fluorescence of the cultures was measured by flow cytometry. For each sample 30,000 events were counted and the mean fluorescence of the population determined to give relative Gfp fluorescence.

other hand for bacteria grown in DMEM *pld* promoter activity decreased throughout the experiment, an effect that was more pronounced in the shaking culture which showed a 19-fold reduction in fluorescence intensity over the experimental timecourse. Analysis of cultures that were then incubated for another 15 hr showed that fluorescence of C231(pSM27) grown in DMEM remained low (data not shown).

DMEM, the tissue culture media used during these experiments is prepared from a 5x DMEM concentrate that consists of a complex mixture of inorganic salts, vitamins, sodium pyruvate and glucose. To generate the final media this is supplemented with fetal calf serum, glutamine, HEPES, sodium bicarbonate and β-mercaptoethanol. In order to determine which of these components were responsible for the decrease in Gfp detected when C231(pSM27) was grown in DMEM, an experiment was performed in which bacteria were grown in BHI broth supplemented with individual and combined components of DMEM. An overnight culture of C231(pSM27) was diluted to an OD_{600} of 0.1 in media containing 1x BHI in addition to one or more tissue culture media components (at the final concentration found in DMEM). 8 ml cultures were incubated in McCartney bottles at 37°C with shaking at 300 rpm. At 0, 2, 4, 6 and 8.5 and 23 hr Gfp fluorescence was determined by flow cytometry (Figure 6.12).

Supplementation of BHI with FCS, β-mercaptoethanol or glutamine had no effect on the Gfp fluorescence detected. Cultures grown in BHI supplemented with HEPES/sodium biocarbonate, 5x DMEM concentrate or all tissue culture components showed different curves. As expected growth in BHI in the presence of all the components of DMEM showed the same pattern of Gfp fluorescence as for bacteria grown in DMEM. HEPES/sodium biocarbonate and 5x DMEM concentrate demonstrated intermediate curves, which showed the usual initial decrease, followed by an increase at the 6 and 8.5 hr timepoints followed by a decrease at 24 hr. These data indicate that the effect of HEPES/sodium bicarbonate and 5X DMEM on reducing Gfp fluorescence may be additive. To determine whether the effect was related to the culture density the optical density of 8.5 hr cultures was measured. Cultures supplemented with FCS, glutamine, HEPES, sodium bicarbonate and β-mercaptoethanol exhibited the same OD_{600} as for bacteria grown in BHI. Culture in DMEM, or BHI supplemented with 5x DMEM or all components increased the rate of bacterial growth 1.5- 3 fold. This indicated

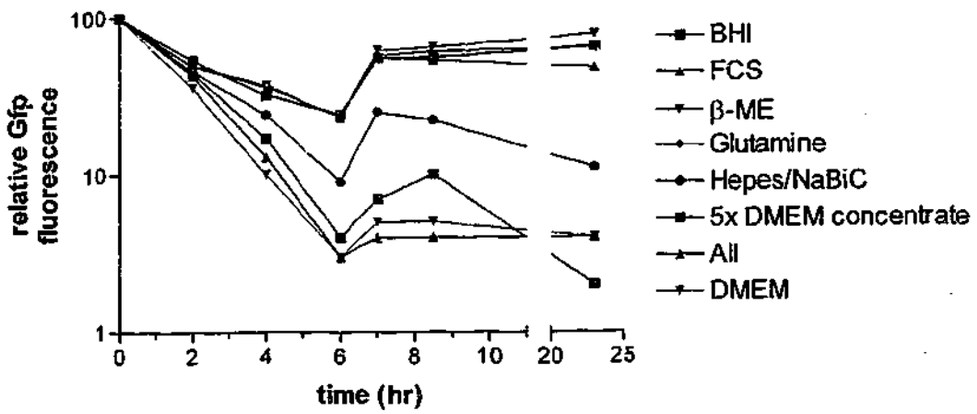

Figure 6.12: Effect of the components of DMEM on Gfp fluorescence detected from C231(pSM27).

An overnight culture of C231(pSM27) was diluted into BHI alone (BHI), DMEM alone (DMEM) or BHI supplemented with 10% FCS (FCS), 50 μM β-mercaptoethanol (β-ME), 2 mM glutamine, 25 mM HEPES/ 0.19% sodium bicarbonate (Hepes/NaBiC), 1x DMEM from a 5X DMEM concentrate (5x DMEM) or all of the above components (All). At various time points Gfp fluorescence of the cultures was measured by flow cytometry. For each sample 30,000 events were counted and the mean fluorescence of the population determined to give relative Gfp fluorescence.

that there may be a correlation between detectable Gfp fluorescence and growth rate such that at increased growth rate there was less fluorescence detected.

6.10 Effect of *pld* expression on mammalian cell function

The results of section 6.9 show that *pld* is expressed at a high level by *C. pseudotuberculosis* inside macrophages. Given the properties of this enzyme it is likely that intracellular Pld may affect macrophage function and viability either by destruction of macrophage membranes or by activation of mammalian cell signalling pathways. In order to study the effects of *C. pseudotuberculosis* Pld on macrophage cell function, a Pld negative strain called Toxminus was utilised (Hodgson *et al.*, 1992). Toxminus is a mutant version of the wildtype strain C231, in which the *pld* gene has been inactivated by insertion of an antibiotic resistance gene. Studies have shown that *in vivo,* Toxminus is significantly attenuated (Hodgson *et al.*, 1992). A third strain Toxminus + pTB111 has a *pld* gene supplied in trans on a plasmid and demonstrates a wildtype phenotype (Tachedjian *et al.*, 1995).

To confirm that the strains to be utilised demonstrated the appropriate phenotype with regard to *pld* expression, each was streaked onto a sheep blood agar plate containing supernatant from *R. equi* culture. *R. equi* supernatant contains cholesterol oxidase which together with *C. pseudotuberculosis* Pld causes haemolysis of ovine erythrocytes. On blood agar plates colonies of C231 and Toxminus+pTB111 were surrounded by zones of clearing indicating haemolysis of the sheep erythrocytes while no clearing was observed around colonies of Toxminus (Figure 6.13).

6.10.1 Effect of *pld* on phagocytosis of *C. pseudotuberculosis* by macrophages

Pathogenic bacteria may secrete factors that interact with the host machinery to induce changes in the host environment. For example *Salmonella* secrete factors that interact with host machinery to induce bacterial uptake (Zhou and Galan, 2001). Given that mammalian Pld has been shown to play a role in the phagocytic process in macrophages (Kusner *et al.*, 1996), we were also interested in determining whether expression of *pld* by *C. pseudotuberculosis* influenced its own uptake into macrophages.

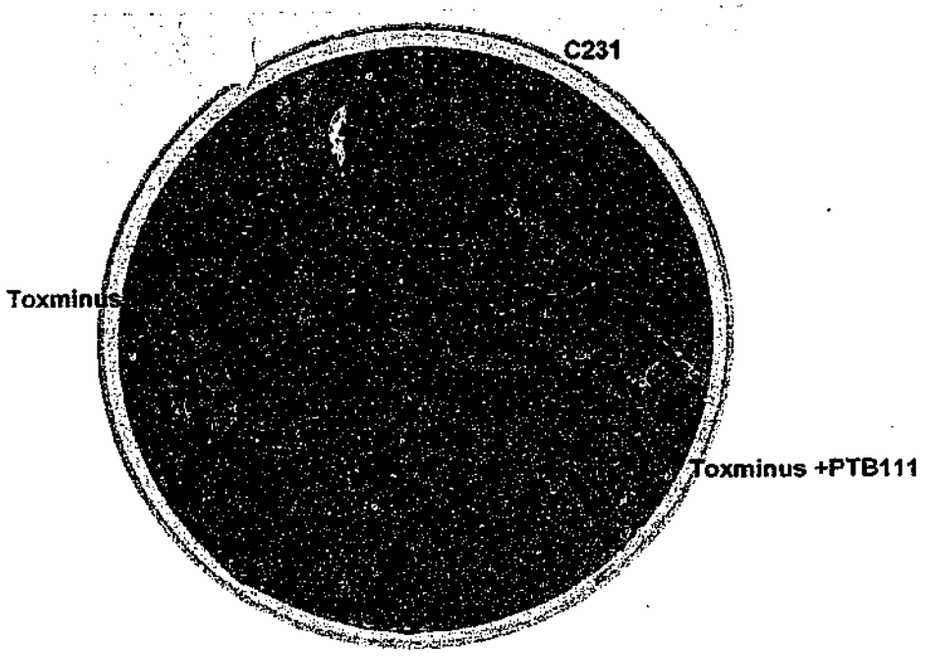

Figure 6.13: Haemolysis of sheep erythrocytes by *pld* expressing strains of *C. pseudotuberculosis*.

C. pseudotuberculosis strains C231, Toxminus and Toxminus +pTB111 were streaked onto agar plates containing 10% each of sheep blood and *R. equi* culture supernatant. Zones of clearing are indicative of synergistic erythrocyte lysis by *C. pseudotuberculosis* Pld and *R. equi* cholesterol oxidase.

J774 macrophages were plated at 7 x 10^5 cells/well in 6 well plates 18 hr prior to infection. Overnight cultures of C231, Toxminus and Toxminus+pTB111 were diluted to OD_{600} of 0.1 and grown for 1 hr. Macrophages were incubated with bacteria at a MOI of 3 for 1 hr, then washed and fresh media containing 100 µg/ml gentamycin added to kill extracellular bacteria. At 3 hr post infection macrophage monolayers were lysed and the extracts sonicated to disrupt bacterial clumps. Extracts were serially diluted, plated on BHI agar plates and incubated at 37°C. Colony counts were performed 48 hr later (Figure 6.14). No difference in the number of bacteria taken up could be observed at either time point, indicating that at least for uninfected macrophages bacterial *pld* expression had no effect on phagocytosis of *C. pseudotuberculosis*.

6.10.2 Effect of *pld* expression on macrophage viability

Treatment of ovine neutrophils with Pld has been shown to reduce their viability (Yozwiak and Songer, 1993). In this study intracellular Pld (delivered as an oil emulsion) was shown to be more toxic to the neutrophils than extracellular Pld (delivered aqueously). Given that this study was performed using relatively large amounts of purified protein its biological significance is not clear. The following experiment aimed to determine whether intracellular *C. pseudotuberculosis* produces a sufficient amount of Pld to alter macrophage viability.

To measure macrophage viability a neutral red assay was established. Neutral red is a dye that is taken up and concentrated in the intracellular compartments of viable cells (Lowik *et al.*, 1993). As the dye is not taken up by dead cells, the amount of dye taken up by a monolayer of cells directly correlates with viable cell number. Upon solubilization of the dye by acidified isopropanol, relative cell number can be determined using a spectrometer.

J774 cells were plated in 24 well plates 18 hr prior to infection. Macrophages were either left untreated or were infected at MOI 4 or 8 with C231, Toxminus and Toxminus + pTB111 for 1 hr. Cells were then washed and incubated in gentamycin containing media for 8.5 hr. The cells were incubated for a further hour in PBS containing neutral red and gentamycin. The cells were then washed extensively to remove residual neutral red and subsequently lysed with acidified isopropanol to solubilize the dye. Absorbance was read at 540nm. Figure 6.15 shows that infection with all three strains of bacteria significantly reduced the viability of the

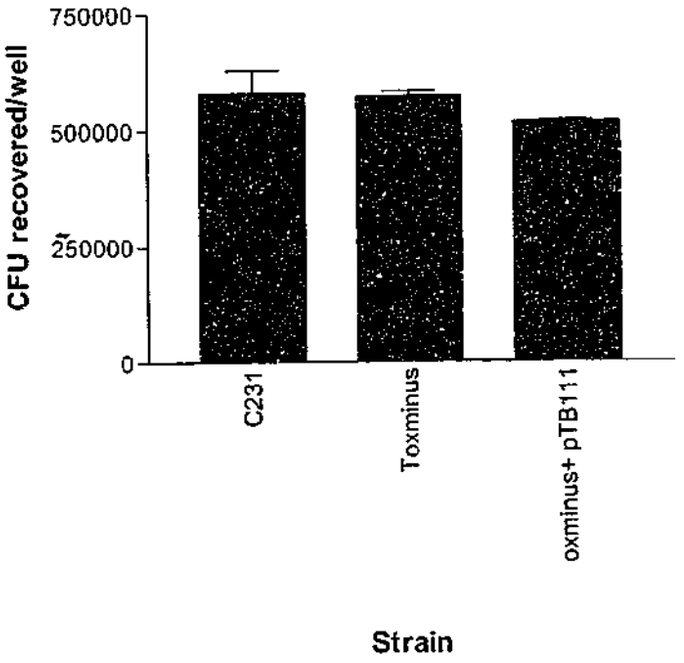

Figure 6.14: Effect of *pld* expression on phagocytosis of *C. pseudotuberculosis* by J774 macrophages.

18 hr prior to infection J774 macrophages were plated in 6 well plates at 7x 10^5 cells/well. Macrophages were infected with exponentially growing C231, Toxminus or Toxminus+ pTB111. Following an incubation period of 1 hr, the macrophages were washed and gentamycin containing media added. At 3 hr post infection the macrophages were lysed and sonicated to disrupt bacterial clumps. Lysates were serially diluted then plated on BHI agar. After 48 hr at 37°C the number of colonies was determined.

macrophages. As expected the percentage of surviving cells was fewer following infection at the higher MOI. At both MOI C231 had a greater effect on macrophage viability than Toxminus. The wildtype phenotype was completely re-established following infection with Toxminus + pTB111 at a MOI 4. At MOI 8 infection with Toxminus + pTB111 caused significantly more cell death than Toxminus, however, the percentage of surviving macrophages was also significantly different from that following infection with C231. These data indicate that the majority of macrophage death following infection with *C. pseudotuberculosis* can be attributed to factors other than Pld, however a small but significant proportion of cell death is directly attributable to the presence of Pld.

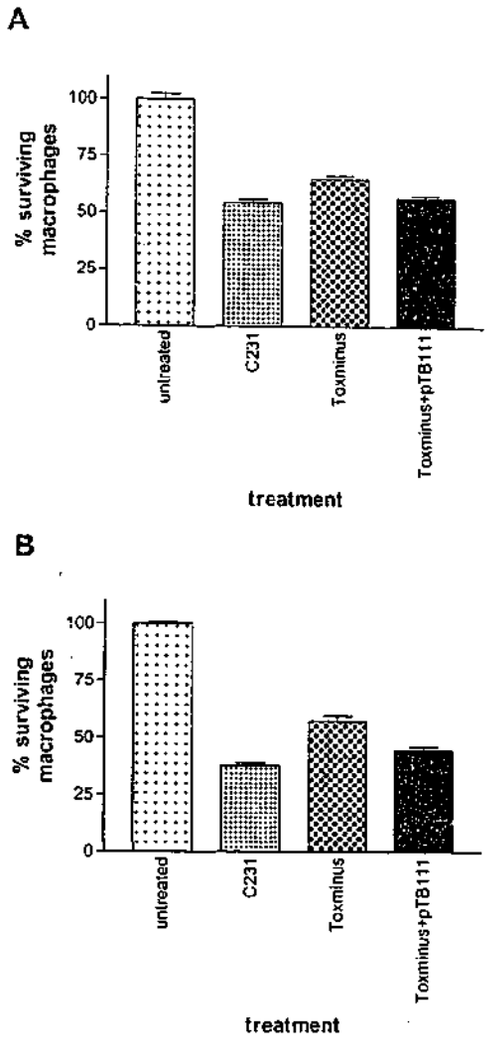

Figure 6.15: Effect of intracellular *pld* expression by *C. pseudotuberculosis* on macrophage viability.

J774 macrophages were infected with C231, Toxminus or Toxminus + pTB111 for 10.5 hr at MOI 4 (A) and MOI 8 (B). Relative cell number was determined using a neutral red viability assay. Results are presented as mean ± S.E.M of 6 replicates. A one-way ANOVA followed by a Tukey's test was performed. At both MOI a significant difference between the number of untreated cells and treatments was observed ($p < 0.001$). At MOI 4 there were significantly more macrophages remaining following infection with Toxminus in comparison to C231 ($P<0.01$), while there was no difference in the percentage of surviving macrophages following infection with C231 or Toxminus+ pTB111. Toxminus was different from Tox+pTB111 at $p<0.05$. At MOI 8 all treatments were different from one another at $p<0.001$ except C231 and pTB111 which were different at $p<0.05$.

6.11 *pld* promoter studies

It is likely that the observed changes in the amount of *pld* mRNA detected during heat shock and culture growth are a result of changes in the level of transcription occurring at the *pld* promoter. A preliminary study of the *pld* promoter was performed in order to gain insights into the mechanisms behind the observed regulatory patterns. In particular a deletion analysis study was performed in order to define the minimum upstream sequence required for density regulation of *pld* expression.

6.11.1 Analysis of the *pld* promoter and identification of putative regulatory elements

Analysis of the *pld* promoter and upstream sequence revealed several features (Figure 6.16). The promoter sequence is AT rich, with the region between +1 and –307 consisting of 66% A or T bases. An estimation of the GC content of the whole genome based on published sequence and that generated during this study, indicates that the GC content of *C. pseudotuberculosis* is approximately 51%. The high AT content of this region is therefore significant and is indicative of promoter regulatory elements. A number of inverted repeats which may act as regulator binding sites can be observed in the upstream region of the *pld* promoter. A region with homology to the UP element consensus (see Section 1.7.1.2) was also identified. The *C. pseudotuberculosis* sequence TGAAAATTATATGAGTTATAAA located at –54 to –74 shows homology to the UP element consensus binding site (NNAAAWWTWTTTTNNNAAANNN where W = A or T and N = any base) at 17 of 22 nucleotides. There is however relatively poor homology at the 3' end of the consensus and the location of the site is suboptimal in regards to the –35 region.

6.11.2 Deletion analysis of the *pld* promoter upstream region: Study 1

A preliminary deletion analysis experiment was performed using pSM20 as the parent vector. Three constructs that contained progressively larger deletions of the 5' region of the *pld* promoter than that found in pSM27 were generated. *pld* promoter fragments were PCR amplified from genomic C231 DNA using primers #77, 98 and 100 as the forward primer and primer #75 as the reverse primer (Figure 6.17). Primer #75 was used as the reverse primer to PCR amplify the *pld* promoter in pSM27 and contains a *Bam*HI site at the 5' end (Figure 6.2). PCR products of the appropriate sizes were generated, then cloned into pGEM®-T easy. The

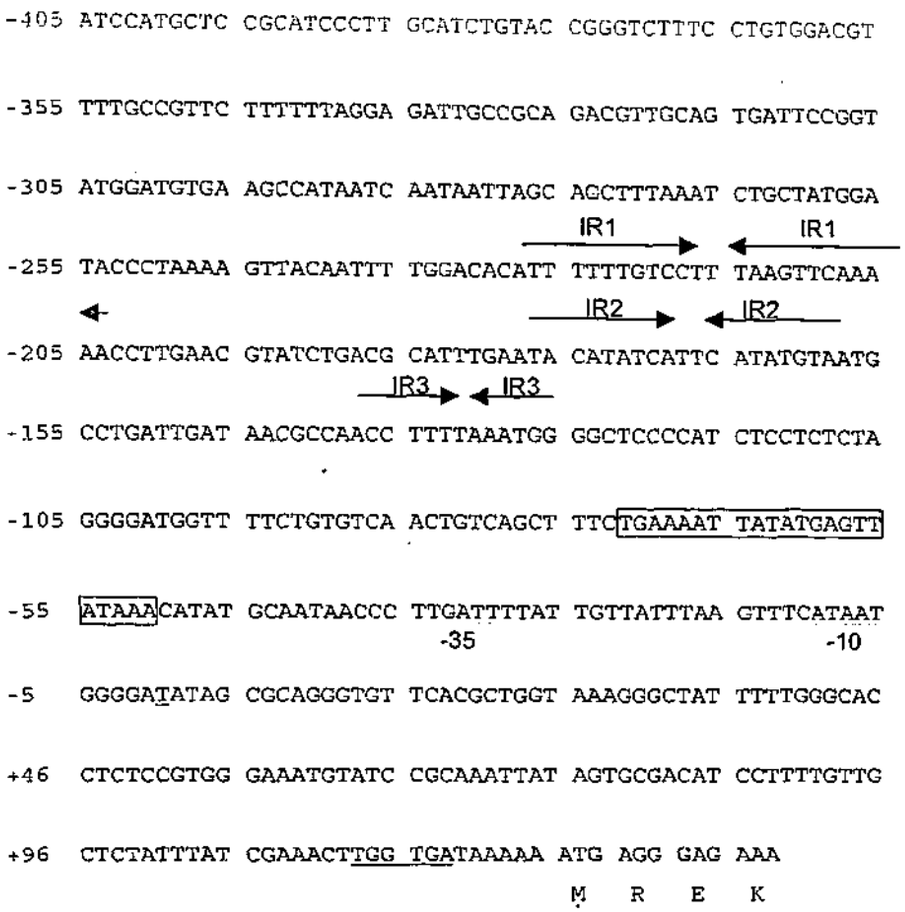

Figure 6.16: Analysis of the *C. pseudotuberculosis pld* promoter

DNA sequence of the *pld* promoter region and first four amino acids. The corresponding protein sequence is shown using the amino acid single letter code. The predicted –35 and –10 regions are doubly underlined, assignment of which is based on the homology of these sequences to *E. coli* σ^{70} –35 and –10 promoter consensus sequences. Downstream is the transcriptional start point (underlined T) which was previously identified using primer extension analysis (Hodgson *et al.*, 1990). Further downstream is a putative RBS (underlined TGGTGA) which lies 6 bases upstream of the start codon. A putative UP element is boxed. Three inverted repeats (with varying degrees of mismatch) are labelled IR1, IR2 and IR3.

```
        pSM27 →
-405 ATCCATGCTC CGCATCCCTT GCATCTGTAC CGGGTCTTTC CTGTGGACGT

-355 TTTGCCGTTC TTTTTTAGGA GATTGCCGCA GACGTTGCAG TGATTCCGGT
                                                    77→
-305 ATGGATGTGA AGCCATAATC AATAATTAGC AGCTTTAAAT CTGCTATGGA

-255 TACCCTAAAA GTTACAATTT TGGACACATT TTTTGTCCTT TAAGTTCAAA
                                                98→
-205 AACCTTGAAC GTATCTGACG CATTTGAATA CATATCATTC ATATGTAATG

-155 CCTGATTGAT AACGCCAACC TTTTAAATGG GGCTCCCCAT CTCCTCTCTA

-105 GGGGATGGTT TTCTGTGTCA ACTGTCAGCT TTCTGAAAAT TATATGAGTT
              100→
-55 ATAAACATAT GCAATAACCC TTGATTTTAT TGTTATTTAA GTTTCATAAT

-5  GGGGATATAG CGCAGGGTGT TCACGCTGGT AAAGGGCTAT TTTTGGGCAC

+46 CTCTCCGTGG GAAATGTATC CGCAAATTAT AGTGCGACAT CCTTTTGTTG

+96 CTCTATTTAT CGAAACTTGG TGATAAAAA ATG AGG GAG
                                     M   R   E
```

Figure 6.17: Location of *pld* promoter primers used for deletion analysis.

Sequence of the *pld* promoter and first three amino acids is shown. The corresponding protein sequence is shown in the amino acid single letter code. The predicted –35 and –10 regions (TTGATT and CATAAT) are doubly underlined. Downstream is the transcriptional start point (underlined T) which was previously identified using primer extension assays (Hodgson *et al.*, 1990). Further downstream is a putative RBS (underlined TGGTGA) which lies 6 bases upstream of the start codon. The sequence corresponding to primers # 77, 98 and 100 is underlined with a dashed lined. These were used in conjunction with primer #75 to generate *pld* promoter inserts of pSM25, pSM32 and pSM33 respectively. The start of the *pld* promoter insert found in pSM27 corresponds to the base at –406.

promoter fragments were released from this vector by digestion with *Sph*I and *Bam*HI and ligated into the promoterless vector, pSM20, which had been prepared in the same manner. The new vectors were titled pSM25, pSM32 and pSM33 (generated using primers # 18, 98 and 100 respectively).

The new constructs were transformed into C231 and gave rise to fluorescent colonies. To determine whether *gfp* expression was regulated in the same manner in the cutback clones as in C231(pSM27), density dependent expression was monitored. Overnight cultures of C231(pSM27), C231(pSM25), C231(pSM32) and C231(pSM33) were diluted to OD_{600} of 0.1

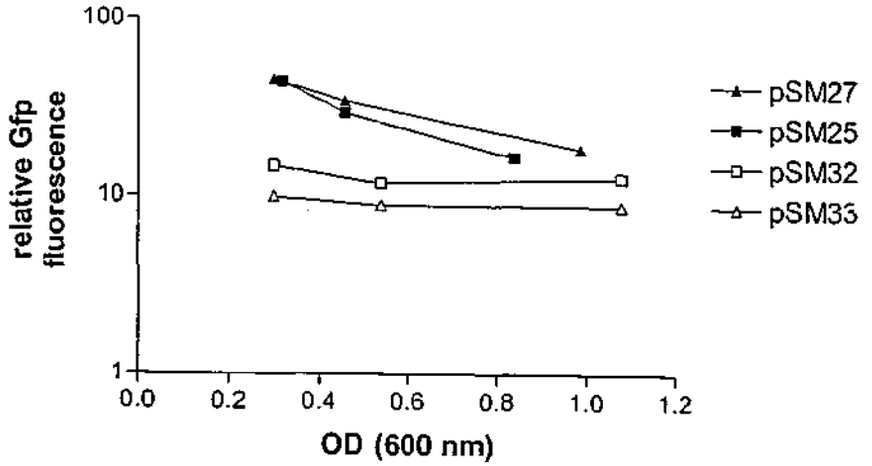

Figure 6.18: Density dependent *pld* expression in *pld* promoter deletion strains.

Overnight cultures of C231(pSM27) (▲), C231(pSM25) (■), C231(pSM32) (□) and C231(pSM33) (Δ) were diluted to an OD_{600} of 0.1 then incubated at 37°C. At various timepoints Gfp fluorescence of the cultures was measured by flow cytometry. For each sample 30,000 events were counted and the mean fluorescence of the population determined.

then incubated at 37°C. At 2, 3, 4 and 4.5 hr post dilution the fluorescence of each population was determined by flow cytometry (Figure 6.18). Expression of *gfp* from the shortened *pld* promoter found in pSM25 was observed to be regulated in the same manner as that observed in pSM27, indicating that the regulatory elements for density dependent expression are contained within the sequence found in pSM25. Gfp detected in C231(pSM32) and C231(pSM33) was lower than that expressed from pSM27 and was unregulated.

This study indicated that the region between –265 and –168 may contain a regulator binding site for density dependent expression of *pld*. In order to further dissect the importance of this region for density dependent expression of *pld*, a second series of deletion constructs were generated. This second series included constructs that contained an intermediate amount of promoter sequence to that found in pSM25 and pSM32. Additionally changes were made to the sequence of the vector backbone. The first series of constructs contained a large amount of sequence between the *pld* promoter and the *gfp* gene, including sequence coding for a RBS. This sequence was present as this vector was originally constructed as the promoter probe

vector to be used in conjunction with DFI. Given that this extraneous sequence could potentially influence either the level of transcription or translation a second series of constructs were generated. In this second series the amount of sequence between the last Pld amino acid and the start of Gfp was reduced from eleven amino acids to four and the vector RBS was removed. Additionally the initiating methionine codon of *gfp* was removed such that the only protein that could be translated was a fusion between the first four amino acids of Pld and Gfp.

6.11.3 Deletion analysis of the *pld* promoter upstream region: Study 2

The promoterless vector pSM20 was first modified to remove extraneous sequence, then *pld* promoter fragments were generated by PCR and inserted into the new vector. The *gfp* gene of pSM20 was PCR amplified using primers #107 and 4b. Primer #107 hybridises to the *gfp* gene starting at the second codon and contains *Bam*HI and *Eco*RI sites at its 5' end. The PCR product was digested with *Bam*H and *Spe*I and ligated into pSM20 that had been prepared in the same manner to generate pSM43. That the original *gfp* gene had been replaced by the new sequence was confirmed by DNA sequencing. Various lengths of the *pld* promoter and upstream region were PCR amplified from C231 genomic DNA as detailed in Figure 6.19. The fragments were then ligated into pSM43 that had been digested with *Bam*HI and *Sph*I to generate pSM46, pSM39, pSM42, pSM47, pSM40 and pSM41. pSM46 contains the longest *pld* promoter fragment, while the others are progressively shorter. Gfp fluorescence of overnight cultures of the cutback strains was measured by flow cytometry. Only the two longest constructs gave rise to fluorescent bacteria (Figure 6.20).

6.11.4 Density regulation of *pld* promoter expression in cutback constructs

The three longest constructs along with pSM27 were analysed for density regulation of *pld* promoter activity. Overnight cultures of C231(pSM27), C231(pSM39), C231(pSM42) and C231(pSM46) were diluted to an OD_{600} of 0.1 and grown at 37°C. Fluorescence of each culture was measured by flow cytometry over a 7 hr period (Figure 6.21). As shown in the previous section C231(pSM42) was not fluorescent. The remaining three constructs showed identical patterns of Gfp fluorescence, with an initial decrease in fluorescence intensity followed by increasing fluorescence intensity after reaching an OD_{600} of approximately 1. This indicates that all the regions required for cell density dependent expression of *pld* are

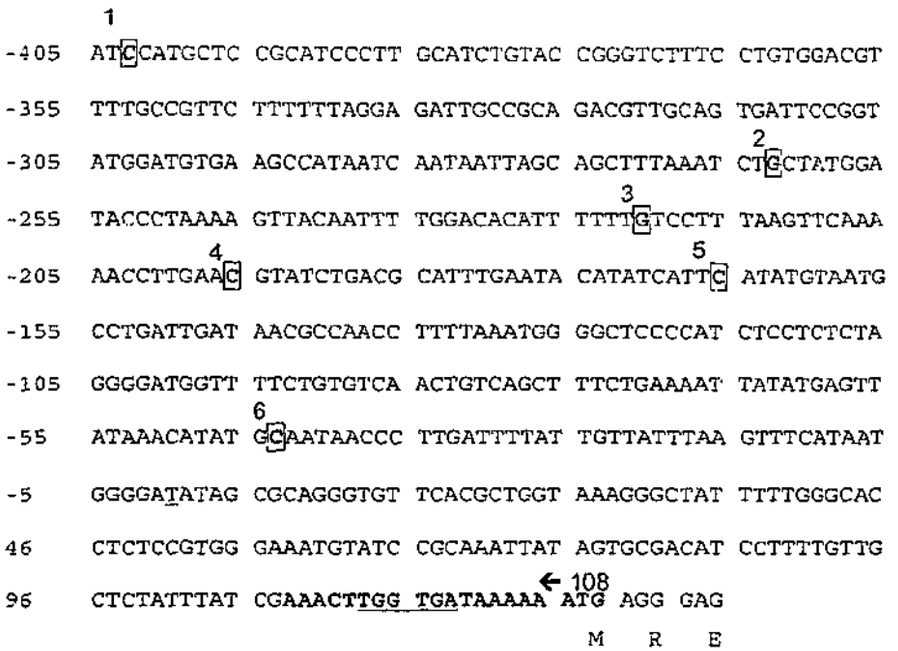

Figure 6.19: DNA sequence of the *pld* promoter region and first three amino acids.

The corresponding amino acid sequence is shown in single letter code. The predicted –35 and –10 regions are doubly underlined. Assignment is based on the homology of these sequences to *E. coli* σ^{70} –35 and –10 promoter consensus sequences. Downstream is the transcriptional start point (underlined T) which was previously identified using primer extension assays (Hodgson *et al.*, 1990). Further downstream is a putative RBS (underlined TGGTGA) which lies 6 bases upstream of the start codon. The start of each cutback generated in this study is indicated by a boxed base. These have been numbered 1 through 6 from longest to shortest and correspond to the inserts found in pSM46, pSM39, pSM42, pSM47, pSM40 and pSM41 respectively. The *pld* promoter fragments were generated by PCR. Sequence corresponding to the reverse primer #108 is bolded. The forward primers #18b, 77, 109, 110b, 98b and 100b hybridised to sequence commencing at 1 to 6 respectively.

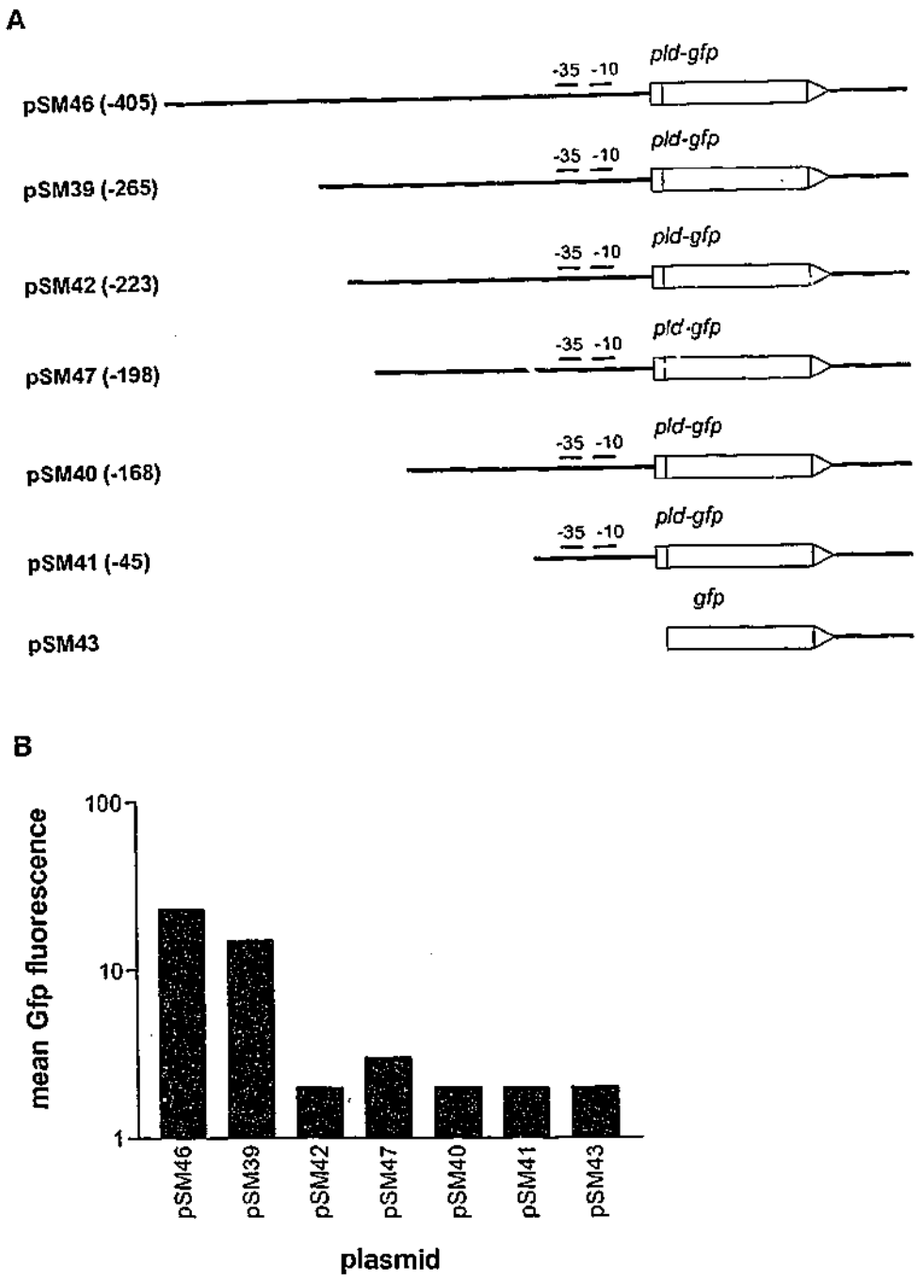

Figure 6.20: Deletion analysis of the *pld* promoter.

A schematic representation of promoter deletion derivatives of pSM46 (A). For each clone the 5' end of the fragment relative to the transcription initiation site is listed. pSM43 is a promoterless construct. For each clone Gfp fluorescence of an overnight culture was measured (B). For each sample 30,000 events were counted and the mean fluorescence of the population determined.

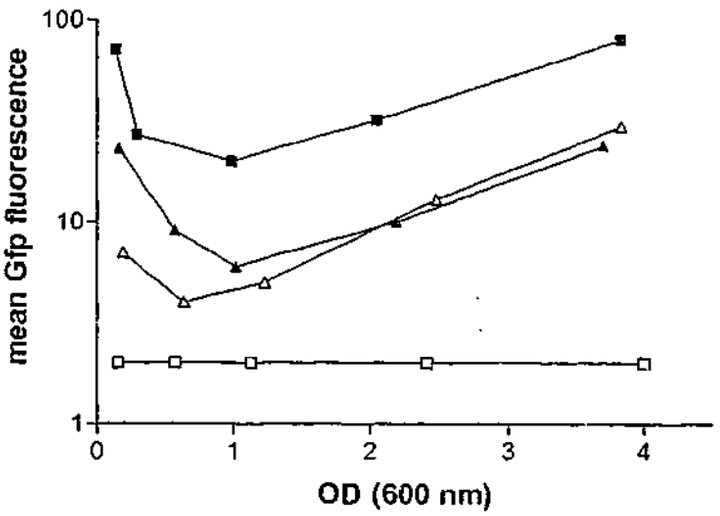

Figure 6.21: Density dependent *pld* expression in *pld* promoter deletion strains.

Overnight cultures of C231(pSM27) (■), C231(pSM39) (▲), C231(pSM42) (□) and C231(pSM46) (Δ) were diluted to an OD_{600} of 0.1 then incubated at 37°C. At various timepoints Gfp fluorescence of the cultures was measured by flow cytometry. For each sample 30,000 events were counted and the mean fluorescence of the population determined.

contained within the promoter sequence found in pSM39. A second observation that is apparent from this experiment is that pSM27 and pSM46 which contain the same amount of upstream sequence have different fluorescence intensities, even though both contain the same amount of promoter sequence.

6.12 Discussion

The primary aim of the work described in this chapter was to gain an understanding of *pld* regulation and the mechanisms underlying that regulation. The observation that *pld* was expressed by *C. pseudotuberculosis* inside macrophages led to studies with the secondary aim of elucidating the functional significance of this finding.

Analysis of the heat shock response of *pld* showed that its expression was rapidly reduced at 43°C such that very little *pld* mRNA could be detected by 20 min post initiation of the heat

shock. This is indicative of a reduction in the rate of transcription from the *pld* promoter and rapid degradation of the *pld* mRNA already present in the cell. Thus it appears that, as for other bacterial mRNA species, the half-life of *pld* mRNA is relatively short. The maximal repression of *pld* expression was maintained from 20 min for the remainder of the experiment (up to 60 min).

Many virulence determinants are regulated by more than one environmental factor. For example, expression of Invasin, the primary invasive factor of *Yersinia pseudotuberculosis* is controlled by a number of environmental signals including temperature, growth phase, nutrient availability, pH and osmolarity (Nagel *et al.*, 2001). In order to allow us to more easily monitor the effects of environmental stimuli on *pld* expression, two reporter constructs were generated using the promoter probe vector with a *gfp* gene (pSM20) that was described in Chapter 4. Two reporter plasmids were constructed that differed only in the amount of 3' promoter sequence incorporated into the vector. pSM28 contained the *pld* promoter sequence up until the base before the *pld* promoter RBS and as such the bacteria's ribosomal apparatus would be expected to use the RBS coded for in the downstream vector sequence. For some promoters sequence downstream of the translational start point can influence promoter activity (Gralla and Collado-Vides, 1996). For this reason a second construct, pSM27, was generated that in addition to the promoter sequence contained in pSM28 contained the *pld* RBS and sequence coding for the first four amino acids of Pld in frame with the *gfp* gene. This was cloned into the same sites as the *pld* promoter fragment in pSM28. Upon association of the cell's ribosomal machinery to the *pld* RBS site it would be expected that the translated protein would be a fusion between the four Pld amino acids, the intermediate sequence of 11 amino acids and the Gfp protein. Upon transformation into C231 both constructs gave rise to detectable Gfp however fluorescence from pSM27 was consistently greater. Observations from Chapter 4 regarding the choice of RBS sequence for the promoter probe vector indicate that the RBS sequence can have a large influence on the amount of Gfp synthesised, presumably by influencing the level of translation. The observation made here for Gfp expression from the two promoter constructs also supports that observation and suggests that the presence of the native *pld* RBS leads to a greater level of translation than the one found in the promoter probe vector. Alternatively the presence of two RBSs in pSM27 may have led to translation initiation from both sites, thus leading to an enhanced rate of translation.

Attempts to monitor the heat shock response using C231(pSM27) and flow cytometry were unsuccessful in that the large decrease in *pld* mRNA observed by Northern analysis could not be detected as a substantial decrease in Gfp fluorescence. Although a small decrease was observed by flow cytometry this did not correlate with either the timing or magnitude of the decrease observed during Northern analysis. As outlined in the results section, this most probably reflected the reduced cell growth of *C. pseudotuberculosis* at 43°C and subsequent accumulation of Gfp in the cells. With the use of any reporter a factor that must be taken into account is the half-life of that protein. Although there are advantages of the long half-life of Gfp such as the potential to see the transient up-regulation of gene expression, there are situations such as this, in which Gfp protein stability could obscure the regulation that is occurring. Attempts to alleviate this problem by generating a vector containing a Gfp protein with reduced half-life were unsuccessful. The vector that was constructed contained a mutant *gfp* gene that in *E. coli* coded for a protein with a half-life of 40 min compared to over 24 hrs for the parent molecule (Andersen *et al.*, 1998). When a *gfp* gene with this type of modification was inserted downstream of either the *pld* or *srp* promoters it was not possible to detect any fluorescence in *C. pseudotuberculosis*. It is possible that the presence of the peptide tag prevents correct processing of Gfp into a fluorescent form by *C. pseudotuberculosis*. Alternatively, because of the destabilising tail the protein may be more efficiently degraded in *C. pseudotuberculosis* than in *E. coli*, and thus cannot be detected. Andersen *et al* (1998) have produced mutant Gfp proteins that have a variety of C-terminal peptide tags and varying half-lives. It is possible that some of these modifications may be better suited to monitoring gene expression in *C. pseudotuberculosis*.

While it was not possible to monitor the heat shock effect on *pld* expression using the reporter construct pSM27, it was possible to demonstrate a density dependent regulation of *pld* expression. This was visualised using flow cytometry as a decrease in detectable Gfp until an OD_{600} of approximately 0.8 was reached, followed by a rapid increase. This phenomenon was confirmed by northern analysis, which showed that *pld* expression gradually increased with increasing cell density. In terms of correlating the Northern data and the flow cytometry data, it is likely that the decrease in fluorescence initially observed by flow cytometry corresponds to low expression levels observed at the first two northern time points. This is likely to represent a slower rate of Gfp production than cell growth, such that the Gfp already present in the cells is dissipated during cell division. At the later timepoints *pld* promoter activity is

increased, thus leading to the rate of Gfp synthesis being greater than the rate of cell division hence an increase in fluorescence is detected. A further analysis of density dependent expression of *pld* by *C. pseudotuberculosis* indicated that the level of *pld* mRNA increased further at higher cell densities than those shown in Figure 6.7 (data not shown).

Serendipitously the Northern and flow cytometry experiments designed to monitor *pld* expression in relation to heat shock were performed at different cell densities, thus allowing characterisation of the two modes of regulation. Had the flow cytometry experiment been performed at a high cell density it may not have been possible to observe the density dependent expression of *pld*. Additionally by performing the heat shock array experiments and Northern analysis at a higher cell density meant that it was possible to detect thermoregulation of *pld*, which may not have been within the limits of detection of the array system if the experiment had been performed at a lower cell density, when *pld* expression was lower.

Pld expression at all cell densities could be down-regulated by heat shock. This indicates that the mechanism of *pld* thermoregulation is able to override the density dependent increase in *pld* expression. At all densities tested, heat shock at 43°C was able to reduce *pld* expression to a similar basal level. Interestingly the experiment showed that even in a low density culture where *pld* expression is not high the level can be further reduced.

Thermoregulation of gene expression generally occurs either at the transcriptional or translational level and as such changes in DNA structure, changes in mRNA conformation and protein conformation have all been implicated (Hurme and Rhen, 1998). Changes in DNA structure at the promoter may be as a result of altered supercoiling or changes in the degree of curvature, both of which may be temperature regulated. Alterations in supercoiling change the local DNA topology which may affect the ability of regulatory proteins to interact with the DNA. Similarly the degree of DNA bending may be altered by temperature changes. This may affect the interaction between bound DNA or if for example DNA bending occurs to allow interaction between a distal regulator and more proximal sites, this interaction may be destroyed at elevated temperatures. Alternatively temperature may alter the conformation of DNA regulatory proteins causing them to alter their interactions with the DNA in a manner that may either enhance or weaken the interaction. The fact that expression from the *pld*

promoter is repressed at 43°C suggests that either an activator of *pld* expression is no longer able to bind to the appropriate regulatory sequence or a repressor that was previously unable to bind to the DNA now can.

Attempts to demonstrate that *pld* expression was regulated by quorum sensing were unsuccessful as determined by the inability of CM from a high density culture to induce *pld* expression. This indicates that either *pld* is not regulated by quorum sensing, the autoinducer is an unstable molecule subject to degradation and is therefore not active in the CM or there is insufficient autoinducer in the CM due to collection of the CM from a culture of an inappropriate density. Given that CM from a number of different culture densities was tested in this study it seems unlikely that an autoinducer containing culture would not have been tested. The approach attempted here to demonstrate a quorum sensing system of regulation has successfully been utilised for the ComQXPA quorum sensing system in *B. subtilis*. In this study CM from a late log phase culture was inoculated at low density with an appropriate reporter strain. Following a three hr incubation, the reporter protein could be detected in the cells grown in the CM but not in those grown in regular media (Tran *et al.*, 2000). The relative stability of the autoinducer is likely to be dependent on its sequence and structure. That at least some autoinducers are fairly stable is supported by the observation that an octapeptide autoinducer of virulence genes in *Staphlococcus aureus*, remains biologically active even after being heated at 100°C for 10 min (Ji *et al.*, 1995).

pld expression was observed to increase gradually with increasing cell density. This phenomenon is not typical of quorum sensing in which gene expression is typically "switched on" when a certain cell density is reached. It therefore seems likely that quorum sensing is not the mechanism by which cell density expression of *pld* occurs. Other mechanisms by which *pld* expression may be regulated include in response to nutrient depletion in the media or changes in bacterial growth rate The first of these is however unlikely as when *C. pseudotuberculosis* was inoculated into CM from a high density culture (which would be at least partially nutrient depleted) *pld* expression was not maintained, but showed the typical decrease observed in a low density culture. Many genes are growth phase regulated, for example ribosomal proteins and fatty acid biosynthetic enzymes. *pld* expression may be regulated by similar environmental cues.

Analysis of the *pld* promoter region and the observation that expression is modulated by several environmental factors suggests that the mechanisms of regulation are likely to be complex. The *pld* promoter –35 and –10 regions (TTGATT and CATAAT) show 4/6 and 5/6 base pair homology to the consensus *E. coli* σ^{70} –35 and –10 regions (TTGACA and TATAAT). In the *pld* promoter, spacing between the two regions is 18 bp. The *pld* promoter is constitutively activated in *E. coli*. Additionally all of the *pld* promoter cutbacks generated were fluorescent in *E. coli*. This indicates that only the –35 and –10 regions are required for promoter activity in this species. This is probably as a result of the *pld* promoter being recognised by the *E. coli* σ^{70} due to its high homology to the consensus sequence. Interestingly there was no regulation of expression from the promoter in *E. coli* indicating that the same regulatory controls do not exist in *E. coli* and in *C. pseudotuberculosis*. Given that the *pld* promoter exhibits high homology to the σ^{70} consensus it is perhaps surprising that the promoter deletions lose activity rather than gain it. That the *C. pseudotuberculosis* σ^{70} recognises a similar consensus sequence is supported by the observation that genes under the control of the *srp* promoter are constitutively and highly expressed in *C. pseudotuberculosis*. The *srp* promoter, which was used extensively as a control during this study, contains consensus –35 and –10 regions separated by 17 bp. Analysis of the consensus motif in other gram-positive bacteria indicate that the consensus is fairly similar to that observed in *E. coli*. In *B. subtilis* the –35 and –10 region consensus is the same as in *E. coli* however in many promoters an additional conserved motif of TG which is centred around –16 is observed (Voskuil and Chambliss, 1998). This motif has been observed in a large proportion of gram-positive promoters (Voskuil and Chambliss, 1998) but was not present in the *pld* promoter. In *C. glutamicum* 33 promoters were analysed to derive a consensus of ttGcca and TA.aaT centred around –35 and –10 respectively (Pátek *et al.*, 1996). In general the –35 region is less well conserved than in *E. coli* and other gram-positive bacteria such as *Bacillus* and *Lactobacillus*. The *pld* promoter showed greater homology to the *E. coli* and *B. subtilis* consensus than the *C. glutamicum* consensus. Given the high homology of the *pld* promoter to the consensus and the fact that it is not constitutively expressed suggests that it must be subject to repression. Additionally given that even pSM41 which contains the least amount of promoter sequence (from –45) does not gain fluorescence, suggests that it is likely that a repressor site is located downstream of –45.

Systematic studies of regulator binding sites have been performed for some bacteria for which a large database of defined promoters exists. Such studies compile information regarding where regulator binding sites are located in relation to promoter features such as the –35 and –10 regions. Given the homology of the *pld* promoter –10 and –35 regions to the *E. coli* σ^{70} consensus it is likely that *pld* promoter is recognised by a sigma factor from the σ^{70} family. Extensive analysis of regulatory sites of *E. coli* σ^{70} promoters indicates that the majority of activator sites are found between –80 and –30 while repressor sites are typically found downstream of –30 (Gralla and Collado-Vides, 1996). It is therefore quite likely that an operator site for a repressor is located downstream of –45. At this location a repressor is likely to sterically interfere with the interaction between RNAP and the DNA, thus preventing transcription. That the *pld* promoter deletion constructs do fluoresce in *E. coli* suggests that either the equivalent repressor proteins do not exist in this species or are sufficiently different so as to be unable to bind to the regulator binding site. It would be interesting to perform DNA fingerprinting experiments to investigate protein binding regions of the promoter region.

The two *pld* promoter deletion analyses performed during this study provided conflicting data in regards to the amount of unregulated basal *pld* promoter activity that was observed. In both studies with vectors containing less than 168 bp of upstream sequence (from the transcription initiation site), there was an abolition of density dependent expression. In the first study, in which pSM20 was used as the base plasmid, these clones demonstrated low level but unregulated promoter activity. In comparison when pSM43 was used as the base vector, clones containing *pld* promoter fragments of an equivalent length were not fluorescent. In order to explain the phenomenon several hypotheses may be proposed.

The only difference in the sequence of pSM27 (first study) and pSM46 (second study) is found between the sequence coding for the first four codons of the *pld* gene and the start of the *gfp* gene (Figure 6.22). In pSM27, 34 bp of sequence is inserted between these two features. This sequence codes for 11 amino acids and includes a second RBS. All of this sequence is derived from the pSM20 vector backbone. In comparison, the same region in pSM46 contains 12 bp, corresponding to four amino acids. In addition the initiating methionine codon has been removed from the start of the *gfp* gene. In the first hypothesis, the extra sequence that is found in pSM27 may have promoter properties, leading to the basal fluorescence that is observed from pSM32 and pSM33. The extra sequence present in these vectors is also present

A

```
1  tatttatcga aacttggtga taaaaaatga gggagaaagt ggatcctcta gatttaagaa
                                   M   R  E  K   V  D  P  L   D  L  R
61  ggagatatac tatgagtaaa ggagaagaac
     R  R  Y   T  M  S  K   G  E  E
```

B

```
1  tatttatcga aacttggtga taaaaaatga gggagaaagg atccgaattc agtaaaggag
                                   M   R  E  K   G  S  E  F   S  K  G
61  aagaactttt cactggagtt gtcccaattc
     E  E  L   F  T  G  V   V  P  I
```

C

```
     -35                        -10    -35                      -10
1  tatttatcga aacttggtga taaaaaatga gggagaaagt ggatcctcta gatttaagaa
                                   M   R  E  K   V  D  P  L   D  L  R
61  ggagatatac tatgagtaaa ggagaagaac
     R  R  Y   T  M  S  K   G  E  E
```

Figure 6.22: Sequence upstream of the *gfp* gene in pSM27 and pSM46.
DNA sequence from pSM27 (A) and pSM46 (B) is shown. Predicted RBSs are boxed. Protein sequence is shown using the single letter amino acid code. Sequence in blue corresponds to GFP and that in red to Pld. Sequence that differs between the two plasmids is shown in black. The *Bam*HI site that was used for cloning of the *pld* promoter fragments is bolded. Sequence of pSM27 that shows homology to *E. coli* σ^{70} –35 and –10 regions are boxed or bolded (C).

in pSM20. Given that no fluorescence is detected from pSM20 (Chapter 4) this would suggest that a second promoter would include sequence from the 3' end of the *pld* promoter insert. In Figure 6.22C several regions with low homology to the –35 and –10 consensus sequences are identified in the sequence of pSM27 in the vicinity of the *gfp* gene. These regions show relatively low homology to the *E. coli* σ^{70} –35 and –10 consensus sequences with matches at only three to four base pairs of the six base pair consensus. In addition with the second –35 and –10 consensus there is unlikely to be sufficient downstream sequence for both a transcription initiation site and a RBS prior to the start of the sequence coding for Gfp.

For some promoters regulator binding sites have been found 3' of the transcription initiation site, either in the untranslated region or in the coding region (Gralla and Collado-Vides, 1996). If a *pld* promoter regulatory element binding site occurs in the vicinity of the region coding for the start of the *pld* gene, then it is conceivable that sequence immediately downstream of this region could alter the local structure at this site, thus potentially disrupting the interaction between the regulator and the DNA. Given that sequence between the Pld amino acids and *gfp* are different in the two series of constructs then the assumption would be that the site was disrupted in one but not the other.

As a third hypothesis, it is possible that regulation could be occurring at the translational level. Assuming that the sequence differences between pSM27 and pSM46 have not altered the level of transcription, then the other point at which regulation could be occurring is at the translational level. Low level fluorescence is detected from pSM32 and pSM33 which is indicative of promoter activity. When comparing the fluorescence of pSM27 and pSM46 which both contain a full-length promoter it is apparent that approximately two to three-fold more Gfp is detected in strains containing pSM27 as pSM46. If this is not as a result of different levels of transcription, then this may be as a result of a greater level of translation from the mRNA generated from pSM27 than pSM46. The rate of translation may vary between the two as a result of better translation from the vector RBS than the *pld* RBS or an additive effect of having two RBS present in pSM27. Only a relatively small amount of Gfp fluorescence is detected from the pSM27 cutbacks, pSM32 and pSM33. Assuming that a two to three-fold further reduction in fluorescence from the equivalent pSM46 derived cutbacks occurs, then the level of fluorescence may be reduced below a level detectable by flow cytometry.

In order to test these hypotheses a number of experiments could be performed. In order to determine where protein translation is starting from, N terminal sequencing of purified proteins could be performed. This would indicate whether translation was being initiated at one or two points. Although it is not possible to detect Gfp fluorescence from the shorter deletion strains in the second study it may be possible to detect mRNA by reverse transcription PCR. This would indicate whether transcription is actually occurring and thus indicate whether the observed differences between the first and second studies was at the transcriptional or translation level. To determine whether there is a second promoter present in pSM27, a construct could be generated in which sequence 5' of the *pld* transcription initiation site was removed. If Gfp fluorescence could be detected from this construct this would be indicative of a second promoter within pSM27.

Regardless of the construct used, both studies showed that a considerable amount of upstream sequence was required for density dependent expression. Of the cutbacks generated during the second study pSM39 which contained sequence from –265 was fully fluorescent whereas pSM42 (starting from –233) had no promoter activity (as indicated by Gfp expression). This is indicative of the region between –265 and –223 being involved in promoter regulation, for example as a regulator binding site. Interestingly the region between –229 to –206 contains an imperfect inverted repeat (5'-TTTTTGTCCTTTAAGTTCAAAAA-3', centred around the twelfth basepair of the sequence) indicating that it may act as such a site. It seems likely that this region may be involved in activation of the *pld* promoter. Given the distance of this region from the promoter it is likely that this site acts in concert with regulators that bind at more proximal sites. This prediction stems from the observation that promoters rarely contain remote regulatory elements in isolation, rather remote regulators typically act in concert with regulators at more proximal sites to modulate the rate of transcription (Gralla and Collado-Vides, 1996). That this site is clearly important for promoter activation is however demonstrated by the fact that regulated expression does not occur without it. Given the distance of this region from the core promoter region, one way in which regulator bound in this region may interact with promoter elements is through DNA bending, thus bringing remote sites into contact with the promoter.

The data generated from this study suggests that the normal state for the *pld* promoter is to be repressed, presumably by binding of a repressor protein in the vicinity of the promoter consensus region. As cell density increases the repression is released perhaps by a mechanism that involves interaction with proteins bound to the DNA at remote upstream sequences. Within intracellular bacteria activation is presumably induced (or repression released) in a manner that is not dependent on the density of the bacteria.

Using the reporter strain C231(pSM27) it was possible to show that *pld* is highly expressed by *C. pseudotuberculosis* resident in macrophages. Although a small decrease in Gfp fluorescence was observed between bacteria used to infect the macrophages and those resident within the macrophage at two hours post infection, *pld* expression was essentially unregulated during the experiment. In fact a greater drop was observed in C231(pSM27) grown in DMEM, which suggests that the decrease observed in macrophage derived bacteria at the two hr timepoint may be as a result of an initial decrease in expression in DMEM before the bacteria were phagocytosed by the macrophages. That no density dependent expression was observed suggests that the density dependent regulation mechanism is not operational within the macrophage or other factors dominate. It is likely that different regulatory systems are activated by the intracellular environment, leading to activation of the *pld* promoter. Alternatively, perhaps the microenvironment of the phagosome appears as a high density environment to the bacteria, thus leading to *pld* expression.

While using C231(pSM27) to monitor *pld* expression by intracellular bacteria, *pld* expression was also monitored in the tissue culture medium, DMEM. Surprisingly the amount of detectable fluorescence from strains carrying both the *pld* and *srp* promoter decreased over the timecourse of the experiment. Additionally expression from the *pld* promoter construct did not appear to increase when higher cell densities were reached. Analysis of this phenomenon indicated that both the media utilised and the manner in which the bacteria were cultured were responsible for this effect. While expression from the *srp* promoter was not affected by the media type in shaking cultures, the effect of DMEM on *pld* expression was dramatic. In comparison to expression of *pld* in BHI there was no increase in detectable fluorescence at the later time-points (in both shaking and non-shaking DMEM cultures). Further dissection of this effect showed that the DMEM components responsible for this effect were likely to be the buffering agents HEPES and sodium bicarbonate, and the 5x DMEM concentrate that forms

the basis of the media. Although an increase in fluorescence was seen in these media at the 6 hr timepoint, this eventually decreased again. This effect may in part be growth related, as growth in these media was increased 1.5-3 fold at the 8.5 hr timepoint. It is not entirely clear as to whether the observed decrease in Gfp fluorescence from the *pld* promoter from cultures grown in DMEM is caused by a change in the level of transcription. Measurement of *pld* mRNA in macrophage derived bacteria and DMEM grown bacteria indicate that *pld* expression is high and does not differ between the two cultures. This is in conflict with the data generated using C231(pSM27). Although the experimental timepoints from the two experiments differ, the DMEM culture analysed by RT-PCR was a high density culture.

The only cofactor required for Gfp fluorescence is O_2, which dehydrogenates the α,β bond of amino acid 66, leading to formation of the fluorophore (Kendall and Badminton, 1998). Once the fluorophore has formed, there is no further requirement for O_2, hence Gfp retains its fluorescence properties regardless of the O_2 supply (Tsien, 1998). The growth of *C. pseudotuberculosis* in DMEM was greater than in BHI. Given that the cultures were grown in an environment with a limited O_2 supply (sealed bottle) or in an environment in which the cultures are not aerated (tissue culture incubator), then the O_2 supply will become exhausted in the faster growing culture at an earlier timepoint. It is therefore possible that Gfp is being produced but not turned into an activated form due to the lack of O_2. That the effect appears more severe in strains carrying pSM27 in comparison to pSM22 could be explained by the observation that the period when there is sufficient oxygen for Gfp to be fluorescent occurs when the *pld* promoter is not active. By the time the promoter becomes active there is reduced oxygen, hence a fluorescent form of Gfp is not formed. An additional phenomenon that may explain the discrepancy between the RNA and fluorescence results is that under anaerobic conditions Gfp can undergo photoconversion from the green fluorescent species to a red fluorescent species (Elowitz *et al.*, 1997), potentially making Gfp undetectable using the standard flow cytometry conditions for detection of Gfp. In order to determine whether *pld* expression is truly regulated by DMEM components a study looking at *pld* mRNA levels should be performed.

Pld expressed by intracellular bacteria was shown to have a small but significant contribution to a *C. pseudotuberculosis* induced reduction in macrophage viability. Several mechanisms may be proposed to explain this observation. Firstly this effect may have been mediated by a

reduction in the integrity of the macrophage plasma membrane as a result of the sphingomyelinase activity of Pld. The plasma membrane of eukaryotic cells is asymmetric in regards to phospholipid content, such that the phospholipids phosphatidylserine, phosphatidylethanolamine and phosphatidylinositol are primarily located in the inner layer of the membrane while phosphatidylcholine and sphingomyelin are predominantly located in the outer layer (Devaux, 1991). Given that the Pld substrate is located in the outer layer of the plasma membrane, its effect may be mediated by an extracellular rather than intracellular mechanism. That is to say that following the death of a macrophage, the cellular contents are released. Pld now in the media is then able to attack sphingomyelin located in the outer membrane of still viable macrophages. Alternatively, sphingomyelin is also a major phospholipid component of murine (J774) phagosomal membranes (Desjardins *et al.*, 1994). Therefore the effect of Pld on macrophage viability may also be mediated through reducing the integrity of intracellular compartments, potentially allowing the escape of bacteria from this restricted area. As a third mechanism, the actions of Pld within the macrophage may be primarily mediated through disruption of mammalian signalling pathways.

As outlined in Chapter 1, mammalian cells possess two Pld proteins that are primarily involved in cell signalling rather than phospholipid degradation. That bacterial phopholipases may modulate these pathways to the advantage of the pathogen has been demonstrated for the two phospholipase C genes of *Listeria monocytogenes* (Schwarzer *et al.*, 1998). The bacterial Plc's stimulate the host cells to produce diacylglycerol. In the subsequent signalling cascade the production of a number of adhesion molecules increases. The presence of these molecules on the endothelial cell surface increases the likelihood of circulating leukocytes adhering to the infected epithelial cells and thus increases the chance of transmission into other cell types such as macrophages. In order to gain insights into whether the Pld of *C. pseudotuberculosis* modulates host signalling cascades the Toxminus strain could be utilised. With this strain the perturbations in host signalling cascades that are directly attributable to Pld could be identified. This type of study may give insights into how *C. pseudotuberculosis* manipulates the host response. Potentially some of these changes may occur at the level of transcription, hence a microarray study of changes in host gene expression following infection with Toxminus and wildtype strains could be very informative.

We were unable to demonstrate a role for Pld in phagocytosis of *C. pseudotuberculosis*. In the macrophage infection model the MOIs utilised correspond to extremely low density cultures, hence expression of Pld would be expected to be low. Potentially pre-treatment of the cells with Pld prior to addition of the bacteria may have produced a different result.

This study has shown that the *pld* gene of *C. pseudotuberculosis* is regulated by multiple environmental stimuli. Possible implications for *pld* regulation during infection are discussed further in the General Discussion (Chapter 7).

Chapter 7

General Discussion

The premise that genes up regulated in the *in vivo* environment may be intimately involved in the virulent phenotype has led to the development and application of techniques such as Signature Tagged Mutagenesis (STM), *in vivo* expression technology (IVET) and DFI. These technologies were specifically developed to provide insights into disease pathogenesis by identifying genes that are either essential for virulence (STM), up-regulated *in vivo* (IVET) or up-regulated under a given set of conditions (DFI). Application of these technologies to a variety of bacterial species has led to an increase in the number of virulence factors that have been identified and an increase in our knowledge of bacterial pathogenesis. The validity of these approaches is demonstrated by the observation that previously described virulence factors are identified, in addition to novel ones.

We have a reasonable knowledge of *C. pseudotuberculosis* disease pathogenesis at the physical level. However, the lack of molecular information regarding gene regulation in this pathogen means that there is little understanding of how this is controlled at the molecular level. The work described in this thesis has begun to address this lack of knowledge by developing and then applying methods to identify regulated genes of *C. pseudotuberculosis* and to then determine their importance during infection.

Both DFI and transcriptional profiling using macroarrays can be applied in the absence of genome sequence information and hence could be utilised with *C. pseudotuberculosis*. Both technologies relied on the development of library technology. The ability to construct a library directly in *C. pseudotuberculosis* for DFI was particularly beneficial as selective pressures from passaging the library through *E. coli* were removed. The relative processivity of DFI meant that larger libraries could be more readily screened than with the DNA array approach for which it was necessary to extract DNA from every clone that was to be arrayed. However, the relative sizes of the libraries analysed did not correlate with the number of regulated clones identified from each. Although utilising a smaller library, many more regulated genes were identified using the array approach than DFI. The relative lack of regulated genes identified using DFI probably reflects that: (a) it can only be used to identify up-regulated genes, (b) only those genes that were not expressed during *in vitro* growth and then subsequently induced following infection were identified and (c) DFI is an enrichment tool rather than a selection tool, hence not every regulated promoter will be identified. Transcriptional profiling on the other hand has the ability to identify both up and down

regulated genes within a single experiment and is able to detect more subtle changes in gene expression.

While performing this study it became apparent that many genes were regulated by more than one stimulus. This was particularly apparent for the major virulence determinant *pld*, which is thermoregulated, expressed in a cell density dependent manner and highly expressed during macrophage infection. The observation that virulence factors are often regulated by a number of environmental factors has significant implications for the choice of the most suitable *in vitro* control. There has been little discussion of this important consideration in the scientific literature. When identifying macrophage-regulated genes by either DFI or transcriptional profiling the *in vitro* control was a relatively high-density culture grown in DMEM. DMEM was chosen, as this was the media in which the macrophages were grown. During this study a number of genes were identified as being expressed in a cell density dependent manner. This included genes such as *fas*, *pld* and *fagC*. This observation has implications for choice of the most appropriate *in vitro* control. For example if a low density culture had been chosen instead of a high density culture then a different subset of genes would have been identified. At a simplistic level the purpose of the *in vitro* control is to provide a nutrient rich environment in which genes required for virulence are not expressed. However, environmental factors such as oxygen availability, pH and nutrient availability are likely to alter during the course of growth of an *in vitro* culture. These changes may affect bacterial gene expression in the *in vitro* culture thus altering the interpretation of what comprise *in vivo* induced genes. Although *C. pseudotuberculosis* can survive for long periods of time within the environment there is no evidence that significant bacterial replication occurs *ex vivo*. Therefore increasing cell density may be a trigger for expression of virulence genes. This would be analogous to quorum sensing control of virulence gene expression that occurs in a number of bacterial pathogens such that virulence gene expression is induced at high culture density.

The ability to look at gene expression under multiple conditions within a single experiment is an advantage that array technology has over other screening technologies. Thus by incorporating multiple control conditions into a single experiment it may be possible to generate a better understanding of the different stimuli that regulate a given gene. This may lead to a better understanding of *in vivo* regulated genes and the precise environmental

changes that occurs *in vivo* to induce their expression. Although no doubt technically difficult to perform, experiments that measure relative gene expression in relevant biological niches may ultimately be the best way to identify virulence factors. For example a comparison of *C. pseudotuberculosis* gene expression in bacteria isolated from the skin as opposed to abscesses in the lymph nodes may identify genes that are required for survival within the lymph node.

Using DFI two macrophage induced genes were identified. The first of these was a non ribosomal peptide synthase that we termed *cps*. The second coded for the beta chain of propionyl coA carboxylase (*pccB*). Further analysis of *pccB* indicated that its expression was down-regulated in the presence of Tween-80. This indicated that what was initially observed as macrophage induction was strictly speaking a release of Tween-80 repression. This observation again demonstrates how the choice of *in vitro* control conditions can influence which genes are identified as being regulated. Tween-80 is a commonly used supplement in bacterial growth media, especially for corynebacterial and mycobacterial species. The demonstration in this study that Tween-80 may regulate gene expression has significance for studies such as this one in which only one of the tested conditions contains Tween-80 or for when investigating transcriptional regulation of genes involved in functions such as lipid biosynthesis.

cps was highly up-regulated during macrophage infection and to a lesser degree during heat shock. *cps* was one of the few genes that was expressed poorly in both DMEM and BHI at 37°C suggesting that it may be important during macrophage infection. Based upon the small amount of sequence currently available for the *cps* gene it is not clear as to length or type of peptide it will synthesise. Sequence prediction of the peptide synthesised by *cps* may be possible upon obtaining the full length gene and the function of the peptide might be ascertained by generating a mutant strain and attempting to identify phenotypic changes.

The number of studies utilising DNA microarrays to monitor changes in bacterial gene expression has increased rapidly within the past few years as a result of the genome sequences for many bacterial species having been completed and a relative increase in the availability of microarray robotic technology and chip scanning devices. However, at the commencement of this study, in early 1999, the use of microarrays for monitoring global transcriptional regulation in bacteria had not yet been reported. Although the macroarray approach used in

this study is now considered a "low tech" approach it was successfully utilised as a tool for identifying regulated genes, of *C. pseudotuberculosis*, for further study. The approach was readily established using standard laboratory techniques and simple hand-held replication devices. Additionally the use of a random library meant that this approach was possible even in the absence of genome sequence. For array studies performed either at the micro or macro scale, purified PCR products are usually arrayed. In this study we demonstrated that it is possible to successfully use plasmid DNA on an array, thus removing a time-consuming and costly step from the array construction process. Now that microarray technology has become more accessible, the techniques developed for this study could be readily applied to the micro scale. In particular the same principles in terms of library construction, RNA preparation and array hybridisation apply. A major advantage of adapting the approach to the micro scale would be an increase in the number of clones that could be simultaneously analysed and the degree of quantiation that can be achieved with the data generated. This study has clearly shown that expression array analysis can be very usefully applied to organisms for which comprehensive genome information is not available.

Even with a relatively small array and a limited number of environmental conditions (i.e. heat, density, macrophage induced), it was apparent that some genes are co-ordinately regulated. For example these genes coding for the heat shock proteins responded in a coordinate manner to heat shock and macrophage infection. Additionally the *pccB* gene identified by DFI demonstrated a similar pattern of regulation as for *fas*. The observation that *pccB* expression was down regulated in the presence of Tween-80 (which is able to act as a lipid source) together with a pattern of regulation that is similar to a fatty acid biosynthesis gene suggests that *pccB* may also play a role in fatty acid biosynthesis. It is likely that the identification of more genes that show a similar pattern of regulation to *pld* will lead to the characterisation of further virulence factors.

In this study, *pld*, the major virulence determinant of *C. pseudotuberculosis* was identified as being regulated. This was a novel observation as *pld* had not previously been shown to be regulated. The key findings were that *pld* was expressed in a cell density manner such that expression increased in correlation with culture density, expression was down-regulated by heat shock and expression by intracellular bacteria was high and essentially unregulated. Heat regulation occurred at all cell densities indicating that the mechanism of heat regulation is

dominant over density dependent expression. The observation that Pld is a major virulence factor indicates that it must be expressed *in vivo*. It is thought that *in vivo* expression of *pld* leads to increased local vascular permeability thus enhancing dissemination of bacteria from the site of infection to the lymph node. It is not known whether *pld* is expressed *in vivo* under all situations and in all locations. The observation that *pld* is expressed by intracellular bacteria suggests that Pld may enhance the formation of abscesses within the lymph node. Abscess formation involves cycles of phagocytosis, bacterial replication within the phagocyte, then phagocytolysis. The observation that *pld* expressed by intracellular bacteria has a direct effect on macrophage viability further supports this hypothesis.

It is not known whether naturally infected animals demonstrate an increased temperature following infection however experimentally infected sheep experience a transient temperature increase in the first day post infection (Pepin *et al.*, 1991b). It has also been postulated that in the early stages of infection *C. pseudotuberculosis* replicates extracellularly (Batey, 1986b). It could be envisioned that *pld* expression may not occur during the early stages of infection as a result of low extracellular bacterial density and repression of *pld* expression by heat shock. This may be a way for the pathogen to replicate without causing excessive tissue damage that would result in the recruitment of immune cells before a sufficient number of bacteria were present to mount a successful infection.

To address the premise that genes that are up-regulated *in vivo*, or in models that mimic aspects of *in vivo* infection, are likely to contribute to virulence, mutant strains of several regulated genes were made and tested for attenuation in the macrophage infection model. Both *sodC* and *met* mutants showed attenuation in the macrophage infection model but not in *in vitro* culture. As it is not clear if mutation of the *met* gene caused the attenuated phenotype it cannot be convincingly answered as to whether the original premise is correct. This study has however shown that lack of attenuation within the macrophage infection model does not necessarily correlate with lack of attenuation *in vivo*. *pld* and *fagC* were both regulated under a variety of conditions. Neither was identified as being macrophage induced, an outcome of using a high-density culture as the *in vitro* control, while searching for macrophage regulated genes. Even though both genes are expressed during macrophage infection, the mutant strains Toxminus and CPfagC were not attenuated in this infection model. However, other studies have shown that *pld* and *fag* mutants are significantly attenuated in the natural hosts, in

particular the ability for abscesses to form in the lymph nodes of infected animals is severely restricted (Hodgson *et al.*, 1992;Hodgson *et al.*, 1994;McNamara *et al.*, 1994;Simmons *et al.*, 1998;Hodgson *et al.*, 1999;Billington *et al.*, 2002). The differences in attenuation in the two models may partially reflect that the macrophage-pathogen interaction is only a part of the host-pathogen interaction. Additionally macrophages within a monolayer will not be exposed to the full gamut of extracellular signals that an *in vivo* macrophage would be subjected to. Although the mouse has clearly been shown to be an effective model for studying the *C. pseudotuberculosis* –host interaction (Jolly, 1965a;Hard, 1972;Batey, 1986a), species differences may also come into play, as in this study a mouse cell line was used to study an interaction with an ovine pathogen. These observations suggest that for strains in which a putative virulence factor has been insertionally inactivated, attenuation should be measured in an animal infection model before a role in virulence is either assigned or disregarded. It remains to be determined whether the *sodC* and *met* mutant strains are attenuated *in vivo*.

This study has shown that useful data regarding gene regulation of a bacterial pathogen can be obtained using fairly simple *in vitro* models. In order to further advance this work the application of screening technologies to whole animal experiments should be performed. Given that transposon mutagenesis methods have not been developed for *C. pseudotuberculosis*, approaches such as STM are currently inapplicable. Although technical difficulties are likely to be encountered, there is no reason why the technologies described in this thesis cannot be used to identify *in vivo* regulated genes of *C. pseudotuberculosis*. In this respect DFI may be particularly advantageous as relatively few bacteria need to be recovered from an infected animal for the DFI process to be applied. Due to the long half-life of Gfp, DFI might also be suited to pathogens for which recovery of bacteria is difficult or takes considerable time. In contrast the requirement of large amounts of intact RNA for DNA array probing may make this approach more problematic.

For some genes regulation occurs at the post-transcriptional level. Genes whose products are regulated at either the translational level or at the protein level may not be identified using the approaches utilised in this study and as such will appear as unregulated genes. In order to circumvent this problem transcriptome and proteome analyses can be performed simultaneously (Yoshida *et al.*, 2001;Eymann *et al.*, 2002). However due to the relative complexities involved in proteomic analysis by two-dimensional electrophoresis it is in

general less comprehensive than global transcriptional profiling (Eymann *et al.*, 2002). Thus although not practical at the moment, future studies correlating gene transcriptional data with protein data should be performed for *C. pseudotuberculosis*.

In conclusion, the work described in this thesis has advanced our knowledge of genetic control in *C. pseudotuberculosis*. In addition to identifying new genes that may be involved in *C. pseudotuberculosis* pathogenesis this study has further elucidated the role and regulation of Pld, the major virulence determinant of *C. pseudotuberculosis*. Further study of the role of regulated genes within the *in vivo* environment will aid in the understanding of the CLA disease process, ultimately leading to better diagnosis, treatment and vaccination regimes.

Appendix 1: Glossary of flow cytometry and FACS analysis terms

dot plot: A graphical means of representing two-parameter data. Each axis of the plot displays values of one parameter. Each dot represents an event.

event: A unit of data representing one particle or cell.

FACS: fluorescence activated cell sorter

FL1: The FACSCalibur flow cytometer used in this study is set up such that FL1 is a measure of emitted light in the green range of the spectrum (515 to 545 nm). In the experiments described in this thesis FL1 data was collected on a logarithmic scale with linear values of 1 to 10,000. Thus an event that has an absolute fluorescence of 100 is five times more fluorescent than one with an absolute fluorescence of 20.

flow cytometry: Analysis of a biological sample by detection of the light-absorbing or fluorescing properties of cells or subcellular fractions passing in a stream of single events through a laser beam.

forward scatter (FSC): A measurement of the light scattering that occurs as a particle passes through the illuminating beam. FSC is related to the size of the particle.

gate: a numerical or graphical boundary (region) that defines a subset of data. An analysis gate excludes or includes specific stored data from subsequent analyses. For example a gate was often set on the basis of FSC and SSC such that only single celled bacteria were subsequently analysed. For sorting experiments logical gates were set such that only those events that lay within two or more specified regions were collected during the sorting process.

histogram: A graphical means of presenting single-parameter data. The horizontal axis of the graph represents the increasing signal intensity of the parameter and the vertical axis represents the number of events (counts).

Linear scale: The scale on which the values increase linearly and output is directly proportional to the input.

Logarithmic scale: The scale on which the values increase logarithmically. This scale is used when the instrument is set to LOG. The scale ranges from 1 to 10,000. A logarithmic scale was used to collect fluorescent data.

mean fluorescence: the average fluorescence intensity of a population of cells.

region: A group of particles that have been selected on the basis of their flow cytometric parameters. In general a boundary is drawn around a sub-population to isolate events for analysis. Regions can be combined to create logical gates.

side scatter (SSC): A measurement of the light that deflects from a particle at an angle of 90° from the direction of the illuminating beam. SSC is a measure of granularity of the particle .

sort: The process by which a FACS is used to enrich for a population of events that possess specified characteristics, for example fluorescence or size.

The definitions given above were adapted from those previously given (Givan, 1992) and those found in the CELLQuest Software User's Guide (Becton Dickinson Immunocytometry systems).

Reference List

Abu Kwaik, Y. and Pederson, L. L. (1996). The use of differential display-PCR to isolate and characterise a *Legionella pneumophila* locus induced during the intracellular infection of macrophages. *Mol Microbiol* **21**:543-556.

Adler, H., Adler, B., Peveri, P., Werner, E. R., Wachter, H., Peterhans, E., and Jungi, T. W. (1996). Differential regulation of inducible nitric oxide synthase production in bovine and caprine macrophages. *J Infect Dis* **173**:971-978.

Amann, E. and Brosius, J. (1985). "ATG vectors" for regulated high-level expression of cloned genes in *Escherichia coli*. *Gene* **40**:183-190.

Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P., Givskov, M., and Molin, S. (1998). New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* **64**:2240-2246.

Ang, S., Lee, C. Z., Peck, K., Sindici, M., Matrubutham, U., Gleeson, M. A., and Wang, J. T. (2001). Acid-induced gene expression in *Helicobacter pylori:* study in genomic scale by microarray. *Infect Immun* **69** :1679-1686.

Arfin, S. M., Long, A. D., Ito, E. T., Tolleri, L., Riehle, M. M., Paegle, E. S., and Hatfield, G. W. (2000). Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factor. *J Biol Chem* **275**:29672-29684.

Baker, S. H., Jin, S., Aldrich, H. C., Howard, G. T., and Shively, J. M. (1998). Insertion mutation of the form I cbbL gene encoding ribulose bisphosphate carboxylase/oxygenase (RuBisCO) in *Thiobacillus neapolitanus* results in expression of form II RuBisCO, loss of carboxysomes, and an increased CO2 requirement for growth. *J Bacteriol* **180**:4133-4139.

Baldwin, D., Crane, V., and Rice, D. (1999). A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr Opin Plant Biol* 2:96-103.

Barker, L. P., Brooks, D. M., and Small, P. L. C. (1998). The identification of *Mycobacterium marinum* genes differentially expressed in macrophage phagosomes using promoter fusions to green fluorescent protein. *Mol Microbiol* **29**:1167-1177.

Barrios, H., Valderrama, B., and Morett, E. (1999). Compilation and analysis of sigma(54)-dependent

promoter sequences. *Nucleic Acids Res* **27**:4305-4313.

Barry, C. E. (2001). Interpreting cell wall "virulence factors" of *Mycobacterium tuberculosis*. *Trends Microbiol* **9**:237-241.

Bassler, B. L. (1999). How bacteria talk to each other: regulation of gene expression by quorum sensing. *Curr Opin Microbiol* **2**:582-587.

Batey, R. G. (1986a). Aspects of pathogenesis in a mouse model of infection of *Corynebacterium pseudotuberculosis*. *Aust J Exp Biol Med Sci* **64**:237-249.

Batey, R. G. (1986b). Pathogenesis of caseous lymphadenitis in sheep and goats. *Aust Vet* 3:269-272.

Bathe, B., Kalinowski , J., and Puhler, A. (1996). A physical and genetic map of the *Corynebacterium glutamicum* ATCC 13032 chromosome. *Mol Gen Genet* **252**:255-65.

Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S., and Small, P. M. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**:1520-1523.

Bertucci, F., Bernard, K., Loriod, B., Chang, Y. C., Granjeaud, S., Birnbaum, D., Nguyen, C., Peck, K., and Jordan, B. R. (1999). Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum Mol Genet* **8**:1715-1722.

Billington, S. J., Esmay, P. A., Songer, J. G., and Jost, B. H. (2002). Identification and role in virulence of putative iron acquisition genes from *Corynebacterium pseudotuberculosis*. *FEMS Microbiol Lett* **208**:41-45.

Billman-Jacobe, H., McConville, M. J., Haites, R. E., Kovacevic, S., and Coppel, R. L. (1999). Identification of a peptide synthetase involved in the biosynthesis of glycopeptidolipids of *Mycobacterium smegmatis*. *Mol Microbiol* **33**:1244-1253.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453-1474.

Bogdan, J. R., Newlands-Monteith, C. F., and Ellis, J. A. (1997). Nitric oxide production following in vitro stimulation of ovine pulmonary alveolar macrophages. *Vet Immunol Immunopathol* **56**:299-310.

Bong, C. T., Fortney, K. R., Katz, B. P., Hood, A. F., San-Mateo, L. R., Kawula, T. H., and Spinola, S. M. (2002). A superoxide dismutase C mutant of *Haemophilus ducreyi* is virulent in human volunteers. *Infect Immun* **70**:1367-1371.

Booth, B. A., Boesman-Finkelstein, M., and Finkelstein, R. A. (1984). *Vibrio cholerae* hemagglutinin/protease nicks cholera enterotoxin. *Infect Immun* **45**:558-560.

Braun, V. and Killmann, H. (1999). Bacterial solutions to the iron-supply problem. *Trends Biochem Sci* **24**:104-109.

Brogden, K. A. and Engen, R. L. (1990). Alterations in the phospholipid composition and morphology of ovine erythrocytes after intravenous inoculation of *Corynebacterium pseudotuberculosis*. *Am J Vet Res* **51**:874-877.

Bucca, G., Hindle, Z., and Smith, C. P. (1997). Regulation of the *dnaK* operon of *Streptomyces coelicolor* A3(2) is governed by HspR, an autoregulatory repressor protein. *J Bacteriol* **179**:5999-6004.

Buchmeier, N. A. and Heffron, F. (1990). Induction of *Salmonella* stress proteins upon infection of macrophages. *Science* **248**:730-732.

Campbell, J. W. and Cronan, J. E. (2001). *Escherichia coli* FadR positively regulates transcription of the fabB fatty acid biosynthetic gene. *J Bacteriol* **183**:5892-90.

Carne, H. R. and Onon, E. O. (1978). Action of *Corynebacterium ovis* exotoxin on endothelial cells of blood vessels. *Nature* **271**:246-248.

Chakrabortty, A., Das, S., Majumdar, S., Mukhopadhyay, K., Roychoudhury, S., and Chaudhuri, K. (2000). Use of RNA arbitrarily primed-PCR fingerprinting to identify *Vibrio cholerae* genes differentially expressed in the host following infection. *Infect Immun* **68**:3878-3887.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green Fluorescent protein as a marker for gene expression. *Science* **263** :802-805.

Challis, G. L. and Ravel, J. (2000). Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiol Lett* **187**:111-4.

Chevalier, J., Pommier, M. T., Cremieux, A., and Michel, G. (1988). Influence of Tween 80 on the mycolic acid composition of three cutaneous corynebacteria. *J Gen Microbiol* **134** :2457-2461.

Chuang, S.-E., Daniels, D. L., and Blattner, F. R. (1993). Global regulation of gene expression in *Escherichia coli*. *J Bacteriol* **175**:2026-2036.

Cohen, P., Bouaboula, M., Bellis, M., Baron, V., Jbilo, O., Poinot-Chazel, C., Galiegue, S., Hadibi, E. H., and Casellas, P. (2000). Monitoring cellular responses to *Listeria monocytogenes* with oligonucleotide arrays. *J Biol Chem* **275**:11181-11190.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. 3., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., and Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537-44.

Colley, W. C., Sung, T. C., Roll, R., Jenco, J., Hammond, S. M., Altshuller, Y., Bar-Sagi, D., Morris, A. J., and Frohman, M. A. (1997). Phospholipase D2, a distinct phospholipase D isoform with novel regulatory properties that provokes cytoskeletal reorganization. *Curr Biol* **7**:191-201.

Collier, L., Balows, A., and Sussman, M. Diphtheria and other coryneform infections. Microbiology and microbial infections. 3. 1998 . Great Britian, Arnold.

Connor, K. M., Quirie, M. M., Baird, G., and Donachie, W. (2000). Characterization of United Kingdom isolates of *Corynebacterium pseudotuberculosis* using pulsed-field gel electrophoresis. *J Clin Microbiol* **38**:2633-2637.

Corish, P. and Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Eng* **12**:1035-1040.

Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**:33-38.

Correia, A., Martin, J. F., and Castro, J. M. (1994). Pulsed-field gel electrophoresis analysis of the genome of amino acid producing corynebacteria: chromosome sizes and diversity of restriction patterns. *Microbiology* **140**:2841-7.

D'Orazio, M., Folcarelli, S., Mariani, F., Colizzi, V., Rotilio, G., and Battistoni, A. (2001). Lipid modification of the Cu,Zn superoxide dismutase from *Mycobacterium tuberculosis*. *Biochem J* **359**:17-22.

Daffe, M. and Draper, P. (1998). The envelope layers of mycobacteria with reference to their

pathogenicity. *Adv Microb Physiol* **39**:131-203.

Dai, X. and Rothman-Denes, L. B. (1999). DNA structure and transcription. *Curr Opin Microbiol* **2**:126-130.

Deichsel, H., Friedel, S., Detterbeck, A., Coyne, C., Hamker, U., and MacWilliams, H. K. (1999). Green fluorescent proteins with short half-lives as reporters in *Dictyostelium discoideum*. *Dev Genes Evol* **209**:63-68.

Desjardins, M., Celis, J. E., van Meer, G., Dieplinger, H., Jahraus, A., Griffiths, G., and Huber, L. A. (1994). Molecular characterization of phagosomes. *J Biol Chem* **269**:32194-32200.

Detweiler, C. S., Cunanan, D. B., and Falkow, S. (2001). Host microarray analysis reveals a role for the *Salmonella* response regulator phoP in human macrophage cell death. *Proc Natl Acad Sci U S A* **98**:5850-5855.

Devaux, P. F. (1991). Static and dynamic lipid asymmetry in cell membranes. *Biochemistry (Mosc)* **30**:1163-1173.

Devescovi, G., Aguilar, C., Majolini, M. B., Marugg, J., Weisbeek, P., and Venturi, V. (2001). A siderophore peptide synthetase gene from plant-growth-promoting *Pseudomonas putida* WCS358. *Syst Appl Microbiol* **24**:321-330.

Dickinson, C. G. and Bull, L. B. (1931). Studies on infection by and resistance to the Preisz-Nocard bacillus. II. Susceptibility of the guinea-pig, and the distribution of lesions after cutaneous and subcutaneous inoculation and ingestion. *Aust J Exp Biol Med Sci* **8**:83.

Dong, Q., Hyde, D., Herra, C., Kean, C., Murphy, P., O'Morain, C. A., and Buckley, M. (2001). Identification of genes regulated by prolonged acid exposure in *Helicobacter pylori*. *FEMS Microbiol Lett* **196**:245-9.

Dorrell, N., Mangan, J. A., Laing, K. G., Hinds, J., Linton, D., Al-Ghusein, H., Barrell, B. G., Parkhill, J., Stoker, N. G., Karlyshev, A. V., Butcher, P. D., and Wren, B. W. (2001). Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res* **11**:1706-1715.

Drazek, E. S., Hammack, C. A., and Schmitt, M. P. (2000). *Corynebacterium diphtheriae* genes required for acquisition of iron from haemin and haemoglobin are homologous to ABC haemin transporters. *Mol Microbiol* **36**:68-84.

Dussurget, O., Stewart, G., Neyrolles, O., Pescher, P., Young, D., and Marchal, G. (2001). Role of *Mycobacterium tuberculosis* copper-zinc superoxide dismutase. *Infect Immun* **69**:529-533.

Dutt, M. J. and Lee, K. H. (2000). Proteomic analysis. *Curr Opin Biotechnol* **11**:176-179.

Eckmann, L., Smith, J. R., Housley, M. P., Dwinell, M. B., and Kagnoff, M. F. (2000). Analysis by high density cDNA arrays of altered gene expression in human intestinal epithelial cells in response to infection with the invasive enteric bacteria Salmonella. *J Biol Chem* **275**:14084-14094.

Eggleton, D. G., Haynes, J. A., Middleton, H. D. , and Cox, J. C. (1991a). Immunisation against ovine caseous lymphadenitis: correlation between *Corynebacterium pseudotuberculosis* toxoid content and protective efficacy in combined clostridial-corynebacterial vaccines. *Aust Vet J* **68**:322-325.

Eggleton, D. G., Middleton, H. D., Doidge, C. V. , and Minty, D. W. (1991b). Immunisation against ovine caseous lymphadenitis: comparison of *Corynebacterium pseudotuberculosis* vaccines with and without bacterial cells. *Aust Vet J* **68**:317-319.

Ellis, T. M., Sutherland, S. S., Wilkinson, F. C., Mercy, A. R., and Paton, M. W. (1987). The role of *Corynebacterium pseudotuberculosis* lung lesions in the transmission of this bacterium to other sheep. *Aust Vet J* **64**:261-263.

Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J., and Leibler, S. (1997). Photoactivation turns green fluorescent protein red. *Curr Biol* **7**:809-812.

Elsinghorst, E. A. (1994). Measurement of invasion by gentamicin resistance. *Methods Enzymol* **236**:405-420.

Engelhorn, M. and Geiselmann, J. (1998). Maximal transcriptional activation by the IHF protein of *Escherichia coli* depends on optimal DNA bending by the activator. *Mol Microbiol* **30**:431-441.

Estrem, S. T., Gaal, T., Ross, W., and Gourse, R. L. (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proc Natl Acad Sci U S A* **95**:9761-9766.

Even, S., Lindley, N. D., and Cocaign-Bousquet, M. (2001). Molecular physiology of sugar catabolism in *Lactococcus lactis* IL1403. *J Bacteriol* **183**:3817-3824.

Exton, J. H. (1997). New developments in phospholipase D. *J Biol Chem* **272**:15579-15582.

Eymann, C., Homouth, G., Scharf, C., and Hecker, M. (2002). *Bacillus subtilis* function genomics: Global characterisation of the stringent response by proteome and transcriptome analysis. *J Bacteriol*

184:2500-2520.

Ezaz-Nikpay, K., Uchino, K., Lerner, R. E., and Verdine, G. L. (1994). Construction of an overproduction vector containing the novel *srp* (sterically repressed) promoter. *Protein Sci* **3**:132-138.

Farrant, J. L., Sansone, A., Canvin, J. R., Pallen, M. J., Langford, P. R., Wallis, T. S., Dougan, G., and Kroll, J. S. (1997). Bacterial copper- and zinc-cofactored superoxide dismutase contributes to the pathogenesis of systemic salmonellosis. *Mol Microbiol* **25**:785-796.

Fassler, J. S. and Gussin, G. N. (1996). Promoters and basal transcription machinery in eubacteria and eukaryotes: concepts, definitions, and analogies. *Methods Enzymol* **273**:3-29.

Fernandez, S., Rojo, F., and Alonso, J. C. (1997). The *Bacillus subtilis* chromatin-associated protein Hbsu is involved in DNA repair and recombination. *Mol Microbiol* **23**:1169-1179.

Fischer, W., Puls, J., Buhrdorf, R., Gebert, B., Odenbreit, S., and Haas, R. (2001). Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol Microbiol* **42**:1337-1348.

Fislage, R. (1998). Differential display approach to quantitation of environmental stimuli on bacterial gene expression. *Electrophoresis* **19**:613-616.

Florell, S. R., Coffin, C. M., Holden, J. A., Zimmermann, J. W., Gerwels, J. W., Summers, B. K., Jones, D. A., and Leachman, S. A. (2001). Preservation of RNA for functional genomic studies: a multidisciplinary tumor bank protocol. *Mod Pathol* **14**:116-28.

Foussard, M., Cabantous, S., Pedelacq, J., Guillet, V., Tranier, S., Mourey, L., Birck, C., and Samama, J. (2001). The molecular puzzle of two-component signaling cascades. *Microbes Infect* **3**:417-424.

Freitag, N. E. and Jacobs, K. E. (1999). Examination of *Listeria monocytogenes* intracellular gene expression by using green fluorescent protein of *Aequorea victoria*. *Infect Immun* **67**:1844-1852.

Gharbi, S., Belaich, A., Murgier, M., and Lazdunski, A. (1985). Multiple controls exerted on in vivo expression of the *pepN* gene in *Escherichia coli*: studies with *pepN-lacZ* operon and protein fusion strains. *J Bacteriol* **163**:1191-1195.

Givan, A. L. (1992). Flow Cytometry First Principles. Wiley-Liss, Inc., USA.

Gonzales, T. and Robert-Baudouy, J. (1996). Bacterial aminopeptidases: properties and functions. *FEMS Microbiol Rev* **18**:319-344.

Gourse, R. L., Ross, W., and Gaal, T. (2000). UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol* **37**:687-695.

Gralla, J. and Collado-Vides, J. (1996). Organization and function of transcriptional regulatory elements.1232-1245.

Grandvalet, C., de'Crecy-Lagard, V., and Mazodier, P. (1999). The ClpB ATPase of *Streptomyces albus* G belongs to the HspR heat shock regulon. *Mol Microbiol* **31**:521-532.

Grandvalet, C., Servant, P., and Mazodier, P. (1997). Disruption of *hspR*, the repressor gene of the *dnaK* operon in *Streptomyces albus G*. *Mol Microbiol* **23**:77-84.

Granjeaud, S., Bertucci, F., and Jordan, B. R. (1999). Expression profiling: DNA arrays in many guises. *Bioessays* **21**:781-790.

Guckenberger, M., Kurz, S., Aepinus, C., Theiss, S., Haller, S., Leimbach, T., Panzner, U., Weber, J., Paul, H., Unkmeir, A., Frosch, M., and Dietrich, G. (2002). Analysis of the heat shock response of *Neisseria meningitidis* with cDNA- and oligonucleotide-based DNA microarrays. *J Bacteriol* **184**:2546-2551.

Guedon, E., Renault, P., Ehrlich, S. D., and Delorme, C. (2001). Transcriptional pattern of genes coding for the proteolytic system of *Lactococcus lactis* and evidence for coordinated regualtion of key enzymes by peptide supply. *J Bacteriol* **183**:3614-3622.

Guerreiro, H., Croda, J., Flannery, B., Mazel, M., Matsunaga, J., Galvao Reis, M., Levett, P. N., Ko, A. I., and Haake, D. A. (2001). Leptospiral proteins recognized during the humoral immune response to leptospirosis in humans. *Infect Immun* **69**:4958-68.

Gulig, P. A., Caldwell, A. L., and Chiodo, V. A. (1992). Identification, genetic analysis and DNA sequence of a 7.8-kb virulence region of the *Salmonella typhimurium* virulence plasmid. *Mol Microbiol* **6**:1395-1411.

Hanawa, T., Fukuda, M., Kawakami, H., Hirano, H., Kamiya, S., and Yamamoto, T. (1999). The *Listeria monocytogenes* DnaK chaperone is required for stress tolerance and efficient phagocytosis with macrophages. *Cell Stress Chaperones* **4**:118-128.

Handfield, M., Brady, L. J., Progulske-Fox, A., and Hillman, J. D. (2000). IVIAT: a novel method to identify microbial genes expressed specifically during human infections. *Trends Microbiol* **8**:336-339.

Hanna, M. N., Ferguson, R. J., Li, Y. H., and Cvitkovitch, D. G. (2001). *uvrA* is an acid-inducible gene

involved in the adaptive response to low pH in *Streptococcus mutans*. *J Bacteriol* **183**:5964-73.

Hard, G. C. (1969a). Electron microscopic examination of *Corynebacterium ovis*. *J Bacteriol* **97**:1480-1485.

Hard, G. C. (1969b). Electron microscopic study of the differentiation of mouse peritoneal macrophages stimulated by *Corynebacterium ovis* infection. *Lab Invest* **21**:309-315.

Hard, G. C. (1970). Adoptive transfer of immunity in experimental *Corynebacterium ovis* infection. *J Comp Pathol* **80**:329-334.

Hard, G. C. (1972). Examination by electron microscopy of the interaction between peritoneal phagocytes and *Corynebacterium ovis*. *J Med Microbiol* **5**:483-491.

Havarstein, L. S., Coomaraswamy, G., and Morrison, D. A. (1995). An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* **92**:11140-11144.

Hayashi, K., Nakazawa, M., Ishizaki, Y., Hiraoka, N., and Obayashi, A. (1986). Regulation of inter- and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol. *Nucleic Acids Res* **14**:7617-7631.

Hayashi, S. and Wu, H. C. (1990). Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**:451-71.

Helmann, J. D., Wu, M. F., Kobel, P. A., Gamo, F. J., Wilson, M., Morshedi, M. M., Navre, M., and Paddon, C. (2001). Global transcriptional response of *Bacillus subtilis* to heat shock. *J Bacteriol* **183**:7318-7328.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., and Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**:400-403.

Hodgson, A. L., Krywult, J., Corner, L. A., Rothel, J. S., and Radford, A. J. (1992). Rational attenuation of *Corynebacterium pseudotuberculosis:* potential cheesy gland vaccine and live delivery vehicle. *Infect Immun* **60**:2900-2905.

Hodgson, A. L. M., Bird, P., and Nisbet, I. T. (1990). Cloning, nucleotide sequence, and expression in *Escherichia coli* of the Phospholipase D gene from *Corynebacterium pseudotuberculosis*. *J Bacteriol* **172**:1256-1261.

Hodgson, A. L. M., Carter, K., Tachedjian, M., Krywult, J., Corner, L. A., McColl, M., and Cameron,

A. (1999). Efficacy of an ovince caseous lymphadenitis vaccine formulated using a genetically inactive form of *Corynebacterium pseudotuberculosis* phospholipase D. *Vaccine* **17**:802-808.

Hodgson, A. L. M., Tachedjian, M., Corner, L. A., and Radford, A. J. (1994). Protection of sheep against Caseous Lymphadenitis by use of a single oral dose of live recombinant *Corynebacterium pseudotuberculosis*. *Infect Immun* **62**:5275-5280.

Hoevel, T., Holz, H., and Kubbies, M. (1999). Cisplatin-digoxigenin mRNA labelling for nonradioactive detection of mRNA hybridized onto nucleic acid cDNA arrays. *Biotechniques* **27**:1064-1067.

Hoffman, P. S., Butler, C. A., and Quinn, F. D. (1989). Cloning and temperature-dependent expression in *Escherichia coli* of a *Legionella pneumophila* gene coding for a genus-common 60-kilodalton antigen. *Infect Immun* **57**:1731-1739.

Hsu, T. Y., Renshaw, H. W., Livingston, C. W., Augustine, J. L., Zink, D. L., and Gauer, B. B. (1985). *Corynebacterium pseudotuberculosis* exotoxin: fatal hemolytic anemia induced in gnotobiotic neonatal small ruminants by parenteral administration of preparations containing exotoxin. *Am J Vet Res* **46**:1206-1211.

Hu, Y. and Coates, A. R. (2001). Increased levels of sigJ mRNA in late stationary phase cultures of *Mycobacterium tuberculosis* detected by DNA array hybridisation. *FEMS Microbiol Lett* **202**:59-65.

Hurme, R. and Rhen, M. (1998). Temperature sensing in bacterial gene regulation--what it all boils down to. *Mol Microbiol* **30**:1-6.

Ichikawa, J. K., Norris, A., Bangera, M. G., Geiss, G. K., van't Wout, A. B., Bumgarner, R. E. , and Lory, S. (2000). Interaction of *Pseudomonas aeruginosa* with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs. *Proc Natl Acad Sci U S A* **97**:9659-9664.

Israel, D. A., Salama, N., Krishna, U., Rieger, U. M., Atherton, J. C., Falkow, S., and Peek, R. M. (2001). *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci U S A* **98**:14625-14630.

Jager, W., Peters-Wendisch, P. G., Kalinowski, J., and Puhler, A. (1996). A *Corynebacterium glutamicum* gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins. *Arch Microbiol* **166**:76-82.

Ji, G., Beavis, R. C., and Novick, R. P. (1995). Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. *Proc Natl Acad Sci U S A* **92**:12055-12059.

Jolly, R. D. (1965a). The pathogenesis of experimental *Corynebacterium ovis* infection in mice. *N Z Vet J* **13**:141-147.

Jolly, R. D. (1965b). The pathogenic action of the exotoxin of *Corynebacterium ovis*. *J Comp Pathol* **75**:417-429.

Kansal, R. G., Gomez-Flores, R., and Mehta, R. T. (1998). Change in colony morphology influences the virulence as well as the biochemical properties of the *Mycobacterium avium* complex. *Microb Pathog* **25**:203-214.

Katayama, S., Matsushita, O., Jung, C. M., Minami, J., and Okabe, A. (1999). Promoter upstream bent DNA activates the transcription of the *Clostridium perfringens* phospholipase C gene in a low temperature-dependent manner. *EMBO J* **18**:3442-3450.

Katayama, S., Matsushita, O., Tamai, E., Miyata, S., and Okabe, A. (2001). Phased A-tracts bind to the alpha subunit of RNA polymerase with increased affinity at low temperature. *FEBS* **509**:235-238.

Keiler, K. C., Waller, P. R., and Sauer, R. T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**:990-993.

Kendall, J. M. and Badminton, M. N. (1998). *Aequorea victoria* bioluminescence moves into an exciting new era. *Trends Biotechnol* **16**:216-224.

Khelef, N., DeShazer, D., Friedman, R. L., and Guiso, N. (1996). In vivo and in vitro analysis of *Bordetella pertussis* catalase and Fe-superoxide dismutase mutants. *FEMS Microbiol Lett* **142**:231-235.

Khodursky, A. B.,.Peter, B. J., Cozzarelli, N. R., Botstein, D., Brown, P. O., and Yanofsky, C. (2000). DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**:12170-12175.

Kiatpapan, P., Kobayashi, H., Sakaguchi, M., Ono, H., Yamashita, M., Kaneko, Y., and Murooka, Y. (2001). Molecular characterization of *Lactobacillus plantarum* genes for beta-ketoacyl-acyl carrier protein synthase III (*fabH*) and acetyl coenzyme A carboxylase (*accBCDA*), which are essential for fatty acid biosynthesis. *Appl Environ Microbiol* **67**:426-33.

Kimura, E., Abe, C., Kawahara, Y., and Nakamatsu, T. (1996). Molecular cloning of a novel gene,

dtsR, which rescues the detergent sensitivity of a mutant derived from *Brevibacterium lactofermentum*. *Biosci Biotechnol Biochem* 60:1565-1570.

Kimura, E., Abe, C., Kawahara, Y., Nakamatsu, T., and Tokuda, H. (1997). A *dtsR* gene-disrupted mutant of *Brevibacterium lactofermentum* requires fatty acids for growth and efficiently produces L-glutamate in the presence of an excess of biotin. *Biochem Biophys Res Commun* 234:157-161.

Kohler, P. and Marahiel, M. A. (1997). Association of the histone-like protein HBsu with the nucleoid of *Bacillus subtilis*. *J Bacteriol* 179:2060-2064.

Konkel, M. E. and Tilly, K. (2000). Temperature-regulated expression of bacterial virulence genes. *Microbes Infect* 2:157-166.

Kothekar, V., Raha, K., and Prasad, H. K. (1998). Molecular dynamics simulation of interaction of histone-like protein of *Mycobacterium tuberculosis* (Hlpmt) and histone of *Clostridium pasteurianum*. (DBHclopa) with 35 based paired GC rich U-bend DNA. *J Biomol Struct Dyn* 16:223-235.

Kremer, L., Baulard, A. R., and Besra, G. S. (2000). Genetics of mycolic acid biosynthesis . *In* Molecular Genetics of Mycobacteria. pp. 173-190, ASM Press, USA.

Kuehnel, M. P., Goethe, R., Habermann, A., Mueller, E., Rohde, M., Griffiths, G., and Valentin-Weigand, P. (2001). Characterization of the intracellular survival of *Mycobacterium avium ssp. paratuberculosis*: phagosomal pH and fusogenicity in J774 macrophages compared with other mycobacteria. *Cell Microbiol* 3:551-566.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., and Danchin, A. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249-256.

Kusner, D. J., Hall, C. F., and Schlesinger, L. S. (1996). Activation of phospholipase D is tightly coupled to the phagocytosis of *Mycobacterium tuberculosis* or opsonized zymosan by human macrophages. *J Exp Med* 184:585-595.

Kusner, D. and Adams, J. (2000). ATP-induced killing of virulent *Mycobacterium tuberculosis* within human macrophages requires Phospholipase D. *J Immunol* 164:379-388.

Langford, P. R., Loynds, B. M., and Kroll, J. S. (1996). Cloning and molecular characterization of

Cu,Zn superoxide dismutase from *Actinobacillus pleuropneumoniae*. *Infect Immun* **64**:5035-5041.

Latimer, E., Simmers, J., Sriranganathan, N., Roop, R. M., Schurig, G. G., and Boyle, S. M. (1992). *Brucella abortus* deficient in copper/zinc superoxide dismutase is virulent in BALB/c mice. *Microb Pathog* **12**:105-113.

Lennartz, M. R. (1999). Phospholipases and phagocytosis: the role of phospholipid-derived second messengers in phagocytosis. *Int J Biochem Cell Biol* **31**:415-430.

Li, M. S., Monahan, I. M., Waddell, S. J., Mangan, J. A., Martin, S. L., Everett, M. J., and Butcher, P. D. (2001). cDNA-RNA subtractive hybridization reveals increased expression of mycocerosic acid synthase in intracellular *Mycobacterium bovis* BCG. *Microbiology* **147**:2293-2305.

Li, S. J. and Cronan, J. E. (1992). The genes encoding the two carboxyltransferase subunits of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* **267**:16841-16847.

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C. C., and Kain, S. R. (1998). Generation of destabilized green fluorescent protein as a transcription reporter. *J Biol Chem* **273**:34970-34975.

Linder, R. and Bernheimer, A. W. (1997). Oxidation of macrophage membrane cholesterol by intracellular *Rhodococcus equi*. *Vet Microbiol* **56**:269-276.

Lissemore, J. L., Jankowski, J. T., Thomas, C. B., Mascotti, D. P., and deHaseth, P. L. (2000). Green fluorescent protein as a quantitative reporter of relative promoter activity in *E. coli*. *Biotechniques* **28**:82-89.

Liu, Y. and Tsinoremas, N. F. (1996). An unusual gene arrangement for the putative chromosome replication origin and circadian expression of *dnaN* in *Synechococcus sp.* strain PCC 7942. *Gene* **172**:105-109.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* **14**:1675-1680.

Lonetto, M., Gribskov, M., and Gross, C. A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* **174**:3843-3849.

Lowik, C. W., Alblas, M. J., van de Ruit, M., Papapoulos, S. E., and van der Pluijm, G. (1993). Quantification of adherent and nonadherent cells cultured in 96-well plates using the supravital stain

neutral red. *Anal Biochem* **213**:426-433.

Lucchini, S., Thompson, A., and Hinton, J. C. (2001). Microarrays for microbiologists. *Microbiology* **147**:1403-1414.

Lund, P. A. (2001). Microbial Molecular Chaperones. *Adv Microb Physiol* **44**: 93-140

Ma, Y., Stern, R. J., Scherman, M. S., Vissa, V. D., Yan, W., Jones, V. C., Zhang, F., Franzblau, S. G., Lewis, W. H., and McNeil, M. R. (2001). Drug targeting *Mycobacterium tuberculosis* cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose. *Antimicrob Agents Chemother* **45**:1407-16.

MacMicking, J., Xie, Q. W., and Nathan, C. (1997). Nitric oxide and macrophage function. *Annu Rev Immunol* **15**:323-350.

Madiraju, M. V., Qin, M. H., Yamamoto, K., Atkinson, M. A., and Rajagopalan, M. (1999). The *dnaA* gene region of *Mycobacterium avium* and the autonomous replication activities of its 5' and 3' flanking regions. *Microbiology* **145** :2913-2921.

Magnuson, K., Jackowski, S., Rock, C. O., and Cronan, J. E. (1993). Regulation of fatty acid biosynthesis in *Escherichia coli*. *Microbiol Rev* **57**:522-542.

Mahan, M. J., Slauch, J. M., and Mekalanos, J. J. (1993). Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* **259**:686-688.

Mahan, M. J., Tobias, J. W., Slauch, J. M., Hanna, P. C., Collier, R. J., and Mekalanos, J. J. (1995). Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. *Proc Natl Acad Sci U S A* **92**:669-673.

Mangan, J. A., Monahan, I. M., Wilson, M. A., Schnappinger, D., Schoolnik, G. K., and Butcher, P. D. (1999). The expression profile of *Mycobacterium tuberculosis* infecting the human monocytic cell line THP-1 using whole genome microarray analysis. Proceedings of The Microarray Meeting: Technology, Application and Analysis.

Matz, M. V. and Lukyanov, S. A. (1998). Different strategies of differential display: areas of application. *Nucleic Acids Res* **26**:5537-5543.

McNamara, P. J., Bradley, G. A., and Songer, J. G. (1994). Targeted mutagenesis of the phospholipase D gene results in decreased virulence of *Corynebacterium pseudotuberculosis*. *Mol Microbiol* **12**:921-

930.

McNeil, M., Daffe, M., and Brennan, P. J. (1990). Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls. *J Biol Chem* **265**:18200-18206.

Merrick, M. J. (1993). In a class of its own-the RNA polymerase sigma factor sigma 54 (sigma N). *Mol Microbiol* **10**:903-909.

Micka, B., Groch, N., Heinemann, U., and Marahiel, M. A. (1991). Molecular cloning, nucleotide sequence, and characterization of the *Bacillus subtilis* gene encoding the DNA-binding protein HBsu. *J Bacteriol* **173**:3191-3198.

Mierau, I., Kunji, E. R., Leenhouts, K. J., Hellendoorn, M. A., Haandrikman, A. J., Poolman, B., Konings, W. N., Venema, G., and Kok, J. (1996). Multiple-peptidase mutants of *Lactococcus lactis* are severely impaired in their ability to grow in milk. *J Bacteriol* **178**:2794-2803.

Mills, A. E., Mitchell, R. D., and Lim, E. K. (1997). *Corynebacterium pseudotuberculosis* is a cause of human necrotising granulomatous lymphadenitis. *Pathology (Phila)* **29**:231-233.

Miyoshi, S. and Shinoda, S. (2000). Microbial metalloproteases and pathogenesis. *Microbes Infect* **2**:91-98.

Mogk, A., Tomoyasu, T., Goloubinoff, P., Rudiger, S., Roder, D., Langen, H., and Bukau, B. (1999). Identification of thermolabile *Escherichia coli* proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J* **18**:6934-6949.

Moore, R. J., Rothel, L., Krywalt, J., Radford, A. J., Lund, K., and Hodgson, A. L. M. (2000). Foreign gene expression in *Corynebacterium pseudotuberculosis*: development of a live vaccine vector. *Vaccine* **18**:487-497.

Moore, R. J., Stewart, D. J., Lund, K., and Hodgson, A. L. (2001). Vaccination against ovine footrot using a live bacterial vector to deliver basic protease antigen. *FEMS Microbiol Lett* **194**:193-196.

Mootz, H. D. and Marahiel, M. A. (1999). Design and application of multimodular peptide synthetases. *Curr Opin Biotechnol* **10**:341-348.

Morett, E. and Segovia, L. (1993). The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J Bacteriol* **175**:6067-6074.

Motohashi, K., Watanabe, Y., Yohda, M., and Yoshida, M. (1999). Heat-inactivated proteins are

rescued by the DnaK.J-GrpE set and ClpB chaperones. *Proc Natl Acad Sci U S A* **96**:7184-7189.

Muckle, C. A. and Gyles, C. L. (1983). Relation of lipid content and exotoxin production to virulence of *Corynebacterium pseudotuberculosis* in mice. *Am J Vet Res* **44** :1149-1153.

Nagel, A., Fleming, J. T., and Sayler, G. S. (1999). Reduction of false positives in prokaryotic mRNA differential display. *Biotechniques* **26**:641-648.

Nagel, G., Lahrz, A., and Dersch, P. (2001). Environmental control of invasin expression in *Yersinia pseudotuberculosis* is mediated by regulation of RovA, a transcriptional activator of the SlyA/Hor family. *Mol Microbiol* **41**:1249-1269.

Nairn, M. E. and Robertson, J. P. (1974). *Corynebacterium pseudotuberculosis* infection of sheep: role of skin lesions and dipping fluids. *Aust Vet J* **50**:537-542.

Narberhaus, F. (1999). Negative regulation of bacterial heat shock genes. *Mol Microbiol* **31**:1-8.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**:1-6.

Norman, E., Dellagostin, O. A., McFadden, J., and Dale, J. W. (1995). Gene replacement by homologous recombination in *Mycobacterium bovis* BCG. *Mol Microbiol* **16**:755-760.

Norrander, J., Kempe, T., and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**:101-106.

Oh-Hashi, Y., Kawaguchi, A., Seyama, Y., and Okuyama, H. (1986). Temperature- and growth-phase-dependent changes in membrane fatty acid compositions of *Brevibacterium ammoniagenes*. *Arch Biochem Biophys* **248**:440-443.

Oh, M. K., Rohlin, L., Kao, K. C., and Liao, J. C. (2002). Global expression profiling of acetate-grown *Escherichia coli*. *J Biol Chem* **277**:13175-13183.

Olson, S. C. and Lambeth, J. D. (1996). Biochemistry and cell biology of phospholipase D in human neutrophils. *Chem Phys Lipids* **80**:3-19.

Orihuela, C. J., Mills, J., Robb, C. W., Wilson, C. J., Watson, D. A., and Niesel, D. W. (2001): *Streptococcus pneumoniae* PstS production is phosphate responsive and enhanced during growth in the murine peritoneal cavity. *Infect Immun* **69**:7565-7571.

Ozenberger, B. A., Nahlik, M. S., and McIntosh, M. A. (1987). Genetic organization of multiple *fep* genes encoding ferric enterobactin transport functions in *Escherichia coli*. *J Bacteriol* **169**:3638-3646.

Pátek, M., Eikmanns, B. J., Pátek, J., and Sahm, H. (1996). Promoters from *Corynebacterium glutamicum*; cloning, molecular analysis and search for a consensus motif. *Microbiology* **142**:1297-1309.

Paton, M. (1997). The epidemiology of Caseous lymphadenitis in Australia and observations on other production systems. Proceedings of the 101st meating of the US Animal Health Association.

Paton, M. W., Rose, I. R., Hart, R. A., Sutherland, S. S., Mercy, A. R., Ellis, T. M., and Dhaliwal, J. A. (1994). New infection with *Corynebacterium pseudotuberculosis* reduces wool production. *Aust Vet J* **71**:4749.

Paton, M. W., Sutherland, S. S., Rose, I. R., Hart, R. A., Mercy, A. R., and Ellis, T. M. (1995). The spread of *Corynebacterium pseudotuberculosis* infection to unvaccinated and vaccinated sheep. *Aust Vet J* **72**:266-269.

Paustian, M. L., May, B. J., and Kapur, V. (2001). *Pasteurella multocida* gene expression in response to iron limitation. *Infect Immun* **69**:4109-4115.

Peel, M. M., Palmer, G. G., Stacpoole, A. M., and Kerr, T. G. (1997). Human lymphadenitis due to *Corynebacterium pseudotuberculosis*: report of ten cases from Australia and review. *Clin Infect Dis* **24**:185-191.

Pepin, M., Fontaine, J. J., Pardon, P., Marly, J., and Parodi, A. L. (1991a). Histopathology of the early phase during experimental *Corynebacterium pseudotuberculosis* infection in lambs. *Vet Microbiol* **29**:123-134.

Pepin, M., Pardon, P., Lantier, F., Marly, J., Levieux, D., and Lamand, M. (1991b). Experimental *Corynebacterium pseudotuberculosis* infection in lambs: kinetics of bacterial dissemination and inflammation. *Vet Microbiol* **26**:381-392.

Pepin, M., Pittet, J. C., Olivier, M., and Gohin, I. (1994). Cellular composition of *Corynebacterium pseudotuberculosis* pyogranulomas in sheep. *J Leukoc Biol* **56**:666-670.

Phillips, G. J. (2001). Green fluorescent protein-a bright idea for the study of bacterial protein localization. *FEMS Microbiol Lett* **204**:9-18.

Phimister, B. (1999). A note on nomenclature. *Nature Genetics, supplement* **21**:1.

Piddington, D. L., Fang, F. C., Laessig, T., Cooper, A. M., Orme, I. M., and Buchmeier, N. A. (2001). Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. *Infect Immun* **69**:4980-4987.

Piontkowski, M. D. and Shivvers, D. W. (1998). Evaluation of a commercially available vaccine against *Corynebacterium pseudotuberculosis*. *J Am Vet Med Assoc* **212**:1765-1768.

Plum, G. and Clark-Curtiss, J. E. (1994). Induction of *Mycobacterium avium* gene expression following phagocytosis by human macrophages. *Infect Immun* **62**:476-483.

Polgar, L. (2002). The prolyl oligopeptidase family. *Cell Mol Life Sci* **59**:349-362.

Pomposiello, P. J., Bennik, M. H., and Demple, B. (2001). Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. *J Bacteriol* **183**:3890-3902.

Prabhakar, S., Annapurna, P. S., Jain, N. K., Dey, A. B., Tyagi, J. S., and Prasad, H. K. (1998). Identification of an immunogenic histone-like protein (HLPMt) of Mycobacterium tuberculosis. *Tuber Lung Dis* **79**:43-53.

Qin, M. H., Madiraju, M. V., and Rajagopalan, M. (1999). Characterization of the functional replication origin of *Mycobacterium tuberculosis*. *Gene* **233**:121-130.

Qin, M. H., Madiraju, M. V., Zachariah, S., and Rajagopalan, M. (1997). Characterization of the *oriC* region of *Mycobacterium smegmatis*. *J Bacteriol* **179**:6311-6317.

Qoronfleh, M. W., Bortner, C. A., Schwartzberg, P., and Wilkinson, B. J. (1998). Enhanced levels of *Staphylococcus aureus stress* protein GroEL and DnaK homologs early in infection of human epithelial cells. *Infect Immun* **66**:3024-3027.

Qoronfleh, M. W., Weraarchakul, W., and Wilkinson, B. J. (1993). Antibodies to a range of *Staphylococcus aureus* and *Escherichia coli* heat shock proteins in sera from patients with S. aureus endocarditis. *Infect Immun* **61**:1567-1570.

Radostits, O. M., Gay, C. C., Blood, D. C., and Hinchcliff, K. W. (1999). Veterinary Medicine. A Textbook of the diseases of cattle, sheep, pigs, goats and Horses. W.B. Saunders Company Ltd, USA.

Rajagopalan, M., Qin, M. H., Nash, D. R., and Madiraju, M. V. (1995). *Mycobacterium smegmatis dnaA* region and autonomous replication activity. *J Bacteriol* **177**:6527-6535.

Ralph, P. and Nakoinz, I. (1975). Phagocytosis and cytolysis by a macrophage tumour and its cloned cell line. *Nature* **257**:393-394.

Ramakrishnan, L., Federspiel, N. A., and Falkow, S. (2000). Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. *Science* **288**:1436-1439.

Richmond, C. S., Glasner, J. D., Mau, R., Jin, H., and Blattner, F. R. (1999). Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res* **27**:3821-3835.

Riegel, P., Ruimy, R., de'Briel, D., Prevost, G., Jehl, F., Christen, R., and Monteil, H. (1995). Taxonomy of *Corynebacterium diphtheriae* and related taxa, with recognition of *Corynebacterium ulcerans* sp. nov. nom. rev. *FEMS Microbiol Lett* **126**:271-276.

Rimini, R., Jansson, B., Feger, G., Roberts, T. C., de'Francesco, M., Gozzi, A., Faggioni, F., Domenici, E., Wallace, D. M., Frandsen, N., and Polissi, A. (2000). Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. *Mol Microbiol* **36**:1279-1292.

Risoen, P. A., Brurberg, M. B., Eijsink, V. G., and Nes, I. F. (2000). Functional analysis of promoters involved in quorum sensing-based regulation of bacteriocin production in *Lactobacillus*. *Mol Microbiol* **37**:619-628.

Robb, C. W., Orihuela, C. J., Ekkelenkamp, M. B. , and Niesel, D. W. (2001). Identification and characterization of an in vivo regulated D15/Oma87 homologue in *Shigella flexneri* using differential display polymerase chain reaction. *Gene* **262**:169-77.

Rodrigue, A., Quentin, Y., Lazdunski, A., Mejean, V., and Foglino, M. (2000). Two-component systems in *Pseudomonas aeruginosa*: why so many? *Trends Microbiol* **8**:498-504.

Roggenkamp, A., Bittner, T., Leitritz, L., Sing, A., and Heesemann, J. (1997). Contribution of the Mn-cofactored superoxide dismutase (SodA) to the virulence of *Yersinia enterocolitica* serotype O8. *Infect Immun* **65**:4705-4710.

Rose, R. E. (1988). The nucleotide sequence of pACYC184. *Nucleic Acids Res* **16**:355.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, Cold Spring Habor Laboratory Press , USA.

San-Mateo, L. R., Hobbs, M. M., and Kawula, T. H. (1998). Periplasmic copper-zinc superoxide dismutase protects *Haemophilus ducreyi* from exogenous superoxide. *Mol Microbiol* **27**:391-404.

San-Mateo, L. R., Toffer, K. L., Orndorff, P. E. , and Kawula, T. H. (1999). Neutropenia restores virulence to an attenuated Cu,Zn superoxide dismutase-deficient *Haemophilus ducreyi* strain in the swine model of chancroid. *Infect Immun* **67**:5345-5351.

Schmitt, M. P. (1999). Identification of a two-component signal transduction system from *Corynebacterium diphtheriae* that activates gene expression in response to the presence of heme and hemoglobin. *J Bacteriol* **181**:5330-5340.

Schubert, S., Fischer, D., and Heesemann, J. (1999). Ferric enterochelin transport in *Yersinia enterocolitica*: molecular and evolutionary aspects. *J Bacteriol* **181**:6387-6395.

Schummer, M., Ng, W.-I., Nelson. P.S., Bumgarner, R. E., and Hood, L. (1997). Inexpensive handheld device for the construction of high-density nucleic arrays. *Biotechniques* **23**:1087-1092.

Schwarzer, N., Nost, R., Seybold, J., Parida, S. K., Fuhrmann, O., Krull, M., Schmidt, R., Newton, R., Hippenstiel, S., Domann, E., Chakraborty, T., and Suttorp, N. (1998). Two distinct phospholipases C of *Listeria monocytogenes* induce ceramide generation, nuclear factor-kappa B activation, and E-selectin expression in human endothelial cells. *J Immunol* **161**:3010-3019.

Servant, P. and Mazodier, P. (2001). Negative regulation of the heat shock response in *Streptomyces*. *Arch Microbiol* **176**:237-242.

Sheehan, B. J., Langford, P. R., Rycroft, A. N., and Kroll, J. S. (2000). [Cu,Zn]-Superoxide dismutase mutants of the swine pathogen *Actinobacillus pleuropneumoniae* are unattenuated in infections of the natural host. *Infect Immun* **68**:4778-4781.

Shimomura, O., Johnson, F. H., and Saiga, Y. (1962). Extraction, purification and properties of Aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol* **59**:223-39.

Shires, K. and Steyn, L. (2001). The cold-shock stress response in *Mycobacterium smegmatis* induces the expression of a histone-like protein. *Mol Microbiol* **39**:994-1009.

Simmons, C. P. Development of *Corynebacterium pseudotuberculosis* for use as a bacterial vaccine vector. 1997. Melbourne, Australia, University of Melbourne.

Simmons, C. P., Dunstan, S. J., Tachedjian, M., Krywult, J., Hodgson, A. L. M., and Strugnell, R. A. (1998). Vaccine potential of attenuated mutants of *Corynebacterium pseudotuberculosis* in sheep. *Infect Immun* **66**:474-479.

Simmons, C. P., Hodgson, A. L. M., and Strugnell, R. A. (1997). Attenuation and vaccine potential of *aroQ* mutants of *Corynebacterium pseudotuberculosis*. *Infect Immun* **65**:3048-3056.

Smoot, L. M., Smoot, J. C., Graham, M. R., Somerville, G. A., Sturdevant, D. E., Migliaccio, C. A., Sylva, G. L., and Musser, J. M. (2001). Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. *Proc Natl Acad Sci U S A* **98**:10416-10421.

Songer, J. G., Beckenbach, K., Marshall, M. M., Olson, G. B., and Kelley, L. (1988). Biochemical and genetic characterization of *Corynebacterium pseudotuberculosis*. *Am J Vet Res* **49**:223-226.

Stachelhaus, T. and Marahiel, M. A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol Lett* **125**:3-14.

Stewart, G. R., Snewin, V. A., Walzl, G., Hussell, T., Tormay, P., O'Gaora, P., Goyal, M., Betts, J., Brown, I. N., and Young, D. B. (2001). Overexpression of heat-shock proteins reduces survival of *Mycobacterium tuberculosis* in the chronic phase of infection. *Nat Med* **7**:732-737.

Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. *Annu Rev Biochem* **69**:183-215.

Stuible, H. P., Wagner, C., Andreou, I., Huter, G., Haselmann, J., and Schweizer, E. (1996). Identification and functional differentiation of two type I fatty acid synthases in *Brevibacterium ammoniagenes*. *J Bacteriol* **178**:4787-4793.

Sutcliffe, I. C. (2000). Cell envelope composition and organisation in the genus *Rhodococcus*. *Antonie Van Leeuwenhoek* **74**:49-58.

Tachedjian, M., Krywult, J., Moore, R. J., and Hodgson, A. L. (1995). Caseous lymphadenitis vaccine development: site-specific inactivation of the *Corynebacterium pseudotuberculosis* phospholipase D gene. *Vaccine* **13**:1785-1792.

Takahashi, T., Mori, Y., Kobayashi, H., Ochi, M., Kikuchi, N., and Hiramune, T. (1997). Phylogenetic positions and assignment of swine and ovine corynebacterial isolates based on the 16S rDNA sequence. *Microbiol Immunol* **41**:649-655.

Talaat, A. M., Hunter, P., and Johnston, S. A. (2000). Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nat Biotechnol* **18**:679-682.

Tang, S. W., Abubakar, S., Devi, S., Puthucheary, S., and Pang, T. (1997). Induction and characterization of heat shock proteins of *Salmonella typhi* and their reactivity with sera from patients

with typhoid fever. *Infect Immun* **65**:2983-2986.

Tao, H., Bausch, C., Richmond, C., Blattner, F. R., and Conway, T. (1999). Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* **181**:6425-6440.

Tashjian, J. J. and Campbell, S. G. (1983). Interaction between caprine macrophages and corynebacterium pseudotuberculosis: an electron microscopic study. *Am J Vet Res* **44**:690-693.

Tatum, F. M., Detilleux, P. G., Sacks, J. M., and Halling, S. M. (1992). Construction of Cu-Zn superoxide dismutase deletion mutants of *Brucella abortus:* analysis of survival in vitro in epithelial and phagocytic cells and in vivo in mice. *Infect Immun* **60**:2863-2869.

Tonetti, M., Sturla, L., Bisso, A., Zanardi, D., Benatti, U., and De'Flora, A. (1998). The metabolism of 6-deoxyhexoses in bacterial and animal cells. *Biochimie* **80**:923-931.

Tran, L. S., Nagai, T., and Itoh, Y. (2000). Divergent structure of the ComQXPA quorum-sensing components: molecular basis of strain-specific communication mechanism in *Bacillus subtilis*. *Mol Microbiol* **37**:1159-1171.

Trenkle, T., Welsh, J., Jung, B., Mathieu-Daude, F., and McClelland, M. (1998). Non-stoichiometric reduced complexity probes for cDNA arrays. *Nucleic Acids Res* **26**:3883-3891.

Triccas, J. A., Berthet, F.-X., Pelicic, V., and Gicquel, B. (1999). Use of fluorescence induction and sucrose counterselection to identify *Mycrobacterium tuberculosis* genes expressed within host cells. *Microbiology* **145**:2923-2930.

Tsai, J. P. and Shi, W. (2000). Analysis of gene expression in *Treponema denticola* with differential display polymerase chain reaction. *Oral Microbiol Immunol* **15**:305-8.

Tsien, R. Y. (1998). The green fluorescent protein. *Annual Reviews in Biochemistry* **67**:509-544.

Valdivia, R. H. and Falkow, S. (1996). Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhurium* acid-inducible promoters by differential fluorescence induction. *Mol Microbiol* **22**:367-378.

Valentin-Weigand, P., Benkel, P., Rohde, M., and Chhatwal, G. S. (1996). Entry and intracellular survival of group B streptococci in J774 macrophages. *Infect Immun* **64**:2467-2473.

Van-Boxtel, R. M., Lambrecht, R. S., and Collins, M. T. (1990). Effects of colonial morphology and

tween 80 on antimicrobial susceptibility of *Mycobacterium paratuberculosis*. *Antimicrob Agents Chemother* **34**:2300-2303.

Viter, S., Shaw, K. J., and Gennaro, M. L. (1999). HSa of Staphylococcus aureus, a new member of the HU family of bacterial histone-like proteins. *Res Microbiol* **150**:287-290.

Voskuil, M. I. and Chambliss, G. H. (1998). The -16 region of *Bacillus subtilis* and other gram-positive bacterial promoters. *Nucleic Acids Res* **26**:3584-3590.

Walker, J., Jackson, H. J., Eggleton, D. G., Meeusen, E. N., Wilson, M. J., and Brandon, M. R. (1994). Identification of a novel antigen from *Corynebacterium pseudotuberculosis* that protects sheep against caseous lymphadenitis. *Infect Immun* **62**:2562-2567.

Wei, Y., Lee, J. M., Richmond, C., Blattner, F. R., Rafalski, J. A., and LaRossa, R. A. (2001). High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J Bacteriol* **183**:545-556.

Wertheimer, A. M., Verweij, W., Chen, Q., Crosa, L. M., Nagasawa, M., Tolmasky, M. E., Actis, L. A., and Crosa, J. H. (1999). Characterization of the angR gene of *Vibrio anguillarum*: essential role in virulence. *Infect Immun* **67**:6496-6509.

Wiegert, T., Homuth, G., Versteeg, S., and Schumann, W. (2001). Alkaline shock induces the *Bacillus subtilis* sigma regulon. *Mol Microbiol* **41**:59-71.

Wilks, K. E., Dunn, K. L., Farrant, J. L., Reddin, K. M., Gorringe, A. R., Langford, P. R., and Kroll, J. S. (1998). Periplasmic superoxide dismutase in meningococcal pathogenicity. *Infect Immun* **66**:213-217.

Williamson, L. H. (2001). Caseous lymphadenitis in small ruminants. *Vet Clin North Am Food Anim Pract* **17**:359-371.

Wilson, J. A., Doyle, T. J., and Gulig, P. A. (1997). Exponential-phase expression of spvA of the *Salmonella typhimurium* virulence plasmid: induction in intracellular salts medium and intracellularly in mice and cultured mammalian cells. *Microbiology* **143** :3827-3839.

Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O., and Schoolnik, G. K. (1999). Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci U S A* **96**:12833-12838.

Wilson, M. J., Brandon, M. R., and Walker, J. (1995). Molecular and biochemical characterization of a protective 40-kilodalton antigen from *Corynebacterium pseudotuberculosis*. *Infect Immun* **63**:206-211.

Wilson, R. L., Tvinnereim, A. R., Jones, B. D., and Harty, J. T. (2001). Identification of *Listeria monocytogenes* in vivo-induced genes by fluorescence-activated cell sorting. *Infect Immun* **69**:5016-5024.

Wosten, M. M. (1998). Eubacterial sigma-factors. *FEMS Microbiol Rev* **22**:127-150.

Yamamoto, T., Hanawa, T., and Ogata, S. (1994). Induction of *Yersinia enterocolitica* stress proteins by phagocytosis with macrophage. *Microbiol Immunol* **38**:295-300.

Ye, R. W., Tao, W., Bedzyk, L., Young, T., Chen, M., and Li, L. (2000). Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. *J Bacteriol* **182**:4458-4465.

Yoshida, K., Kobayashi, K., Miwa, Y., Kang, C. M., Matsunaga, M., Yamaguchi, H., Tojo, S., Yamamoto, M., Nishi, R., Ogasawara, N., Nakayama, T., and Fujita, Y. (2001). Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis*. *Nucleic Acids Res* **29**:683-692.

Young, D., Lathigra, R., Hendrix, R., Sweetser, D., and Young, R. A. (1988). Stress proteins are immune targets in leprosy and tuberculosis. *Proc Natl Acad Sci U S A* **85**:4267-4270.

Yozwiak, M. L. and Songer, J. G. (1993). Effect of *Corynebacterium pseudotuberculosis* phospholipase D viability and chemotactic responses of ovine neutrophils. *Am J Vet Res* **54**:392-397.

Yu, S., Fiss, E., and Jacobs, W. R. (1998). Analysis of the exochelin locus in *Mycobacterium smegmatis*: biosynthesis genes have homology with genes of the peptide synthetase family. *J Bacteriol* **180**:4676-4685.

Zhang, Y., Praszkier, J., Hodgson, A., and Pittard, A. J. (1994). Molecular analysis and characterisation of a broad-host range plasmid, pEP2. *J Bacteriol* **176**:5718-5728.

Zheng, M., Wang, X., Templeton, L. J., Smulski, D. R., LaRossa, R. A., and Storz, G. (2001). DNA Microarray-Mediated Transcriptional Profiling of the *Escherichia coli Response* to Hydrogen Peroxide. *J Bacteriol* **183**:4562-4570.

Zhou, D. and Galan, J. (2001). *Salmonella* entry into host cells: the work in concert of type III secreted effector proteins. *Microbes Infect* **3**:1293-1298.

Zolkiewski, M. (1999). ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from *Escherichia coli*. *J Biol Chem* **274**:28083-28086.