IIDMM
INSTITUTE OF INFECTIOUS DISEASE AND MOLECULAR MEDICINE

Optimisation of the Mycobacterial Replicon of an *E. coli*-Mycobacterial Shuttle Vector

Sarah Griffin
GRFSAR003

A Dissertation Submitted In Fulfilment Of The Requirements For The Degree of Master In Science

In The Division of Medical Virology, Department of Clinical Laboratory Sciences,

Faculty Of Health Sciences, University Of Cape Town.

Cape Town, June 2007

CONTENTS

Abstract

DNA plasmids which can replicate in two hosts, thus allowing gene expression in both, are termed shuttle vectors. *Escherichia coli*-mycobacterial shuttle vectors, specifically, are used extensively in many areas of mycobacterial research, including elucidation of mycobacterial pathogenesis and the development of recombinant mycobacterial vaccines. These *E. coli*-mycobacterial shuttle vectors provide a convenient mycobacterial expression system since they permit recombinant gene expression within mycobacteria, while all genetic manipulation can be carried out in *E. coli*, an organism which is far easier to manipulate at the genetic level and grows significantly faster. Episomal *E. coli*-mycobacterial shuttle vectors containing the pAL5000-derived mycobacterial replicon are the most widely utilised despite being regularly characterised by instability and poor levels of recombinant protein expression. To date, properties of the recombinant gene and promoter elements within *E. coli*-mycobacterial shuttle vectors have been thought to determine stability and expression levels. This study aimed to investigate whether the mycobacterial replicon may also play a role. The common mycobacterial replicon in episomal *E. coli*-mycobacterial shuttle vectors contains the truncated *rap* gene, encoding an auxiliary replication factor, and the *repA* and *repB* genes which code for essential replication proteins. Specific aspects of this typical mycobacterial replicon were identified as potential targets to improve stability and increase recombinant protein expression levels, and were modified accordingly. The modifications included: (1) incorporation of a high copy number mutation within *repA*, (2) inclusion of the entire *rap* gene and its putative transcription terminator and (3) introduction of an *hsp60* transcription terminator following *repB*. While none of the modifications improved genetic stability, inclusion of the high copy number mutation and the entire *rap* gene both caused a significant increase in copy number. In addition, the *rap* gene modification as well as incorporation of the *hsp60* transcription terminator increased expression of the recombinant gene, the green fluorescent protein, considerably. Although observed increases in copy number provide impetus for including the modified mycobacterial replicon in *E. coli*-mycobacterial shuttle vectors, vector stability, in the absence of antibiotic selection, was compromised upon inclusion of the full length *rap* gene in *Mycobacterium smegmatis* and insertion of the *hsp60* transcription terminator in both *M. smegmatis* and *Mycobacterium bovis* Bacille Calmette-Guèrin (BCG) Δ*panCD*. Interestingly, results suggested a possible correlation between increased recombinant protein expression levels and reduced stability. Further investigation is necessary to determine whether instability *in vitro* may be partially overcome through use of a regulatable promoter to control recombinant protein expression, and codon optimisation of the recombinant gene. If this could be achieved, the modifications identified in this study, which notably increased expression levels, may have widespread application in the development of recombinant mycobacterial vaccines and other mycobacterial research.

Acknowledgements

This project would not have been possible without the consistent supervision I received from Dr Rosamund Chapman. I am enormously grateful to her for her empathy, guidance and excellent support throughout the study. I would like to thank Principal Investigator Professor Anna-Lise Williamson for her insight and advice regarding my work and for the incredible opportunity to have worked on this project. Special thanks to Jolanda Truter for her practical assistance as laboratory manager and to laboratory members Dr Helen Stutz, Nureen Mayat, and former member Dr William Bourn, for their scientific and technical advice. Many thanks to Dr Rosamund Chapman, Professor Anna-Lise Williamson, Lindi Roberts, Jono Griffin, Ryan Morison, Gordon Griffin, Mykl Funke and Nureen Mayat for critical appraisal of this thesis.

On a personal note I would like to thank my parents Gordon and Susan Griffin for offering emotional support, additional financial support and many prayers which (besides the prayer for "Micro" brains) have been much appreciated. Thank you to Mykl Funke and Jono Griffin for their encouragement, and accompaniment to the laboratory after hours as bodyguards. Colossians 3:17

This work was funded in part by The South African Aids Vaccine Initiative (SAAVI) and the Poliomyelitis Research Foundation (PRF).

List of Abbreviations

2YT	2x yeast-tryptone	mM	Millimolar
A	Adenine	mm	Millimetre
BCG	Bacille Calmette-Guèrin	NBT	Nitro blue tetrazolium chloride
BCIP	5-bromo-4-chloro-3-indolyl phosphate	NCBI	National Centre for Biotechnology Information
bp	Base pair	ng	Nanogram
BSA	Bovine serum albumin	nm	Nanometre
C	Cytosine	OADC	Oleic-acid albumin dextrose complex
cfu	Colony-forming unit	°C	Degree celsius
Ct	Threshold cycle	OD	Optical density
C-terminus	Carboxy-terminus	Ω	Ohm
ΔCt	Change in threshold cycle	ORF	Open reading frame
DMSO	Dimethyl sulphoxide	ori	origin of replication
DNA	Deoxyribonucleic acid	oriE	*E. coli* origin of replication
ELISA	Enzyme-linked immunosorbent assay	oriM	Mycobacterial origin of replication
EDTA	Ethylene diamine tetraacetic acid	*p*	Probability
EPI	Expanded Programme for Immunisation	PAGE	Polyacrylamide gel electrophoresis
G	Guanine	PBS	Phosphate-buffered saline
GFP	Green fluorescent protein	PCR	Polymerase chain reaction
HI	High copy number mutation	Pr	Promoter
HIV	Human Immunodeficiency Virus	r	recombinant
H-site	High affinity binding site	rpm	Revolutions per minute
HRP	Horseradish peroxidase	RBS	Ribosome binding site
hsp	Heat shock protein	RNA	Ribonucleic acid
hygR	Hygromycin resistance	RT	Room temperature
IPTG	Isopropyl-β-D-thio-galactopyranoside	SDS	Sodium dodecyl sulphate
kanR	Kanamycin resistance	SIV	Simian Immunodeficiency Virus
kb	Kilobase	T	Thymidine
kDa	KiloDalton	TB	tuberculosis
kV	Kilovolt	Tn	Transposon
λ	Lambda	tt	Transcription terminator
ln	Natural logarithm	U	Unit
L-site	Low affinity binding site	U(-tailed)	Uracil
M	Molar	μΦ	Microfarad
MB	Megabase	μλ	Microlitre
M7H10	Middlebrook 7H10	μM	Micomolar
mg	Milligram	UV	Ultraviolet
ml	Millilitre	Xgal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

1.1. APPLICATIONS OF *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS

DNA plasmids which can replicate in two hosts are called shuttle vectors (FAO, 2007). The term is derived from their ability to 'shuttle' genes between hosts thus permitting gene expression in both. *Escherischia coli*-mycobacterial shuttle vectors are used extensively in mycobacterial research to allow the expression of foreign genes in mycobacteria. This includes research focused on the elucidation of mycobacterial pathogenesis, and the development of vaccines utilising mycobacteria as a delivery vehicle. Dual replicative capacity of *E. coli*-mycobacterial shuttle vectors in both *E. coli* and mycobacterial hosts allows genetic manipulation to be carried out in *E. coli*, and subsequent recombinant gene expression within mycobacteria. Such a system is necessary since there exist a number of difficulties associated with genetic manipulation in mycobacteria. For instance, mycobacteria are generally difficult to culture *in vitro*, as a result of their notably slow growth and tendency to aggregate (Ohara & Yamada, 2001). Genetic manipulation of mycobacteria is further compounded by their highly resilient, hydrophobic cell walls which demand specialised protocols for the introduction and extraction of genetic material (Bloom, 1994). Model organism *E. coli*, in contrast to mycobacteria, is considerably better suited to recombinant purposes as it is far easier to manipulate at the genetic level and grows significantly faster. Thus *E. coli*-mycobacterial shuttle vectors provide a convenient mycobacterial expression system that has wide-ranging application.

1.1.1. ELUCIDATING MYCOBACTERIAL PATHOGENESIS

The main mycobacterial pathogens affecting humans are *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), and *Mycobacterium leprae* which causes Hansen's disease (formerly leprosy) (Connell, 2001). Before the advent of a genetic expression system for mycobacteria, elucidation of mycobacterial pathogenesis was largely restricted to immunological, biochemical and physiological studies (Bloom, 1994). The subsequent development of *E. coli*-mycobacterial shuttle vectors, however, offered the use of molecular genetics as an additional tool for understanding mycobacterial virulence.

In order to study the effect of specific mycobacterial pathogenic genes, they need to be expressed within non-pathogenic mycobacterial hosts. Both non-virulent *Mycobacterium bovis* Bacille Calmette-Guèrin (BCG) and *Mycobacterium smegmatis* fulfil this requirement and are used as the recombinant mycobacterial hosts of choice. BCG shares approximately 99.95%

genetic similarity with pathogenic *M. tuberculosis* and, barring a few deletions, is virtually identical to pathogenic *M. bovis* from which it was originally derived (Garnier *et al*, 2003). Consequently, BCG offers an appealing environment in which to study pathogenic mycobacterial genes as genes are expressed against a near identical background to those of their natural hosts. *M. smegmatis* grows significantly faster than other mycobacteria and, as such, it is often used as a mycobacterial model to provide early predictions of anticipated results for slow-growing mycobacteria (Snapper *et al*, 1988).

Numerous mycobacterial pathogenesis studies have employed *E. coli*-mycobacterial shuttle vectors to express pathogenic mycobacterial genes in BCG and *M. smegmatis*. Such studies have contributed greatly to the understanding of mycobacterial virulence. This increased knowledge of mycobacterial pathogenesis has been exploited in the design of vaccines against mycobacterial infection, in improving the diagnosis of mycobacterial disease, in identifying transmission risks, and in generating new therapeutic drugs.

1.1.2. RECOMBINANT MYCOBACTERIA AS VACCINE VECTORS

BCG and *M. smegmatis* are being pursued at present as live vectors for vaccines against mycobacterial and other pathogens. The reasons why BCG and *M. smegmatis* represent appealing recombinant mycobacterial vaccine vectors are described below.

1.1.2.1. Recombinant BCG Vaccines

There exist numerous advantages to developing BCG-based vaccines. BCG is the current preventative vaccine for TB and forms part of the World Health Organisation (WHO) Expanded Programme on Immunisation (Labidi *et al*, 2001;Kaufmann, 2006) and, as such, networks are already established to transport and administer BCG. From a safety perspective, BCG vaccination can be given at birth and few side effects have been reported from over three billion doses given since 1948. Furthermore it can be produced for just $0.05 per dose (Kaufmann, 2006). This cost-effectiveness makes BCG-based vaccines particularly attractive for developing nations which bear much of the burden of infectious disease. In terms of the elicited immune response, BCG functions as a potent adjuvant and is able to persistently stimulate the immune system as it replicates *in vivo* (Ohara & Yamada, 2001).

Recombinant genes from a number of viral, bacterial, and parasitic pathogens have been successfully expressed in BCG via *E. coli*-mycobacterial shuttle vectors in efforts to develop

vaccines. While it is hoped that rBCG vaccines may eventually be used to provide dual protection against TB and other pathogens, the BCG vaccine in circulation provides limited protection against pulmonary adult TB. Since pulmonary TB in adults is responsible for the majority of TB-related deaths, there is an obvious need for more effective TB vaccines. Two of the five alternative TB vaccine candidates currently in clinical trials utilise *E. coli*-mycobacterial shuttle vectors as recombinant, optimised versions of the existing BCG vaccine (Kaufmann, 2006). Encouragingly, recombinant (r) BCG30 and Δ*ureC* Hly$^+$ rBCG exhibited improved antigenicity and immunogenicity, respectively, compared to standard BCG during phase I trials.

1.1.2.2. Recombinant *M. smegmatis* Vaccines

M. smegmatis is currently being pursued as an additional vaccine vector candidate to BCG. Upon entering the immune system, *M. smegmatis* is processed by macrophages far faster than BCG (Neyrolles *et al*, 2001). In spite of its lack of persistence, recombinant *M. smegmatis* has recently been shown to be capable of eliciting recombinant antigen-specific immune responses (Cayabyab *et al*, 2006;Yu *et al*, 2006). Thus, in cases where a more rapid immune response is desirable over a longer persistent response, *M. smegmatis* appears a promising vaccine vector.

1.2. TYPES OF *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS

Two differing approaches are employed in order to achieve foreign gene expression in mycobacteria. The first involves introduction of the recombinant genes into the mycobacterial host and their subsequent integration into the host chromosome via integrative shuttle vectors. The second utilises self-replicating, episomal shuttle vectors to introduce genes into the host and express them.

1.2.1. INTEGRATIVE *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS

In 1987 a shuttle phasmid capable of both plasmid replication in *E. coli* and phage replication in *M. smegmatis* and BCG was designed (Jacobs, Jr. *et al*, 1987). The phasmid was essentially a chimaera of TM4 mycobacteriophage DNA and an *E. coli* plasmid replicon (oriE), packaged into phage particles. The advantage of the phage delivery system is that it allows efficient infection of mycobacteria characterised by resilient cell walls, which otherwise necessitate the use of specialised protocols for introducing foreign DNA. Integrative shuttle

vectors have been further improved upon by the inclusion of antibiotic resistance genes to allow selection of desired transformants (Snapper *et al*, 1988). In addition, replicons from alternative mycobacteriophages containing integrase genes have been successfully used to enable site-specific integration between the *attP* phage and *attB* mycobacterial attachment sites (Parish & Stoker, 2001). These include use of the L1 mycobacteriophage replicon in the phAE19 series of integrative vectors (Snapper *et al*, 1988), Ms6 in the pEA4 series (Anes *et al*, 1992), as well as L5 in the pMH94, pMH947 and pMV361 series (Lee *et al*, 1991;Marklund *et al*, 1995;Stover *et al*, 1991). Studies have repeatedly shown that integrative shuttle vectors are highly stable. For example, after 30 generations in the absence of antibiotic selection, Dennehy *et al* (2007) showed that 100% of pMV361-derived integrative vectors expressing rotavirus VP6 in BCG retained antibiotic resistance and genetic integrity. Similarly, after 100 days Mederle *et al* (2002) demonstrated antibiotic resistance retention and genetic stability in up to 100% of pEA4-derived integrative vectors expressing SIV-1 *gag* and *nef* in BCG. Vector stability is particularly important during the manufacturing of recombinant vaccines when mycobacteria in which the vector has been lost or mutated may potentially take over cultures. Despite offering the advantage of having high levels of genetic stability, integrative vectors integrate as just single copies which effectively restricts recombinant gene expression levels. For instance, one study reported that only 0.16% of total protein in BCG represented recombinant SIV-1 Gag and Nef when pMV361-derived integrative vectors were utilised (Mederle *et al*, 2002). With gene expression levels of up to 15% of total protein having been described, expression levels attained via integrative vectors fall consistently within the lower end of the range.

1.2.2. EPISOMAL *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS

The first episomal *E. coli*-mycobacterial shuttle vector was developed in 1988 by Snapper *et al*. Briefly, this vector, pYUB12, contained an antibiotic resistance selectable marker in addition to chimaeras of *E. coli* and mycobacterial plasmid replicons. The mycobacterial plasmid replicon was derived from a natural *Mycobacterium fortuitum* plasmid, pAL5000, and its origin of replication is commonly termed the mycobacterial origin of replication (oriM). It has, however, been argued that the ori instead be named AL/Ori and that mycobacterial replicon may be a misnomer since the replicon does not facilitate mycobacterial replication but rather plasmid replication, within mycobacteria (Labidi *et al*, 2001). In accordance with standard nomenclature, this manuscript refers to the pAL5000-derived replicon as the mycobacterial replicon. The replicative capacities of other natural mycobacterial plasmids

have been exploited to generate alternative episomal *E. coli*-mycobacterial shuttle vectors, however, those based on the pAL5000 mycobacterial replicon remain the most extensively used in mycobacterial research (Table 1.1). Thus, this review concentrates on further optimisation of pAL5000 mycobacterial replicon-based vectors.

Table 1.1. Examples of types of episomal *E. coli*-mycobacterial expression vectors

Replicon	Organism derived from	Vector example	Reference
pAL5000	*Mycobacterium fortuitum*	pYUB12	(Snapper *et al*, 1990)
		pMV261	(Stover *et al*, 1991)
		pCG59	(Guilhot *et al*, 1992)
		pPE207	(Marklund *et al*, 1995)
		pCG76	(Guilhot *et al*, 1994)
		p16R1	(Garbe *et al*, 1994)
		pMD132	(Donnelly-Wu *et al*, 1993)
		pMR001	(Baulard *et al*, 1995)
		pMJ10	(Pelicic *et al*, 1997)
		pIJK-1	(Matsuo *et al*, 1990)
		p19PS	(Stover *et al*, 1993)
		pJAM2	(Triccas *et al*, 1998)
		pCB119	(Dennehy *et al*, 2007)
pLR7	*Mycobacterium intracellulare*	pMB351	(Beggs *et al*, 1995)
pMSC262	*Mycobacterium scrofulaceum*	pYT937	(Goto *et al*, 1991)
pJAZ38	*Mycobacterium fortuitum*	pJAZ40	(Gavigan *et al*, 1997)
pMF1	*Mycobacterium fortuitum*	pBP10	(Bachrach *et al*, 2000)
pCLP	*Mycobacterium celatum*	pCL4D	(Picardeau *et al*, 2000)
ϕD29	Mycobacteriophage D29	pBL415	(David *et al*, 1992)
ϕD29	Mycobacteriophage D29	phAE77	(Bardarov *et al*, 1997)
ϕTM4	Mycobacteriophage TM4	phAE94	(Bardarov *et al*, 1997)
pNG2	*Corynebacterium*	pEP2	(Radford & Hodgson, 1991)
pNG2	*Corynebacterium*	pEP3	(Radford & Hodgson, 1991)
IS900	Artificial	pUS702	(England *et al*, 1991)

Adapted from similar Table in Parish & Stoker, 2001

The independent replicative capacity of episomal *E. coli*-mycobacterial shuttle vectors means that they are not limited to single copy number. This allows levels of recombinant gene expression to reach significantly greater levels than those achieved via integrative vectors. The highest recombinant gene expression level of 15% of total mycobacterial protein was attained for β-lactamase expressed from an episomal pAL5000-based vector (Langermann *et al*, 1994). In comparison to an equivalent integrative vector, recombinant SIV-1 Gag and Nef were expressed at a seven-fold greater level in an episomal vector based on the pAL5000 mycobacterial replicon (Mederle *et al*, 2002). Although capable of higher recombinant gene expression levels, episomal vectors have frequently been noted as being significantly less stable than integrative vectors. For example, the same study by Mederle *et al* (2002) showed

that as few as 0.1% of episomal vectors were genetically stable while all integrative vector equivalents maintained 100% stability.

1.3. OPTIMISATION OF EPISOMAL *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS BY COMPONENT

An ideal *E. coli*-mycobacterial shuttle vector would allow for the expression of high levels of recombinant antigen while retaining stability. To this end, attempts have been made to optimise various components of episomal *E. coli*-mycobacterial shuttle vectors including the selectable marker, the promoter region, the recombinant antigen, and the regions of the pAL5000 replicon utilised.

1.3.1. THE SELECTABLE MARKER

To facilitate the identification of recombinant mycobacteria, a selectable marker of sorts was required to be present in *E. coli*-mycobacterial shuttle vectors (Jacobs, Jr. *et al*, 1987). Antibiotic resistance markers were explored for this purpose because of their well-documented efficacy in *E. coli* and other bacterial expression systems. While a naturally occurring mycobacterial plasmid harbouring an antibiotic resistance marker would have been ideal based on its functionality in mycobacteria, no such antibiotic resistant plasmid was available. Accordingly, an artificial antibiotic marker was required which needed to meet two conditions. First, the antibiotic needed to be effective against mycobacteria as a number of widely used antibiotics are unable to permeate the mycobacterial cell surface (Labidi *et al*, 2001) and a number of pathogenic mycobacteria have developed antibiotic resistance. In addition, the antibiotic needed to be highly stable to ensure that it remained effective for the duration of the extended time required for cultivation of slow-growing mycobacteria.

Snapper *et al* (1988) were the first to show that chloramphenicol and kanamycin could be used to identify *M. smegmatis* and BCG transformants. The chloramphenicol acetyltransferase (*cat*) gene and two kanamycin resistance genes; aminoglycoside 3′ phosphotransferase (*aph*) and neomycin/ kanamycin phosphotransferase II (*neo*) from transposons (Tn) 5 and 903 respectively, were investigated. Chloramphenicol resistance was subsequently deemed an unsuitable selectable marker as there was a high risk of transfer of chloramphenicol resistance between mycobacteria. Similar concerns about transferal of antibiotic resistance to pathogenic mycobacteria, resulted in kanamycin-specific aminoglycoside 3′

phosphotransferase being favoured over neomycin/kanamycin phosphotransferase II which targets both neomycin and kanamycin (Labidi *et al*, 2001). In 1991, Radford and Hodgson used the hygromycin resistance gene, *hyg*, in an *E. coli*-mycobacterial shuttle vector to successfully identify *M. smegmatis* and BCG recombinants. A number of alternative resistance markers such as apramycin, gentamicin and streptomycin have since been used but have either lowered transformation efficiency or proved unreliable (Paget & Davies, 1996;Pelicic *et al*, 1997;Guilhot *et al*, 1994). Additionally, the use of counter-selectable markers, such as *sacB*, *rpsL* and *katG*, has been pursued as an appealing system for gene replacement since it allows simultaneous selection of homologous recombination and vector loss (Pelicic *et al*, 1996;Sander *et al*, 1995;Norman *et al*, 1995). At present, *hyg* and *aph* are the antibiotic resistant genes used most extensively as selectable markers within *E. coli*-mycobacterial shuttle vectors. Consequently, expression vectors used in this study include the kanamycin resistance gene *aph* to reflect a frequently utilised antibiotic marker.

1.3.2. THE PROMOTER REGION

Promoter elements which can function efficiently in mycobacteria are required within *E. coli*-mycobacterial shuttle vectors to regulate recombinant gene expression (Dennehy & Williamson, 2005). The choice of promoter in *E. coli*-mycobacterial shuttle vectors depends on the type of regulation desired, as promoters affect the strength and timing of gene expression. Since foreign gene expression by *E. coli*-mycobacterial shuttle vectors is typically low, strong promoter activity may offer a strategy for improving levels of expression. On the other hand, an upper limit for optimal expression may exist, above which expression may cause a metabolic burden which would be disadvantageous during *in vitro* growth. The strength of promoter activity may be limited accordingly. The immune response against mycobacteria is elicited during their uptake and processing by macrophages. Thus, for recombinant mycobacterial vaccines specifically, strong expression of the recombinant antigen at the time of exposure to the immune system may improve the elicited immune response. A promoter which is inducible *in vivo* may fulfil this requirement.

An extensive range of possible promoters have been explored for use in *E. coli*-mycobacterial shuttle vectors. Some of those more commonly utilised are summarised in Table 1.2.

Table 1.2. Source references of promoters commonly utilised in *E. coli*-mycobacterium shuttle vectors

Promoter	Source	Mycobacterium expressed in	Reference
α antigen	*Mycobacterium kansasii*	BCG; *M. smegmatis*	(Matsuo *et al*, 1990)
groES/EL1 operon	*Streptomyces albus*	BCG	(Winter *et al*, 1991)
hsp60	BCG	BCG	(Stover *et al*, 1991)
hsp70	BCG	BCG	(Aldovini & Young, 1991)
pAN ORF2	*Mycobacterium paratuberculosis*	BCG	(Murray *et al*, 1992)
19kDa antigen	*Mycobacterium tuberculosis*	BCG	(Stover *et al*, 1993)
18kDa antigen	*Mycobacterium leprae*	BCG; *M. smegmatis*	(Dellagostin *et al*, 1993)
β–lactamase	*Mycobacterium fortuitum*	BCG; *M. smegmatis*	(Timm *et al*, 1994)
mtrA	*Mycobacterium tuberculosis*	BCG	(Via *et al*, 1996)
acetamidase	*Mycobacterium smegmatis*	*M. smegmatis*	(Parish *et al*, 1997)

1.3.2.1. Heat Shock Protein Promoters

Heat shock promoters *hsp60* and *hsp70* regulate the respective expression of heat shock proteins Hsp60 and Hsp70 in mycobacteria. Although heat shock proteins are constitutively expressed, during conditions of extreme temperature, pH or salt concentrations, levels of expression are up-regulated in order to elicit an adaptive stress response (Young *et al*, 1988). As the heat shock promoters are able to drive high levels of protein expression, they were chosen as promoters to drive gene expression in *E. coli*-mycobacterial shuttle vectors. During foreign gene expression, the promoters displayed strong promoter activity in mycobacteria and the *hsp60* promoter specifically, demonstrated greater promoter activity than the 18kDa antigen promoter in both *M. smegmatis* and BCG (Dellagostin *et al*, 1993;Dhandayuthapani *et al*, 1995;da Cruz *et al*, 2001). Clearly the responsiveness of these promoters to the environment represents an advantage upon entering the *in vivo* environment of the immune system when the induction of high levels of gene expression is desirable (Dennehy & Williamson, 2005). In an *in vitro* context, however, high expression levels in response to fluctuating growth conditions may be disadvantageous. This is because high expression levels of a foreign antigen can be toxic to the mycobacterial cell. Furthermore, the *hsp60* promoter has been shown to contribute directly to plasmid instability in a number of studies (Haeseleer, 1994;Kumar *et al*, 1998;Al Zarouni & Dale, 2002). Nonetheless heat shock promoters remain extensively utilised within *E. coli*-mycobacterial shuttle vectors. The *hsp60* promoter specifically, is used commonly in the pMV series of *E. coli*-mycobacterial shuttle vectors. As a result of its widespread application, the *hsp60* promoter was utilised as the promoter within expression vectors for this study.

1.3.2.2. 18kDa Promoter

The 18kDa promoter has a high degree of relatedness to heat shock protein promoters and drives expression of the 18kDa antigen in *M. leprae* (Dellagostin *et al*, 1995). As with other heat shock protein promoters, such as *hsp60* and *hsp70* promoters described above, the 18kDa promoter was explored for use as a regulator of gene expression in *E. coli*-mycobacterial shuttle vectors due to its suspected capacity for inducing high expression levels *in vivo*. The 18kDa promoter was shown to have weak activity during *in vitro* growth of *M. smegmatis* and BCG and to be induced in macrophages *in vivo*. Although *in vivo* expression levels induced by the 18kDa promoter are lower than levels induced under the control of the *hsp60* promoter in a similar context, regulation of lower expression levels *in vitro* by the 18kDa promoter may see it utilised increasingly in *E. coli*-mycobacterial shuttle vectors of recombinant mycobacterial vaccines.

1.3.2.3. *mtrA* Promoter

Within *M. tuberculosis*, *mtrA* and *mtrB* promoters constitute a two-component signal transduction regulator. Constitutive expression under the control of the *mtrA* promoter is required for *in vitro* growth of *M. tuberculosis* (Zahrt & Deretic, 2000). The *mtrA* promoter was initially of interest to researchers investigating *in vivo* genetics in an attempt to better understand *M. tuberculosis* virulence. Studies which introduced the *mtrA* promoter into BCG as a non-virulent control for comparison, revealed that the *mtrA* promoter is inducible *in vivo* in BCG. Further investigation by Via *et al* (1996) compared *mtrA-gfp* and *hsp60-gfp* promoter fusions. Constructs in BCG were used to infect macrophages and fluorescence levels were monitored by epifluorescence microscopy. Together with subsequent flow cytometric analysis, the *hsp60-gfp* fluorescence levels were not significantly increased *in vivo* whereas the *mtrA-gfp* fluorescence levels increased by a log fluorescent unit to reach levels similar to *hsp60-gfp*, following macrophage infection. Relative to regulation by the *hsp60* promoter, the *mtrA* promoter displayed significantly lower activity during *in vitro* growth of *M. smegmatis*. Accordingly, the *mtrA* promoter is a strong candidate for the promoter of choice within *E. coli*-mycobacterial shuttle vectors intended for recombinant mycobacterial vaccines.

1.3. 3. RECOMBINANT GENE(S)

To date, an extensive variety of recombinant genes has been expressed off episomal *E. coli*-mycobacterial shuttle vectors mainly for development of recombinant vaccines or genetic investigation. Recombinant genes from a number of pathogens have been successfully

expressed in efforts to develop vaccines. These include viral genes from HIV (Aldovini & Young, 1991), SIV (Yasutomi *et al*, 1993), and measles virus (Fennelly *et al*, 1995); bacterial genes from *Streptococcus pneumoniae* (Langermann *et al*, 1994), *Borrelia burgdorferi* (Stover *et al*, 1993) and *Bordetella* (Nascimento *et al*, 2000); as well as parasitic genes from *Plasmodium falciparum* (Matsumoto *et al*, 1998) and Leishmania (Connell *et al*, 1993). Reporter genes have been expressed in numerous recombinant mycobacterial studies. In terms of vaccine development reporter genes are useful as models for pathogenic genes and can be used in gene fusions for monitoring and tracking pathogenic antigens. In the context of genetic elucidation, reporter genes have been utilised to characterise genes and study their regulation. A range of reporter genes have been utilised; these include *gfp* encoding the green fluorescent protein (GFP) (Kremer *et al*, 1995), *luc* encoding luciferase (Jacobs, Jr. *et al*, 1993), *lacZ* encoding β-galactosidase (Timm *et al*, 1994), *phoA* encoding alkaline phosphatase (Lim *et al*, 1995) and *mbp* which codes for MalE (Himmelrich *et al*, 2000). Because GFP is known to be highly stable and has been utilised extensively in literature, it was selected as a model antigen for use in expression vectors in this study.

1.3.3.1. Codon Optimisation

Despite comparable copy numbers, recombinant gene expression levels among similar vectors have revealed considerable variation. For example, using the same recombinant expression vector in BCG, β-galactosidase was expressed at levels 200-fold greater than HIV-1 gp120 (Fuerst *et al*, 1991). This observation suggested that the reasons for differing expression levels exist with the genes themselves. Some recombinant genes were suspected to be toxic to mycobacteria. Variable gene translation efficiencies were also thought to affect expression levels. Since mycobacteria are typically G-C rich, they have a codon bias for residues of similar constitution (Mulder *et al*, 1997). As such, foreign genes with significantly lower G-C contents may incur a metabolic burden on mycobacterial hosts which would reduce expression efficiency. Codon optimisation of foreign genes has therefore been employed as a strategy to increase compliance with the mycobacterial codon bias (Kanekiyo *et al*, 2005). Kanekiyo *et al* (2005) showed significantly stronger immune responses to a rBCG vaccine expressing codon optimised Gag p24 than non-optimised p24. They suggested that the improved immunogenicity was due to an increase in recombinant antigen expression levels as a result of codon optimisation.

1.3.3.2. Secretion Signals

As discussed, one of the reasons for recombinant gene expression is the generation of vaccine candidates against pathogens. In terms of vaccine design, it has been suggested that because most recombinant proteins are located in the cytosol of BCG, their exposure to the immune system, and consequently the immune response specific to recombinant antigens may be restricted (Dennehy & Williamson, 2005). It was hypothesised that this obstacle may be overcome by secreting recombinant proteins from the mycobacterial vaccine vector. Therefore, the α-antigen gene for a secretory protein from *Mycobacterium kansasii* was fused to a recombinant gene (Matsuo *et al*, 1990). This allowed for the combined extracellular secretion of the α-antigen and the recombinant protein. Subsequent studies fused recombinant genes to the 18kDa gene encoding the secretory protein from *M. leprae* or the 19kDa *M. tuberculosis* signal peptide to improve presentation of recombinant proteins to the immune system (Dellagostin *et al*, 1993;Stover *et al*, 1993). In addition, fusion of recombinant genes to 85A secretion signal has been utilised to increase levels of recombinant gene expression (Al-Zarouni & Dale, 2002).

1.3.4. THE MYCOBACTERIAL REPLICON

1.3.4.1. Detailed Functional Analysis of the Mycobacterial Replicon

The ability of *E. coli*-mycobacterial shuttle vectors to replicate in mycobacteria is most commonly conferred by the pAL5000-derived mycobacterial replicon.

Sequencing of the entire pAL5000 revealed five open reading frames (ORFs) (Fig. 1.1) (Rauzier *et al*, 1988). Numerous investigations were carried out into the specific functions of each ORF product and, while the functions of ORFs 3 and 4 remain unknown, ORFs 1, 2 and 5; since named the *repA*, *repB* and *rap* genes respectively, were revealed to function in plasmid replication.

The RepA protein appears to have homology to the replicase group of proteins and has been demonstrated as a requirement for replication (Stolt & Stoker, 1996a). RepB has also been shown to be an essential protein for replication (Villar & Benitez, 1992). For this reason, the mycobacterial replicon commonly included in *E. coli*-mycobacterial shuttle vectors includes both *repA* and *repB*. The helix-turn-helix motif of RepB resembles that seen in the structure of regulatory proteins which bind to DNA (Rauzier *et al*, 1988). A single transcriptional start site was identified for the *repA* and *repB* genes which overlap by a nucleotide suggesting

their coupled expression (Stolt *et al*, 1999). This hypothesis was supported by the identification of a putative Shine Delgarno sequence upstream of *repB* which was positioned within the C-terminus of *repA*. The region separating *repA* and *rap* contains the oriM. RepB was shown to bind to two upstream regions with varying specificities termed High and Low affinity sites (H- and L-sites) accordingly. It has been proposed that the L-site is in fact the oriM. The H-site is riddled with repeats and palindromic sequences and overlaps the promoter for *repA* and *repB* which led to speculation that RepB auto-regulates its own expression (Stolt & Stoker, 1996b). RepA and host factors were shown to stimulate RepB binding to the ori provided *repA* and *repB* were expressed as a single transcript (Basu *et al*, 2002). The structure of RepB was further shown to be influenced by the coupled expression of *repA* and *repB* (Basu *et al*, 2004).

The *rap* gene lies in the opposite orientation to *repA* and *repB* (Rauzier *et al*, 1988). The protein it encodes was shown to have an auxiliary function with respect to plasmid replication. Studies suggested that Rap was non-essential to replication. It was alternatively proposed that Rap could be replaced by a host factor to enable replication (Stolt *et al*, 1999). Since *rap* was revealed to be dispensable for replication, it is typically truncated in the mycobacterial replicon of *E. coli*-mycobacterial shuttle vectors. This may be a significant oversight however as Stolt and Stoker (1996a) demonstrated that mycobacteria transformed with vectors carrying the full length *rap* formed colonies faster and of higher copy number than those harbouring vectors containing the truncated *rap* in *M. smegmatis* and BCG. The authors suggested that two pairs of direct repeats within *rap* may be responsible for increased plasmid copy number, which in turn improved stability. The *rap* transcriptional start site was identified 37bp downstream of the single *repA-repB* start site (Stolt *et al*, 1999). This finding suggested that *repA-repB* and *rap* may have overlapping promoters.

Since RepA and RepB are required for replication, their availability limits the copy number of *E. coli*-mycobacterial shuttle vectors which has been determined to average five copies per mycobacterial cell (Ranes *et al*, 1990). *E. coli*-mycobacterial shuttle vectors of greater copy number are desirable for two main reasons. First to increase the efficiency of plasmid extraction during genetic manipulation and second to improve recombinant gene expression. The latter is of particular importance for recombinant mycobacterial vaccines where improved expression has been shown to elicit an improved immune response (Himmelrich *et al*, 2000).

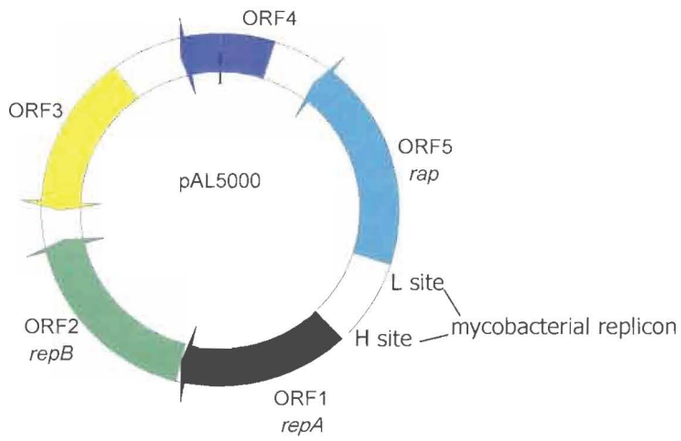

Figure 1.1. Annotation of the entire *Mycobacterium fortuitum* pAL5000 plasmid.

1.3.4.2. Recent Optimisation of the Mycobacterial Replicon

The mycobacterial replicon region was originally selected rather primitively, based on the position of unique restriction sites and, besides two recent studies which modified the mycobacterial replicon to achieve increased copy number, the mycobacterial replicon has been little modified since (Stover *et al*, 1993). Typical *E. coli*-mycobacterial shuttle vectors have been estimated to have a copy number of approximately five (Ranes *et al*, 1990). A higher copy number phenotype is desirable since it increases yields obtained during vector isolation. Furthermore, it is logical that as copy number increases, so recombinant gene expression increases. Although not always the case, several vaccine studies have shown that increased levels of recombinant protein result in an improved immune response. High copy number phenotypes have been attained using differing strategies. One employed strategy utilised random mutagenesis combined with positive selection to achieve an approximately seven-fold increase in copy number in both *M. smegmatis* and BCG (Bourn *et al*, 2007). Sequencing revealed this phenotype to be the result of an Alanine deletion within the C-terminus of RepA. It was postulated that the mutation modified RNA folding, which influenced *repA-repB* expression. The second strategy involved overexpression of RepB from the *M. smegmatis* chromosome which yielded a ten-fold increase in copy number (Mo *et al*, 2007).

1.4. RATIONALE FOR STUDY

Considering recent advances in functional genomics, the fact that the mycobacterial replicon has not been altered may be a significant oversight as *E. coli*-mycobacterial shuttle vectors continue to exhibit low expression and stability levels. Accordingly, this study aimed to further optimise the mycobacterial replicon. The mycobacterial replicon found in the majority of *E. coli*-mycobacterial shuttle vectors, including the pMV and pCB episomal vector series, was selected as the starting point for the study, since findings would have the widest relevance. This mycobacterial replicon was searched for possible causes of poor expression and instability and the following specific elements within the mycobacterial replicon were identified as possible targets for improvement (Fig. 1.2):

- the truncated *rap* gene
- the lack of transcription terminators following the *rap* and *repB* genes
- the low plasmid copy number

The *rap* gene is typically truncated in *E. coli*-mycobacterial shuttle vectors as it encodes an auxiliary factor which is not essential for replication (Stolt *et al*, 1999). As discussed, however, it was demonstrated that vectors carrying the full length *rap* grew faster and had higher copy numbers than vectors harbouring truncated *rap* in *M. smegmatis* and BCG (Stolt & Stoker, 1996a). Consequently, this study aimed to evaluate preferential inclusion of the entire *rap* gene over its common truncated form. This modification would be carried out for the dual purpose of improving stability and increasing copy number.

As indicated in Figure 1.2, the truncated *rap* gene is positioned immediately upstream of the oriE. Therefore, its transcription is likely unterminated which may cause transcriptional read-through into the oriE. In mycobacteria, termination of transcription is typically initiated by a stop codon at the end of a gene and a downstream transcription terminator (Unniraman *et al*, 2002). Where neither a stop codon nor a terminator is present, as is the case for truncated genes such as *rap*, transcription is likely non-terminated and read-through into neighbouring genes occurs (Mehta *et al*, 2006). This results in the production of extended mRNA transcripts. Subsequent translation of extended mRNA transcripts may lead to the production of nonsense proteins and contribute to metabolic stress and instability (Stutz *et al*, 2005). Thus, this study aimed to introduce sequence with homology to a transcription terminator,

downstream of the *rap* gene to eliminate possible instability caused by transcriptional read-through into the oriE.

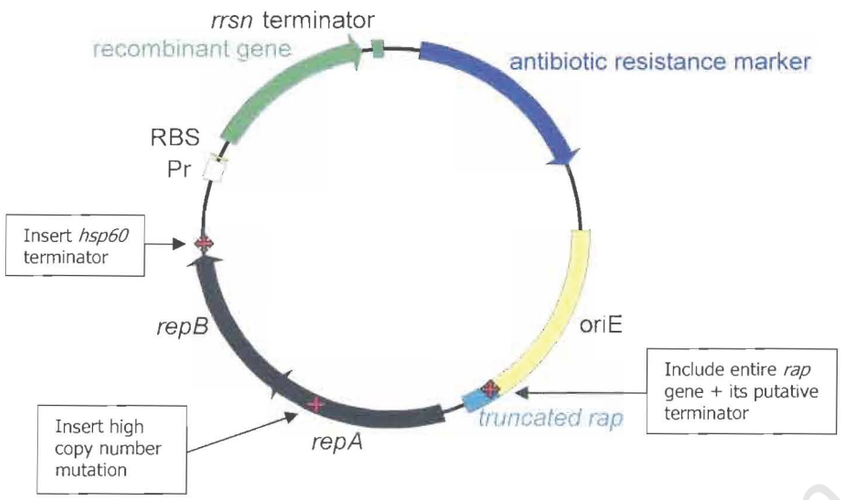

Figure 1.2. Targets for modification of the mycobacterial replicon typical in *E. coli*-mycobacterial shuttle vectors.

Targets include: ✥ the truncated *rap* gene and its lack of terminator, ✥ the absence of a transcription terminator downstream of the *repB* gene, and ✥ the low plasmid copy number

At the 3′ end of the mycobacterial replicon, the *repB* gene ends upstream of the promoter region and RBS for the expressed antigen. The fact that the entire *repB* gene is present implies that its translation is likely to be halted by a stop codon. As is the case for the truncated *rap* gene, no terminator appears to be positioned downstream of *repB*. As a result, transcription of *repB* may be inefficiently terminated and interfere with the promoter region and RBS of the expression cassette. Thus, this study aimed to insert an *hsp60* mycobacterial terminator downstream of *repB* to prevent possible interference with the promoter region and RBS of the expression cassette and consequently improve stability.

As mentioned, most pAL5000-based vectors have a copy number of approximately five, however, this has been increased seven-fold following a specific Alanine deletion within the *repA* gene (Bourn *et al*, 2007). This study aimed to incorporate the high copy number mutation in *repA* for improved recombinant gene expression levels and greater ease of vector isolation.

1.5. STUDY AIMS

In summary, the objectives of this study were:

- To construct a modified expression vector characterised by a full length *rap* gene and its putative downstream terminator, a terminator downstream of *repB* and a high copy number mutation within *repA*.

- To construct a series of intermediate expression vectors with which to evaluate the modified vector.

- To transform the modified expression vector and intermediate expression vectors into *M. smegmatis* and BCG for evaluation.

- To evaluate expression levels of a model recombinant gene in expression vectors.

- To evaluate copy number of the expression vectors.

- To evaluate fitness of the expression vectors in terms of growth rate and stability.

2.1. MODIFICATION OF THE MYCOBACTERIAL REPLICON OF AN *E. COLI*-MYCOBACTERIAL SHUTTLE VECTOR

Plasmid pHIGH100 (obtained from Dr William Bourn) and pRC100 (obtained from Dr Rosamund Chapman) functioned as the derivative *E. coli*-mycobacterial shuttle vectors from which modified vectors for the study could be generated. The mycobacterial replicon of pHIGH100 is characterised by a number of desirable modifications: the high copy number mutation, *repA*, *repB*, full length *rap* and truncated ORF3 and ORF4 (Fig. 2.1). Accordingly the pHIGH100 mycobacterial replicon was utilised as the template for bioinformatic searches and subsequent PCR. Plasmid pRC100 provided the desired vector backbone which comprised the model GFP recombinant antigen, a kanamycin resistance gene, and an oriE.

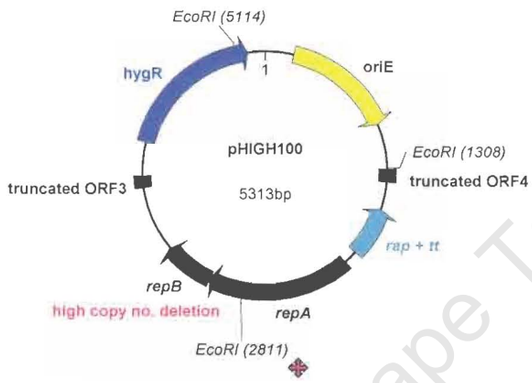

Figure 2.1. Structure of the plasmid pHIGH100

2.1.1. BIOINFORMATIC SCANNING OF THE MYCOBACTERIAL REPLICON FOR TERMINATORS

DNA sequence downstream of the *rap* and *repB* genes in plasmid pHIGH100 [GenBank Accession No: EF216316] was scanned for putative intrinsic transcription terminators using the DNAMAN bioinformatics software package [Lynnon Biosoft, version 4.0.0.1.]. Parameters

were set to identify terminators with a stem consisting of a minimum of seven paired base pairs and a loop consisting of a maximum of 35 unpaired base pairs.

2.1.2. SEQUENCING OF THE MYCOBACTERIAL REPLICON OF THE PCR TEMPLATE

As the GenBank pHIGH100 and pAL5000 nucleotide sequences were known to differ from that of the pHIGH100 used in our laboratory (Personal communication from Dr William Bourn), the nucleotide sequence of the mycobacterial replicon of plasmid pHIGH100 was confirmed prior to amplification. The sequencing strategy is depicted in Figure 2.2.

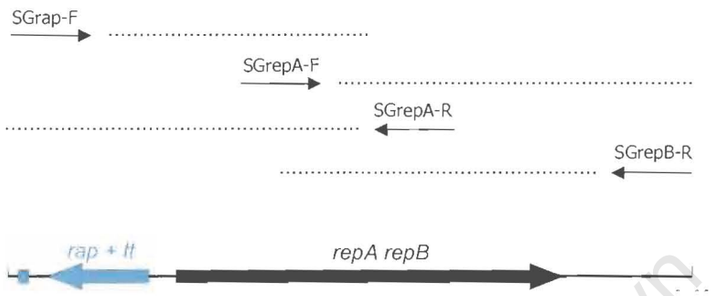

Figure 2.2. Strategy for sequencing the mycobacterial replicon of PCR template pHIGH100.

Utilised primers are represented as arrows. The various fragments of sequence produced by individual sequencing reactions are indicated by the dotted lines. A consensus sequence was generated by the alignment of overlapping sequence files.

Primers were designed with the aid of Primer Design software package [Freeware, version 1] (Table 2.1) and sequencing reactions were prepared using the Big Dye Terminator v3.1 cycle sequencing kit [Applied Biosystems]. Cycle sequencing was performed in the GeneAmp® PCR System 9700 thermocycler [Applied Biosystems] by the Genome Platform of the Department of Human Genetics, University of Cape Town. Sequences were edited using the Chromas v2.3 software package [Technelysium] and aligned with the GenBank pHIGH100 and pAL5000 [GenBank Accession No: M23557] sequences using the DNAMAN bioinformatics software package [Lynnon Biosoft, version 4.0.0.1].

Table 2.1. Primers used in this study

Primer	Orientation	Use	Sequence
			MluI
* SGrap-F	Forward	Sequencing of PCR Template pHIGH100 PCR Amplification Stage 1 + 3 Sequencing of PCR Product from Stage 1	5' ACG CGT GTT GGA TTC GGA GAC AAG C 3'
* SGrepA-R	Reverse	Sequencing of PCR Template pHIGH100 PCR Amplification Stage 1 Sequencing of PCR Product from Stage 1	5' TCG TTG CAC GGA A(A)T TCG 3'
* SGrepA-F	Forward	Sequencing of PCR Template pHIGH100 PCR Amplification Stage 2 Sequencing of PCR Product from Stage 2	5' TTC GAT TCC GTC AGG TTG TG 3'
			KpnI
* SGrepB-R	Reverse	Sequencing of PCR Template pHIGH100 PCR Amplification Stage 2 + 3 Sequencing of PCR Product from Stage 2	5' GGT ACC GAA ACA TGC GCA ACG AAC 3'
T7	Forward	Sequencing of pSG100 insert	5' TAA TAC GAC TCA CTA TAG GG 3'
SP6	Reverse	Sequencing of pSG100 insert	5' CG ATT TAG GTG ACA CTA TAG 3'

* Primers were designed with the aid of Primer Design Software Package [Freeware, Version 1.] and synthesised at the DNA Synthesis Service, Department of Molecular and Cell Biology, University of Cape Town

Restriction enzyme sites introduced for subsequent cloning are underlined

Additional nucleotide introduced to disrupt *EcoRI* restriction site is in parentheses

2.1.3. PCR AMPLIFICATION OF THE MYCOBACTERIAL REPLICON CONTAINING FULL LENGTH *RAP* AND THE HIGH COPY NUMBER MUTATION

A three-stage PCR strategy was employed to amplify the high copy number mycobacterial replicon from pHIGH100 and simultaneously modify it for subsequent cloning (Fig. 2.3). PCR cycling was carried out using the GeneAmp® PCR System 2400 thermocycler [Applied Biosystems]. The various primer pairs utilised for PCR are tabulated in Table 2.1.

2.1.3.1. Amplification Stage One

A 1.372 kb fragment including *rap*, a putative *rap* terminator and a portion of *repA* was amplified from pHIGH100. Approximately 5ng of pHIGH100 plasmid DNA was used as template in a final reaction volume of 50µl containing: 1µM each of primer SGrap-F and SGrepA-R, 0.2mM of the four deoxy-nucleotides, 0.125U Pfu ® DNA polymerase [Promega], 1x Pfu reaction buffer of 4mM $MgSO_4$ [Promega] and 3µl DMSO. PCR cycling conditions consisted of an initial denaturing step at 95°C for 3 minutes followed by 25 cycles of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds, and elongation at 72°C for 2 minutes and 50 seconds increasing by 5 seconds per cycle, and finally a 7 minute extension step at 72°C. To facilitate subsequent cloning, the forward primer SGrap-F

introduced an *MluI* site downstream of the likely *rap* terminator. The reverse primer SGrepA-R introduced an additional, arbitrary nucleotide within the *EcoRI* site to disrupt it. This was done to preserve the uniqueness of a second *EcoRI* site, positioned at the 3′ end of the promoter region of the expression cassette, to enable uncomplicated future cloning of promoter elements in our laboratory. Three negative controls were included; namely a template-free contamination control and two non-specific product controls using only the reverse and forward primers respectively. Following amplification, the PCR product was initially verified according to size by gel electrophoresis and subsequently confirmed by sequencing using the PCR amplification primer pair SGrap-F and SGrepA-R (See 2.1.2. for sequencing details).

2.1.3.2. Amplification Stage Two

During stage two a 1.171 kb fragment was amplified from pHIGH100 which overlapped with the product amplified in stage one by approximately 150 bp. The PCR product spanned part of *repA* including the high copy number mutation, the entire *repB* gene and some flanking sequence downstream of *repB*. Conditions were consistent with those outlined for stage one except that primers SGrepA-F and SGrepB-R were used and elongation time was decreased to 2 minutes and 20 seconds. To assist in later cloning, a *KpnI* site was introduced by the reverse primer SGrepB-R. As in stage one, a template-free contamination control and two single primer non-specific product controls were included. PCR product size was confirmed by gel electrophoresis initially and subsequently sequenced using the PCR amplification primer pair SGrepA-F and SGrepB-R (See 2.1.2. for sequencing details).

2.1.3.3. Amplification Stage Three

The final modified mycobacterial replicon amplified during stage three was characterised by the high copy number mutation, the full length *repA*, *repB* and *rap* genes, the putative *rap* terminator and flanking restriction sites. Aside from the following noted deviations, amplification conditions were consistent with those outlined for stage one. A mixture of 5ng of each product from stages one and two served as template. Forward primer SGrap-F and reverse primer SGrepB-R were used. Elongation time was increased to 4 minutes. While a contamination control was included, single primer specificity controls were unnecessary since templates and primers utilised in stage three had been tested previously either in stage one or two. Because equal concentrations of product from stage one and two were used as template, the *EcoRI* site was disrupted only in the half of the yield which had been amplified

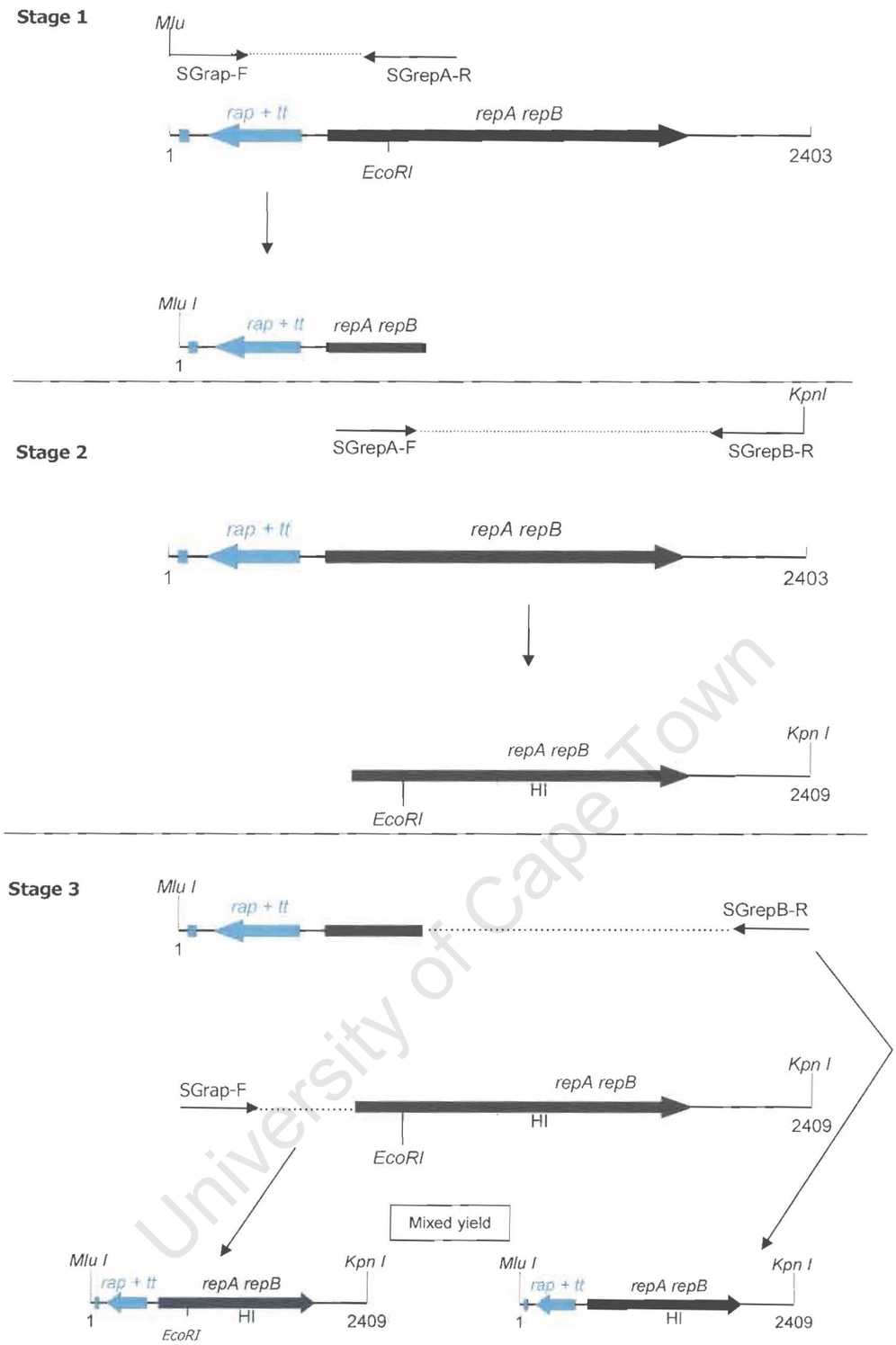

Figure 2.3. Stylised Diagram indicating the PCR amplification strategy.

from the stage one PCR product. Conversely, the second half of the yield contained an intact *EcoRI* site since the product from stage two acted as its template. The resulting PCR product size was confirmed by gel electrophoresis. Sequencing was not carried out at this stage as the amplified fragment was sequenced following subsequent cloning.

2.1.4. CONSTRUCTION OF THE MODIFIED *E. COLI*-MYCOBACTERIAL SHUTTLE VECTOR

General recombinant procedures used in cloning are outlined here. Vectors and DNA fragments were digested using restriction enzymes [Roche Diagnostics] as per manufacturer's instruction. Resulting DNA fragment sizes were determined by electrophoretic resolution on a 1% agarose gel alongside either a 1 kb marker [Promega] or λDNA [Roche Diagnostics] digested with *PstI*. Desired fragments were purified from agarose gels using the QIAquick Gel Extraction Kit [Qiagen] or from reaction mixes using the QIAquick Purification Kit [Qiagen]. Following quantitation on the Nanodrop Spectrophotometer [Inqaba Biotec], fragments were ligated using the Rapid Ligation Kit [Roche] according to the supplied protocol. *E. coli* JM109 cells (obtained from the BioVac Institute) were made competent using the method by Inoue *et al.* (1990). Ligations were transformed into competent *E. coli* cells according to standard protocol (Sambrook & Russell, 2001). Except for pGEM-T Easy cloning, initial positive transformants were selected by plating transformed cells onto 25ml of 2YT (16g tryptone, 10g yeast extract, 5g NaCl per litre of water) solid medium supplemented with 50μg/ml kanamycin and propagated at 37 °C. Selected colonies were used to inoculate 1ml of 2YT liquid medium supplemented with 50μg/ml kanamycin and incubated at 37°C overnight with shaking for aeration. Small amounts of plasmid DNA were isolated from initial positive transformants with a kanamycin resistant phenotype using the alkaline lysis method of Ish-Horowicz and Burke (1981) and used to conduct secondary screening by restriction digest analysis. The Nucleobond AX PC-Kit [Machery-Nagel] was used according to the supplied protocol to isolate large amounts of plasmid DNA from 100ml cultures of confirmed positive transformants.

2.1.4.1. Construction of Sub-Clone pSG100

An *EcoRI* digest was performed to isolate the *EcoRI*-free half of the 2.409 kb final PCR product yield from stage three. The uncut, *EcoRI*-free PCR product was gel purified. The resultant isolated fragment therefore consisted of a mycobacterial replicon modified to lack an *EcoRI* site and contain the full *rap* gene, its putative terminator, the high copy number

mutation and the complete *repA* and *repB* genes. To generate sub-clone pSG100, this modified mycobacterial replicon fragment was A-tailed using Taq polymerase [Supertherm] and cloned into the pGEM-T Easy vector™ [Promega pGEM-T Easy Vector System I] at a 1:1 insert to vector ratio according to the protocol supplied by the manufacturer. Transformation mixes were plated onto 2YT media supplemented with 80μg/ml Xgal, 0.5mM IPTG and 100μg/ml ampicillin. The identification of desired transformants as white colonies was facilitated by selection for β-galactosidase insertional inactivation and ampicillin resistance following overnight propagation at 37°C. A *SalI* digest was used to confirm insertion of the mycobacterial replicon fragment into pGEM-T Easy in the correct orientation. The insert in the pSG100 sub-clone was sequenced before its use in constructing the modified vector pSG300 using primers SGrepA-R, SGrepA-F and universal primers SP6 and T7 (See section 2.1.2. for sequencing details).

2.1.4.2. Construction of Sub-Clone pSG200

Sub-clone pSG200 was generated to introduce a transcription terminator following the *repB* gene within the mycobacterial replicon. Oligonucleotides were designed to create a *M. bovis hsp60* transcription terminator similar in sequence to that published by Kanekiyo *et al.* (2005) (Fig. 2.4). Complimentary oligonucleotides were designed to include flanking *KpnI* and *XbaI* overhangs with phosphorylated 5' ends to assist cloning. In addition, an internal *XhoI* site was incorporated to allow confirmation of the presence of the terminator via restriction mapping. Oligonucleotides were annealed by combining equal concentrations of complementary oligonucleotides in STE Buffer (10mM Tris pH 8.0, 50mM NaCl, 1mM EDTA). The reaction mix was heated to 95°C and allowed to cool gradually to prevent the formation of secondary structure. To generate pSG200 the resulting annealed oligonucleotides containing the *hsp60* transcription terminator were ligated at a 5:1 insert to vector ratio to the backbone of vector pRC100 which was isolated following a sequential *KpnI XbaI* digest. The pRC100 vector backbone contained a gene encoding GFP, a kanamycin resistance gene, an oriE and an unmodified mycobacterial replicon containing a truncated *rap* gene, the *repA* gene, the *repB* gene and lacking the high copy number mutation. Successful insertion of the terminator was confirmed by an *XhoI* restriction digest.

Figure 2.4. Annealed oligonucleotides with overhangs.

Restriction sites *KpnI*, *XhoI* and *XbaI* are highlighted.

2.1.4.3. Construction of Modified Vector pSG300

The *E.coli*-mycobacterial shuttle vector, pSG300, was generated to contain the fully modified mycobacterial replicon. The optimised mycobacterial replicon consisted of the high copy number mutation, the entire *rap* gene, its putative terminator, the complete *repA* and *repB* genes and an *hsp60* transcription terminator following the *repB* gene. A *KpnI MluI* sequential digest of pSG100 preceded isolation of the modified mycobacterial replicon fragment consisting of the high copy number mutation and full length *rap*. A *KpnI MluI* sequential digest of pSG200 was used to isolate the vector backbone containing the *hsp60* transcription terminator. The modified mycobacterial replicon was ligated at a 1:1 insert to vector ratio to the modified backbone at the *KpnI* and *MluI* restriction sites to create the vector pSG300. This optimised expression vector was mapped using the *XhoI* restriction site.

2.2. CONSTRUCTION OF INTERMEDIATE *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS

Three intermediate *E. coli*-mycobacterial shuttle vectors were constructed for use in specific evaluation of each of the modifications made to the mycobacterial replicon in the optimised *E. coli*-mycobacterial shuttle vector.

2.2.1. CONSTRUCTION OF INTERMEDIATE VECTOR pSG400

Intermediate vector pSG400 was constructed to evaluate the effects of incorporating the high copy number mutation into the mycobacterial replicon. This could be determined by directly comparing pSG400, which had the high copy number mutation, to unmodified vector pRC100 which lacked the mutation but was otherwise identical. To construct pSG400, a high copy number fragment containing *repA*, *repB* and truncated *rap*, was obtained following a sequential digest of intermediate vector pSG600 (Details of pSG600 construction are given

below) with *KpnI* and *MluI*. A backbone fragment containing the *hsp60* promoter, GFP, the kanamycin resistance gene and oriE was isolated from similarly digested pRC100. The two fragments were ligated at a 3:1 insert to vector ratio to produce intermediate construct pSG400. Restriction enzyme mapping with *KpnI* and *MluI* was performed to confirm the conservation of sites and additional confirmatory restriction analyses were carried out using *SalI* and *BglII*.

2.2.2. CONSTRUCTION OF INTERMEDIATE VECTOR pSG500

Intermediate vector pSG500 was constructed to determine the effect of including the full length *rap* gene in the mycobacterial replicon. Since pSG500 differed from pSG400 in that it contained the entire *rap* gene, but was otherwise identical, the effect of inclusion of the full *rap* could be determined by directly comparing the vectors. Vector pSG500 was constructed using optimised vector pSG300 and unmodified vector pRC100. Following a sequential digest of pSG300 with *KpnI* and *MluI*, the high copy number fragment containing full length *rap*, *repA* and *repB* was obtained. A fragment containing kanamycin resistance and oriE was gel purified from similarly digested pRC100. The two fragments were ligated at a 3:1 insert to vector ratio to yield intermediate construct pSG500. Restriction enzyme mapping using *KpnI* and *MluI* was performed to confirm the conservation of sites and a further *SalI* restriction analysis was carried out to confirm plasmid structure.

2.2.3. CONSTRUCTION OF INTERMEDIATE VECTOR pSG600

Intermediate vector pSG600 was generated for the purpose of evaluating the effect of introducing the *hsp60* terminator following *repB* within the mycobacterial replicon. Plasmid pSG600 differed from pSG400 in that it contained the *hsp60* terminator but was otherwise identical. Therefore, these vectors could be directly compared to appraise the effect of including the *hsp60* terminator. Vector pSG600 was created from modified vector pSG300 and unmodified vector pRC100. Following a sequential digest of pSG300 with *BglII* and *HpaI*, a fragment containing *repA* and the high copy number mutation, *repB*, the *hsp60* terminator, the *hsp60* promoter and GFP was isolated. A fragment containing kanamycin resistance, oriE and truncated *rap* was gel purified from similarly digested pRC100. The two fragments were ligated at a 3:1 insert to vector ratio to yield intermediate construct pSG600. Restriction enzyme mapping using *BglII* and *HpaI* was used to confirm conservation of restriction sites and *XhoI* endonuclease digestion was used to confirm the presence of the *hsp60* terminator.

2.3. TRANSFORMATION OF *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS INTO MYCOBACTERIA

Modified expression vector pSG300 and intermediate expression vectors; pRC100, pSG400, pSG500 and pSG600, were transformed into *M. smegmatis* mc^2155 (Obtained from the BioVac Institute) and *M. bovis* BCG Pasteur Δ*panCD* $mc^2$6000 strain (Obtained from Dr William Jacobs Jr) competent cells by electroporation. Mycobacterial electrocompetent cells were prepared according to the method described by Parish and Stoker (2001). Electrocompetent mycobacteria were electroporated with plasmid DNA using the Gene Pulser™ [BioRad] at setting 1.8kV, 25μF and 1000Ω. In order to obtain colonies, electroporated cells were plated onto 35ml of Middlebrook 7H10 solid medium (M7H10) supplemented with oleic-acid albumin dextrose complex (OADC) [BioLab] and 10ug/ml kanamycin [Sigma]. Plates were protected from dehydration by being sealed in plastic bags and subsequently incubated at 37°C for two to three days for *M. smegmatis* or three to four weeks for BCG Δ*panCD* colony formation. Colonies were used to inoculate Middlebrook (M7H9) liquid medium supplemented with OADC and kanamycin. *M. smegmatis* and BCG Δ*panCD* cultures were further supplemented with tween80 or 0.25ul/ml tyloxapol detergent respectively, to reduce cell clumping. In addition, BCG Δ*panCD* cultures were supplemented with 0.048mg/ml $_D$-Pantothenic Acid [Sigma] and 0.05mg/ml Hygromycin B [Roche]. Cultures were incubated at 37°C with gentle rolling in order to minimise cell clumping until logarithmic growth phase of OD_{600} 0.4-1.0 was reached. For *M. smegmatis* this took one to two days and for BCG 10 to 25 days. Plasmid DNA was isolated from transformants using the High Pure Plasmid Isolation Kit [Roche Diagnostics] according to the method described by Parish and Stoker (2001). Because mycobacterial plasmid DNA yield is typically low, isolated plasmid DNA was used to transform *E. coli* for subsequent verification. Recombinant plasmids isolated from *E. coli* were again mapped with restriction enzymes to confirm their general structure. In addition, a sequential *XbaI HpaI* restriction digest was carried out specifically to confirm the integrity of the GFP gene.

2.4. EVALUATION OF THE MODIFIED *E. COLI*-MYCOBACTERIAL SHUTTLE VECTOR

2.4.1. EVALUATION OF GFP EXPRESSION LEVELS

2.4.1.1. GFP Fluorescence Study

Relative levels of colony GFP fluorescence were estimated for recombinant *M. smegmatis* and BCG Δ*panCD*. Colonies were viewed and photographed on a Chromato-vue TM-20 Transilluminator [UVP] at 254nm (UV light) to visualise GFP fluorescence as a preliminary indication of GFP expression levels.

2.4.1.2. GFP Western Blot Analysis

Relative levels of GFP expression were determined via duplicate Western blots for recombinant *M. smegmatis* and BCG Δ*panCD*.

Log phase mycobacterial cultures (OD_{600} 0.5-0.8) were harvested by centrifugation (5000rpm, 10 minutes, 4°C) using the GP Centrifuge [Beckman]. Pellets were washed in 10ml phosphate-buffered saline (PBS pH7.4) and centrifuged as before. Cells were resuspended in 1ml PBS and transferred to FastRNA tubes [Bio 101] containing 0.1mm zirconia-silica beads [BioSpec]. Cell were lysed in the FastPrep machine [Bio 101] by four 30 second repeats of vigorous shaking at speed setting 6. Tubes were cooled on ice between repeats. Following benchtop centrifugation in the Eppendorf 5415R (14000rpm, 10 minutes, RT) the supernatant was collected and stored in aliquots at –80°C. Protein concentration of all samples was determined using the Dc BioRad Quantitation Kit Assay according to supplier's instruction.

Protein samples shown in Table 2.2. were separated by SDS-PAGE on 12.5% denaturing gels (Sambrook & Russell, 2001) alongside Kaleidoscope Molecular Weight Marker [BioRad]. Gels were run in triplicate. Equal loading was demonstrated by silver staining one of the three identical SDS-PAGE gels (Merrill *et al*, 1981). A TransBlot semi-dry transfer cell [BioRad] was used to transfer protein (15V, 40 min) from the remaining two SDS-PAGE gels onto nitrocellulose membrane. Membranes were pre-incubated in block buffer (1 hour, RT) and then probed with a mouse monoclonal anti-GFP primary antibody [Roche] diluted 1:1000 in block buffer (1 hour, RT). Blots were washed three times in wash buffer (3x 10 minutes, RT) after which they were hybridised with anti-rabbit alkaline phosphatase conjugated secondary antibody [Sigma] diluted 1:5000 in block buffer (1 hour, RT). Following a second set of

washes, blots were developed using Nitro blue tetrazolium chloride/ 5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP) [Roche].

Table 2.2. Protein samples used for GFP Western Blot Analysis

Use	Source	Characteristic	Form	Amount
Positive Control	r*M. smegmatis* or rBCG Δ*panCD* pHS207*	High GFP-expresser	Cell lysate	2μg
Positive Control + MW Standard	r*E. coli* expressing GFP	GFP protein	Purified protein	15ng
Negative Control	r*M. smegmatis* or rBCG Δ*panCD* pHS200*	Non GFP-expresser	Cell lysate	2μg
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pRC100	Unknown	Cell lysate	2μg
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pSG300	Unknown	Cell lysate	2μg
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pSG400	Unknown	Cell lysate	2μg
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pSG500	Unknown	Cell lysate	2μg
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pSG600	Unknown	Cell lysate	2μg

MW- Molecular Weight

*Plasmids pHS207 and pHS200 were obtained from Dr Helen Stutz

2.4.1.3. GFP Capture ELISA

GFP expression levels for recombinant *M. smegmatis* and BCG Δ*panCD* were quantitated by GFP Capture ELISA performed in triplicate.

Cell lysate was isolated and its protein content quantitated as in 2.4.1.2. Protein samples used and their experimental functions are outlined in Table 2.3. For quantification purposes, purified recombinant GFP protein [Clontech] was used to establish a linear standard curve. Samples were diluted in PBST containing 1% milk powder and incubated in wells of commercially bought 96-well plates pre-coated with anti-GFP antibody [Pierce] (1 hour, 37°C). Wells were washed four times with PBST using the ELx50 Autostrip Washer [Bio-Tek Instruments] and incubated with a goat polyclonal anti-GFP HRP-conjugated antibody [Abcam] (1 hour, 37°C). Wells were incubated with 100μl substrate [TMB Microwell Peroxidase Substrate System, KPL] (30 minutes, RT, in the dark) prepared following manufacturer's protocol to detect the HRP. The detection reaction was stopped by addition of 100μl of 0.5M H_2SO_4 per well. Absorbancies were read in a PR2100 ELISA plate reader [Sanofi Diagnostic Pasteur] at 450nm. For each sample, the amount of GFP was extrapolated from the standard curve and corrected by the dilution factor to determine the percentage of GFP expressed per total protein sample. Statistica 7 Software Package [StatSoft] was used to perform statistical analysis on ELISA results.

Table 2.3. Protein samples used for GFP Capture ELISA

Use	Source	Characteristic	Form
Positive Control	r*M. smegmatis* or rBCG Δ*panCD* pHS207*	High GFP-expresser	Cell lysate
Positive Control + MW Standard	r*E. coli* expressing GFP	GFP protein	Purified protein
Negative Control	r*M. smegmatis* or rBCG Δ*panCD* pHS200*	Non GFP-expresser	Cell lysate
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pRC100	Unknown	Cell lysate
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pSG300	Unknown	Cell lysate
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pSG400	Unknown	Cell lysate
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pSG500	Unknown	Cell lysate
Experimental	r*M. smegmatis* or rBCG Δ*panCD* pSG600	Unknown	Cell lysate

MW- Molecular Weight

*Plasmids pHS207 and pHS200 were obtained from Dr Helen Stutz

2.4.2. EVALUATION OF VECTOR COPY NUMBER

2.4.2.1. Restriction Digest Analysis

To determine relative differences in plasmid copy number, recombinant *M. smegmatis* and BCG Δ*panCD* total DNA was digested to compare plasmid fragment to chromosomal background intensities.

Total DNA was extracted from 100ml stationary phase cultures of recombinant *M. smegmatis* and BCG Δ*panCD*. Harvested cultures were washed in 20ml MTB extraction buffer (267mM monosodium glutamate, 50mM Tris-HCl, 25mM EDTA; pH 7.4) and incubated in 12ml of MTB extraction buffer supplemented with 100mg lysozyme powder for 5-6 hours at 37°C. Samples were incubated for a further 14-18 hours at 45°C following the addition of 10µl RNase (10mg/ml), 400µl proteinase K (10mg/ml) and 1.4ml proteinase K buffer (100mM Tris-HCl, 50mM EDTA, 5% SDS; pH 7.8). The sample was extracted with an equal volume of phenol/chloroform/isoamylalcohol [25:24:1] and total DNA precipitated in the aqueous phase by addition of 0.7 volume isopropanol and 350µl of 5M Sodium Acetate. Precipitated total DNA was resuspended in 1ml TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0) and extracted a second time with an equal volume of phenol/chloroform/isoamylalcohol. Samples were washed twice with an equal volume of chloroform and total DNA precipitated in the aqueous phase by addition of 0.7 volume isopropanol and 350µl of 5M Sodium Acetate. Total DNA was washed with 0.7 volume of 70% ethanol and suspended in 400µl TE buffer. Total DNA concentrations were estimated by gel quantitation against known concentrations of λDNA [Roche].

Eight μg of total DNA from each sample was subjected to a double *MluI XbaI* restriction digest. Two-fold dilutions were prepared in TE buffer. A similar restriction digest was carried out for 300ng of corresponding plasmid DNA. The digested plasmid DNA and two-fold dilutions of digested total DNA were separated on 1% agarose by gel electrophoresis. Following electrophoresis the gel was uniformly stained in a 0.05% ethidium bromide solution. SynGene GeneTools Image Analysis Software Package [Syngene, version 3.00.22] was used to estimate plasmid fragment intensities relative to chromosomal background intensities.

2.4.2.2. Qualitative Real-Time PCR

Qualitative real-time PCR was used to determine relative plasmid copy numbers in recombinant *M. smegmatis* and BCG Δ*panCD* using probes with 16S chromosome- and GFP plasmid-specificity respectively.

A 5'-VIC labelled TAMRA probe and flanking primers specific to the 16S gene in *M. tuberculosis* were available for use (Obtained from Dr Mark Nicol). To ensure that this probe-primer set designed for *M. tuberculosis* would be able to amplify the 16S genes of *M. smegmatis* and BCG, probe and primers were tested using the online *in silico* simulation of PCR amplification tool (Bikandi *et al*, 2004). The specificity of the probe-primer set for the 16S chromosomal genes in *M. smegmatis* and BCG was verified by sequence homology searches using the NCBI BLAST algorithm (Altschul *et al*, 1997). The Primer Express software package [Applied Biosystems] was employed for design of a 5'-FAM labelled TAMRA probe and flanking primers specific to the GFP gene in plasmids pRC100 and pSG300-600. The DNAMAN bioinformatics software package [Lynnon Biosoft, version 4.0.0.1] was used to determine the degree of cross specificity.

All real-time PCR amplification reactions had a final volume of 25μl and included 1x TaqMan Universal PCR Master Mix [Applied Biosystems] and, in various combinations, 0.08μM of 16S probe, 0.3μM of each 16S primer, 0.3μM of GFP probe and 0.8μM of each GFP primer. Amplification conditions were as follows: an incubation step at 50°C for 2 minutes, a Taq polymerase (Master Mix component) activation step at 95°C for 10 minutes, 40 cycles of a denaturation step at 95°C for 15 seconds and a combined annealing-extension step at 60°C for 1 minute. PCR amplification was carried out using the ABI Prism 7000 [Applied Biosystems]. To determine the linear range for amplification of the 16S gene, ten-fold

dilutions of *M. smegmatis* chromosomal DNA [Obtained from Dr Karen Shires] were used as template to plot a 16S standard curve of threshold cycle (Ct) versus the log of the amount of chromosomal DNA. Similarly, ten-fold dilutions of pRC100 plasmid DNA were used to establish the linear range for amplification of the GFP gene. Template-free controls as well as probe-primer cross-specificity controls were included. Approximately 20ng of recombinant *M. smegmatis* and BCG total DNA samples (See 2.4.2.1. for extraction details) were used as templates for multiplex real-time PCR performed in duplicate using both the chromosomal 16S and the plasmid GFP probe-primer sets. Results were analysed using the ABI Prism 7000 SDS software package [Applied Biosystems, Version 1]. Baseline was arbitrarily set at a level higher than background. The difference in threshold cycle (ΔCt) between the 16S threshold cycle (Ct) and the GFP Ct was used to determine the fold difference in plasmid copy number for each recombinant sample relative to the unmodified pRC100 recombinant.

2.4.3. EVALUATION OF VECTOR FITNESS

2.4.3.1. Growth Rate Study

As a preliminary indication of growth rate, relative colony sizes were estimated for recombinant *M. smegmatis* and BCG Δ*panCD* respectively. Care was taken to compare colony sizes between plates with similar cell counts since a high density of colony forming units may limit colony size. Growth rates were more accurately determined by measuring the optical density of triplicate recombinant *M. smegmatis* and BCG Δ*panCD* cultures over time. 100ml cultures were inoculated with vigorously growing cells to an initial OD_{600} of 0.005 (See 2.3. for culturing conditions). Measurements were recorded every 3 hours for *M. smegmatis* cultures and every 24 hours for BCG Δ*panCD* cultures until stationary phase was reached. Cultures with an OD_{600} exceeding 1 ODU were diluted to fall within the linear range.

2.4.3.2. Stability Study

Recombinant *M. smegmatis* and BCG Δ*panCD* cultures were grown in the absence of antibiotic selection for 43 generations to assess stability in terms of GFP fluorescence and antibiotic resistance retention.

Recombinant BCG cultures were grown in duplicate in the absence of kanamycin antibiotic selection. At an OD_{600} of approximately 1, cultures were sub-cultured. At each sub-culture time point OD_{600} was recorded and glycerol stocks were stored at –20°C. Based on the recorded optical densities, the approximate generation of each culture was calculated

according to the following equation: Generation = [ln (final OD_{600}) – ln (starting OD_{600})] / ln 2 (Widdel, 2007). At each subculture time point, Ziehl-Neelsen staining was carried out on glycerol stocks to verify that cultures were not contaminated (Sleigh & Timburg, 1998). Glycerol stocks were used to plate cultures onto media with and without kanamycin selection. Colonies were viewed on a Chromato-vue TM-20 Transilluminator [UVP] at 254nm (UV light) to ascertain the percentage of colony forming units which maintained GFP fluorescence. Colony counts from plates with and without kanamycin selection were used to calculate the percentage of colony forming units which retained antibiotic resistance.

CHAPTER 3: RESULTS

3.1. MODIFICATION OF THE MYCOBACTERIAL REPLICON OF AN *E. COLI*-MYCOBACTERIAL SHUTTLE VECTOR

The pAL5000-derived mycobacterial replicon typically utilised in episomal *E. coli*-mycobacterial shuttle vectors was modified in an attempt to improve vector stability and increase recombinant antigen expression levels. The modifications included :

- Incorporation of the entire *rap* gene, commonly truncated in *E. coli*-mycobacterial shuttle vectors, to determine the effect on vector stability, copy number and recombinant protein expression levels
- Introduction of a high copy number mutation in an effort to increase recombinant antigen expression levels
- Inclusion of a transcription terminator downstream of *rap* to determine whether vector stability may be improved by preventing transcriptional read-through into the oriE
- Inclusion of a transcription terminator downstream of *repB* to establish if vector stability may be improved by preventing transcriptional read-through into the promoter region, upstream of the recombinant antigen expression cassette.

In this study, episomal *E. coli*-mycobacterial shuttle vectors pHIGH100 and pRC100 were selected as the derivative plasmids from which to generate modified expression vectors. While both vectors' mycobacterial replicons are based on pAL5000, pHIGH100 was utilised as the template for bioinformatic searches and subsequent PCR because it contained many of the elements desired within the modified mycobacterial replicon, including the high copy number mutation and the full length *rap* gene (See Fig. 2.1). Plasmid pRC100 provided expression vectors with a backbone which contained the model GFP recombinant antigen, a kanamycin resistance gene, and an oriE.

3.1.1. BIOINFORMATIC SCANNING OF THE MYCOBACTERIAL REPLICON FOR TERMINATORS

The mycobacterial replicon within pHIGH100 further differs from the mycobacterial replicon commonly found in *E. coli*-mycobacterial shuttle vectors in that it contains the downstream flanking regions of the *rap* and *repB* genes. Accordingly these regions in pHIGH100 were analysed using bioinformatics to identify potential intrinsic terminators for *rap* and *repB*. The stem-loop/hairpin search tool of the DNAMAN bioinformatics software package [Lynnon Biosoft, version 4.0.0.1.] was used to scan for terminators comprising a minimum stem

length of seven paired base pairs and a maximum loop of 35 unpaired base pairs. These parameters were relatively lenient, considering that a stem length containing a minimum of 12 bp had been defined previously to identify terminators in prokaryotes bioinformatically (Petrillo *et al*, 2006).

3.1.1.1. Identification of a Putative *rap* Gene Transcription Terminator

A putative transcription terminator was identified 69 bp after the *rap* gene stop codon in pHIGH100 (Figure 3.1). The terminator has a poly-U tailed stem-loop structure. This sequence was consequently included to effectuate transcriptional termination downstream of *rap*.

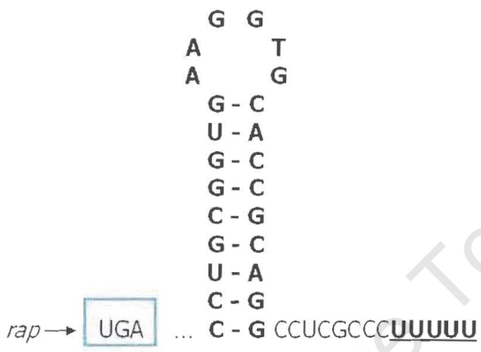

Figure 3.1. Putative poly-U tailed transcription terminator identified for *rap*.
Arrow indicates direction of *rap* transcription, UGA likely stop codon is boxed and ellipses represent 69 bp.

3.1.1.2. Absence of a *repB* Terminator

No homology to intrinsic terminator sequences was identified in the region following the *repB* gene. Consequently, the *M. bovis hsp60* terminator was introduced directly upstream of the *hsp60* promoter, to prevent transcriptional read-through into the promoter of the expression cassette. Approximately 466bp of DNA downstream of the *repB* gene was included as this region contains numerous direct and inverted repeats that could play a role in plasmid replication and stability.

3.1.2. SEQUENCING OF THE MYCOBACTERIAL REPLICON OF THE PCR TEMPLATE

The mycobacterial replicon of plasmid pHIGH100 served as the template for PCR amplification for construction of the modified mycobacterial replicon. Prior to amplification, the nucleotide sequence of the pHIGH100 mycobacterial replicon was confirmed since the GenBank pHIGH100 and pAL5000 nucleotide sequences were known to differ from that of the pHIGH100 used in our laboratory (Personal communication from Dr William Bourn).

3.1.2.1. Identification of an Inconsistency Between Template and Published pAL5000 Mycobacterial Replicon Sequence

Sequencing of the mycobacterial replicon of the pHIGH100 PCR template revealed an inconsistency with the GenBank pAL5000 mycobacterial replicon sequence. Within *repA*, there exists a triple direct repeat in the GenBank pAL5000 sequence. One of these 12 bp repeats appeared deleted in the pHIGH100 template such that only a single direct repeat remained (Fig. 3.2). As expected, the pHIGH100 high copy number deletion was absent in the pAL5000 mycobacterial replicon.

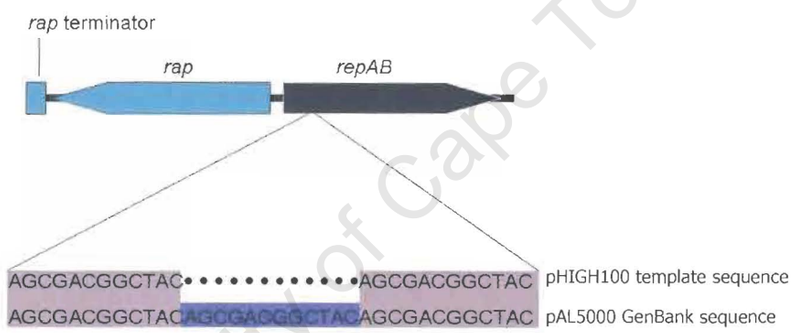

Figure 3.2. Deletion in *repA* of the pHIGH100 template.

An alignment of the triple direct repeat is shown; grey shading indicates 100% sequence homology, blue shading indicates the sequence of the deleted repeat; ellipses represent the corresponding deletion in pHIGH100.

3.1.3. PCR AMPLIFICATION OF THE MYCOBACTERIAL REPLICON CONTAINING FULL *RAP* AND THE HIGH COPY NUMBER MUTATION

Plasmid pHIGH100 contained three of the modifications required within the final modified mycobacterial replicon; the high copy number mutation, the full length *rap* gene and the putative *rap* terminator. However, the presence of a second *EcoRI* site within *repA* of the pHIGH100 mycobacterial replicon was undesirable since vectors in our laboratory exploit a unique *EcoRI* site positioned at the 3' end of the promoter region of the expression cassette which allows promoter elements to be interchanged. Thus, a three-stage PCR strategy (See Fig. 2.2) was employed in order to isolate the modified mycobacterial replicon from pHIGH100, introduce flanking restriction sites to facilitate subsequent cloning, and preserve the uniqueness of the *EcoRI* site upstream of the promoter.

For all three PCR reactions, products were successfully amplified (Fig. 3.3). With the exception of the PCR amplification stage two forward primer control (Fig 3.3.B, lane 4), control reactions did not yield PCR products. The resulting non-specific product from the PCR stage two forward primer control was approximately 1 kb in size, and thus distinguishable from the desired product of 1.171 kb. In addition, the non-specific product was absent from the complete PCR reaction containing the desired product. Sequencing confirmed error-free amplification of PCR products from stages one and two.

3.1.4. CONSTRUCTION OF AN *E. COLI*-MYCOBACTERIAL SHUTTLE VECTOR CONTAINING THE MODIFIED MYCOBACTERIAL REPLICON

The fully modified vector pSG300 was constructed to contain a mycobacterial replicon characterised by the full length *rap* gene and its putative terminator, *repA* including the high copy number mutation, *repB*, and an *hsp60* transcription terminator downstream of *repB*. The cloning strategy employed to generate this final vector, outlined in Figure 3.4, was based on the construction of two sub-clones; pSG100 and pSG200. A series of restriction enzyme digests confirmed successful generation of sub-clones pSG100 and pSG200 and modified expression vector pSG300 (Table 3.1. and Figure 3.5).

3.1.4.1. Construction of Sub-Clone pSG100

The *EcoRI*-free 2.409 kb PCR product of the mycobacterial replicon required for construction of the modified vector was cloned into pGEM-T Easy to construct sub-clone pSG100 (Fig. 3.4). The desired half of the PCR stage three yield containing the disrupted *EcoRI* restriction

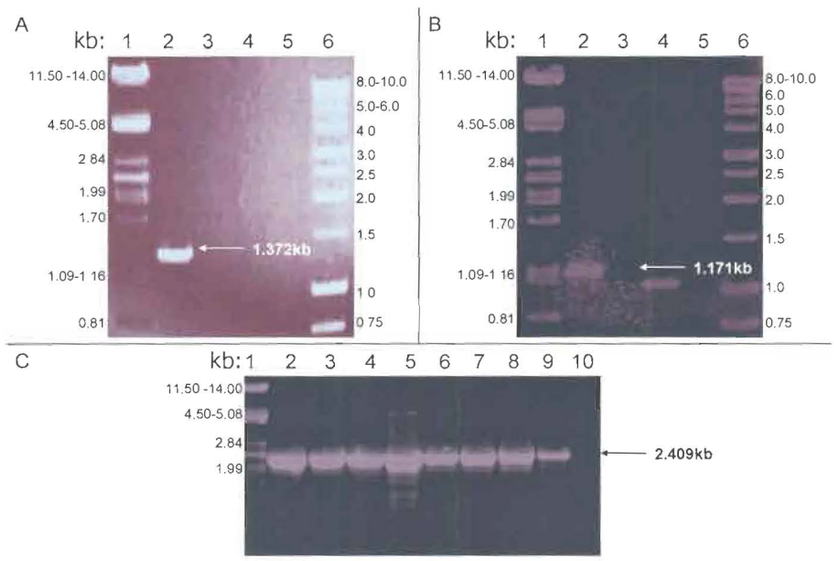

Figure 3.3. PCR Amplification of the Mycobacterial Replicon.

A. PCR Stage one. Amplification of a 1.372 kb portion from the mycobacterial replicon of pHIGH100 containing part of *repA*, the entire *rap* gene and the putative *rap* terminator. To assist in future cloning, primer SGrap-F introduced a flanking *MluI* restriction site downstream of the likely *rap* terminator, and primer SGrepA-R introduced an arbitrary nucleotide within the *EcoRI* site of *repA* to disrupt it. Lane 1- λDNA/*PstI* molecular weight marker, Lane 2- 1.372 kb product, Lane 3- No template control, Lane 4- Forward primer only specificity control, Lane 5- Reverse primer only specificity control, Lane 6- Promega 1 kb molecular weight marker. **B. PCR Stage two.** Primers SGrepA-F and SGrepB-R were used to amplify a 1.171 kb portion from the mycobacterial replicon of pHIGH100 which overlapped with the PCR product amplified in stage one by 150 bp and included the high copy number mutation within *repA*, *repB* and some downstream sequence of *repB*. To facilitate future cloning, SGrepB-R introduced a flanking *KpnI* restriction site downstream of *repB*. Lane 1- λDNA/*PstI* molecular weight marker, Lane 2- 1.171 kb product, Lane 3- No template control, Lane 4- Forward primer only specificity control, Lane 5- Reverse primer only specificity control, Lane 6- Promega 1kb molecular weight marker. **C. PCR Stage three.** Equal concentrations of product from stage one and two served as template for primers SGrap-F and SGrepB-R to amplify the final 2.409 kb modified mycobacterial replicon, characterised by full length *rap* and its putative terminator, *repA*, the high copy number mutation, *repB*, some sequence downstream of *repB* and flanking restriction sites with half of the yield containing the intact *EcoRI* site and the other half containing the disrupted *EcoRI* site. Lane 1- λDNA/*PstI* molecular weight marker, Lane 2-10- 2.409 kb final mycobacterial replicon product, Lane 10- No template control.

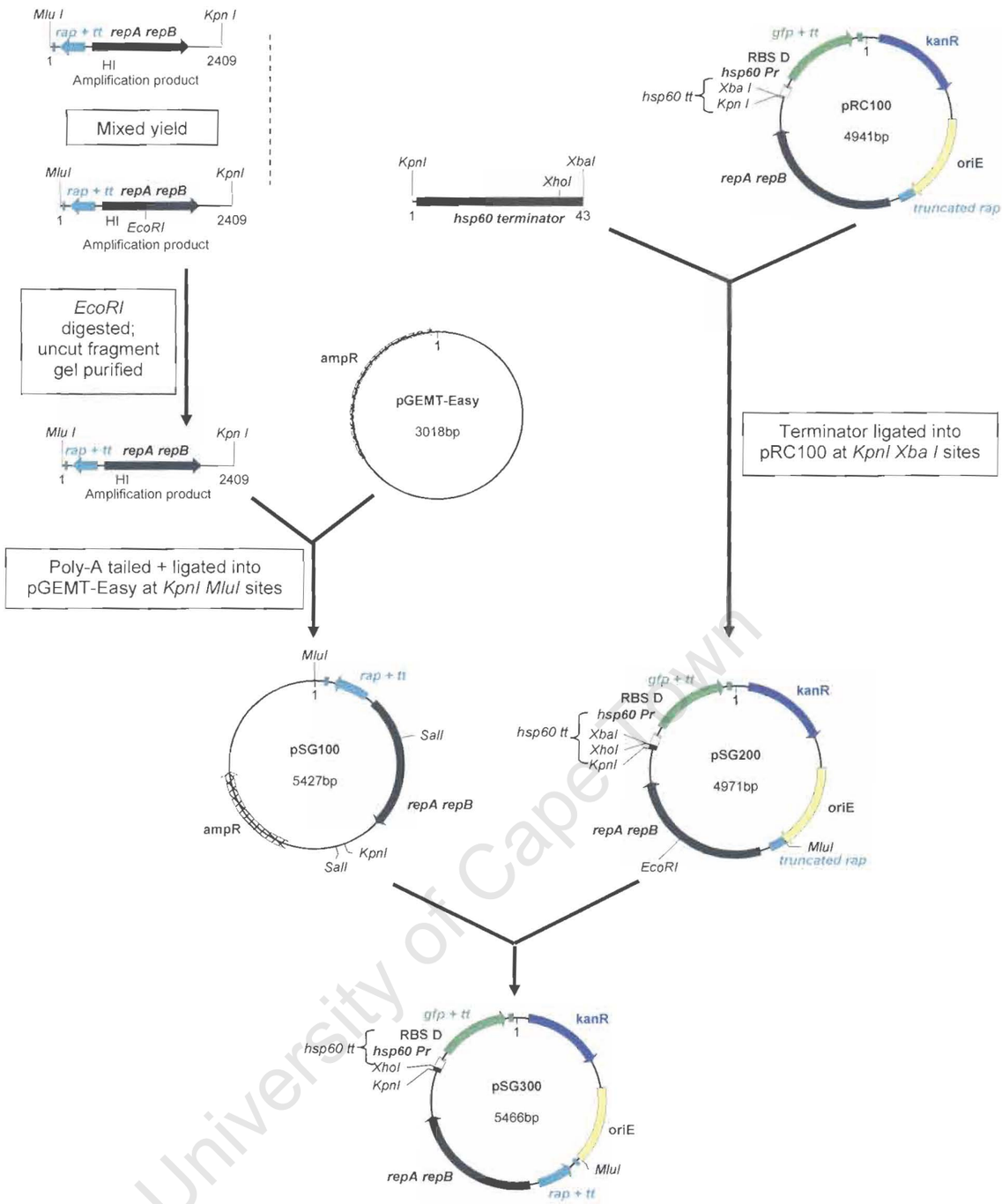

Figure 3.4. Overview of the cloning strategy used to generate modified vector pSG300.

Table 3.1. Confirmatory restriction digests following transformation of *E. coli* and mycobacteria

Expression vector	Plasmid size (bp)	Enzyme	Enzyme sites	Product sizes	Carried out following transformation of:		
					E. coli	*M. smegmatis*	BCG Δ*panCD*
Sub-clones							
pSG100	5424	*EcoRI*	52, 2479	2427, 3000	√		
		KpnI MluI	61, 2468, 2523	55, 2407, 2965	√		
		SalI	1434, 2499	1065, 4359	√		
pSG200	4971	*KpnI MluI*	2010, 3922	1912, 3059	√		
		XhoI	149, 2408, 3959	1161, 1551, 2259	√		
Modified vector							
pSG300	5466	*XhoI*	149, 2752, 4454	1161, 1702, 2603	√	√	√
		BglII HpaI	3126, 5355	2229, 3237	√		
		HpaI SnaBI	4590, 5355	765, 4701	√		
		HpaI XbaI	4460, 5355	895, 4571		√	
		KpnI MluI	2010, 4417	2407, 3059	√		
		MluI XbaI	2010, 4460	2450, 3016	√		
		NcoI	3588, 4801	1213, 4253	√		√
		SacI	2636, 3161, 4178	525, 1017, 3924	√		
Intermediate vectors							
pRC100	4941	*BglII*	2787, 3924	1142, 3799			√
		HpaI SnaBI	4065, 4830	765, 4176	√		
		MluI XbaI	2010, 3935	1925, 3016	√		
		NcoI	3246, 4276	1030, 3911			√
pSG400	5092	*BglII*	2782, 4075	1293, 3799	√	√	√
		HpaI SnaBI	4216, 4981	765, 4327	√		
		HpaI XbaI	4086, 4981	765, 4327		√	
		KpnI MluI	2010, 4073	2063, 3029	√		
		MluI XbaI	2010, 4086	2076, 3016	√		
		NcoI	3244, 4427	1183, 3909			√
		SacI	2292, 2817, 3834	525, 1107, 3550	√		
		SalI	2702, 4069	1367, 3572	√	√	
pSG500	5436	*BglII*	3126, 4419	1293, 4143	√		√
		HpaI SnaBI	4560, 5325	765, 4671	√		
		HpaI XbaI	4430, 5325	895, 4541		√	
		KpnI MluI	2010, 4417	2407, 3029	√		
		MluI XbaI	2010, 4430	2420, 3016	√		
		NcoI	3588, 4771	1183, 4253			√
		SalI	3046, 4566	1520, 3916	√		
pSG600	5122	*BglII HpaI*	2782, 5011	2229, 2893	√		
		HpaI SnaBI	4246, 5011	765, 4357	√		
		HpaI XbaI	4116, 5011	895, 4227		√	
		KpnI MluI	2010, 4073	2063, 3059	√		
		MluI XbaI	2010, 4116	2106, 3016	√		
		NcoI	3244, 4457	1213, 3909			√
		XhoI	149, 2408, 4110	1161, 1702, 2259	√	√	√

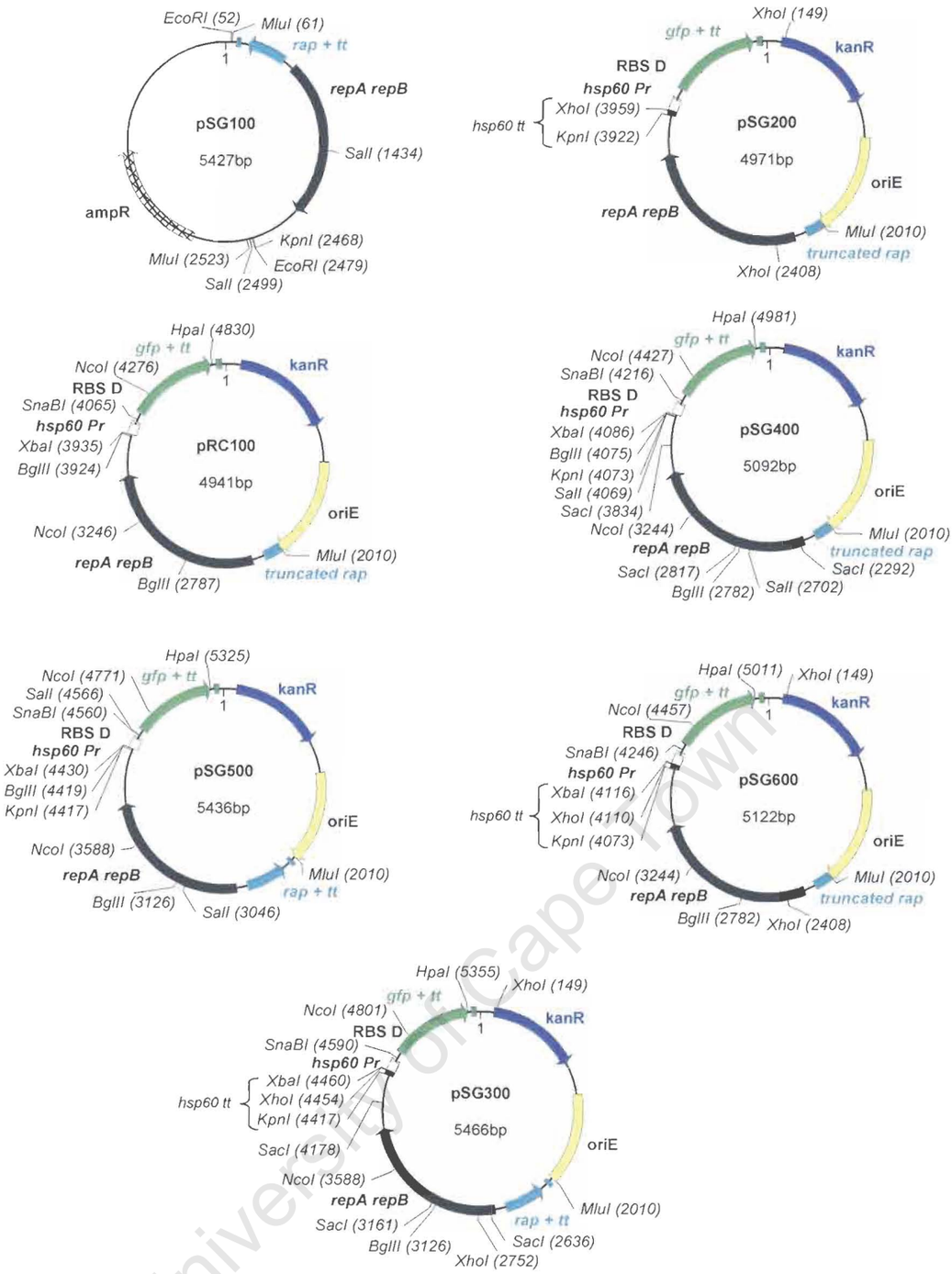

Figure 3.5. Restriction maps of sub-clones, intermediate *E. coli*-mycobacterial shuttle vectors and modified *E. coli*-mycobacterial shuttle vectors.

site was successfully isolated from the mixed population by an *EcoRI* restriction digest (Fig. 3.6). Sequencing of the mycobacterial replicon insert in pGEM-T Easy confirmed the presence of the modifications in sub-clone pSG100; namely the high copy number mutation, the entire *rap* gene, a putative *rap* terminator and a disrupted *EcoRI* site.

3.1.4.2. Construction of Sub-Clone pSG200

Sub-clone pSG200 was generated to provide both the vector backbone and the *hsp60* terminator downstream of *repB* for construction of the modified vector. Oligonucleotides with sequence homology to the *M. bovis hsp60* terminator were cloned downstream of the *repB* gene in unmodified expression vector pRC100 to generate pSG200 with a vector backbone consisting of the kanamycin resistance gene, the oriE, the *gfp* gene and the *rrsn* terminator (Fig. 3.4).

3.1.4.3. Construction of Modified Vector pSG300

The final modified vector pSG300 was constructed by ligation of the pSG100 modified mycobacterial replicon to the *hsp60* terminator and vector backbone of pSG200 (Fig. 3.4). Modified vector pSG300 thus contained a mycobacterial replicon with all of the desired modifications; the full length *rap* gene and its putative terminator, the high copy number mutation within *repA*, and the *hsp60* terminator downstream of *repB*.

3.2. CONSTRUCTION OF INTERMEDIATE *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS

To evaluate the collective contribution of the various modifications made to the mycobacterial replicon, fully modified expression vector pSG300 could have been directly compared to pRC100, harbouring an unmodified mycobacterial replicon. However, in order for any observed difference in vector stability, copy number or recombinant antigen expression level to be accurately attributed to a specific modification, a series of intermediate expression vectors were required which differed from pSG300 by one or more modification. Three intermediate expression vectors pSG400, pSG500, and pSG600 were constructed for this purpose. The various modifications included in each vector are tabulated in Table 3.2. Cloning strategies are outlined in Figure 3.7, 3.8, and 3.9. respectively. Restriction enzyme digests confirmed successful generation of intermediate expression vectors pSG400, pSG500 and pSG600 in *E. coli* (Table 3.1. and Figure 3.5).

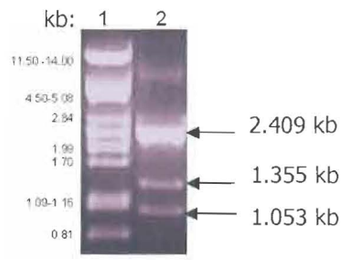

Figure 3.6. *EcoRI* digest of the final stage three PCR yield.

The desired half of the PCR stage three yield containing the disrupted *EcoRI* restriction site was isolated from the mixed population of 2.409 kb products by an *EcoRI* restriction digest. The half of the PCR products containing the intact site was digested to yield two fragments of 1.053 kb and 1.355 kb while the 2.409 kb fragment remained undigested and was therefore easily gel purified. Lane 1- λDNA/*PstI* molecular weight marker, Lane 2- *EcoRI* digested PCR products from stage three. Arrows indicate digested fragment sizes.

Table 3.2. Differences between mycobacterial replicons of vectors used in evaluating modified expression vector

Vector	High copy number mutation	Full length *rap* gene	*hsp60* transcription terminator
pRC100	x	x	x
pSG400	√	x	x
pSG500	√	√	x
pSG600	√	x	√
pSG300	√	√	√

Final modified *E. coli*-mycobacterial shuttle vector is shown in boldface

3.3. TRANSFORMATION OF *E. COLI*-MYCOBACTERIAL SHUTTLE VECTORS INTO MYCOBACTERIA

Since this study forms part of a larger initiative to develop mycobacterial-based vaccines for Southern Africa, mycobacterial strains were selected accordingly. BCG Pasteur Δ*panCD* is a pantothenic acid-deficient mutant. This particular strain of BCG was utilised in the present study because it exhibits a similar growth rate to wild-type BCG Pasteur, if supplemented with

pantothenic acid. Since the availability of pantothenic acid is very limited within the phagosome environment, the growth of this strain is expected to be severely impaired, and hence represent a safer candidate vaccine strain for use in immunocompromised individuals (Sambandamurthy *et al*, 2002). *M. smegmatis* mc^2155 is an efficient plasmid transformation mutant which has been used extensively in mycobacterial studies. The strain was chosen for use in this study as it displays a comparable growth rate to wild-type *M. smegmatis* and, in addition, is currently being pursued as a live bacterial vaccine vector.

The modified and intermediate *E. coli*-mycobacterial shuttle vectors were successfully transformed into *M. smegmatis* and BCG Δ*panCD* to allow evaluation of recombinant mycobacteria in terms of vector stability, copy number and recombinant antigen expression levels. Restriction enzyme digests verified both general plasmid structure and the specific GFP antigen region (Table 3.1. and Figure 3.5).

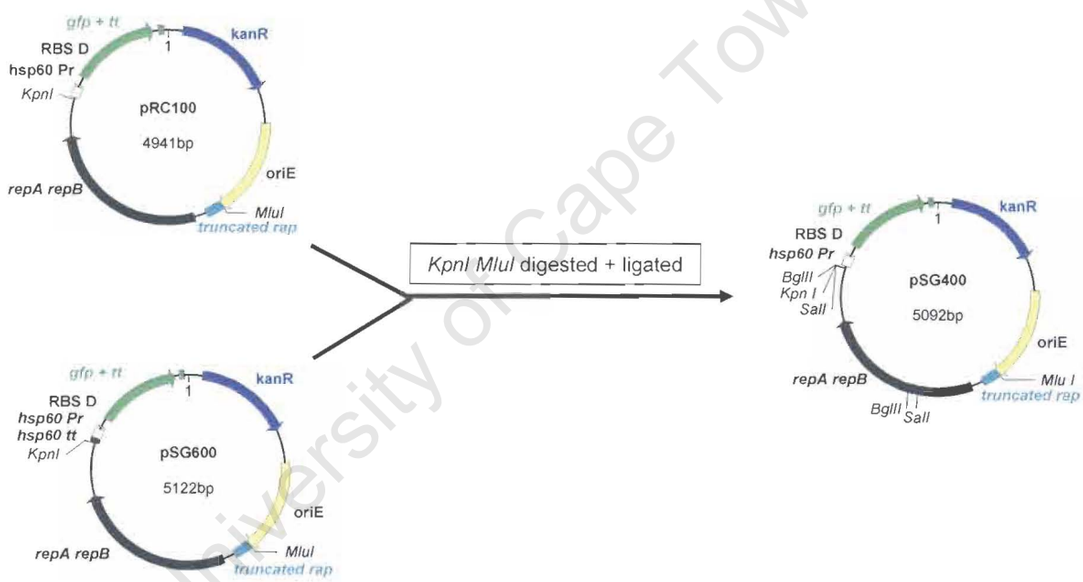

Figure 3.7. Cloning strategy used to generate intermediate construct pSG400.

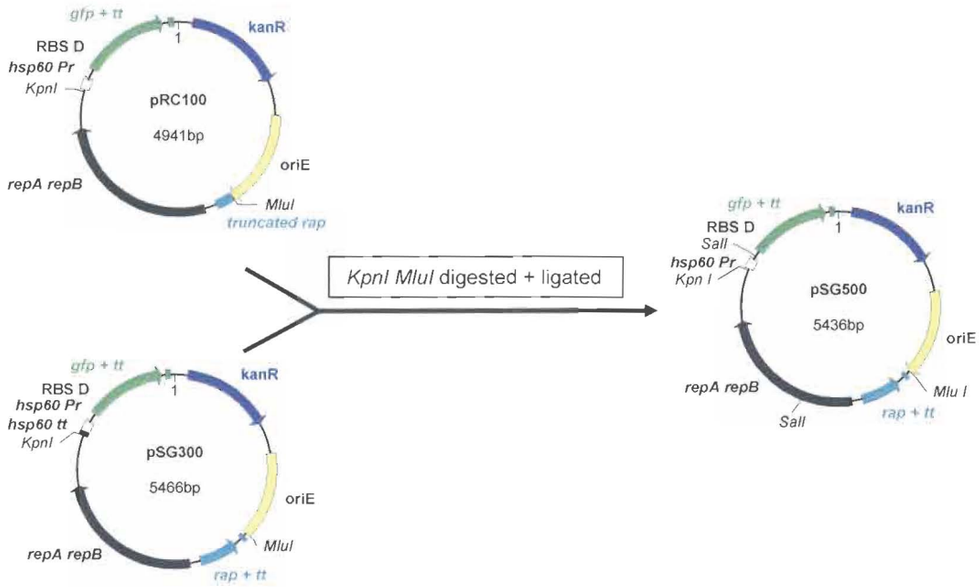

Figure 3.8. Cloning strategy used to generate intermediate construct pSG500.

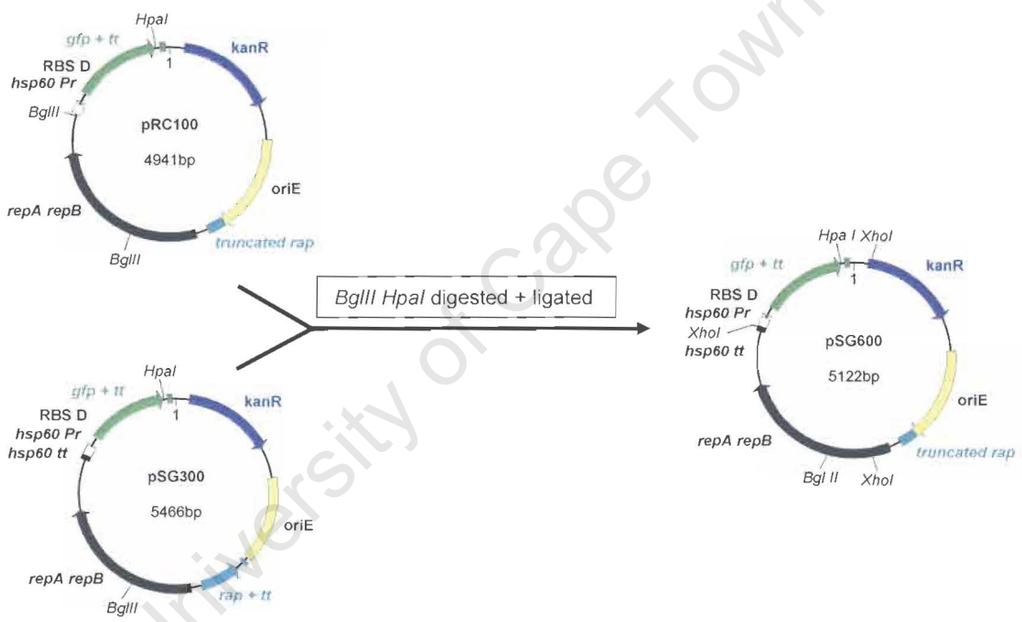

Figure 3.9. Cloning strategy used to generate intermediate construct pSG600.

3.4. EVALUATION OF THE MODIFIED *E. COLI*-MYCOBACTERIAL SHUTTLE VECTOR

3.4.1. EVALUATION OF GFP EXPRESSION LEVELS

Recombinant antigen expression levels are typically low in *E. coli*-mycobacterial shuttle vectors. In addition to increasing recombinant antigen expression levels through optimisation of the recombinant gene itself, an increase in plasmid copy number has been shown to increase recombinant antigen expression levels. Two of the modifications made to the mycobacterial replicon in this study; inclusion of the full length *rap* gene and introduction of a high copy number mutation, were expected to increase levels of recombinant antigen expression indirectly through an increase in plasmid copy number. It was therefore necessary to evaluate the modified expression vectors in terms of recombinant antigen expression levels. GFP was selected as the recombinant antigen in this study for three reasons. First, GFP is considered a highly stable protein and it is thus suited to *in vitro* analysis. Second, GFP has fluorescent properties which allow it to be visualised under UV light and finally, it has been used extensively in mycobacteria as a marker for gene expression (Kremer *et al*, 1995;Dhandayuthapani *et al*, 1995).

3.4.1.1. GFP Fluorescence Study

The ability of GFP to fluoresce under UV light was exploited to provide a preliminary, crude indication of whether the final modified expression vector exhibited increased recombinant antigen expression and to determine which modifications appeared responsible. *M. smegmatis* and BCG Δ*panCD* were transformed with final modified vector pSG300, unmodified vector pRC100 and partially modified intermediate vectors pSG400, pSG500 and pSG600. Recombinant *M. smegmatis* and BCG Δ*panCD* colonies were viewed under UV light to visualise relative levels of GFP fluorescence. Fluorescent r*M. smegmatis* and rBCG Δ*panCD* colonies are shown in Figures 3.10. and 3.11. respectively. Due to the non-qualitative nature of this experiment, relative GFP fluorescence levels were recorded simply as either 'low' or 'high'. Results are tabulated in Table 3.3. for r*M. smegmatis* colonies and in Table 3.4. for rBCG Δ*panCD* colonies.

Observed differences in fluorescence do not represent definitive differences in GFP expression levels, as there are limitations to this study. Foremost, fluorescence intensities were not evaluated qualitatively and, as such, are purely observational. Second, although conditions for colony growth; such as time, temperature, media and environment, were uniform, colony

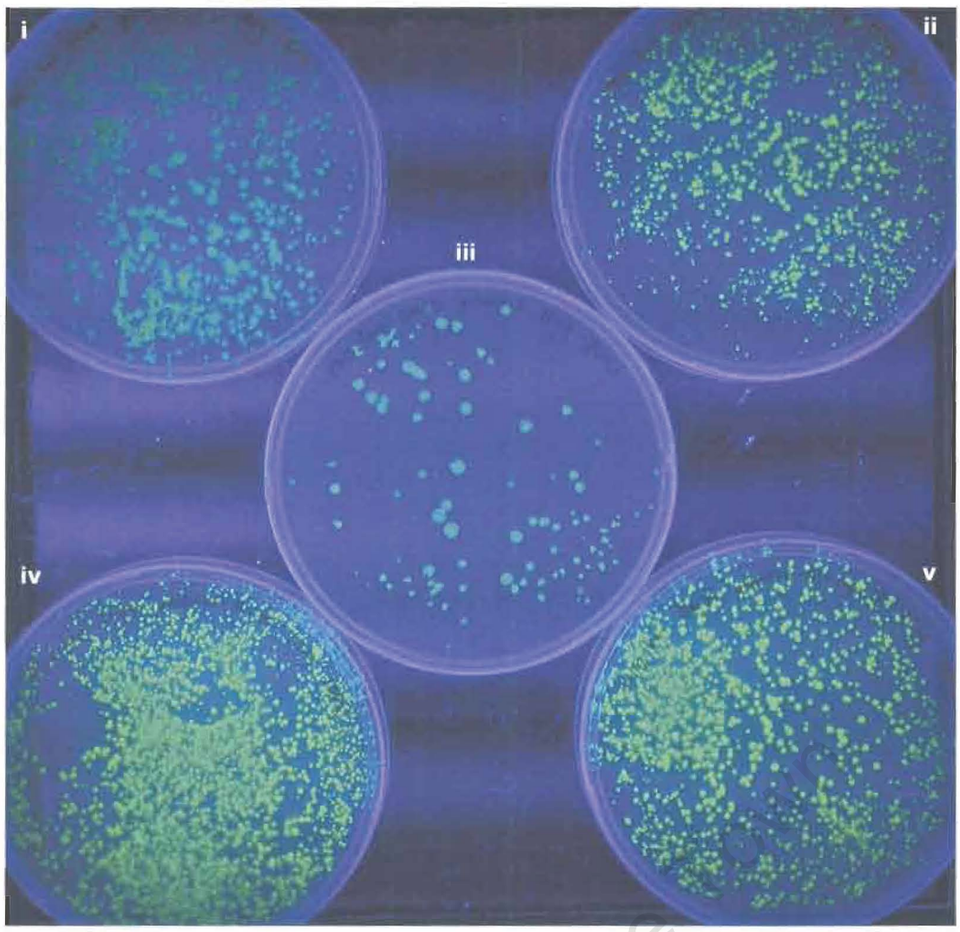

Figure 3.10. GFP Fluorescence of r*M. smegmatis* colonies viewed under UV light.

i- r*M. smegmatis* [pSG400] colonies; expression vector pSG400 contains the high copy number mutation, **ii**- r*M. smegmatis* [pSG500] colonies; expression vector pSG500 contains the high copy number mutation and the entire *rap* gene including its putative terminator, **iii**- r*M. smegmatis* [pRC100] colonies; pRC100 is the unmodified expression vector, **iv**- r*M. smegmatis* [pSG600] colonies; expression vector pSG600 contains the high copy number mutation and the *hsp60* terminator following *repB*, **v**- r*M. smegmatis* [pSG300] colonies; fully modified expression vector pSG300 contains the high copy number mutation, the entire *rap* gene including its putative terminator, and the *hsp60* terminator following *repB*.

Table 3.3. Comparison of GFP Fluorescence of recombinant *M. smegmatis* colonies

E. coli-mycobacterial shuttle vector	GFP Fluorescence
pRC100	Low
pSG400	Low
pSG500	High
pSG600	High
pSG300	**High**

Modified expression vector is shown in boldface

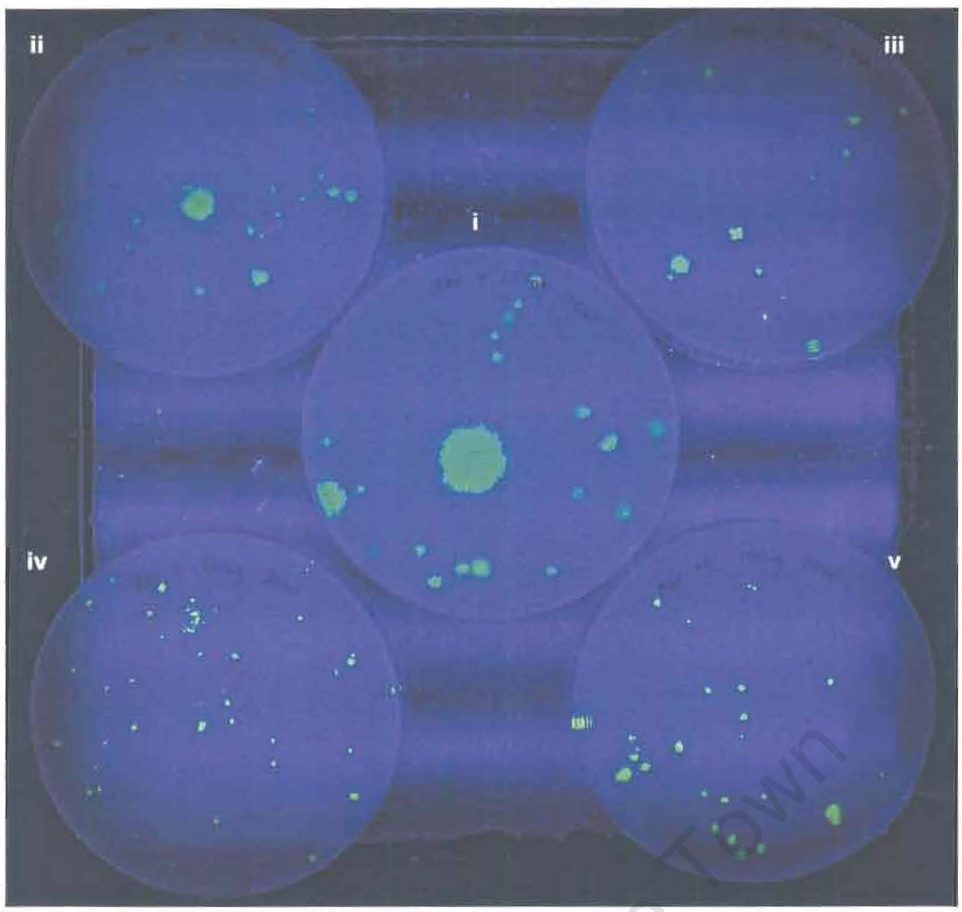

Figure 3.11. GFP Fluorescence of rBCG Δ*panCD* colonies viewed under UV light.

i- rBCG Δ*panCD* [pSG400] colonies; expression vector pSG400 contains the high copy number mutation, **ii**- rBCG Δ*panCD* [pSG500] colonies; expression vector pSG500 contains the high copy number mutation and the entire *rap* gene including its putative terminator, **iii**- rBCG Δ*panCD* [pRC100] colonies; pRC100 is the unmodified expression vector, **iv**- rBCG Δ*panCD* [pSG600] colonies; expression vector pSG600 contains the high copy number mutation and the *hsp60* terminator following *repB*, **v**- rBCG Δ*panCD* [pSG300] colonies; fully modified expression vector pSG300 contains the high copy number mutation, the entire *rap* gene including its putative terminator, and the *hsp60* terminator following *repB*.

Table 3.4. Comparison of GFP Fluorescence of recombinant BCG Δ*panCD* colonies

E. coli-mycobacterial shuttle vector	GFP Fluorescence
pRC100	Low
pSG400	Low
pSG500	Low
pSG600	High
pSG300	**High**

Modified expression vector is shown in boldface

growth rates may have varied among recombinants. Accordingly, colonies which may have been compared during different phases of growth, could be characterised by different rates of gene expression and protein synthesis. Seventy five percent of *M. smegmatis* proteins, for example, have been shown to vary in expression level between log phase and stationary phase (Blokpoel *et al*, 2005). Thus, observed differences in colony fluorescence may reflect different growth phases rather than genuine differences in expression levels.

In *M. smegmatis*, GFP fluorescence of colonies harbouring the fully modified vector pSG300 (Fig. 3.10.v) were distinctly brighter than those transformed with the unmodified vector pRC100 (Fig. 3.10.iii). Unexpectedly, colonies containing the high copy number intermediate vector pSG400 (Fig. 3.10.i) did not appear to fluoresce with greater intensity than unmodified r*M. smegmatis* [pRC100] colonies. Both r*M. smegmatis* [pSG500] and [pSG600] colonies (Fig. 3.10.ii and iv), containing high copy number intermediate vectors with full length *rap* and its putative terminator, and the *hsp60* terminator after *repB* respectively, exhibited increased GFP fluorescence relative to unmodified r*M. smegmatis* [pRC100] colonies.

As in *M. smegmatis*, rBCG Δ*panCD* colonies containing the modified expression vector pSG300 (Fig. 3.11.v) fluoresced with notably greater intensity than unmodified rBCG Δ*panCD* [pRC100] colonies (Fig. 3.11.iii). Recombinant BCG Δ*panCD* colonies containing the high copy number intermediate vector pSG400 (Fig. 3.11.i) again showed an unexpected similarity in fluorescence intensity to those transformed with unmodified vector pRC100. In contrast to *M. smegmatis* results, however, rBCG Δ*panCD* [pSG500] colonies (Fig. 3.11.ii) harbouring the full length *rap* and its putative terminator in addition to the high copy number mutation did not appear to fluoresce with increased intensity in comparison to unmodified rBCG Δ*panCD* [pRC100] colonies. Colonies containing pSG600 (Fig. 3.11.iv), the intermediate expression vector characterised by the high copy number mutation and the *hsp60* terminator after *repB*, displayed a clear increase in fluorescence relative to unmodified rBCG Δ*panCD* [pRC100] colonies.

3.4.1.2. GFP Western Blot Analysis

Since preliminary GFP fluorescence results suggested that some of the modifications made to the mycobacterial replicon conferred an increase in GFP expression levels, relative GFP expression levels were investigated by more accurate, qualitative Western blot analysis. Figure 3.12. and 3.13, respectively, depict Western blot results of r*M. smegmatis* and rBCG

Δ*panCD* protein samples probed with GFP-specific antibody. Protein samples utilised as well as their experimental functions are outlined in figure legends.

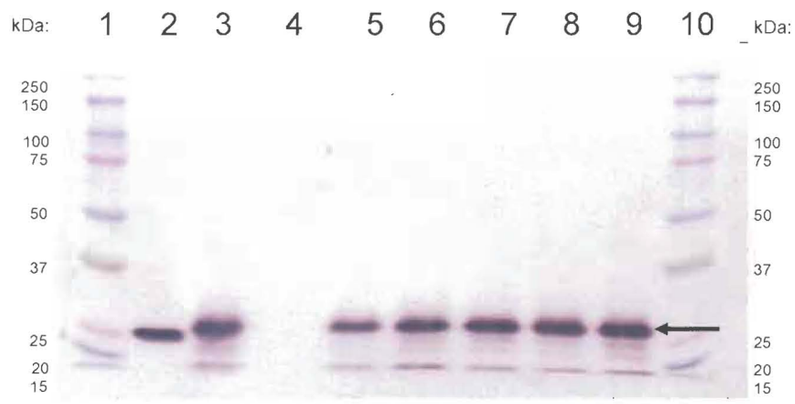

Figure 3.12. GFP Western Blot Analysis of r*M. smegmatis* protein samples.

Cell lysate was obtained from log phase r*M. smegmatis* cultures and loaded into lanes as follows: **Lane 1**- 10μl Kaleidoscope Protein Marker, **Lane 2**- 15ng purified GFP (positive control), **Lane 3**- 2μg r*M. smegmatis* [pHS207] protein (high GFP-expressing positive control), **Lane 4**- 2μg r*M. smegmatis* [pHS200] protein (GFP-free negative control), **Lane 5**- 2μg r*M. smegmatis* [pRC100] protein (unmodified vector), **Lane 6**- 2μg r*M. smegmatis* [pSG400] protein (intermediate vector with high copy number mutation), **Lane 7**- 2μg r*M. smegmatis* [pSG500] protein (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), **Lane 8**- 2μg r*M. smegmatis* [pSG600] protein (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), **Lane 9**- 2μg r*M. smegmatis* [pSG300] protein (fully modified vector), **Lane 10**- 10μl Kaleidoscope Protein Marker. A GFP-specific antibody was used to probe membranes. The arrow indicates protein bands of approximately 27 kDa, correlating to the size of GFP.

Positive and negative controls gave anticipated results in the Western blot analysis of r*M. smegmatis* samples. The GFP-specific antibody recognised a protein band of approximately 27 kDa, correlating to the size expected for GFP, in purified GFP (Fig. 3.12. Lane 2) and high GFP-expressing r*M. smegmatis* [pHS207] (Fig. 3.12. Lane 3) positive controls. GFP-free r*M. smegmatis* [pHS200] (Fig. 3.12. Lane 4), utilised as a negative control, did not react with the GFP antibody. The 27 kDa band corresponding to GFP was detected in all experimental r*M. smegmatis* protein samples (Fig. 3.12. Lane 5-9). While absent from the purified GFP positive control (Fig. 3.12. Lane 2), an additional slew of smaller bands was consistently detected in all GFP-expressing r*M. smegmatis* protein samples (Fig. 3.12. Lane 3 and 5-9). Although GFP

is characterised as highly stable (Yang *et al*, 1996), these additional bands are likely degraded GFP. Degradation may have occurred during protein extraction as additional bands are present in all recombinant *M. smegmatis* samples expressing GFP, but not in the GFP standard. Alternately, the smaller bands may represent GFP that has been prematurely truncated. The *gfp* gene which was available for inclusion in vectors in this study was originally modified for use in *Arabidopsis thaliana*. It has a G-C content of just 45%, differing greatly from mycobacteria which are typically G-C rich. Accordingly, A-T rich residues may place a significant metabolic load on *M. smegmatis* and BCG in which such codons are rare (Table 3.5).

Table 3.5. Differences in codon usage between mycobacteria and GFP utilised in this study

Residue	Amino Acid Composition	Percentage Frequency in:		
		M. smegmatis	*M. bovis*	*A. thaliana*-derived GFP
Phenylalanine	TTT	5	21	42
Leucine	TTA	0	1	11
Isoleucine	ATT	4	15	50
Lysine	AAA	18	25	50

Calculated using the Codon Usage Analysis Tool of The Institute for Genomic Research (TIGR, 2007)

The relative intensity of the GFP protein band in experimental samples revealed differences in levels of GFP expression between vectors. Recombinant *M. smegmatis* [pSG300] containing the fully modified expression vector (Fig. 3.12. Lane 9) showed the most marked increase in GFP band intensity compared to r*M. smegmatis* [pRC100] containing the unmodified vector (Fig. 3.12. Lane 5). Recombinant samples containing the intermediate expression vectors (Fig. 3.12. Lane 6-8) all appeared to exhibit an increased GFP band intensity relative to unmodified r*M. smegmatis* [pRC100]. The lowest increase in relative GFP band intensity was observed for r*M. smegmatis* [pSG400] (Fig. 3.12. Lane 6) harbouring the intermediate vector with the high copy number mutation. A greater increase in relative band intensity was demonstrated by r*M. smegmatis* [pSG500] (Fig. 3.12. Lane 7) which contained the intermediate vector characterised by the high copy number mutation in addition to the full length *rap* and its putative terminator. In turn, an even greater increase in GFP band intensity relative to unmodified r*M. smegmatis* [pRC100] was evident for r*M. smegmatis* [pSG600] (Fig. 3.12. Lane 8) which contained the intermediate vector modified to include the high copy number mutation and the *hsp60* terminator after *repB*. Duplicate Western blots supported the reproducibility of results. Silver staining of replicate SDS-PAGE gels confirmed that all r*M. smegmatis* protein samples were loaded in equal amounts of 2μg.

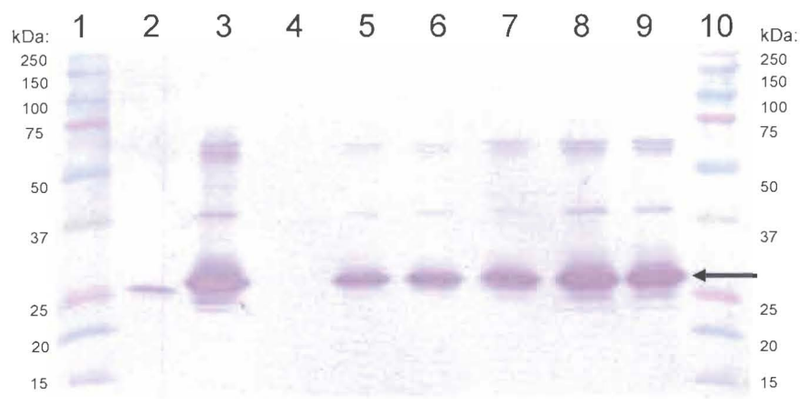

Figure 3.13. GFP Western Blot Analysis of rBCG Δ*panCD* protein samples.

Cell lysate was obtained from log phase rBCG Δ*panCD* cultures and loaded into lanes as follows: **Lane 1**- 10μl Kaleidoscope Protein Marker, **Lane 2**- 15ng purified GFP (positive control), **Lane 3**- 2μg rBCG Δ*panCD* [pHS207] protein (high GFP-expressing positive control), **Lane 4**- 2μg rBCG Δ*panCD* [pHS200] protein (GFP-free negative control), **Lane 5**- 2μg rBCG Δ*panCD* [pRC100] protein (unmodified vector), **Lane 6**- 2μg rBCG Δ*panCD* [pSG400] protein (intermediate vector with high copy number mutation), **Lane 7**- 2μg rBCG Δ*panCD* [pSG500] protein (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), **Lane 8**- 2μg rBCG Δ*panCD* [pSG600] protein (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), **Lane 9**- 2μg rBCG Δ*panCD* [pSG300] protein (fully modified vector), **Lane 10**-10μl Kaleidoscope Protein Marker. A GFP-specific antibody was used to probe membranes. The arrow indicates protein bands of approximately 27 kDa, correlating to the size of GFP.

GFP Western results for rBCG Δ*panCD* were largely consistent with r*M. smegmatis* results. Positive and negative controls behaved as anticipated in the Western blot analysis of rBCG Δ*panCD* samples. The GFP-specific antibody recognised a protein band of approximately 27 kDa, correlating to the size expected for GFP, in purified GFP (Fig. 3.13. Lane 2) and high GFP-expressing rBCG Δ*panCD* [pHS207] (Fig. 3.13. Lane 3) positive controls. The GFP-free rBCG Δ*panCD* [pHS200] (Fig. 3.13. Lane 4), which was utilised as a negative control, did not react with the antibody to GFP. The 27 kDa band corresponding to GFP was detected in all experimental rBCG Δ*panCD* protein samples (Fig. 3.13. Lane 5-9). Aside from the 27 kDa GFP band detected, a number of additional bands were consistently detected in all GFP-expressing rBCG Δ*panCD* protein samples (Fig. 3.13. Lane 3 and 5-9). In addition to the 27 kDa band anticipated for GFP, additional larger bands were seen in all GFP-expressing BCG Δ*panCD* samples but were absent from the GFP standard. This is unlikely to be a result of non-specific

protein binding, as no such bands are seen in the negative control, which lacks the *gfp* gene but is otherwise identical to other BCG Δ*panCD* samples. Furthermore, the additional bands vary in intensity proportionally to the 27 kDa band. One possible explanation for the additional band positioned slightly above the 50 kDa molecular weight is that it represents a 54 kDa GFP dimer. *In vitro* dimerisation of 27 kDa GFP subunits has been previously noted (Hofmann *et al*, 2002). The band of approximately 37 kDa cannot be explained. In contrast to *M. smegmatis*, relative to the 15ng standard, GFP expression levels in the BCG Δ*panCD* Western blot appeared generally higher than levels determined by ELISA. This may be explained by the fact that although the concentration of the purified GFP used as a standard was stated as equal between aliquots by the supplier, subsequent dot blot analysis revealed an inconsistency between aliquots utilised in the r*M. smegmatis* and BCG Δ*panCD* Western blots (Data not shown). This adds further weight to interpreting Western blot results qualitatively as opposed to semi-quantitatively.

The relative intensity of the GFP protein band in experimental samples revealed differences in levels of GFP expression among vectors. Recombinant BCG Δ*panCD* [pSG300] containing the fully modified expression vector (Fig. 3.13. Lane 9) showed the most considerable increase in GFP band intensity compared to rBCG Δ*panCD* [pRC100] which contained the unmodified vector (Fig. 3.13. Lane 5). Recombinant samples containing the intermediate expression vectors (Fig. 3.13. Lane 6-8) all exhibited a greater GFP band intensity than unmodified rBCG Δ*panCD* [pRC100]. The least substantial increase in relative GFP band intensity was observed for rBCG Δ*panCD* [pSG400] (Fig. 3.13. Lane 6) harbouring the intermediate vector modified to high copy number. A further increase in relative band intensity was demonstrated by rBCG Δ*panCD* [pSG500] (Fig. 3.13. Lane 7) which contained the intermediate vector with the high copy number mutation and the full length *rap* and its putative terminator. Relative to unmodified rBCG Δ*panCD* [pRC100], an even greater increase in GFP band intensity was evident for rBCG Δ*panCD* [pSG600] (Fig. 3.13. Lane 8) which contained the intermediate vector modified to include the high copy number mutation and the *hsp60* terminator after *repB*. Duplicate Western blots supported result reproducibility. Silver staining of replicate SDS-PAGE gels confirmed that all rBCG Δ*panCD* protein samples were loaded in equal amounts of 2μg.

3.4.1.3. GFP Capture ELISA

Seeing as qualitative Western blot analysis confirmed relative differences in GFP expression levels among recombinant mycobacteria, levels of GFP expression were further analysed by quantitative GFP capture ELISA. The amount of GFP expressed by r*M. smegmatis* and rBCG Δ*panCD* protein samples is shown in Figure 3.14. and 3.15. respectively. Protein samples utilised as well as their experimental functions are outlined in figure legends.

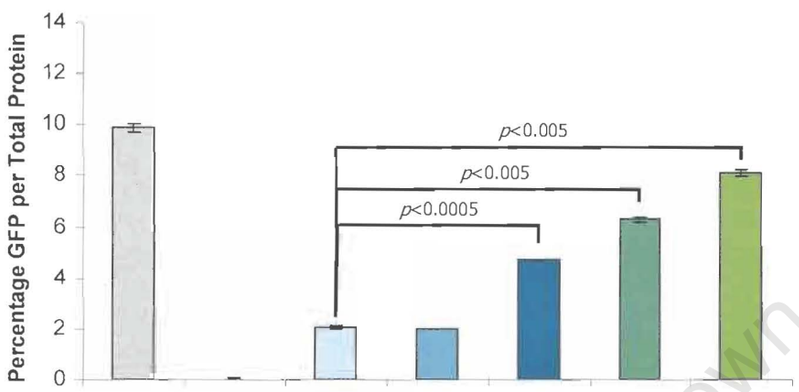

Figure 3.14. GFP Expression Levels of r*M. smegmatis* Samples Based on Capture ELISA.

Protein obtained from log phase r*M. smegmatis* cultures was used for GFP capture ELISA. Levels of GFP in samples were quantitated by extrapolation against a standard curve of known amounts of purified GFP protein. The amount of GFP expressed by r*M. smegmatis* protein samples is represented by bars coloured as follows: □ - r*M. smegmatis* [pHS207] (high GFP-expressing positive control), ■ - r*M. smegmatis* [pHS200] (GFP-free negative control), □ - r*M. smegmatis* [pRC100] (unmodified vector), ■ - r*M. smegmatis* [pSG400] (intermediate vector with high copy number mutation), ■ - r*M. smegmatis* [pSG500] (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), ■ - r*M. smegmatis* [pSG600] (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), ■ - r*M. smegmatis* [pSG300] (fully modified vector). Error bars represent the degree of variability between triplicate samples. Statistical significance of differences in expression levels is indicated above pairwise lines joining compared samples.

For both r*M. smegmatis* and rBCG Δ*panCD* capture ELISA experiments, the included positive and negative controls gave expected results. High amounts of GFP were detectable in positive controls r*M. smegmatis* and rBCG Δ*panCD* containing the high GFP-expressing vector pHS207

(□ in Fig. 3.14. and Fig. 3.15. respectively). No significant level of GFP was detected in negative controls r*M. smegmatis* and rBCG Δ*panCD* containing the non GFP-expressing vector pHS200 (■ in Fig. 3.14. and Fig. 3.15. respectively). The GFP standard curve was shown to be linear up to an amount of 10ng of GFP protein. For quantitation accuracy all samples were diluted as necessary, to fall within this linear range. Depicted error bars show little variability between triplicate samples suggesting that these results are reproducible.

In *M. smegmatis*, the fully modified vector pSG300 (■ in Fig. 3.14) expressed the greatest amount of GFP of all r*M. smegmatis* experimental samples. At a concentration of 8.1% of total protein, this represented a four-fold increase in expression compared to the 2.1% of total protein expressed by unmodified r*M. smegmatis* [pRC100] (□ in Fig. 3.14) with statistical significance of $p<0.005$. This result supports trends identified in fluorescence and Western blot studies in which r*M. smegmatis* [pSG300] exhibited the greatest GFP expression levels. Recombinant *M. smegmatis* [pSG400] (■ in Fig. 3.14) containing the intermediate vector characterised by the high copy number mutation behaved similarly to unmodified r*M. smegmatis* [pRC100], expressing GFP at a concentration of 2.0% per total protein. This confirms preliminary GFP fluorescence observations in which both r*M. smegmatis* [pRC100] and [pSG400] were recorded as low fluorescers. Interestingly, this capture ELISA datum disagreed with the qualitative GFP Western blot analysis which suggested that its GFP expression levels were greater than r*M. smegmatis* [pRC100] levels. A GFP expression level of 4.7% of total protein was determined for r*M. smegmatis* [pSG500] (■ in Fig. 3.14) containing the intermediate vector with the high copy number mutation and the full length *rap* and putative terminator. This correlated to a two-fold increase in GFP expression compared to unmodified r*M. smegmatis* [pRC100] with statistical significance of $p<0.005$, supporting both fluorescence and Western blot studies. Recombinant *M. smegmatis* [pSG600], harbouring the intermediate vector modified to high copy number and incorporating the *hsp60* terminator after *repB*, expressed GFP at a concentration of 6.3% per total protein (■ in Fig. 3.14). This corresponds to a three-fold increase in expression relative to unmodified r*M. smegmatis* [pRC100] at a statistical significance of $p<0.0005$. This result supports both preliminary fluorescence observations recording r*M. smegmatis* [pSG600] as a high fluorescer and subsequent Western blot findings that r*M. smegmatis* [pSG600] had the second greatest level of expression.

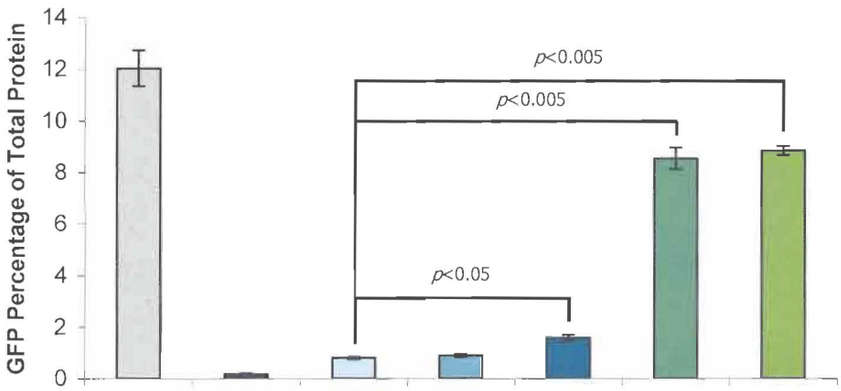

Figure 3.15. GFP Expression Levels of rBCG Δ*panCD* Samples Based on Capture ELISA.

Protein obtained from log phase rBCG Δ*panCD* cultures was used for GFP capture ELISA. Levels of GFP in samples were quantitated by extrapolation against a standard curve of known amounts of purified GFP protein. The amount of GFP expressed by rBCG Δ*panCD* protein samples is represented by bars coloured as follows: □ - rBCG Δ*panCD* [pHS207] (high GFP-expressing positive control), ■ - rBCG Δ*panCD* [pHS200] (GFP-free negative control), □ - rBCG Δ*panCD* [pRC100] (unmodified vector), ■ - rBCG Δ*panCD* [pSG400] (intermediate vector with high copy number mutation), ■ - rBCG Δ*panCD* [pSG500] (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), ■ - rBCG Δ*panCD* [pSG600] (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), ■ - rBCG Δ*panCD* [pSG300] (fully modified vector). Error bars represent the degree of variability between triplicate samples. Statistical significance of differences in expression levels is indicated above pairwise lines joining compared samples.

While gross trends observed for capture ELISA of rBCG Δ*panCD* samples were consistent with those noted for r*M. smegmatis* samples, fold differences in expression as well as actual levels of expression differed. The fully modified vector rBCG Δ*panCD* [pSG300] (■ in Fig. 3.15) expressed the greatest amount of GFP of rBCG Δ*panCD* experimental samples. At a concentration of 8.8% of total protein, this represents a fourteen-fold increase in expression compared to the 0.8% expressed by unmodified rBCG Δ*panCD* [pRC100] (□ in Fig. 3.15) with statistical significance of $p<0.005$. This result supports trends identified in fluorescence and Western blot studies in which rBCG Δ*panCD* [pSG300] exhibited the greatest level of GFP expression. Recombinant BCG Δ*panCD* [pSG400] (■ in Fig. 3.15) containing the intermediate vector characterised by the high copy number mutation behaved similarly to

unmodified rBCG Δ*panCD* [pRC100], expressing GFP at 0.9% of total protein. This confirmed preliminary GFP fluorescence observations in which both rBCG Δ*panCD* [pRC100] and [pSG400] were recorded as low fluorescers. In contrast, this capture ELISA result contradicted the qualitative GFP Western blot analysis which suggested that rBCG Δ*panCD* [pSG400] GFP expression levels were greater than rBCG Δ*panCD* [pRC100] levels. A GFP expression level of 1.6% of total protein was determined for rBCG Δ*panCD* [pSG500] (■ in Fig. 3.15) containing the intermediate vector with the high copy number mutation and the full length *rap* and putative terminator. This correlated to a two-fold increase in GFP expression compared to unmodified rBCG Δ*panCD* [pRC100] with statistical significance of $p<0.05$, supporting both fluorescence and Western blot studies. Recombinant BCG Δ*panCD* [pSG600], harbouring the intermediate vector modified to high copy number and incorporating the *hsp60* terminator after *repB*, expressed GFP at a concentration of 8.5% per total protein (■ in Fig. 3.15). This corresponded to a thirteen-fold increase in expression relative to unmodified rBCG Δ*panCD* [pRC100] at a statistical significance of $p<0.005$. This datum supports both preliminary fluorescence observations recording rBCG Δ*panCD* [pSG600] as a high fluorescer and subsequent Western blot findings that rBCG Δ*panCD* [pSG600] had the second greatest level of expression.

A disadvantage to the capture ELISA technique is that it measures *in vitro* levels of extracted protein. As such it depends on the premise that all proteins are isolated with equal efficiencies. Furthermore, protein samples are subject to degradation during isolation. Consequently, *in vivo* analysis of GFP expression levels is ideal and the fluorescent property of GFP permits this (Kremer *et al*, 1995). In this study, attempts to quantitate whole cell GFP expression levels using fluorimetry were abandoned due to technical problems with the fluorimeter which prevented a sensitive standard curve from being produced (Data not shown).

3.4.2. EVALUATION OF VECTOR COPY NUMBER

Determination of plasmid copy number was of interest for two reasons: first, to verify that incorporation of the high copy number mutation alone caused an increase in vector copy number, since GFP expression results revealed no notable improvement in expression, and second, to further investigate the effect of including the full length *rap* gene on copy number (Stolt & Stoker, 1996a).

3.4.2.1. Restriction Digest Analysis

To determine relative differences in plasmid copy number, r*M. smegmatis* and rBCG Δ*panCD* total DNA was digested using restriction enzymes, separated by gel electrophoresis, and subsequently analysed by densitometry. Gel electrophoresis images of separated total DNA samples and corresponding fold differences in the copy number of plasmids are depicted in Figure 3.16. and 3.17. respectively for r*M. smegmatis*, and in Figure 3.18. and Figure 3.19. respectively for rBCG Δ*panCD*.

As expected, the lowest plasmid copy number was observed for r*M. smegmatis* [pRC100] (□ - in Fig. 3.17) harbouring the unmodified vector. Despite GFP expression results revealing no notable improvement in expression caused by incorporation of the high copy number mutation, an approximate two-fold increase in plasmid copy number relative to unmodified r*M. smegmatis* [pRC100] was observed for r*M. smegmatis* [pSG400] (■ - in Fig. 3.17), containing the intermediate vector modified to include the high copy number mutation. A similar two-fold relative increase in copy number was revealed for r*M. smegmatis* [pSG600] (■ - in Fig. 3.17) containing the intermediate vector with high copy number mutation and the *hsp60* terminator downstream of *repB*. In comparison to the unmodified r*M. smegmatis* [pRC100], fully modified r*M. smegmatis* [pSG300] (■ - in Fig. 3.17) and r*M. smegmatis* [pSG500] (■ - in Fig. 3.17) containing the intermediate vector with high copy number mutation in addition to the full length *rap* gene and its putative terminator demonstrated an increase in plasmid copy number of approximately four- and six-fold respectively.

While trends observed for plasmid copy numbers of rBCG Δ*panCD* samples were consistent with those noted for r*M. smegmatis* samples, actual fold differences in copy numbers varied. As expected, the lowest plasmid copy number was observed for rBCG Δ*panCD* [pRC100] (□ - in Fig. 3.19) harbouring the unmodified vector. Although incorporation of the high copy number mutation did not improve GFP expression, an approximate two-fold increase in plasmid copy number relative to unmodified rBCG Δ*panCD* [pRC100] was observed for rBCG Δ*panCD* [pSG400] (■ - in Fig. 3.19), containing the intermediate vector modified to include the high copy number mutation. Similarly, a two-fold relative increase in copy number was noted for rBCG Δ*panCD* [pSG600] (■ - in Fig. 3.19) containing the intermediate vector with high copy number mutation and the *hsp60* terminator downstream of *repB*. As suggested by Stolt and Stoker (Stolt & Stoker, 1996a), vectors containing the full length *rap* gene showed an increase in plasmid copy number. Relative to the unmodified rBCG Δ*panCD* [pRC100], fully

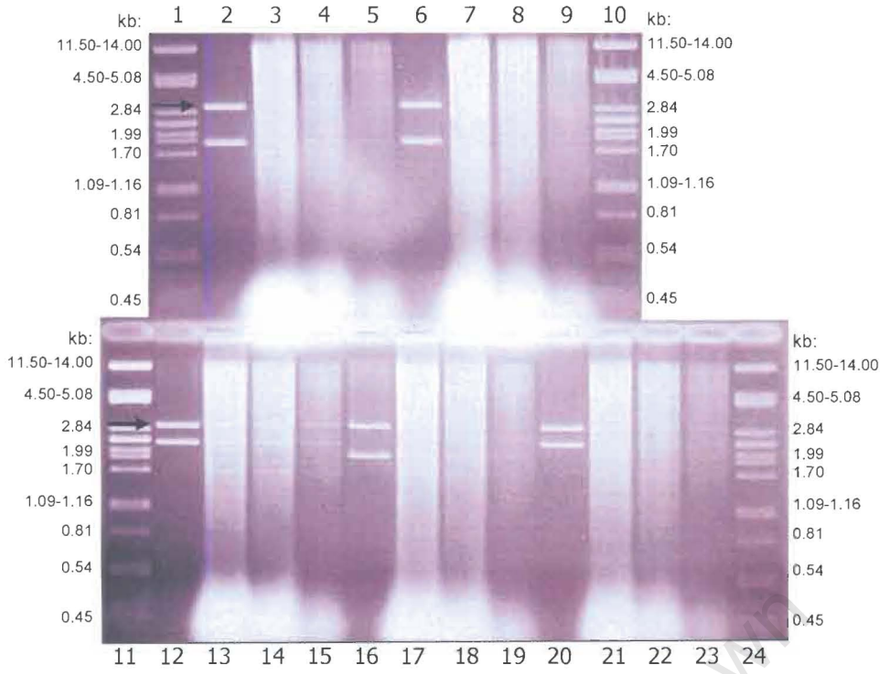

Figure 3.16. Restriction Digest Analysis of r*M. smegmatis* total DNA for copy number determination.

Total DNA was extracted from stationary phase r*M. smegmatis* cultures and digested with *MluI* and *XbaI* restriction enzymes. Doubling dilutions of restriction digest reactions were loaded alongside similarly digested corresponding plasmid DNA and molecular weight marker as follows: **Lane 1**, **10**, **11** + **24**- λDNA/*PstI* molecular weight marker, **Lane 2**- 300ng pRC100 plasmid DNA control, **Lane 3**- 4μg r*M. smegmatis* [pRC100] total DNA, **Lane 4**- 2μg r*M. smegmatis* [pRC100] total DNA, **Lane 5**- 1μg r*M. smegmatis* [pRC100] total DNA, **Lane 6**- 300ng pSG400 plasmid DNA control, **Lane 7**- 4μg r*M. smegmatis* [pSG400] total DNA, **Lane 8**- 2μg r*M. smegmatis* [pSG400] total DNA, **Lane 9**- 1μg r*M. smegmatis* [pSG400] total DNA, **Lane 12**- 300ng pSG500 plasmid DNA control, **Lane 13**- 4μg r*M. smegmatis* [pSG500] total DNA, **Lane 14**- 2μg r*M. smegmatis* [pSG500] total DNA, **Lane 15**- 1μg r*M. smegmatis* [pSG500] total DNA, **Lane 16**- 300ng pSG600 plasmid DNA control, **Lane 17**- 4μg r*M. smegmatis* [pSG600] total DNA, **Lane 18**- 2μg r*M. smegmatis* [pSG600] total DNA, **Lane 19**- 1μg r*M. smegmatis* [pSG600] total DNA, **Lane 20**- 300ng pSG300 plasmid DNA control, **Lane 21**- 4μg r*M. smegmatis* [pSG300] total DNA, **Lane 22**- 2μg r*M. smegmatis* [pSG300] total DNA, **Lane 23**- 1μg r*M. smegmatis* [pSG300] total DNA. Arrows depict the 3.016 kb plasmid fragment used for analysis of all samples.

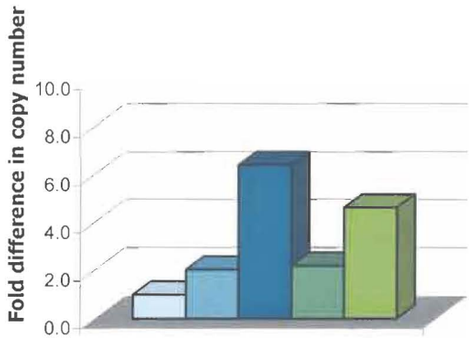

Figure 3.17. Relative fold difference between plasmid copy number of r*M. smegmatis* samples.

Bars are coloured to represent fold-differences between r*M. smegmatis* samples as follows: □ - r*M. smegmatis* [pRC100] (unmodified vector), ■ - r*M. smegmatis* [pSG400] (intermediate vector with high copy number mutation), ■ - r*M. smegmatis* [pSG500] (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), ■ - r*M. smegmatis* [pSG600] (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), ■ - r*M. smegmatis* [pSG300] (fully modified vector).

modified rBCG Δ*panCD* [pSG300] (■ - in Fig. 3.19) and rBCG Δ*panCD* [pSG500] (■ - in Fig. 3.19) containing the intermediate vector modified to high copy number in addition to the full length *rap* gene and its putative terminator exhibited an approximately four-fold increase in plasmid copy number.

The obvious limitation to this analysis is that the accuracy of the ratios of plasmid DNA to chromosomal DNA relies on the premise that each is extracted with equivalent efficiencies, which may not necessarily be the case. In addition, ratios may be skewed if background smears of chromosomal DNA interfere with densitometric analysis, however, no degradation was visible in uncut total DNA samples analysed by gel electrophoresis. Because the restriction digest analysis was intended as a pilot for further investigation into copy number, it was performed singularly. As such, results cannot be interpreted independently, but did provide impetus for more conclusive copy number analysis.

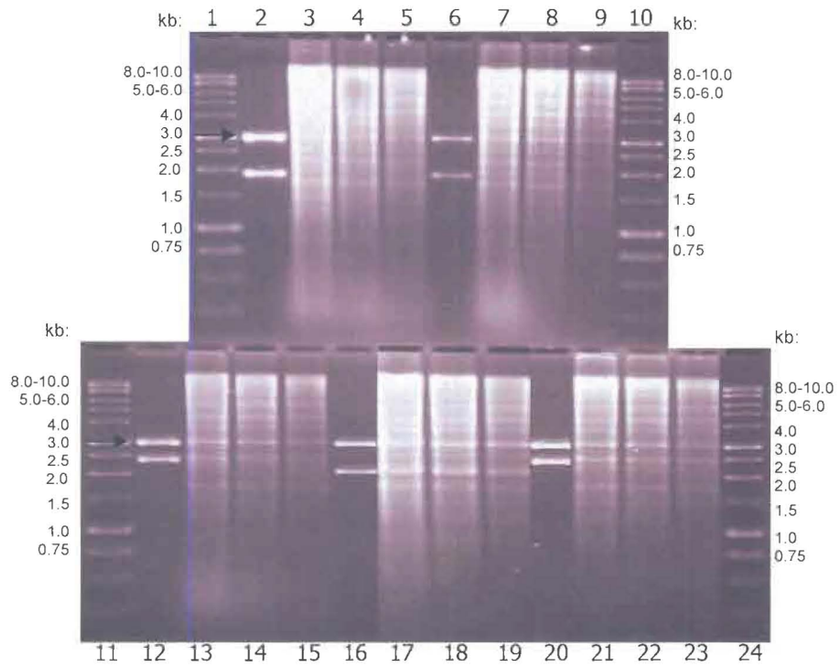

Figure 3.18. Restriction Digest Analysis for rBCG Δ*panCD* plasmid copy number determination.

Total DNA was extracted from stationary phase rBCG Δ*panCD* cultures and digested with *MluI* and *XbaI* restriction enzymes. Doubling dilutions of restriction digest reactions were loaded alongside similarly digested corresponding plasmid DNA and molecular weight marker as follows: **Lane 1**, **10**, **11** + **24**- λDNA/*PstI* molecular weight marker, **Lane 2**- 300ng pRC100 plasmid DNA control, **Lane 3**- 4μg rBCG Δ*panCD* [pRC100] total DNA, **Lane 4**- 2μg rBCG Δ*panCD* [pRC100] total DNA, **Lane 5**- 1μg rBCG Δ*panCD* [pRC100] total DNA, **Lane 6**- 300ng pSG400 plasmid DNA control, **Lane 7**- 4μg rBCG Δ*panCD* [pSG400] total DNA, **Lane 8**- 2μg rBCG Δ*panCD* [pSG400] total DNA, **Lane 9**- 1μg rBCG Δ*panCD* [pSG400] total DNA, **Lane 12**- 300ng pSG500 plasmid DNA control, **Lane 13**- 4μg rBCG Δ*panCD* [pSG500] total DNA, **Lane 14**- 2μg rBCG Δ*panCD* [pSG500] total DNA, **Lane 15**- 1μg rBCG Δ*panCD* [pSG500] total DNA, **Lane 16**- 300ng pSG600 plasmid DNA control, **Lane 17**- 4μg rBCG Δ*panCD* [pSG600] total DNA, **Lane 18**- 2μg rBCG Δ*panCD* [pSG600] total DNA, **Lane 19**- 1μg rBCG Δ*panCD* [pSG600] total DNA, **Lane 20**- 300ng pSG300 plasmid DNA control, **Lane 21**- 4μg rBCG Δ*panCD* [pSG300] total DNA, **Lane 22**- 2μg rBCG Δ*panCD* [pSG300] total DNA, **Lane 23**- 1μg rBCG Δ*panCD* [pSG300] total DNA. Arrows depict the 3.016 kb plasmid fragment used for analysis of all samples.

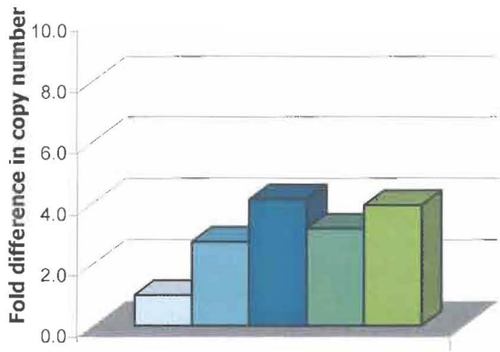

Figure 3.19. Relative fold difference between plasmid copy number of rBCG Δ*panCD* samples.

Bars are coloured to represent fold-differences between rBCG Δ*panCD* samples as follows: □ - rBCG Δ*panCD* [pRC100] (unmodified vector), ■ - rBCG Δ*panCD* [pSG400] (intermediate vector with high copy number mutation), ■ - rBCG Δ*panCD* [pSG500] (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), ■ - rBCG Δ*panCD* [pSG600] (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), ■ - rBCG Δ*panCD* [pSG300] (fully modified vector).

3.4.2.2. Qualitative Real-Time PCR

Since the pilot restriction digest analysis identified differences in plasmid copy number among recombinants, fold differences were confirmed by qualitative real-time PCR. Probe-primer sets with specificity to regions of the 16S chromosomal gene and GFP plasmid gene respectively, were utilised simultaneously in mulitplex PCR with r*M. smegmatis* and rBCG Δ*panCD* total DNA as template. The ΔCt between the 16S and GFP Ct was calculated for each recombinant sample and used to estimate plasmid copy number relative to the unmodified recombinant. Relative differences are tabulated in Table 3.6. and Table 3.7. for r*M. smegmatis* and rBCG Δ*panCD* respectively. Trends in fold differences calculated by real-time PCR were consistent with those determined by restriction enzyme analysis. In comparison, actual fold differences in vector copy number were generally slightly higher when real-time PCR was used. This may be explained by linearity differences between the techniques, since ethidium bromide used to visualise total DNA in the restriction digest analysis is known to have a linear threshold (Koekemoer *et al*, 1998), while all results generated by real-time PCR fell within a determined linear range.

A significant increase in vector copy number of approximately 3.5-fold was seen relative to the unmodified vector pRC100 in r*M. smegmatis* by pSG400, the high copy number vector,

and pSG600, the high copy number vector containing the *hsp60* transcription terminator downstream of *repB*. Although the fully modified r*M. smegmatis* [pSG300] showed an even greater increase in relative copy number of approximately five-fold, this was not significantly different to fold differences calculated for r*M. smegmatis* [pSG400] and [pSG600], due to variability between duplicate results. Recombinant *M. smegmatis* [pSG500], containing the high copy number vector which included the full length *rap* gene, displayed the greatest increase in copy number compared to unmodified r*M. smegmatis* [pRC100] of approximately 10-fold.

Table 3.6. Relative fold difference in copy number among recombinant *M. smegmatis* plasmids

E. coli-mycobacterial shuttle vector	Fold difference in plasmid copy number
pRC100	1.00
pSG400	3.54 ± 1.28
pSG500	10.41 ± 3.77
pSG600	3.78 ± 1.37
pSG300	**5.35 ± 1.94**

Modified expression vector is shown in boldface

In BCG Δ*panCD*, an approximately two-fold and 2.5-fold increase in copy number was observed for the high copy number vector pSG400 and the high copy number vector harbouring the *hsp60* terminator following *repB*, respectively, relative to the unmodified vector pRC100. The greatest increases in vector copy number were displayed by rBCG Δ*panCD* [pSG500], containing the high copy number vector with the full length *rap*, and the fully modified rBCG Δ*panCD* [pSG300] which showed approximately four- and five-fold relative increases respectively.

Table 3.7. Relative fold difference in copy number among recombinant BCG Δ*panCD* plasmids

E. coli-mycobacterial shuttle vector	Fold difference in plasmid copy number
pRC100	1.00
pSG400	2.39 ± 0.58
pSG500	4.48 ± 1.08
pSG600	2.72 ± 0.66
pSG300	**5.43 ± 1.31**

Modified expression vector is shown in boldface

3.4.3. EVALUATION OF VECTOR FITNESS

In a number of studies, episomal *E. coli*-mycobacterial shuttle vectors have proven unstable (Mederle *et al*, 2002;Medeiros *et al*, 2002;Haeseleer, 1994;Haeseleer *et al*, 1993;Dennehy *et al*, 2007). Typically, recombinant instability is attributed to the toxicity or metabolic burden of the recombinant gene expressed. This study aimed to investigate whether the mycobacterial replicon may also play a role in plasmid stability. Two modifications were made to the mycobacterial replicon with the aim of improving stability; inclusion of the full length *rap* gene and its putative terminator, and introduction of the *hsp60* terminator downstream of *repB*. Therefore modified vectors were evaluated in terms of fitness.

Vector fitness can be measured in a number of ways. First, the fitness of vectors may be reflected in recombinant mycobacterial growth rate. Thus the less stress placed on the host cell, the better the growth rate. Conversely, the greater the stress placed on the host cell, the worse the growth rate. Fitness can be ascertained either by colony size or by standard growth curves measured by changes in liquid culture absorbancy over time. Second, the fitness of vectors may be reflected by their stability. That is, the lower the metabolic burden of vectors, the greater their stability in their mycobacterial hosts, and vice versa.

3.4.3.1. Growth Rate Study

3.4.3.1.1. Colony Size

Since the size of colonies reflects their growth rate, colony sizes were analysed as a preliminary, crude indication of whether the final modified expression vector exhibited increased relative fitness and which modifications appeared responsible. It must be stated that colony sizes do not strictly reflect growth rates directly. Particularly in dealing with mycobacteria, because of their tendency to clump, there exists a risk of two or more colonies clumping to form a single unit which would give an inaccurate apparent colony size. Therefore, colony size determination merely encouraged a more accurate investigation into growth rates.

M. smegmatis and BCG Δ*panCD* were transformed with final modified vector pSG300, unmodified vector pRC100 and partially modified intermediate vectors; pSG400, pSG500 and pSG600. Due to the non-qualitative nature of this experiment, relative colony sizes were recorded simply as either 'small' or 'large'. Results are tabulated in Table 3.8. for r*M. smegmatis* colonies and in Table 3.9. for rBCG Δ*panCD* colonies. Vector fitness was not

improved by any of the modifications made. In fact, some modifications resulted in an apparent reduction in fitness.

Table 3.8. Comparison of colony size of recombinant *M. smegmatis* colonies

E. coli-mycobacterial shuttle vector	Colony size
pRC100	Large
pSG400	Large
pSG500	Small
pSG600	Small
pSG300	**Small**

Modified expression vector is shown in boldface

In *M. smegmatis*, relatively large colony sizes were observed for the unmodified r*M. smegmatis* [pRC100] and r*M. smegmatis* [pSG400] containing the intermediate vector converted to high copy number. In contrast, relatively small colony sizes were noted for r*M. smegmatis* colonies harbouring the fully modified vector pSG300, r*M. smegmatis* [pSG500] containing the intermediate vector with high copy number mutation and full length *rap* gene and its putative terminator and r*M. smegmatis* [pSG600] harbouring the intermediate vector with high copy number mutation and the *hsp60* terminator downstream of *repB*. These colony size observations suggested an interesting correlation between colony size and expression levels since r*M. smegmatis* samples with larger colony sizes had GFP expression levels less than or equal to 2.0% of total protein while samples with smaller colony size had expression levels greater than or equal to 4.7% of total protein.

Table 3.9. Comparison of colony size of recombinant BCG Δ*panCD* colonies

E. coli-mycobacterial shuttle vector	Colony size
pRC100	Large
pSG400	Large
pSG500	Large
pSG600	Small
pSG300	**Small**

Modified expression vector is shown in boldface

As in r*M. smegmatis*, relatively large colony sizes were observed for the unmodified rBCG Δ*panCD* [pRC100] and rBCG Δ*panCD* [pSG400] containing the intermediate vector converted to high copy number. In addition, rBCG Δ*panCD* [pSG500], containing the intermediate vector with high copy number mutation and full length *rap* gene and its putative terminator, exhibited a large colony size contrasting with observations for the same recombinant in r*M. smegmatis*. Relatively small colony sizes were noted for rBCG Δ*panCD* colonies harbouring the fully modified vector pSG300 and rBCG Δ*panCD* [pSG600], harbouring the intermediate vector with high copy number mutation and the *hsp60* terminator downstream of *repB*, as observed in *M. smegmatis*. These colony size observations support the possible correlation between colony size and expression levels noted for *M. smegmatis* in that rBCG Δ*panCD* samples with larger colony sizes had GFP expression levels less than or equal to 1.6% of total protein while samples with smaller colony size had expression levels greater than or equal to 8.5% of total protein.

3.4.3.1.2. Growth Curves

Standard growth curves were carried out in triplicate for r*M. smegmatis* and rBCG Δ*panCD* to confirm preliminary indications of different growth rates based on colony size for evaluation of relative differences in vector fitness. Results are shown in Figure 3.20. for r*M. smegmatis* and in Figure 3.21. for rBCG Δ*panCD*. Although the risk of absorbancies being affected by mycobacterial clumping is especially high for growth curve studies (Meyers *et al*, 1998), depicted error bars show little variability between triplicate samples implying that results are reproducible.

Little difference in growth rate was observed for r*M. smegmatis* cultures during mid-log phase of growth, as indicated by the similarity of mid-log phase gradients. However, differences in culture early log phases were noted. Recombinant *M. smegmatis* [pRC100] (- in Fig. 3.20), containing the unmodified vector, and r*M. smegmatis* [pSG400] (- in Fig. 3.20), containing the intermediate vector converted to high copy number, exhibited the fastest relative growth rates during early log phase. A lower early log phase growth rate was observed for r*M. smegmatis* [pSG500] (- in Fig. 3.20) containing the intermediate vector with high copy number mutation and full length *rap* gene and its putative terminator. In comparison, the lowest log phase growth rates were demonstrated by r*M. smegmatis*

[pSG600] (- in Fig. 3.20), harbouring the intermediate vector with high copy number mutation and the *hsp60* terminator downstream of *repB*, and r*M. smegmatis* [pSG300] (- in Fig. 3.20) containing the fully modified vector. During stationary phase the highest final OD was recorded for r*M. smegmatis* [pRC100] and r*M. smegmatis* [pSG400] while relatively lower endpoint ODs were recorded for r*M. smegmatis* [pSG500], r*M. smegmatis* [pSG600] and r*M. smegmatis* [pSG300]. The growth curve findings suggest a possible correlation between rate of growth in early log phase and GFP expression levels. During early log phase specifically, lowest GFP-expressers r*M. smegmatis* [pRC100] and r*M. smegmatis* [pSG400] exhibited faster growth rates while highest expressers r*M. smegmatis* [pSG600] and r*M. smegmatis* [pSG300] displayed slower growth rates. In addition, intermediary expresser r*M. smegmatis* [pSG500] also exhibited an intermediate growth rate.

As in r*M. smegmatis*, rBCG Δ*panCD* cultures exhibited virtually no difference in mid-log phase growth rates. In contrast, differences were observed during early log phase growth and stationary phase. Recombinant BCG Δ*panCD* [pRC100] ( - in Fig. 3.21), containing the unmodified vector, and rBCG Δ*panCD* [pSG400] ( - in Fig. 3.21), containing the intermediate vector converted to high copy number, exhibited the fastest early log phase growth rate. In comparison, the slowest early log phase growth rates were exhibited by rBCG Δ*panCD* [pSG500] ( - in Fig. 3.21), containing the intermediate vector with high copy number mutation and full length *rap* gene and its putative terminator, rBCG Δ*panCD* [pSG600] (- in Fig. 3.21), harbouring the intermediate vector with high copy number mutation and the *hsp60* terminator downstream of *repB*, and rBCG Δ*panCD* [pSG300] ( - in Fig. 3.21), containing the fully modified vector. During stationary phase the highest final ODs were recorded for rBCG Δ*panCD* [pRC100], rBCG Δ*panCD* [pSG400] and rBCG Δ*panCD* [pSG500] while relatively lower endpoint ODs were recorded for rBCG Δ*panCD* [pSG600] and rBCG Δ*panCD* [pSG300]. The growth curve findings support a possible correlation between early log phase growth rate, endpoint ODs, and GFP expression levels. Specifically, lowest GFP-expressers rBCG Δ*panCD* [pRC100], rBCG Δ*panCD* [pSG400] and rBCG Δ*panCD* [pSG500] exhibited the fastest early log phase growth rates and the highest endpoint absorbancies while highest expressers rBCG Δ*panCD* [pSG600] and rBCG Δ*panCD* [pSG300] reflected the slowest early log phase growth rates and the lowest endpoint absorbancies.

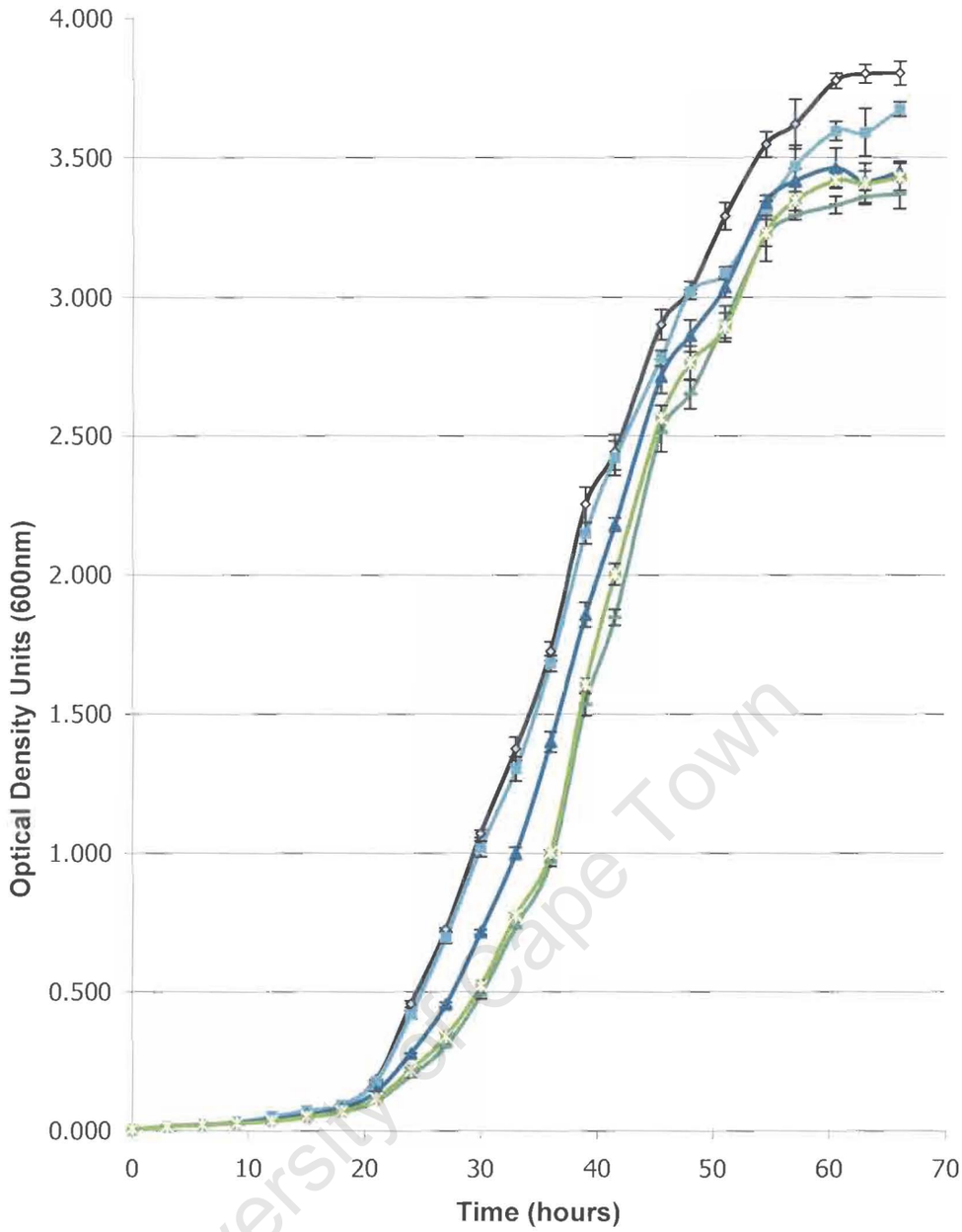

Figure 3.20. Standard growth curves of r*M. smegmatis* cultures.

Coloured lines with markers represent the five r*M. smegmatis* cultures as follows: —◇— - r*M. smegmatis* [pRC100] (unmodified vector), —■— - r*M. smegmatis* [pSG400] (intermediate vector with high copy number mutation), —▲— - r*M. smegmatis* [pSG500] (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), — - r*M. smegmatis* [pSG600] (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), —✕— - r*M. smegmatis* [pSG300] (fully modified vector).

3.4.3.2. Stability Study

While recombinant growth studies provide a good indication of how vectors contribute to metabolic stress, it was necessary to determine the stability of vectors to further evaluate vector fitness. This has particular relevance for recombinant mycobacterial vaccine development where vaccines are required to be stable for up to twelve passages (Brosch *et al*, 2007). To determine relative differences in stability, duplicate cultures were grown in the absence of selection and passaged every seven generations to an endpoint of 42 generations. Due to time considerations, duplicate cultures were passaged six times as this was sufficient time for differences in recombinant stability to manifest. At each passage event, cultures were evaluated in terms of retention of antibiotic resistance and maintenance of GFP fluorescence.

3.4.3.2.1. Antibiotic Resistance Retention

The kanamycin antibiotic resistance of r*M. smegmatis* and rBCG Δ*panCD* cultures at each passage over 42 generations is shown in Figure 3.22. and Figure 3.23. respectively. Variability of up to 30% for r*M. smegmatis* and up to 50% for rBCG Δ*panCD* reflects a lack of reproducibility.

Antibiotic resistance retention was determined by recording antibiotic resistant colony forming units as a percentage of total colony forming units. Mutations are random events and, as such, variation between duplicate cultures is not surprising. Nevertheless, fluctuating mean percentages of antibiotic resistance among passages implies that the protocol of the experiment accounts for some of the variation. The accuracy of this experiment relies heavily on plating equal numbers of colony forming units onto media supplemented with antibiotic and antibiotic-free media. Because mycobacteria have a propensity to clump, equal volumes of culture dilutions may differ in colony count which may explain the poor reproducibility achieved. To improve accuracy, replica plating was considered as an alternative but was rejected because further vector loss may occur during the approximately 20 additional generations required for replica colony formation. The use of Etest® antibiotic strips [AB Biodisk] may represent the best alternative for evaluating antibiotic resistance since it limits incubation time and is independent of the number of colony forming units plated. However, Etest® indications of antibiotic resistance are reliant on the formation of a confluent lawn of colonies, which may be difficult to achieve for mycobacteria which clump, and further, Etest® strips are expensive. Considering the erratic nature of results recorded for this study, only

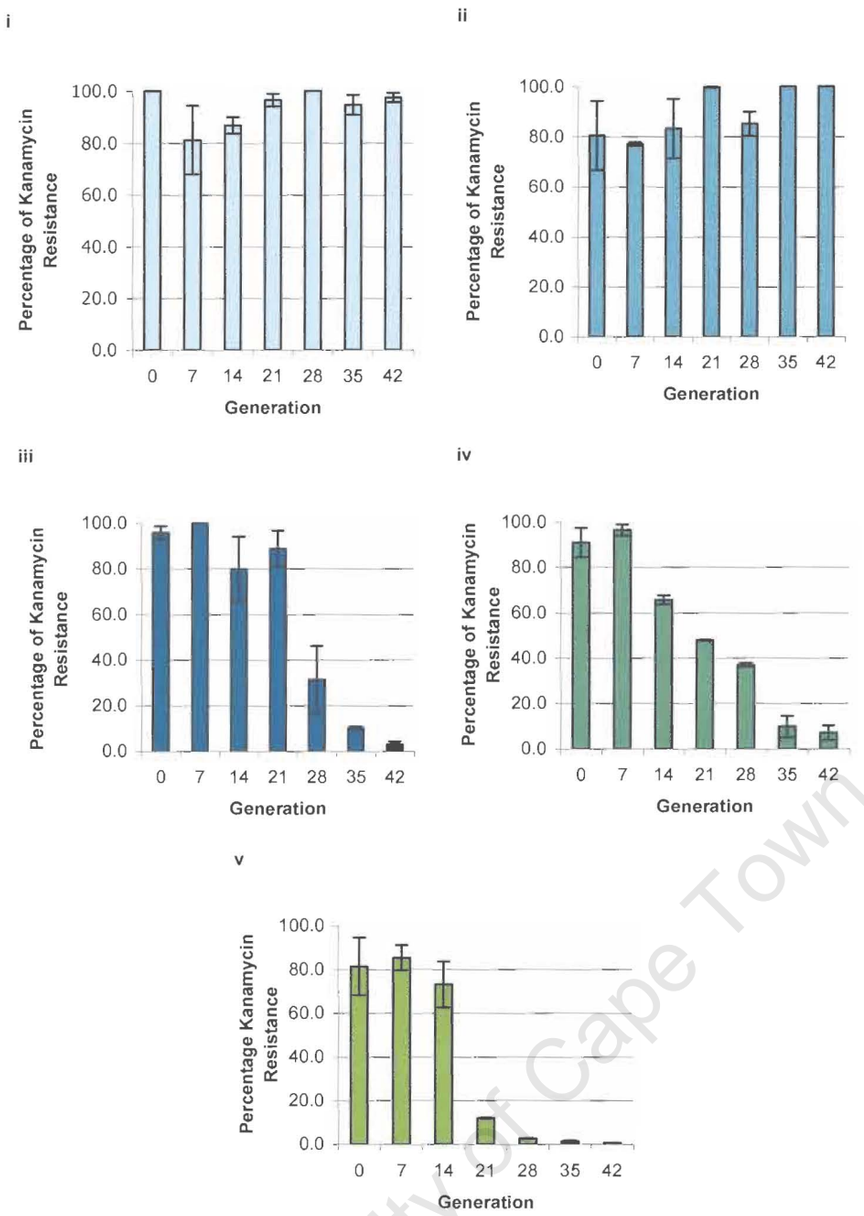

Figure 3.22. Kanamycin Resistance of r*M. smegmatis* cultures at multiple generations.

Duplicate cultures were passaged in the absence of selection over 42 generations. At each passage event, cultures were plated onto media containing and lacking kanamycin. Retention of kanamycin resistance was determined by recording colony counts from media containing kanamycin as a percentage of colony counts from kanamycin-free media. Graph **i**- r*M. smegmatis* [pRC100] (unmodified vector), **ii**- r*M. smegmatis* [pSG400] (intermediate vector with high copy number mutation), **iii**- r*M. smegmatis* [pSG500] (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), **iv**- r*M. smegmatis* [pSG600] (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), **v**- r*M. smegmatis* [pSG300] (fully modified vector). Error bars represent the degree of variability between duplicate samples.

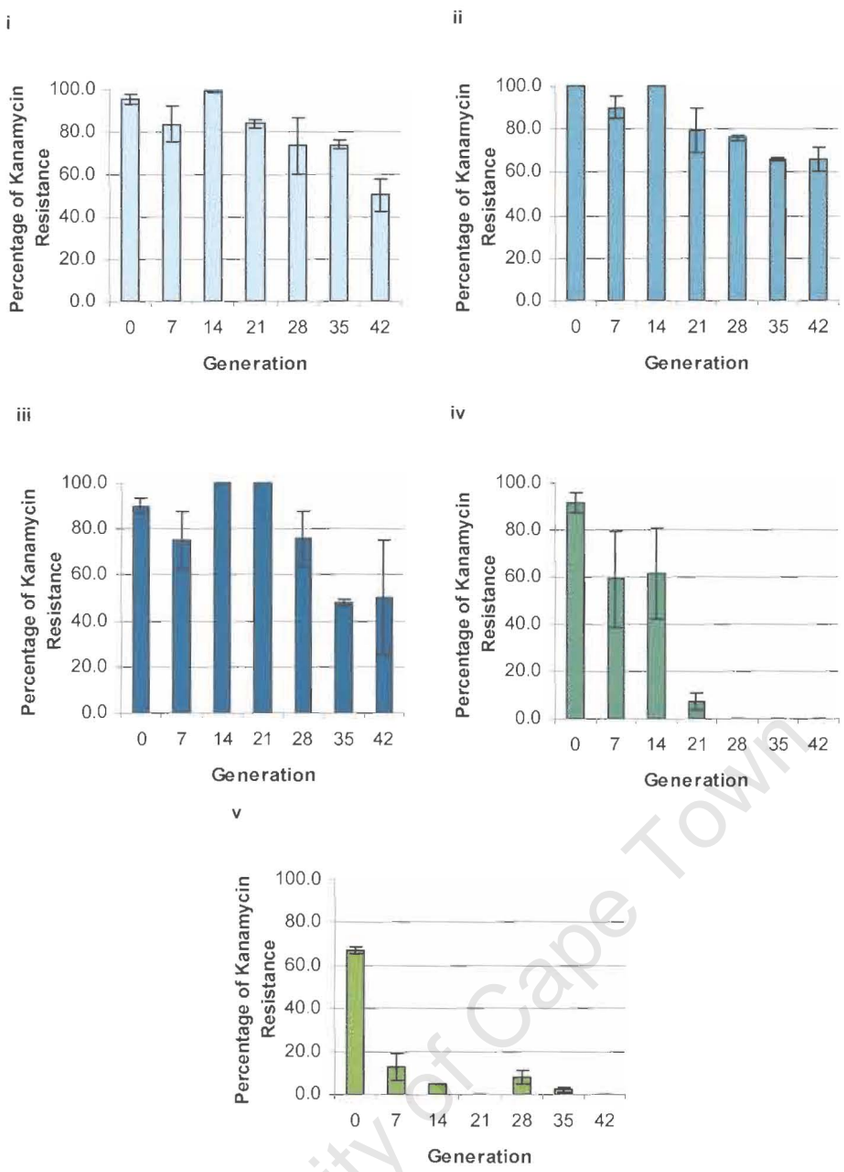

Figure 3.23. Kanamycin Resistance of rBCG Δ*panCD* cultures at multiple generations.

Duplicate cultures were passaged over 42 generations. At each passage event, cultures were plated onto media containing and lacking kanamycin. Retention of kanamycin resistance was determined by recording colony counts from media containing kanamycin as a percentage of colony counts from kanamycin-free media. Graph **i**- rBCG Δ*panCD* [pRC100] (unmodified vector), **ii**- rBCG Δ*panCD* [pSG400] (intermediate vector with high copy number mutation), **iii**- rBCG Δ*panCD* [pSG500] (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), **iv**- rBCG Δ*panCD* [pSG600] (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), **v**- rBCG Δ*panCD* [pSG300] (fully modified vector). Error bars represent the degree of variability between duplicate samples.

average of at least 90% kanamycin resistance was observed for the first two passages after which resistance gradually decreased to reach an endpoint of approximately 65%. Recombinant BCG Δ*panCD* [pSG500] behaved similarly to the unmodified recombinant, except for an apparent reduction in kanamycin resistance to 75% during the first passage. Recombinant rBCG Δ*panCD* [pSG600] (Fig. 3.23.iv), which contained the high copy number mutation in addition to the *hsp60* terminator following *repB*, and rBCG Δ*panCD* [pSG300] (Fig. 3.23.v), containing the fully modified vector, were significantly less stable. For rBCG Δ*panCD* [pSG600], an initial kanamycin resistance of 90% decreased to approximately 60% during the first and second passages. Resistance decreased dramatically to 10% during the third passage after which no kanamycin resistance was observed. Recombinant BCG Δ*panCD* [pSG300] showed initial kanamycin resistance of only 65%. By just the first passage, average kanamycin resistance had decreased significantly to an average of 15%. Recorded percentages of antibiotic resistance during subsequent passages were erratic, varying between 0 and 15%.

3.4.3.2.2 GFP Fluorescence Maintenance

The maintenance of GFP fluorescence by r*M. smegmatis* and rBCG Δ*panCD* cultures at each passage over 42 generations is shown in Figure 3.24. and 3.25. respectively. Without exception, colonies formed on media containing kanamycin displayed GFP fluorescence (Data not shown). This suggested that observed instability was due to whole plasmid loss rather than genetic rearrangements, although this was not verified by plasmid mapping. Fluorescent colony forming units were therefore represented as the percentage of total colony forming colony forming units on kanamycin-free plates. Duplicate results appear highly reproducible, as implied by the small sized error bars which represent variability of less than 10% in all cases. This is likely due to the fact that fluorescent colony forming units were scored as a percentage of total colony forming units on the same antibiotic-free plate at each passage, and thus did not depend on the plating equivalent colony forming unit counts between plates with and without antibiotic for accuracy.

In r*M. smegmatis*, both the unmodified pRC100 vector (Fig. 3.24.i) and the high copy number intermediate pSG400 vector (Fig. 3.24.ii) conferred the most stability in terms of GFP fluorescence maintenance. In both cases, the percentage of GFP fluorescence was consistently greater than 85% at each passage. Recombinant *M. smegmatis* [pSG500] (Fig. 3.24.iii), harbouring the high copy number vector modified to include the entire *rap* gene and

its putative terminator, was comparatively less stable. Close to 100% of colony forming units fluoresced for the first passage only, after which the percentage of GFP fluorescence decreased gradually per passage to reach a final GFP fluorescence of less than 5%. The least stability was observed for r*M. smegmatis* [pSG600] (Fig. 3.24.iv), containing the intermediate vector with the high copy number mutation in addition to the *hsp60* terminator following *repB*, and the fully modified r*M. smegmatis* [pSG300] (Fig. 3.24.v). More than 90% of colony forming units exhibited GFP fluorescence during the first passage only, with the percentage fluorescence decreasing significantly to just over 40% at the second passage. A further drastic reduction in GFP fluorescence to less than 10% was noted during the third passage and fluorescence continued to decrease so that virtually no colonies had maintained GFP fluorescence by the end passage. Interestingly, these observed differences in GFP fluorescence maintenance showed a potential correlation with results for studies of growth rates and expression levels. Specifically, recombinants which were the most stable, also grew the fastest and exhibited the lowest levels of GFP expression.

Recombinant BCG Δ*panCD* cultures displayed different trends in the maintenance of GFP fluorescence to r*M. smegmatis* cultures. Unmodified rBCG Δ*panCD* [pRC100] (Fig. 3.25.i), high copy number rBCG Δ*panCD* [pSG400] (Fig. 3.25.ii) and intermediate rBCG Δ*panCD* [pSG500] (Fig. 3.25.iii), which contained the high copy number vector including the entire *rap* gene and its putative terminator, were the most stable in terms of GFP fluorescence maintenance. All three recombinant cultures maintained a GFP fluorescence percentage of at least 85% for the duration of the stability study. In comparison, intermediate rBCG Δ*panCD* [pSG600] (Fig. 3.25.iv), containing the high copy number mutation and the *hsp60* terminator following *repB*, and fully modified rBCG Δ*panCD* [pSG300] (Fig. 3.25.v) were the least stable. During the very first passage, less than 40% of rBCG Δ*panCD* [pSG600] colonies were fluorescing. By the third passage, merely 5% of colonies displayed GFP fluorescence after which no GFP fluorescence was observed. Similarly, just over 50% of rBCG Δ*panCD* [pSG300] colonies were fluorescing at the first passage. Only 5% of colonies exhibited GFP fluorescence by the third passage after which no GFP fluorescence was detected. These stability results supported the possible correlation between stability, growth rate and expression level noted from data of r*M. smegmatis* fluorescence maintenance. Again, recombinants which were most stable also showed the fastest growth rates and had the lowest levels of GFP expression while less stable recombinants seemed to grow slower and exhibited higher expression levels.

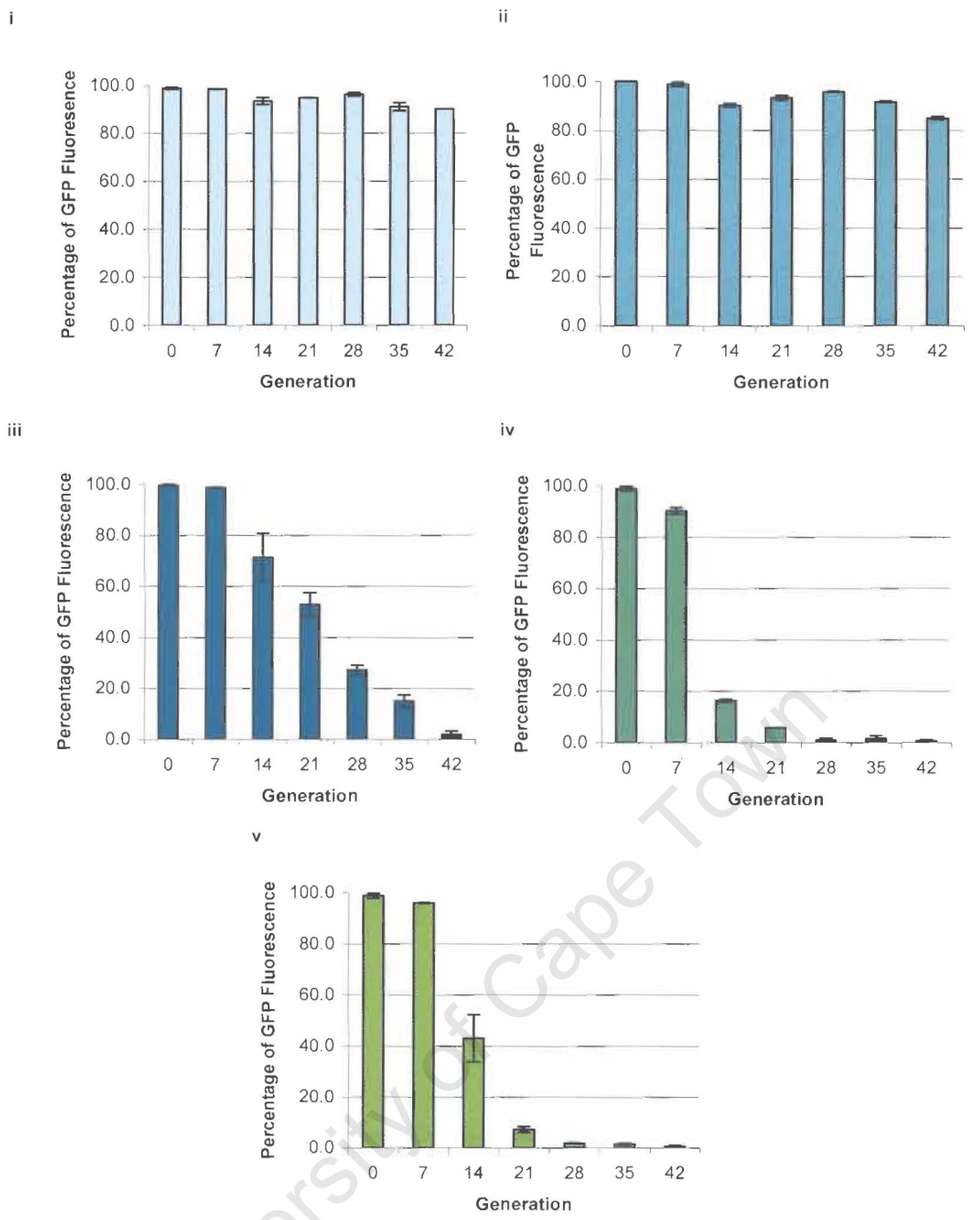

Figure 3.24. GFP Fluorescence of r*M. smegmatis* cultures at multiple generations.

Duplicate cultures were passaged over 42 generations. At each passage event, cultures were plated onto media lacking kanamycin. Maintenance of GFP fluorescence was determined by recording counts of GFP fluorescing colonies as a percentage of total colony counts following visualisation of colonies under UV light. Graph **i**- r*M. smegmatis* [pRC100] (unmodified vector), **ii**- r*M. smegmatis* [pSG400] (intermediate vector with high copy number mutation), **iii**- r*M. smegmatis* [pSG500] (intermediate vector containing full length *rap* and its putative terminator and high copy number mutation), **iv**- r*M. smegmatis* [pSG600] (intermediate vector containing *hsp60* terminator following *repB* and high copy number mutation), **v**- r*M. smegmatis* [pSG300] (fully modified vector). Error bars represent the degree of variability between duplicate samples.

CHAPTER 4: DISCUSSION

4.1. GENERAL AIM

This study aimed to evaluate a series of modifications made to the mycobacterial replicon most commonly used in *E. coli*-mycobacterial shuttle vectors, in terms of their effect on gene expression, plasmid copy number and plasmid stability. These modifications included: (1) incorporation of a high copy number mutation within *repA*, (2) inclusion of the entire *rap* gene and its putative transcription terminator and (3) introduction of an *hsp60* transcription terminator following *repB*.

4.2. MODIFICATION OF THE MYCOBACTERIAL REPLICON OF AN *E. COLI*-MYCOBACTERIAL SHUTTLE VECTOR

Although bioinformatic analysis did not reveal any intrinsic terminator sequences after *repB*, the disadvantage to most bioinformatics software is that searches for transcription terminators are confined to identifying typical transcription structures. Considering the 99.95% sequence similarity between *M. bovis* and *M. tuberculosis* (Garnier *et al*, 2003), the possibility that a non-typically structured terminator may nevertheless exist within this region is highly unlikely however, since statistically less than 1% of terminators within *M. tuberculosis* have a non-typical structure (Unniraman *et al*, 2002). The *hsp60* transcription terminator inserted downstream of *repB* consists of a 14 bp stem and a 5 bp loop. Lacking a poly-U tail, such a terminator is said to have an I-type structure which ninety percent of all identified termination sequences within *M. tuberculosis* have been predicted to form.

The sequence identified by bioinformatics as a putative transcription terminator for the *rap* gene, has a 13 bp stem, a 6 bp loop and a 5 bp poly-U tail. Such an L-type terminator has been calculated as representing approximately one tenth of termination structures found in *M. tuberculosis* (Unniraman *et al*, 2002). Based on the exceptionally high sequence homology between *M. tuberculosis* and BCG, L-type terminators may be equally common in BCG. In early literature, Stolt and Stoker (1996a) identified the stem of this putative *rap* terminator as a palindromic inverted repeat. This is the first report to further characterise the inverted repeats as the stem of a putative poly-U tailed stem-loop *rap* terminator.

4.3. EVALUATION OF THE MODIFIED *E. COLI*-MYCOBACTERIAL SHUTTLE VECTOR

4.3.1. EVALUATION OF GFP EXPRESSION LEVELS

Two of the modifications made to the mycobacterial replicon were hypothesised to cause an increase in levels of recombinant gene expression. Specifically, inclusion of both the high copy number mutation and the full length *rap* gene have been previously demonstrated to facilitate an increase in vector copy number. In turn, it was suspected that an increase in copy number would be reflected by an increase in recombinant antigen expression levels. Therefore, expression levels of the recombinant GFP protein were evaluated in this study. Levels of GFP in r*M. smegmatis* and rBCG Δ*panCD* were analysed using three techniques: colony fluorescence, which served as a pilot evaluation; Western blotting, which provided a qualitative analysis; and capture ELISA which quantitated expression levels.

Surprisingly, the high copy number mutation did not seem to correlate to an increase in expression in rBCG Δ*panCD.* Similarly, neither the fluorescence study nor the capture ELISA study revealed the expected increase in GFP expression in r*M. smegmatis* due to the high copy number mutation. In contrast, Western blot analysis of r*M. smegmatis* lysate showed an apparent increase in GFP expression as a result of introducing the high copy number mutation. This discrepancy in results for r*M. smegmatis* harbouring the high copy number mutation may be explained by slight differences in sampling conditions between the unmodified and modified *M. smegmatis* Western blot samples. Since ELISA is generally considered a more sensitive and more accurate technique than Western blot, GFP expression levels are probably best reflected by ELISA results for this study. Thus, besides the Western blot analysis for r*M. Smegmatis*, GFP expression results obtained using the various techniques confirmed that the high copy number mutation did not confer the anticipated increase in recombinant GFP expression levels. Accordingly, this finding raised doubts as to whether an increase in plasmid copy number leads to a corresponding increase in GFP expression.

As suspected, preferential inclusion of the full length *rap* gene over its regularly truncated form caused a significant increase in GFP expression levels. This result was observed for r*M. smegmatis* and rBCG Δ*panCD* using all techniques, with the exception of the fluorescence study for rBCG Δ*panCD* colonies. Despite an obvious increase in GFP expression levels evaluated using either Western blot analysis or capture ELISA, the fact that no visual increase

in fluorescence intensity of rBCG Δ*panCD* was noted following inclusion of full length *rap*, may be explained by the greater sensitivity of detection permitted by the Western blot and ELISA techniques compared to the crude fluorescence study. Interestingly, the *rap* gene modification yielded higher levels of GFP expression in r*M. smegmatis* than in rBCG Δ*panCD* suggesting a possible difference between *M. smegmatis* and BCG Δ*panCD* at transcriptional, translational, or indeed post-translational level. A number of studies have also shown that gene expression and protein levels vary among mycobacteria. For instance, six percent of genes were shown to be differentially expressed even in genetically similar *M. bovis* and *M. tuberculosis* (Rehren *et al*, 2007) and Jungblut *et al* (1999) demonstrated that 10% of proteins common to both *M. tuberculosis* and BCG were present at differing levels. Concerning GFP specifically, Roberts *et al* (2004) reported up to two-fold greater levels of GFP expression in r*M. smegmatis* than in r*M. tuberculosis* transformed with identical expression vectors. Possible reasons for differential gene expression between *M. smegmatis* and BCG may include the fact that the *M. smegmatis* genome (6.9 MB) is considerably larger than the BCG genome (4.4 MB) (TIGR, 2007), and that *M. smegmatis* is an environmental organism whereas BCG is derived from *M. bovis* which is an intracellular pathogen.

Incorporation of the *hsp60* terminator upstream of the *hsp60* promoter appeared to increase GFP expression in both mycobacteria investigated. This result was unexpected as the modification was carried out for the purpose of improving stability by preventing any possible read-through of *repB* into the promoter region upstream of the recombinant gene. Accordingly, since the terminator was not introduced downstream of the *gfp* gene, it was not expected to affect GFP expression levels. The increase in GFP expression may be explained by one of two hypotheses. First, enhanced termination of *repB* may also increase the transcription activity of *repB*. A study by Chu *et al* (1997) showed that including a stem-loop transcription terminator downstream of human Alu elements significantly increased their expression. As RepB is a rate-limiting factor in plasmid replication, an increase in its supply may increase plasmid replication which would increase GFP expression proportionally. Recently, Mo *et al* (2007) found that over-expression of *repB* led to an increase in plasmid copy number. The second hypothesis is that prevention of read-through interference with the promoter region of the recombinant *gfp* gene enhanced efficiency of GFP expression. Incidentally, in the present study a striking 11-fold increase in levels of GFP expression was determined for rBCG Δ*panCD* compared to only a three-fold increase for r*M. smegmatis*,

further illustrating that there exist differences at the transcriptional or translational level between *M. smegmatis* and BCG Δ*panCD*.

4.3.2. EVALUATION OF VECTOR COPY NUMBER

It was important to ascertain whether plasmid copy numbers were influenced by the modifications made, since this could account for the observed changes in GFP expression levels. First, the high copy number mutation did not seem to yield the expected increase in GFP expression which could be explained if the high copy number genotype did not facilitate the intended increase in copy number in this study. Second, although the full length *rap* and its putative terminator were included primarily for improved vector stability, literature suggested that in addition it may increase plasmid copy number (Stolt & Stoker, 1996a). An increase in copy number would explain why vectors harbouring the full length *rap* gene had higher GFP expression levels. Lastly, the addition of the *hsp60* terminator downstream of *repB* may have enhanced transcription, and therefore expression, of *repB*. As RepB is essential for plasmid replication, an increase in its abundance has been reported to increase copy number (Mo *et al*, 2007) which could explain the observed increase in GFP expression following inclusion of the terminator.

Relative differences in copy number were initially determined by means of a qualitative restriction digest analysis of total DNA and were subsequently confirmed by qualitative real-time PCR. A large degree of variation was observed between duplicate fold differences in copy number. This variation may represent actual variation in plasmid copy number, since a two-fold difference in plasmid copy number has been previously reported for duplicate results of the same sample (Bourn *et al*, 2007).

Interestingly, the high copy number mutation yielded a significant increase in copy number, which raises the question as to why double the plasmid copies per bacterium was not reflected by a proportional two-fold increase in GFP expression levels. Perhaps one explanation for this is that the total DNA extraction protocol specified that cultures be harvested during late stationary phase, whereas all protein was extracted during mid log phase. Thus, GFP expression levels may have represented a corresponding increase during stationary phase. This would be surprising, however, as total DNA and protein samples were also harvested during stationary and log phase respectively by Bourn *et al* (2007) who reported a seven-fold increase in both expression and copy number.

Of all the modifications evaluated, inclusion of the full length *rap* gene and its putative transcription terminator within vectors facilitated the greatest increase in plasmid copy number, implying that the *rap* gene does appear to play a role in copy number determination. This supports evidence by Stolt and Stoker (1996a) which showed that *rap* influenced copy number and offers an explanation as to why vectors harbouring full length *rap* exhibited greater GFP expression levels than those containing truncated *rap*. Interestingly, the fold difference in copy number following inclusion of the entire *rap* gene was significantly greater for r*M. smegmatis* than rBCG Δ*panCD* which suggests that the hypothesis that corresponding differences in levels of GFP expression recorded for these mycobacteria are due to differing copy numbers is correct. The five- to ten-fold increase in copy number for r*M. smegmatis* and rBCG Δ*panCD* respectively, was only reflected as a two-fold increase in GFP expression level. This non-direct proportionality between copy number and expression level phenotypes may explain why no significant increase in GFP expression level was observed by including the high copy number mutation despite the observed doubling in copy number.

Insertion of the *hsp60* transcription terminator downstream of the *repB* gene in expression vectors had no effect on plasmid copy number. This suggests that the observed increase in GFP expression is not the result of an increase in vector replication due to an increase in supply of RepB replication protein. Rather, incorporation of the *hsp60* terminator facilitates an increase in the expression of recombinant *gfp* by an alternative mechanism. For example, the terminator may enhance the efficiency of GFP expression by preventing interference with the promoter region caused by read-through of the upstream *repB* gene. It must be emphasised, however, that this proposed explanation is purely speculative and cannot be substantiated.

4.3.3. EVALUATION OF VECTOR FITNESS

As discussed in the Introduction (Section 1.3.3.1), much of the onus for varied stability of recombinant mycobacteria has traditionally been placed on the recombinant gene itself. This study aimed to investigate whether the mycobacterial replicon may also affect stability. Consequently, two modifications to the mycobacterial replicon were made with the goal of improving stability; namely inclusion of the full length *rap* gene and putative terminator, and introduction of the *hsp60* terminator downstream of *repB*. Vector fitness was evaluated through growth rate and stability studies.

The high copy number mutation did not reduce vector fitness in either r*M. smegmatis* or rBCG Δ*panCD*. This result confirms findings by Bourn *et al* (2007) which showed that mycobacteria transformed with the high copy number expression vector were consistently stable. Similarly, preferential inclusion of the entire *rap* gene did not affect fitness in rBCG Δ*panCD*. Conversely, when full length *rap* was included within expression vectors in r*M. smegmatis*, a notable reduction in vector fitness was observed. In both r*M. smegmatis* or rBCG Δ*panCD* fitness was severely compromised following the introduction of the *hsp60* terminator downstream of *repB*. Interestingly, a possible inverse correlation between vector fitness and GFP expression levels became evident. Compromised vector fitness was only observed in recombinant mycobacteria characterised as high GFP-expressers while fitness was maintained in recombinant mycobacteria which exhibited lower levels of GFP expression. Such a correlation could exist due to high levels of recombinant antigen expression which confer a metabolic load on the mycobacterial cell due to antigen toxicity or rare codon usage, thus decreasing vector fitness. As stated in Section 3.2.1.2, the *gfp* gene utilised in this study was originally modified for use in *Arabidopsis thaliana* and thus contains codons which are rarely found in mycobacteria.

In addition, much has been reported on the instability of recombinant antigens expressed by the *hsp60* promoter in particular (Haeseleer, 1994;Kumar *et al*, 1998;Al Zarouni & Dale, 2002). Further work is necessary to determine whether the *hsp60* promoter incurs greater instability when regulating higher levels of recombinant antigen expression. The observed decrease in stability with increasing recombinant antigen expression may be overcome through use of an inducible promoter which would allow propagation of recombinant cultures to high cell density *in vitro* while retaining stability. Such a promoter could then be induced to express high levels of recombinant antigen *in vitro*, immediately prior to harvesting, or *in vivo*, on exposure of recombinant mycobacterial vaccines to the immune system. Two studies provide proof of concept for replacing the *hsp60* promoter with inducible promoters to overcome instability. The studies showed that HPV-11 L1 genes (Pupa, 2003), and the rotavirus VP6 gene (Dennehy *et al*, 2007), were highly unstable when expressed by the *hsp60* promoter within episomal shuttle vectors in BCG. In both cases, stability was improved dramatically when the *hsp60* promoter was substituted by either the 18kDa or the *mtrA* inducible promoter.

4.4. GENERAL CONCLUSIONS

In summary, this study evaluated three modifications to the mycobacterial replicon in episomal *E. coli*-mycobacterial shuttle vectors in terms of their effect on plasmid copy number, expression levels and stability: (1) incorporation of a high copy number mutation within *repA*, (2) inclusion of the entire *rap* gene and its putative transcription terminator and (3) introduction of an *hsp60* transcription terminator following *repB*.

Incorporation of a previously characterised high copy number mutation (Bourn *et al*, 2007) increased plasmid copy number significantly in both *M. smegmatis* and BCG Δ*panCD* which has application for improving yields of plasmid isolated from recombinant mycobacteria.

The full length *rap* gene and its putative transcription terminator, commonly truncated in expression vectors, were included in the vector in an attempt to improve stability and increase GFP expression. This modification conferred a significant increase in copy number in *M. smegmatis* and BCG Δ*panCD* and would also be relevant to mycobacterial research requiring improved yields of recombinant mycobacterial plasmids. In BCG Δ*panCD*, the modification was stable and is therefore useful for developing rBCG Δ*panCD*-based vaccines, as recombinant vaccine efficacy has been shown to be enhanced by an increase in expression levels of recombinant antigen. Although a significant increase in GFP expression was achieved in *M. smegmatis*, it was accompanied by a considerable reduction in stability, in the absence of antibiotic selection.

An *hsp60* transcription terminator was included downstream of the *repB* gene with the aim of improving vector stability. Yet, in *M. smegmatis* and BCG Δ*panCD*, this modification resulted in a severe decrease in stability. This reduction in stability, however, was accompanied by a significant increase in GFP expression. This finding raised the possibility of a correlation between higher recombinant antigen expression levels and lower levels of stability. *E. coli*-mycobacterial shuttle vectors which demonstrated high levels of recombinant antigen expression, but were unstable in *M. smegmatis* and BCG Δ*panCD*, may still have application in recombinant mycobacterial vaccines utilising an inducible promoter to regulate expression levels temporally.

While expression studies of vectors utilising the recombinant GFP protein revealed encouraging trends due to inclusion of the full length *rap* gene as well as incorporation of the

hsp60 transcription terminator, future work is required to establish whether similar results can be achieved when different recombinant genes are expressed, and whether vector stability could be improved through codon optimisation of the recombinant gene for mycobacteria. Nevertheless, this study serves to confirm that modifications to the typically utilised mycobacterial replicon play a noteworthy role in levels of episomal *E. coli*-mycobacterial shuttle vector stability and recombinant gene expression levels.

References

Al Zarouni,M. & Dale,J.W. (2002) Expression of foreign genes in Mycobacterium bovis BCG strains using different promoters reveals instability of the hsp60 promoter for expression of foreign genes in Mycobacterium bovis BCG strains. *Tuberculosis.(Edinb.)*, **82**, 283-291.

Aldovini,A. & Young,R.A. (1991) Humoral and cell-mediated immune responses to live recombinant BCG-HIV vaccines. *Nature*, **351**, 479-482.

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W., & Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389-3402.

Anes,E., Portugal,I., & Moniz-Pereira,J. (1992) Insertion into the Mycobacterium smegmatis genome of the aph gene through lysogenization with the temperate mycobacteriophage Ms6. *FEMS Microbiol.Lett.*, **74**, 21-25.

Basu,A., Chatterjee,S., & Das Gupta,S.K. (2004) Translational coupling to an upstream gene promotes folding of the mycobacterial plasmid pAL5000 replication protein RepB and thereby its origin binding activity. *J.Bacteriol.*, **186**, 335-342.

Basu,A., Chawla-Sarkar,M., Chakrabarti,S., & Das Gupta,S.K. (2002) Origin binding activity of the Mycobacterial plasmid pAL5000 replication protein RepB is stimulated through interactions with host factors and coupled expression of repA. *J.Bacteriol.*, **184**, 2204-2214.

Bikandi,J., San,M.R., Rementeria,A., & Garaizar,J. (2004) In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics.*, **20**, 798-799.

Blokpoel,M.C., Smeulders,M.J., Hubbard,J.A., Keer,J., & Williams,H.D. (2005) Global analysis of proteins synthesized by Mycobacterium smegmatis provides direct evidence for physiological heterogeneity in stationary-phase cultures. *J.Bacteriol.*, **187**, 6691-6700.

Bloom,B. (1994) *Tuberculosis: Pathogenesis, Protection, and Control*, ASM.

Bourn,W.R.J.Y., Stutz,H., Warren,R.M., Williamson,A.L., & van Helden PD (2007) Creation and Characterisation of a high-copy-number version of the pAL5000 mycobacterial replicon. *Tuberculosis.(Edinb.)*, (in press).

Brosch,R., Gordon,S.V., Garnier,T., Eiglmeier,K., Frigui,W., Valenti,P., Dos,S.S., Duthoy,S., Lacroix,C., Garcia-Pelayo,C., Inwald,J.K., Golby,P., Garcia,J.N., Hewinson,R.G., Behr,M.A., Quail,M.A., Churcher,C., Barrell,B.G., Parkhill,J., & Cole,S.T. (2007) Genome plasticity of BCG and impact on vaccine efficacy. *Proc.Natl.Acad.Sci.U.S.A*, **104**, 5596-5601.

Cayabyab,M.J., Hovav,A.H., Hsu,T., Krivulka,G.R., Lifton,M.A., Gorgone,D.A., Fennelly,G.J., Haynes,B.F., Jacobs,W.R., Jr., & Letvin,N.L. (2006) Generation of CD8+ T-cell responses by a recombinant nonpathogenic Mycobacterium smegmatis vaccine vector expressing human immunodeficiency virus type 1 Env. *J.Virol.*, **80**, 1645-1652.

Chu,W.M., Ballard,R.E., & Schmid,C.W. (1997) Palindromic sequences preceding the terminator increase polymerase III template activity. *Nucleic Acids Res.*, **25**, 2077-2082.

Connell,N.D. (2001) Expression systems for use in actinomycetes and related organisms. *Curr.Opin.Biotechnol.*, **12**, 446-449.

Connell,N.D., Medina-Acosta,E., McMaster,W.R., Bloom,B.R., & Russell,D.G. (1993) Effective immunization against cutaneous leishmaniasis with recombinant bacille Calmette-Guerin expressing the Leishmania surface proteinase gp63. *Proc.Natl.Acad.Sci.U.S.A*, **90**, 11473-11477.

da Cruz,F.W., McBride,A.J., Conceicao,F.R., Dale,J.W., McFadden,J., & Dellagostin,O.A. (2001) Expression of the B-cell and T-cell epitopes of the rabies virus nucleoprotein in Mycobacterium bovis BCG and induction of an humoral response in mice. *Vaccine*, **20**, 731-736.

Dellagostin,O.A., Esposito,G., Eales,L.J., Dale,J.W., & McFadden,J. (1995) Activity of mycobacterial promoters during intracellular and extracellular growth. *Microbiology*, **141 (Pt 8)**, 1785-1792.

Dellagostin,O.A., Wall,S., Norman,E., O'Shaughnessy,T., Dale,J.W., & McFadden,J. (1993) Construction and use of integrative vectors to express foreign genes in mycobacteria. *Mol Microbiol*, **10**, 983-993.

Dennehy,M., Bourn,W., Steele,D., & Williamson,A.L. (2007) Evaluation of recombinant BCG expressing rotavirus VP6 as an anti-rotavirus vaccine. *Vaccine*, **25**, 3646-3657.

Dennehy,M. & Williamson,A.L. (2005) Factors influencing the immune response to foreign antigen expressed in recombinant BCG vaccines. *Vaccine*, **23**, 1209-1224.

Dhandayuthapani,S., Via,L.E., Thomas,C.A., Horowitz,P.M., Deretic,D., & Deretic,V. (1995) Green fluorescent protein as a marker for gene expression and cell biology of mycobacterial interactions with macrophages. *Mol.Microbiol.*, **17**, 901-912.

FAO. Food and Agricultural Organisation of the United Nations. FAO . 2007.
Ref Type: Electronic Citation

Fennelly,G.J., Flynn,J.L., ter,M., V, Liebert,U.G., & Bloom,B.R. (1995) Recombinant bacille Calmette-Guerin priming against measles. *J Infect Dis*, **172**, 698-705.

Garnier,T., Eiglmeier,K., Camus,J.C., Medina,N., Mansoor,H., Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsempe,C., Simon,S., Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L., Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V., & Hewinson,R.G. (2003) The complete genome sequence of Mycobacterium bovis. *Proc.Natl.Acad.Sci.U.S.A*, **100**, 7877-7882.

Guilhot,C., Otal,I., Van,R., I, Martin,C., & Gicquel,B. (1994) Efficient transposition in mycobacteria: construction of Mycobacterium smegmatis insertional mutant libraries. *J.Bacteriol.*, **176**, 535-539.

Haeseleer,F. (1994) Structural instability of recombinant plasmids in mycobacteria. *Res.Microbiol.*, **145**, 683-687.

Haeseleer,F., Pollet,J.F., Haumont,M., Bollen,A., & Jacobs,P. (1993) Stable integration and expression of the Plasmodium falciparum circumsporozoite protein coding sequence in mycobacteria. *Mol.Biochem.Parasitol.*, **57**, 117-126.

Himmelrich,H., Lo-Man,R., Winter,N., Guermonprez,P., Sedlik,C., Rojas,M., Monnaie,D., Gheorghiu,M., Lagranderie,M., Hofnung,M., Gicquel,B., Clement,J.M., & Leclerc,C.

(2000) Immune responses induced by recombinant BCG strains according to level of production of a foreign antigen: malE. *Vaccine*, **18**, 2636-2647.

Hofmann,A., Iwai,H., Hess,S., Pluckthun,A., & Wlodawer,A. (2002) Structure of cyclized green fluorescent protein. *Acta Crystallogr.D.Biol.Crystallogr.*, **58**, 1400-1406.

Inoue,H., Nojima,H., & Okayama,H. (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene*, **96**, 23-28.

Ish-Horowicz,D. & Burke,J.F. (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Res.*, **9**, 2989-2998.

Jacobs,W.R., Jr., Barletta,R.G., Udani,R., Chan,J., Kalkut,G., Sosne,G., Kieser,T., Sarkis,G.J., Hatfull,G.F., & Bloom,B.R. (1993) Rapid assessment of drug susceptibilities of Mycobacterium tuberculosis by means of luciferase reporter phages. *Science*, **260**, 819-822.

Jacobs,W.R., Jr., Tuckman,M., & Bloom,B.R. (1987) Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature*, **327**, 532-535.

Jungblut,P.R., Schaible,U.E., Mollenkopf,H.J., Zimny-Arndt,U., Raupach,B., Mattow,J., Halada,P., Lamer,S., Hagens,K., & Kaufmann,S.H. (1999) Comparative proteome analysis of Mycobacterium tuberculosis and Mycobacterium bovis BCG strains: towards functional genomics of microbial pathogens. *Mol.Microbiol.*, **33**, 1103-1117.

Kanekiyo,M., Matsuo,K., Hamatake,M., Hamano,T., Ohsu,T., Matsumoto,S., Yamada,T., Yamazaki,S., Hasegawa,A., Yamamoto,N., & Honda,M. (2005) Mycobacterial codon optimization enhances antigen expression and virus-specific immune responses in recombinant Mycobacterium bovis bacille Calmette-Guerin expressing human immunodeficiency virus type 1 Gag. *J.Virol.*, **79**, 8716-8723.

Kaufmann,S.H. (2006) Envisioning future strategies for vaccination against tuberculosis. *Nat.Rev.Immunol.*, **6**, 699-704.

Koekemoer,T.C., Downing,T.G., & Oelofsen,W. (1998) An alternative PCR assay for quantifying mitochondrial DNA in crude preparations. *Nucleic Acids Res.*, **26**, 2829-2830.

Kremer,L., Baulard,A., Estaquier,J., Poulain-Godefroy,O., & Locht,C. (1995) Green fluorescent protein as a new expression marker in mycobacteria. *Mol.Microbiol.*, **17**, 913-922.

Kumar,D., Srivastava,B.S., & Srivastava,R. (1998) Genetic rearrangements leading to disruption of heterologous gene expression in mycobacteria: an observation with Escherichia coli beta-galactosidase in Mycobacterium smegmatis and its implication in vaccine development. *Vaccine*, **16**, 1212-1215.

Labidi,A.H., Estes,R.C., David,H.L., & Bollon,A.P. (2001) Mycobacterium recombinant vaccines. *Tunis Med.*, **79**, 65-81.

Langermann,S., Palaszynski,S.R., Burlein,J.E., Koenig,S., Hanson,M.S., Briles,D.E., & Stover,C.K. (1994) Protective humoral response against pneumococcal infection in mice elicited by recombinant bacille Calmette-Guerin vaccines expressing pneumococcal surface protein A. *J.Exp.Med.*, **180**, 2277-2286.

Lee,M.H., Pascopella,L., Jacobs,W.R., Jr., & Hatfull,G.F. (1991) Site-specific integration of mycobacteriophage L5: integration-proficient vectors for Mycobacterium smegmatis,

Mycobacterium tuberculosis, and bacille Calmette-Guerin. *Proc.Natl.Acad.Sci.U.S.A*, **88**, 3111-3115.

Lim,E.M., Rauzier,J., Timm,J., Torrea,G., Murray,A., Gicquel,B., & Portnoi,D. (1995) Identification of mycobacterium tuberculosis DNA sequences encoding exported proteins by using phoA gene fusions. *J.Bacteriol.*, **177**, 59-65.

Marklund,B.I., Speert,D.P., & Stokes,R.W. (1995) Gene replacement through homologous recombination in Mycobacterium intracellulare. *J.Bacteriol.*, **177**, 6100-6105.

Matsumoto,S., Yukitake,H., Kanbara,H., & Yamada,T. (1998) Recombinant Mycobacterium bovis bacillus Calmette-Guerin secreting merozoite surface protein 1 (MSP1) induces protection against rodent malaria parasite infection depending on MSP1-stimulated interferon gamma and parasite-specific antibodies. *J.Exp.Med.*, **188**, 845-854.

Matsuo,K., Yamaguchi,R., Yamazaki,A., Tasaka,H., Terasaka,K., Totsuka,M., Kobayashi,K., Yukitake,H., & Yamada,T. (1990) Establishment of a foreign antigen secretion system in mycobacteria. *Infect Immun*, **58**, 4049-4054.

Medeiros,M.A., Dellagostin,O.A., Armoa,G.R., Degrave,W.M., Mendonca-Lima,L., Lopes,M.Q., Costa,J.F., McFadden,J., & McIntosh,D. (2002) Comparative evaluation of Mycobacterium vaccae as a surrogate cloning host for use in the study of mycobacterial genetics. *Microbiology*, **148**, 1999-2009.

Mederle,I., Bourguin,I., Ensergueix,D., Badell,E., Moniz-Peireira,J., Gicquel,B., & Winter,N. (2002) Plasmidic versus insertional cloning of heterologous genes in Mycobacterium bovis BCG: impact on in vivo antigen persistence and immune responses. *Infect.Immun.*, **70**, 303-314.

Mehta,A., Trotta,C.R., & Peltz,S.W. (2006) Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. *Genes Dev.*, **20**, 939-953.

Merrill,C.R., Goldman,D., Seldman,S.A., & Ebert,M.H. (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science*, **211**, 1437-1438.

Meyers,P.R., Bourn,W.R., Steyn,L.M., van Helden,P.D., Beyers,A.D., & Brown,G.D. (1998) Novel method for rapid measurement of growth of mycobacteria in detergent-free media. *J.Clin.Microbiol.*, **36**, 2752-2754.

Mo,Y., Quanquin,N.M., Vecino,W.H., Ranganathan,U.D., Tesfa,L., Bourn,W.R., Derbyshire,K.M., Letvin,N.L., Jacobs,W.R., Jr., & Fennelly,G. (2007) Genetic Alteration of *Mycobacterium smegmatis* to Improve Mycobacterial Bactofection and DNA Immunisation. *Infect Immun*.

Mulder,M.A., Zappe,H., & Steyn,L.M. (1997) Mycobacterial promoters. *Tuber.Lung Dis.*, **78**, 211-223.

Nascimento,I.P., Dias,W.O., Mazzantini,R.P., Miyaji,E.N., Gamberini,M., Quintilio,W., Gebara,V.C., Cardoso,D.F., Ho,P.L., Raw,I., Winter,N., Gicquel,B., Rappuoli,R., & Leite,L.C. (2000) Recombinant Mycobacterium bovis BCG expressing pertussis toxin subunit S1 induces protection against an intracerebral challenge with live Bordetella pertussis in mice. *Infect.Immun.*, **68**, 4877-4883.

Neyrolles,O., Gould,K., Gares,M.P., Brett,S., Janssen,R., O'Gaora,P., Herrmann,J.L., Prevost,M.C., Perret,E., Thole,J.E., & Young,D. (2001) Lipoprotein access to MHC

class I presentation during infection of murine macrophages with live mycobacteria. *J.Immunol.*, **166**, 447-457.

Norman,E., Dellagostin,O.A., McFadden,J., & Dale,J.W. (1995) Gene replacement by homologous recombination in Mycobacterium bovis BCG. *Mol.Microbiol.*, **16**, 755-760.

Ohara,N. & Yamada,T. (2001) Recombinant BCG vaccines. *Vaccine*, **19**, 4089-4098.

Paget,E. & Davies,J. (1996) Apramycin resistance as a selective marker for gene transfer in mycobacteria. *J.Bacteriol.*, **178**, 6357-6360.

Parish,T. & Stoker,N.G. (2001) *Mycobacterium tuberculosis Protocols,* Humana Press.

Pelicic,V., Jackson,M., Reyrat,J.M., Jacobs,W.R., Jr., Gicquel,B., & Guilhot,C. (1997) Efficient allelic exchange and transposon mutagenesis in Mycobacterium tuberculosis. *Proc.Natl.Acad.Sci.U.S.A*, **94**, 10955-10960.

Pelicic,V., Reyrat,J.M., & Gicquel,B. (1996) Expression of the Bacillus subtilis sacB gene confers sucrose sensitivity on mycobacteria. *J.Bacteriol.*, **178**, 1197-1199.

Petrillo,M., Silvestro,G., Di Nocera,P.P., Boccia,A., & Paolella,G. (2006) Stem-loop structures in prokaryotic genomes. *BMC.Genomics*, **7**, 170.

Pupa,D. (2003) Investigation of recombinant BCG as a vector to deliver HPV-11 L1 vaccines. MSc, University of Cape Town.

Ranes,M.G., Rauzier,J., Lagranderie,M., Gheorghiu,M., & Gicquel,B. (1990) Functional analysis of pAL5000, a plasmid from Mycobacterium fortuitum: construction of a "mini" mycobacterium-Escherichia coli shuttle vector. *J.Bacteriol.*, **172**, 2793-2797.

Rauzier,J., Moniz-Pereira,J., & Gicquel-Sanzey,B. (1988) Complete nucleotide sequence of pAL5000, a plasmid from Mycobacterium fortuitum. *Gene*, **71**, 315-321.

Rehren,G., Walters,S., Fontan,P., Smith,I., & Zarraga,A.M. (2007) Differential gene expression between Mycobacterium bovis and Mycobacterium tuberculosis. *Tuberculosis.(Edinb.)*, **87**, 347-359.

Roberts,E.A., Clark,A., McBeth,S., & Friedman,R.L. (2004) Molecular characterization of the eis promoter of Mycobacterium tuberculosis. *J.Bacteriol.*, **186**, 5410-5417.

Sambandamurthy ,V., Wang,X., Chen,B., Russell,R.G., Derrick,S., Collins,F.M., Morris,S.L & Jacobs,W.R., Jr. (2002) A pantothenate auxotroph of Mycobacterium tuberculosis is highly attenuated and protects mice against tuberculosis. *Nat. Medicine*, **8**, 10, 1171-1174.

Sambrook,J. & Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual,* 3rd edn, Cold Spring Harbour Laboratory Press.

Sander,P., Meier,A., & Bottger,E.C. (1995) rpsL+: a dominant selectable marker for gene replacement in mycobacteria. *Mol.Microbiol.*, **16**, 991-1000.

Sleigh,D. & Timburg,M. (1998) *Notes on Medical Bacteriology,* Churchill Livingstone.

Snapper,S.B., Lugosi,L., Jekkel,A., Melton,R.E., Kieser,T., Bloom,B.R., & Jacobs,W.R., Jr. (1988) Lysogeny and transformation in mycobacteria: stable expression of foreign genes. *Proc.Natl.Acad.Sci.U.S.A*, **85**, 6987-6991.

Stolt,P. & Stoker,N.G. (1996a) Functional definition of regions necessary for replication and incompatibility in the Mycobacterium fortuitum plasmid pAL5000. *Microbiology*, **142 (Pt 10)**, 2795-2802.

Stolt,P. & Stoker,N.G. (1996b) Protein-DNA interactions in the ori region of the Mycobacterium fortuitum plasmid pAL5000. *J.Bacteriol.*, **178**, 6693-6700.

Stolt,P., Zhang,Q., & Ehlers,S. (1999) Identification of promoter elements in mycobacteria: mutational analysis of a highly symmetric dual promoter directing the expression of replication genes of the Mycobacterium plasmid pAL5000. *Nucleic Acids Res.*, **27**, 396-402.

Stover,C.K., Bansal,G.P., Hanson,M.S., Burlein,J.E., Palaszynski,S.R., Young,J.F., Koenig,S., Young,D.B., Sadziene,A., & Barbour,A.G. (1993) Protective immunity elicited by recombinant Bacille Calmette-Guerin(BCG) expressing Outer Surface Protein A(OspA) lipoprotein: A candidate lyme disease vaccine. *Journal of Experimental Medicine*, **178**, 197-209.

Stover,C.K., de,I.C., V, Fuerst,T.R., Burlein,J.E., Benson,L.A., Bennett,L.T., Bansal,G.P., Young,J.F., Lee,M.H., Hatfull,G.F., & . (1991) New use of BCG for recombinant vaccines. *Nature*, **351**, 456-460.

Stutz,H., Bourn,W., Shephard,E., & Williamson,A.L. Use of the Green Fluorescent Protein (GFP) to develop a model system for the optimisation of recombinant BCG vaccination strategies. 2005.
Ref Type: Slide

TIGR. The Institute for Genomic Research. TIGR . 2007.
Ref Type: Electronic Citation

Timm,J., Perilli,M.G., Duez,C., Trias,J., Orefici,G., Fattorini,L., Amicosante,G., Oratore,A., Joris,B., Frere,J.M., & . (1994) Transcription and expression analysis, using lacZ and phoA gene fusions, of Mycobacterium fortuitum beta-lactamase genes cloned from a natural isolate and a high-level beta-lactamase producer. *Mol.Microbiol.*, **12**, 491-504.

Unniraman,S., Prakash,R., & Nagaraja,V. (2002) Conserved economics of transcription termination in eubacteria. *Nucleic Acids Res.*, **30**, 675-684.

Via,L.E., Curcic,R., Mudd,M.H., Dhandayuthapani,S., Ulmer,R.J., & Deretic,V. (1996) Elements of signal transduction in Mycobacterium tuberculosis: in vitro phosphorylation and in vivo expression of the response regulator MtrA. *J.Bacteriol.*, **178**, 3314-3321.

Villar,C.A. & Benitez,J. (1992) Functional analysis of pAL5000 plasmid in Mycobacterium fortuitum. *Plasmid*, **28**, 166-169.

Widdel,F. Theory and Measurement of Bacterial Growth. Online Microbiology Notes . 2007.
Ref Type: Electronic Citation

Yang,F., Moss,L.G., & Phillips,G.N., Jr. (1996) The molecular structure of green fluorescent protein. *Nat.Biotechnol.*, **14**, 1246-1251.

Yasutomi,Y., Koenig,S., Haun,S.S., Stover,C.K., Jackson,R.K., Conard,P., Conley,A.J., Emini,E.A., Fuerst,T.R., & Letvin,N.L. (1993) Immunization with recombinant BCG-SIV elicits SIV-specific cytotoxic T lymphocytes in rhesus monkeys. *J.Immunol.*, **150**, 3101-3107.

Young,D., Lathigra,R., Hendrix,R., Sweetser,D., & Young,R.A. (1988) Stress proteins are immune targets in leprosy and tuberculosis. *Proc.Natl.Acad.Sci.U.S.A*, **85**, 4267-4270.

Yu,J.S., Peacock,J.W., Vanleeuwen,S., Hsu,T., Jacobs,W.R., Jr., Cayabyab,M.J., Letvin,N.L., Frothingham,R., Staats,H.F., Liao,H.X., & Haynes,B.F. (2006) Generation of mucosal anti-human immunodeficiency virus type 1 T-cell responses by recombinant Mycobacterium smegmatis. *Clin.Vaccine Immunol.*, **13**, 1204-1211.

Zahrt,T.C. & Deretic,V. (2000) An essential two-component signal transduction system in Mycobacterium tuberculosis. *J.Bacteriol.*, **182**, 3832-3838.