

GENE EXPRESSION IN MYCOBACTERIA:
ATTENUATION OF GENE EXPRESSION BY
ANTISENSE METHODS AND TRANSLATION
ENHANCEMENT BY DOWNSTREAM BOX ELEMENTS.

Gavin John Rush

Thesis presented for the degree of
DOCTOR OF PHILOSOPHY
In the Division of Medical Microbiology
Department of Clinical Laboratory Sciences
Faculty of Medicine
UNIVERSITY OF CAPE TOWN
March 2004

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

CERTIFICATION OF SUPERVISOR

In terms of paragraph 9 of "General regulations for the degree of Doctor of Philosophy," University of Cape Town, I as supervisor of the candidate Gavin John Rush, certify that I approve of the incorporation in this thesis of material that has been submitted for publication.

signature removed

15/03/24

Supervisor

Professor L. M. Steyn

Division of Medical Microbiology

Department of Clinical Laboratory Sciences

University of Cape Town

BRIEF CONTENTS

	Page
Acknowledgements	i
Abbreviations	ii
Abstract	vi
Chapter 1: General introduction	1
Part A: Attenuation of gene expression in mycobacteria	19
Chapter 2: Natural and artificial control of gene expression	20
Chapter 3: Construction of a transcription vector for antisense RNA expression in mycobacteria	38
Chapter 4: Attenuation of MPB83 expression in <i>M. bovis</i>	72
Chapter 5: External guide sequences	84
Chapter 6: Attenuation of gene expression in <i>M. smegmatis</i> by external guide sequences	105
Part B: Translation initiation	123
Chapter 7: Introduction	124
Chapter 8: Translation augmentation by downstream box sequences in <i>E. coli</i> and <i>M. smegmatis</i> in the absence of a Shine Dalgarno sequence	148
Chapter 9: General conclusions	166
References:	172
Appendix A: Bacterial strains, plasmids, and culture conditions	194
Appendix B: General techniques	200
Appendix C: Buffers and solutions	205

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Prof. Lafras Steyn, for his support throughout my research and during the writing of my thesis.

My sincere appreciation goes to all the members of the Medical Microbiology laboratory but especially to Ros, Karen, and Marc for their friendship, encouragement and technical advice in the laboratory.

I am grateful to the Glaxo/MRC TB initiative and the Wellcome Trust for the funding of my projects.

Finally to Nix, little Ben, and the beautiful Poppy: Thank you for the support, patience and understanding throughout this extended journey. Without your help none of this would have been possible.

ABBREVIATIONS

Aa	Amino acids
aDB	Anti-downstream box
AIDS	Acquired immunodeficiency syndrome
ADC	Albumin-dextrose-catalase
ASD	Anti-Shine Dalgarno
ATP	Adenosine triphosphate
BCG	Bacille Calmette-Guérin
bp	Base pairs
BP	Before present
BSA	Bovine serum albumin
C	Carboxy-
°C	Degrees centigrade
CAP	Cold-acclimation protein
CFU	Colony-forming units
Ci	Curie
CIP	Cold-inducible protein
CSP	Cold-shock protein
ΔG	Thermodynamic stability
ΔG°_f	Gibbs free energy
DATP	Deoxyadenosine 5'-triphosphate
DB	Downstream box
DCTP	Deoxycytidine 5'-triphosphate
DEPC	Diethyl pyrocarbonate
DGTP	Deoxyguanosine triphosphate
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
Dnase	Deoxyribonuclease
DTT	Dithiothreitol
DTTP	Deoxythymidine triphosphate
E	Relative expression

EDTA	Ethylenediaminetetra-acetic acid (disodium salt)
EGS	External guide sequence
FMN	Flavin mononucleotide
G	Gram
GFP	Green fluorescent protein
GRDC	Growth rate dependant control
HCl	Hydrochloric acid
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HIV	Human immunodeficiency virus
Hr	Hour
IC	Initiation complex
IF	Initiation factor
IgG	Immunoglobulin G
IPTG	Isopropyl - β -D-thiogalactopyranosid
Kb	kilobase
KDa	Kilodalton
Kcal	Kilocalories
LAM	Lipoarabinomannan
M	Molar
MA	Milliamperes
MCS	Multiple cloning site
MDR	Multi-drug resistant
Mg	Milligram
ml	Millilitre
MM	Millimolar
MOPS	Morpholinepropanesulfonic acid
mRNA	Messenger ribonucleic acid
N-	Amino-
NaAc	Sodium acetate
Ng	Nanogram
OD	Optical density

OI	Opportunistic infection
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PC	Pre-initiation complex
PCR	Polymerase chain reaction
Pg	Picogram
PGRS	Polymorphic G+C-rich sequences
RD	Region of difference
RLU	Relative light units
RNA	Ribonucleic acid
RNase	Ribonuclease
RNase P	Ribonuclease P
Rpm	Revolutions per minute
RRNA	Ribosomal ribonucleic acid
SD	Shine Dalgarno
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SsDNA	Single-stranded deoxyribonucleic acid
T	Time
TAE	Tris-acetate EDTA
TB	Tuberculosis
TBS	Tris-buffered saline
TC	Translation complex
TK	Thimadine kinase
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
TIR	Translation initiation complex
SCC	Short course chemotherapy
μCi	Microcurie
μg	Microgram

μl	Microlitre
μM	Micromolar
UB	Upstream box
UTR	Untranslated region
UV	Ultra-violet
V	Volts
V	Division rate
v/v	Volume/volume
W	Watts
WHO	World health organization
w/v	Weight/volume
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranosid
YT	Yeast-tryptone

Amino acids:

Ala / A	alanine	Met / M	methionine
Arg / R	arginine	Phe / F	phenylalanine
Asn / N	asparagines	Pro / P	proline
Asp / D	aspartic acid	Ser / S	serine
Cys / C	cysteine	Thr / T	threonine
Gln / Q	glutamine	Trp / W	tryptophan
Glu / E	glutamic acid	Tyr / Y	tyrosine
Gly / G	glycine	Val / V	valine
His / H	histidine		
Ile / I	isoleucine		
Leu / L	leucine		
Lys / K	lysine		

Abstract

This study examines gene expression in members of the genus *Mycobacterium* from two perspectives.

First, I demonstrated that gene expression in both slow-growing and rapid-growing mycobacteria could be reduced by RNA interference. Using an antibody to the mycobacterial lipoprotein, MPB83, I was able to show that plasmids expressing 200bp antisense fragments to the *mpb83* gene are able to reduce expression of the MPB83 protein. I then adopted a novel approach of using an External Guide Sequence (EGS) system to demonstrate attenuation in *ImpC* expression in *M. smegmatis*. In essence, the essential enzyme, RNase P, is induced to cleave an mRNA/EGS hybrid with a resultant reduction in gene expression. In the case of *ImpC* expression, the reduction in *impC* mRNA is confirmed by real-time PCR. This system has potential to target essential genes that cannot be deleted by conventional gene knockout systems.

Secondly, I looked at the problem of the translation of mRNA lacking a defined ribosomal binding site, the Shine Dalgarno region. I demonstrated that there is a highly conserved sequence region at the 3'-end of all mycobacterial 16S rRNA. This sequence shares conserved bases with the putative anti-downstream box of *E. coli*. Following optimization of the putative downstream boxes of leaderless mRNA's for binding to the anti-downstream boxes of either *M. smegmatis* or *E. coli*, I was able to show that these messages are more effectively expressed in their respective hosts, i.e. the *E. coli* optimized mRNA is more effectively expressed in *E. coli* than *M. smegmatis* and visa versa. While not providing definitive proof for the existence of an interaction between the mRNA downstream box and ribosomal anti-downstream box, I showed that the optimized sequences have an enhancer effect on gene expression.

Chapter 1

General Introduction

Contents	page
1.1 History of tuberculosis	2
1.2 Tuberculosis and HIV	4
1.3 Molecular Genetics and Biochemistry of mycobacteria	5
1.4 Technology: Changing the face of pathogen research	7
1.4.1. Automated DNA sequencing	8
1.4.2. Making sense of raw data	9
1.4.3. Challenges in bioinformatics	10
1.4.4. Comparative genomics	10
1.5 Treatment of tuberculosis	11
1.6 Determining gene function in mycobacteria	15
1.6.1 Homologous recombination in mycobacteria	15
1.6.2 Study of genes essential to survival of mycobacteria	16
1.7 Project Aims	17

Chapter 1

General introduction

1.1 History of tuberculosis

The mycobacteria were among the first bacteria to be associated with human disease, after Robert Koch discovered the tubercle bacillus in 1882. The earliest evidence of tuberculosis (TB) in humans comes from human skeletal remains from approximately 5000 BC, which contained deformities suggestive of TB (Sager *et al.*, 1972), and from studies on Egyptian mummies. These date TB infection back to about 3400 BC (Zimmerman 1979). Because of the unusually high degree of conservation of their housekeeping genes, it has been suggested that the members of the *Mycobacterium tuberculosis* complex underwent an evolutionary bottleneck at the time of speciation, estimated to have occurred roughly 15,000 – 20,000 years ago (Sreevatsan *et al.*, 1997). Also, it has been speculated that *M. tuberculosis*, the most widespread etiological agent of human tuberculosis has evolved from *Mycobacterium bovis*, the agent of bovine tuberculosis, by specific adaptation of an animal pathogen to the human host (Stead *et al.*, 1995). This hypothesis is based on the property of *M. tuberculosis* to be almost exclusively a human pathogen, whereas *M. bovis* has a much broader host range. However, both hypotheses were proposed before the entire genome sequence of *M. tuberculosis* (Cole *et al.*, 1998) was available and before comparative genomics uncovered several variable genome regions in the members of the *M. tuberculosis* complex (Brosch *et al.*, 2002). Brosch and colleagues (2002) showed that *M. bovis* has undergone numerous deletions relative to *M. tuberculosis*. They speculated that *M. bovis* is the final member of a separate lineage represented by *M. africanum* (RD9), *M. microti* (RD7, RD8, RD9, RD10) and *M. bovis* (RD4, RD5, RD7, RD8, RD9, RD10, RD12, RD13; Brosch *et al.*, 1999) that branched from the progenitor strain of *M. tuberculosis* isolates. Successful loss of DNA may have contributed to clonal expansion and the appearance of more successful pathogens in certain hosts (Figure 1.1).

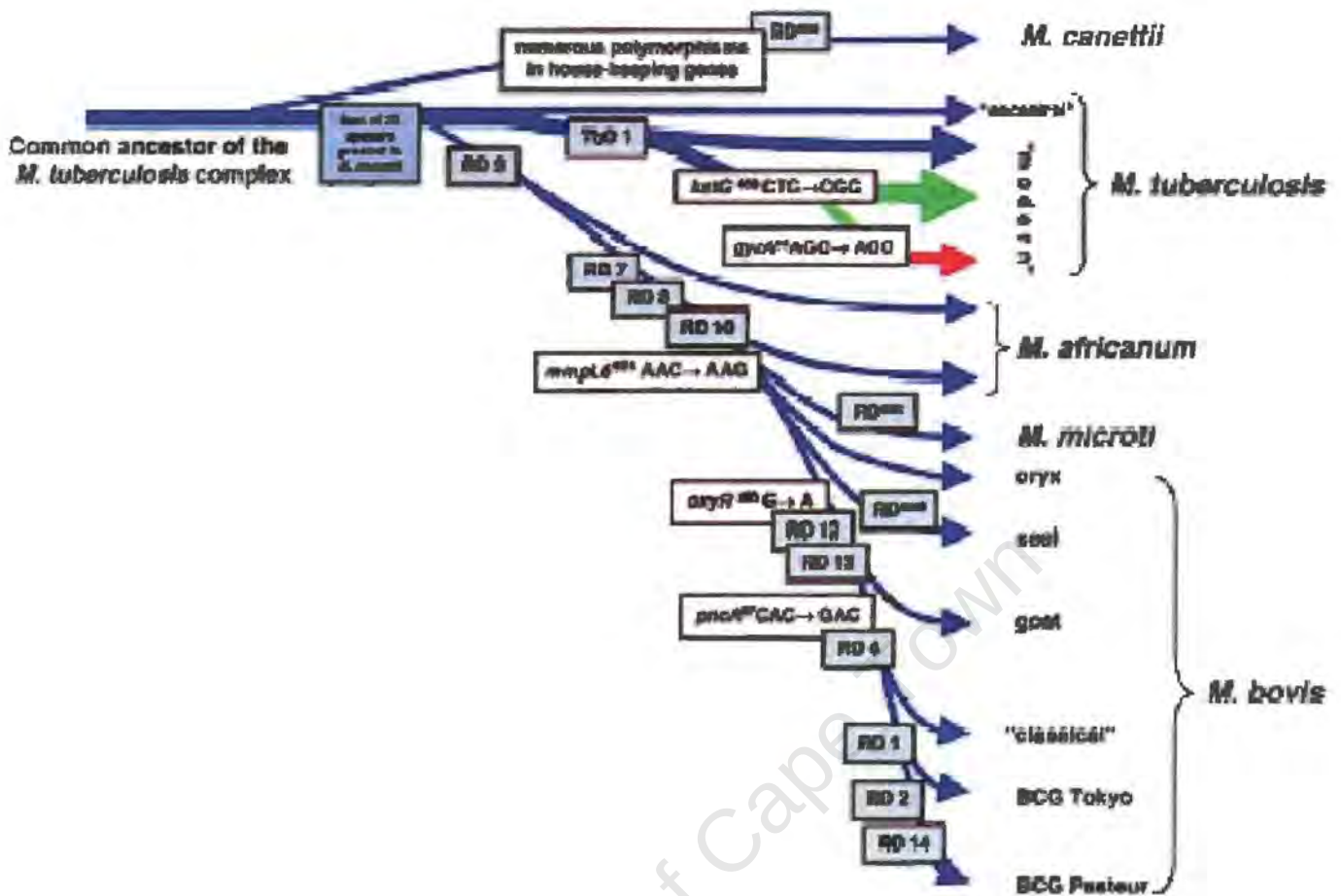


Fig. 1.1. Scheme of the proposed evolutionary pathway of the tubercle bacilli illustrating successive loss of DNA in certain lineages (gray boxes). The scheme is based on the presence or absence of conserved deleted regions and on sequence polymorphisms in five selected genes. Note that the distances between certain branches may not correspond to actual phylogenetic differences calculated by other methods. Blue arrows indicate that strains are characterized by *katG*463. CTG (Leu), *gyrA*95 ACC (Thr), typical for group 1 organisms. Green arrows indicate that strains belong to group 2 characterized by *katG*463 CGG (Arg), *gyrA*95 ACC (Thr). The red arrow indicates that strains belong to group 3, characterized by *katG*463 CGG (Arg), *gyrA*95 AGC (Ser), as defined by Sreevatsan *et al.* (1997). Illustration from Brosch *et al.*, 2002.

Documentation of *M. tuberculosis* complex in humans on both sides of the Atlantic ocean who lived > 3 millenia ago (Martin and Rothschild, 1998; Verano and Ubelaker, 1992; Salo *et al.*, 1994) has lead to a great deal of speculation about its origins. The presence of *M. tuberculosis* complex on both sides of the Atlantic at such early dates suggests that it either arrived with early settlers in North America, or was present in North America when they arrived. The latter hypothesis is supported

by suggestive evidence in the fossils of North American bovids (Martin and Rothschild, 1989). Rothschild and colleagues (2001) provided the first definitive diagnosis of *M. tuberculosis* complex in fossils. They suggest that bovids were the vectors that transported the primordial organism that caused what we today call the white plague: tuberculosis. Their data suggests that tuberculosis in bovids had a Holarctic pattern and that it reached North America at least 20,000 years before present (BP).

Knowledge of the evolutionary history and the antibiotic susceptibilities of the *M. tuberculosis* complex may provide insights that will lead to better control in the future.

In Europe in the nineteenth century pulmonary TB accounted for more deaths than war or cholera or smallpox, measles, whooping cough, scarlet fever, and typhus combined, and yet, it was given less attention than these other infectious diseases (Kochi, 1991). Tuberculosis mortality rates showed a decline from 1838, possibly due to sanitation and labour reform, which occurred around that time. In the twentieth century, these mortality rates continued to decline and levels were relatively low by the 1950s when effective TB chemotherapy was introduced. Similar declines in TB mortality were noted in other developed countries such as the United States, and these have been attributed to improved living conditions (Young and Cole, 1993). While tuberculosis gradually disappeared from the health agenda in the western world it remained a big killer in the developing world (Kochi, 1994). In addition to this, the emergence of human immunodeficiency virus (HIV) has accelerated the mortality due to TB. In 1993, the World Health Organisation (WHO) took an unprecedented step and declared tuberculosis a global emergency, so great was the concern about the modern TB epidemic.

1.2 Tuberculosis and HIV

HIV and TB form a lethal combination, each augmenting the progression of the other. Opportunistic infections (OIs) and rising levels of circulating HIV are hallmarks of

progressing HIV disease (Feinberg 1996). After infection with HIV, a variably paced loss of CD4 T lymphocytes leads to immunodeficiency and increasing susceptibility to a broad range of opportunistic viral, bacterial, fungal, and protozoan pathogens. Considering that up to one third of the human population is currently infected with latent TB, activation of latent TB when co-infected with HIV is a very real possibility. For individuals who are latently infected with TB, the lifelong risk of developing active disease is 10%. This risk increases to 10% per annum of developing active TB in HIV positive individuals (WHO Report 2001). TB is a leading cause of death among people who are HIV-positive. It accounts for about 15% of AIDS deaths worldwide. In Africa, HIV is the single most important factor determining the increase incidence of TB in the last ten years (WHO report 2001). It is estimated that between 2000 and 2020, nearly one billion people will be newly infected with *Mycobacterium tuberculosis*, 200 million people will get sick, and 35 million people will die from TB - if control is not further strengthened (WHO report, 2001).

1.3 Molecular Genetics and Biochemistry of mycobacteria

Research has lead to major insights into cellular immunology and the physiology and biochemistry of the mycobacteria. In contrast, relatively little was known about the genetics or molecular biology of mycobacteria until recently. The lack of progress in molecular biology was primarily because: a) mycobacteria grow very slowly (*M. tuberculosis* has a generation time of 18-24 hrs); b) mycobacteria are hydrophobic and tend to grow in clumps, which makes it difficult to purify individual cells for genetic analysis; c) only recently has naturally occurring genetic exchange between mycobacteria been described (McAdam *et al.*, 1995). Two major advances include the first introduction of foreign DNA into mycobacteria (Jacobs *et al.*, 1987) and the first genetic transformation (Snapping *et al.*, 1988); d) very few genetic markers have been identified in mycobacteria that could be exploited for the development of molecular biologic tools; such as cloning vehicles (Shinnick *et al.*, 1995).

Since its isolation in 1905, the H37Rv strain of *M. tuberculosis* has been used worldwide in biomedical research because it has retained full virulence in animal models of tuberculosis, unlike some clinical isolates; it is also susceptible to drugs and

amenable to genetic manipulation (Cole *et al.*, 1998). In 1998, Cole and colleagues released the 4,411,529 base pair entire genome sequence of *M. tuberculosis* H37Rv. It revealed 3924 genes, and a G + C content of 65.6%, at the time representing the second-largest bacterial sequence available after *E. coli* (Blattner *et al.*, 1997).

The genome is rich in repetitive DNA, particularly insertion sequences, and in new multigene families and duplicated housekeeping genes. The G + C content is relatively constant throughout the genome indicating that horizontally transferred pathogenicity islands of atypical base composition are probably absent. Several regions showing higher than average G + C content were detected; these correspond to sequences belonging to a large family that includes the polymorphic G + C-rich sequences (PGRSs) (Cole *et al.*, 1998).

Since the genome data became available many more virulence genes involved in pathogenicity have been identified. This is but one area where this data will accelerate the study of mycobacterial virulence. By using various database comparisons, Cole and colleagues (1998) attributed precise functions to approximately 40% of the predicted proteins and found some information or similarity for another 44%. The remaining 16% resembled no known proteins and may account for specific mycobacterial functions. These sequences of unknown function will be of particular value in searching for novel pathways and potential drug targets.

Although *M. tuberculosis* is weakly gram-positive, the cell envelope contains an additional layer beyond the peptidoglycan that is exceptionally rich in unusual lipids, glycolipids and polysaccharides (Brennan and Draper 1994; Kolattukudy *et al.*, 1997). Novel biosynthetic pathways generate cell-wall components such as mycolic acids, mycocerosic acid, phenolthiocerol, lipoarabinomannan and arabinogalactan, and several of these may contribute to mycobacterial longevity, trigger inflammatory host reactions and act in pathogenesis (Cole *et al.*, 1998). Little is known about the mechanisms involved in life within the macrophage, or the extent and nature of the virulence factors produced by the bacillus and their contribution to disease.

1.4 Technology: Changing the face of pathogen research

Microbial genome sequencing is driven by the need to understand and control biological pathogens and to exploit extremophiles and their enzymes in bioremediation and industry. It is hard for the traditional bacteriologist to grasp the scale and pace of the venture which has transformed the way research is performed. More than two-dozen microbial genomes have now been completed and, within a decade, genomes from every significant species of bacterial pathogen of humans, animals, and plants will have been sequenced (Pallen 1999). Indeed, we will often have more than one sequence from a species or genus - for example, we already have sequences from two strains of *Helicobacter pylori*, from two strains of *M. tuberculosis* (H37Rv and CDC1551), and from three species of *Pyrococcus*. However, genome sequencing risks becoming expensive molecular stamp-collecting without the tools to mine the data and fuel hypothesis-driven laboratory-based research (Pallen, 1999). Bioinformatics, twinned with the new experimental approaches forming functional genomics, provides some of the needed tools.

Comparative analysis of gene sequences from hundreds of organisms have revolutionised our concepts of biological diversity. The sequence of the first complete bacterial genome (*Haemophilus influenzae*) was published in 1995 and it is expected that complete genome sequence information for 50-100 organisms will be available during the next few years. The exponential increase in the amount of genome sequence data stored in the public databases and the continuous development of novel methods and tools for the analysis of DNA sequences represent new challenges for modern molecular biologists. An understanding of bioinformatic methods is required in order to be able to handle, analyse and interpret large volumes of sequence data that will be generated in the near future. Bioinformatics is thus emerging as a new field of research of relevance for basic sciences as well as for the many projects initiated on the design of new antibiotics and vaccine strategies by the pharmaceutical industries.

1.4.1 Automated DNA sequencing

For the last decade, wet molecular biology work has primarily been associated with DNA sequencing. Genome projects and other large-scale biological research are now producing enormous quantities of biological data that will never be published in the traditional literature. Nucleotide sequences are being added to the database at a rate of more than 210 million base pairs per year, and the database content is doubling in size approximately every 14 months. There are now complete genome sequence data available for all major kingdoms of life, bacteria, archaea and eukaryotes. The sequence databases contained four billion base pairs by late 1999. Much of the progress in bioinformatics is in fact due to the accelerated rate at which sequence data is being produced. To fully appreciate this rate acceleration, here is a brief historical overview of the efficiency of DNA sequencing.

Until 1970, only RNAs could be sequenced, such as for example tRNAs, which were small and easy to obtain in large quantities. The RNAs were usually broken into pieces of a few nucleotides and each fragment was sequenced by chromatographic methods. The first sequence of an intact molecule, a yeast tRNA of 80 bases was sequenced in 1965. DNA sequencing became possible with the discovery of restriction enzymes and DNA polymerises around 1970. For the first time well-defined fragments could be derived from larger molecules and the 5.4kb genome of bacteriophage EX was sequenced in 1976.

A major breakthrough came with the development of the dideoxy chain termination (Sanger) and chemical degradation (Maxam and Gilbert) methods in 1977. The Sanger dideoxy method was successfully used to sequence the human mitochondrial DNA of 16.5 kb in 1981 and the Maxam Gilbert method was used to sequence the 40 kb bacteriophage T7 DNA in 1983. When the dideoxy method was introduced the rate of sequencing was only about 1.5 kb per person per year. With the development of the M13 cloning vectors and oligonucleotide synthesisers the dideoxy method became universally applicable to any DNA fragment. Several automated sequencing machines were developed in the late 1980s to automate gel electrophoresis, raw data

acquisition and base calling. Computer-operated robotic workstations and sophisticated software considerably accelerated the rate at which sequence data could be generated. By the early 1990s the enhanced throughput had increased to approximately 100 kb per person per year. Further important developments were the amplification of tiny quantities of DNA by the polymerase chain reaction (PCR). The combination of PCR and the dideoxy method has allowed the development of cycle sequencing, a process that has tremendously increased the sensitivity of sequencing. The developments in sequencing and bioinformatic technologies enabled the first bacterial genome, the 1.7 Mb genome of *Hemophilus influenzae*, to be completely sequenced by TIGR in 1995.

1.4.2 Making sense of raw data

The raw sequence data is meaningless in the form of naked base pairs. The availability of robust algorithms for predicting coding regions is required in order to convert megabase pairs of sequence data into important biological insights. However, sequence data analysis and interpretations are a painful compromise between what is desired and what is achieved. By combining a variety of computational methods with some experimental work, molecular biologists have been fairly successful at identifying coding and regulatory regions. If a pattern has been recognised, a weight matrix can be used to represent the frequencies of each nucleotide at each position in the site. Consensus sequences and weight matrices are special cases of pattern recognition in which humans play a large part in defining the pattern. A more general approach is to train a program to discriminate between frequencies that are known to contain a particular site or pattern, and those that do not have the site. For example, neural networks have been trained to recognise signals and patterns in sequences. The development of computer methods for the identification of coding regions is currently considered to represent one of the most important areas of research in bioinformatics.

1.4.3 Challenges in bioinformatics

Biological research has long been characterised by a twofold approach (a) field observation, specimen collection and classification (b) laboratory experimentation, often based on hypothesis formed by careful examination of specimen collections. This paradigm in biological research is currently shifting. The new paradigm is based on the availability of the sequences of huge volumes of gene and protein sequences in the databases. As the cloning and sequencing steps will have been accomplished *en masse* for all genes, individual researchers will not need to repeat them for each gene of interest, but will develop hypotheses for new experiments by “mining” the databases. In the new paradigm, the bulk of discoveries might be made by experiments conducted *in silico*, rather than *in vivo* or *in vitro*.

The post-genomic research encompasses many diverse aspects of modern science. Laboratory based functional genomics aims to measure and assess either the messenger RNA (mRNA) levels (transcriptome studies) or the protein content (proteome studies) of cells and tissues. All of these methods, which include microarray and proteomic technologies, have been applied recently to the study of host and pathogen interactions for both bacteria and viruses.

1.4.4 Comparative Genomics

Comparative genomics is a rapidly advancing discipline that is currently being driven by the availability of genome sequences for multiple strains of pathogenic bacterial species and by the advent of DNA micro-array technology. The *Mycobacterium tuberculosis* reference strain, H37Rv, was sequenced in 1998 (Cole *et al.*, 1998). Although originally cultured from a patient with tuberculosis, this strain has been passaged *in vitro* for approximately 100 years. Since then, a recent clinical isolate, described as the “Oshkosh” or CDC1551 strain has been sequenced. In addition, the *M. leprae* genome sequence has been completed (Cole *et al.*, 2001), and the genomes of strains of *M. bovis*, *M. avium*, and *M. smegmatis* have been sequenced. These sequence data should provide a framework to begin elucidating the mechanisms

involved in virulence (Fitzgerald and Musser, 2001). Interestingly, sequence comparison of 26 structural genes in > 800 isolates of members of the *M. tuberculosis* complex (Sreevatson *et al.*, 1997), reveals that the level of synonymous (silent) nucleotide variation in structural genes is greatly restricted and considerably less than present in other pathogens. Whole-genome micro-array technology has also been used for epidemiological analysis of 19 *M. tuberculosis* isolates (Kata-Maeda *et al.*, 2001). An inverse correlation between the percentage of the genome deleted and the percentage of patients infected with that clone who have pulmonary cavitations was observed. Taken together with the large chromosomal regions of difference (RDs) used by Brosch *et al.* (2002) and Behr *et al.* (1999) in their proposed evolutionary pathways of the tubercle bacilli, the data indicate that single-nucleotide polymorphisms are very rare and deletion events are an important source of genome variation in the *M. tuberculosis* complex. Moreover, horizontal gene transfer has been very limited within the *M. tuberculosis* complex.

A new and exciting avenue of research is the study of the *M. tb* transcriptome. In particular, differences in expression profiles of *M. tb* in response to several different environmental conditions. Snappinger *et al.* (2003) began with microarray studies comparing *M. tb* gene expression in log-phase cultures with that of bacteria residing within macrophage phagosomes. A benchmark was provided by the profiles of log-phase cultures compared with cultures exposed to various environmental stressors thought to reflect the environment inside the phagosome. Overlap between the expression profile of the “stressed” cultures and that of bacteria inside the macrophages suggests that the stressors tested could account for changes in 60% of the 601 genes that are differentially expressed in macrophages. These changes in gene expression presumably reflect adaptation of the bacterium to intracellular conditions. Other changes could reflect bacterial mechanisms for actively modifying the phagosome to its own advantage, but remains to be seen (McKinney and Gomez, 2003).

Comparative genomics is still a relatively novel tool and will gain momentum as more clinical isolates are sequenced. Questions for future research include:

- 1) What is the extent of genome diversity within pathogenic bacterial species?
- 2) How has lateral transfer contributed to the evolution of pathogenic bacteria?
- 3) How does the genetic variation between strains result in the differences in host specificity, tissue tropism and virulence?
- 4) How can comparative genomics be used to develop better therapeutics and vaccines? (Fitzgerald and Musser, 2001).

1.5 Treatment of tuberculosis

The continuing high incidence of tuberculosis throughout the world, the increased incidence of tuberculosis in association with the AIDS pandemic, and the emergence of multi-drug-resistant strains of tuberculosis have give new urgency to the need to develop new therapeutic strategies. This emergence of drug-resistant bacteria in general and drug-resistant tuberculosis in particular is rapidly becoming a major public health problem around the world, and the need for finding new treatment modalities for drug-resistant tuberculosis is emphasized by two molecular and conventional epidemiological studies (Alland *et al.*, 1994; Small *et al.*, 1994). Not only was it demonstrated that more than a third of newly diagnosed cases of tuberculosis in New York and San Francisco are the result of recent person-to-person transmission rather than activation of active infections but also that nearly half of the isolates from patients with recently transmitted infections in New York were drug-resistant *M. tuberculosis*. Of these drug-resistant isolates, half were resistant to multiple drugs. Drug-resistant *M. tuberculosis* strains, because of their ability to lead to a long infectious state, are especially active in the recent person-to-person transmissions.

The perceived threat of drug-resistant tuberculosis is enormous. Health warnings have been issued, not just in the popular press but also in principal scientific and medical journals. The spread of MDR has been classified as a global pandemic (Farmer and Kim, 1998) more deadly than AIDS, with the potential to destabilize

society. Drug-resistant mycobacteria are said to be on the rampage, and MDR is thought to have become the norm (Davies, 1996). Two reasons for this are: First, the fraction of patients carrying resistant strains has escalated to levels that put TB control in jeopardy in some parts of the world. About 10% of new TB cases tested are carrying MDR strains in Estonia and Latvia; the rate exceeds 25% among cases that have previously been treated (Dye *et al.*, 2002). The second reason is that people with active TB, including antibiotic-resistant TB, are becoming more mobile internationally.

Antibiotic resistant tuberculosis is on the increase and poses the a major threat to tuberculosis control worldwide. There are at least five reasons in principle, why antibiotic resistance might be found more often among HIV-infected TB patients:

- 1) *M. tuberculosis* strains with lower genetic fitness, manifested as a lower capacity to cause disease after infection (ie. they are less virulent), may appear only in immunosuppressed people. Laboratory experiments have shown that resistant strains are less viable than sensitive strains *in vitro* and often less virulent in guinea pigs (Ordway *et al.*, 1995)
- 2) Tuberculosis among HIV infected individuals is due to recent infections, where a higher fraction of recent infections are drug resistant (CDC 1998).
- 3) There may be shared risk factors for infection with HIV and drug-resistant strains of *M. tuberculosis*, such as injection drug use and hospitalisation. This is likely to be the main reason for the numerous nosocomial outbreaks of MDR tuberculosis among AIDS patients (McCray and Onorato, 2000).
- 4) The treatment of immunosuppressed TB patients might fail because such patients carry a larger number of bacteria. In addition to this, resistance would emerge more readily if the larger population of microorganisms is also genetically more diverse (Dye *et al.*, 2002)
- 5) HIV-infected TB patients may be subjected more often to functional monotherapy (Vernon *et al.*, 1999). There are various mechanisms by which bacteria may be exposed to one drug only, even while a patient is under treatment with the recommended drug combination. For HIV-uninfected patients carrying drug-sensitive strains, bacterial replication is normally suppressed with four drugs during the first 2 months of intensive

phase treatment. Patients then typically switch to a combination of isoniazid and rifampicin during the 4-month continuation phase. If *M. tuberculosis* in AIDS patients continues to replicate in the continuation phase, then replicating organisms might be exposed to rifampicin alone because isoniazid has a shorter half-life and there are no other supporting drugs (Dye *et al.*, 2002). Resistance to rifamycins has been seen repeatedly in HIV-infected TB patients, perhaps for this reason. Another possibility is that the azole antifungal drugs used in AIDS treatment enhance the serum concentrations of rifamycins but not isoniazid, increasing the exposure to the former (Vernon *et al.*, 1999).

Part of the management problem is that drugs to treat resistant strains are expensive and toxic. The cost of treating a patient carrying MDR is typically hundreds of times greater than that for patients carrying fully sensitive strains. In the United States, the bill will normally run into tens of thousands of dollars per MDR patient (Mahmoudi and Iseman, 1993).

Modern chemotherapy of tuberculosis began in 1952 with the development of isoniazid (isonicotinic acid hydrazide), which along with rifampin (rifampicin) are still the mainstays of the treatment of tuberculosis. Streptomycin, which was shown to be effective in the treatment of tuberculosis before 1952, frequently resulted in treatment failure due to rapidly developing resistance to streptomycin. Combinations of streptomycin with p-aminosalicylic acid and isoniazid were then utilised, resulting in a high degree of permanent arrest in tuberculosis patients. The anti-tuberculosis drugs that are currently being used clinically are isoniazid, rifampin, pyrazinamide, and ethambutol as first-line oral drugs; streptomycin, amikacin, kanamycin, and capreomycin as injectable drugs; and ofloxacin, ciprofloxacin, ethionamide, aminosalicylic acid, and D-cycloserine as second-line oral drugs (Iseman, 1993; Mandell and Sande, 1990).

Drug resistance and multi-drug resistance in tuberculosis originate in spontaneous mutations, which occur at predictable rates in tubercle bacilli. The mutations are not linked, and the rise of drug resistant organisms is the result of pre-existing mutations

rather than a result of drug treatment or another mechanism not yet established or understood (Heyn *et al.*, 1994). Thus the emergence of drug resistance is the direct result of the survival of random pre-existing mutations and the selection of the mutation-carrying organisms, due to the killing of the drug-sensitive organisms by effective drugs.

It is highly likely that standard short-course chemotherapy (SCC), based on a combination of cheap and safe-first-line drugs, can prevent the build up of resistance in a population of patients in which most carry drug-sensitive TB. One outstanding epidemiological question is whether first-line drugs alone can defeat MDR by reducing the number of cases (or better still the proportion of MDR) through time. The data we have now suggest they can (Dye *et al.*, 2002).

1.6 Determining gene function in mycobacteria

Sequence database searches have failed to assign any homology to 16% of *M. tuberculosis* proteins. Determining the function of these as well as the other 44% of protein with putative (based on partial homology identity) function could identify novel mycobacterial pathways. These could be exploited in the future to identify novel drug and therapeutic targets.

1.6.1 Homologous recombination in mycobacteria

The “reverse genetics” approach to understanding gene function is to specifically disrupt the gene of interest by exploiting the homologous recombination properties of a cell to replace the functional allele with an inactivated copy (Ruvkun and Ausubel, 1981). Usually the gene is inactivated by the insertion of an antibiotic-resistance marker, such that the recombination event is easily detected on a selective medium. Several groups have used either small linear DNA fragments (Azad *et al.*, 1996; Kapalana *et al.*, 1991; Reyrat *et al.*, 1995), long linear DNA fragments (Balasubramanian *et al.*, 1996), or suicide plasmids (Berthet *et al.*, 1998; Knipfer *et al.*, 1997; Pavelka and Jacobs, 1996; Pelicic *et al.*, 1997; Pelicic *et al.*, 1996a; Pelicic

et al., 1996b; Sander *et al.*, 1995) to achieve allelic exchange in both fast- and slow-growing mycobacteria. However, performing allelic exchange has been relatively cumbersome due to the rarity of double-cross-over events, requiring an extensive screening to isolate a gene-exchange mutant. This method has proven particularly inefficient in slow-growing mycobacteria where homologous recombination is less frequent than illegitimate recombination, i.e. recombination at a site other than the selected gene by an unknown mechanism.

In many cases an antibiotic-resistance marker is not desirable. It may not be known whether a gene is essential, and targeted disruption does not let one ascertain essentiality. The failure to obtain a mutant might be due to the failure of the methodology and not the essentiality of the gene. Furthermore the possibility of polar effect from an inserted antibiotic resistance marker can prevent the disruption of a nonessential gene. Also, there are a limited number of antibiotic resistance genes available for use in mycobacteria and making a marked mutation excludes one antibiotic from further consideration. The most promising counter-selectable system for the slow-growing mycobacteria is *sacB*, which confers sensitivity to sucrose. Methods using *sacB* were used for the targeted disruptions of *ureC* in *M. bovis* BCG (Pelicic *et al.*, 1996b) and *M. tuberculosis* (Pelicic *et al.*, 1997) and the *erp* gene of *M. bovis* BCG and *M. tuberculosis* (Berthet *et al.*, 1998).

Allelic exchange in *M. smegmatis* is relatively easy, and this species does not appear to integrate DNA via illegitimate recombination, whereas slow-growing mycobacteria have a propensity for incorporating exogenous DNA into their genome via illegitimate recombination (Aldovini *et al.*, 1993; Kapalana *et al.*, 1991; McFadden, 1996). Several workers have suggested that the homologous recombination machinery is rather inefficient in the slow-growing mycobacteria. It is generally believed that illegitimate recombination occurs at a higher frequency than does homologous recombination in the slow-growing mycobacteria, but this does not necessarily mean that homologous recombination is defective in these organisms (McFadden 1996). The differences in primary recombination frequencies between slow-growing mycobacteria and *M. smegmatis* likely reflect a difference in DNA entry into the cells, since it is generally agreed that higher electroporation efficiencies are possible with *M. smegmatis* than with the slow growers (Pavelka and Jacobs,

1999). The illegitimate recombination mechanism in slow-growing mycobacteria is not characterised in any detail, but it is hypothesised that the illegitimate recombination machinery may be relatively more sensitive to linear DNA than is the homologous recombination machinery (Pavelka and Jacobs, 1999). In this view, linear DNA might stimulate illegitimate recombination to a much higher degree than homologous recombination.

1.6.2 Study of genes essential to the survival of mycobacteria

Gene disruption/inactivation technology remains an important tool for identifying essential genes for bacterial growth and pathogenesis. Various strategies have been used in bacterial systems to achieve functional inactivation of gene products. Most of these involve gene knockout methods that use insertional techniques, deletions (e.g., allelic replacement), and point mutation (Cheung *et al.*, 1995; Hensel *et al.*, 1995; Akerley *et al.*, 1995; Arigoni *et al.*, 1998). Strains carrying null mutations in genes essential for growth are generally not recoverable for further analysis; thus, conditional lethal mutants are required to successfully identify and characterise genes essential to viability. One method currently used to recover mutants is to provide a functional second copy of the gene, but obtaining statistically significant data can be prohibitively difficult. Parish and co-workers (2001) used the L5-based integrative plasmids along with the L5 excisionase in *M. tuberculosis* to excise and recover these plasmids from the chromosome. This technology also allows further analysis of essential genes.

Attenuation of gene expression by antisense methods has been used successfully in mycobacteria (Rapaport *et al.*, 1996; Parish and Stoker 1997; McDermot *et al.*, 1998; Wilson *et al.*, 1998; Harth *et al.*, 2000). Parish and Stoker (1997) showed that it is possible to alter the phenotype of *M. smegmatis* by antisense methods and more importantly that this could be achieved by using the inducible amidase promoter. By using this conditional system it is now possible to study essential genes using antisense technology.

1.7 Project Aims

Outlined above is a brief review of the history and control of *M. tuberculosis*. Also reviewed is the change in the way research is performed in the 21st century. The pathogen sequence databases and explosion in bioinformatic algorithms has led us to review many of the current methodologies. The thesis is divided into two parts. The first part deals with transcription initiation in general and enhancement of transcription by downstream box (DB) sequences in particular. The second part is dedicated to attenuation of gene expression in mycobacteria.

Genes of unknown function can now be targeted for functional study and open the doors for novel drug targets. In order to harness these exciting avenues of research, new tools need to be developed to aid the study of essential genes. It was for this reason that we embarked on a mission of attenuation of gene expression using antisense sequences. This method then evolved into the attenuation of specific gene expression using external guide sequences (EGSs) based on the work of Dr Sidney Altman, who described RNase P, a ribozyme responsible for the maturation of pre-tRNA to tRNA. This technology has been adapted for use in mycobacteria partly because of the general difficulty in obtaining targeted gene replacements but also with a view of later use as a possible therapeutic agent.

The availability of sequence data has made it possible to design EGSs to any of the 3924 annotated genes of the entire 4,411,529 bp *M. tuberculosis* genome. Any gene can therefore be targeted. EGSs are short oligoribonucleotides designed to bind to a given target mRNA to form a precursor tRNA-like complex. This complex can be recognised by endogenous RNaseP, resulting in specific cleavage of the target mRNA. The result is attenuation of gene expression by irreversible cleavage of the target mRNA.

PART A

Attenuation of Gene Expression in Mycobacteria

University of Cape Town

Chapter 2

Natural and artificial control of gene expression

Contents	page
2.1 Post-transcriptional control of gene expression	21
2.2 Naturally occurring antisense RNAs	21
2.2.1 Plasmid ColE1	22
2.2.2 Plasmid R1	25
2.2.3 IS10	25
2.2.4 Chromosomal antisense RNAs	26
2.3 Mechanism of inhibition of gene expression by antisense control	28
2.3.1 Direct occlusion of the ribosome binding site	28
2.3.2 RNase digestion	29
2.3.3 Alternative mechanisms	29
2.4 Artificial antisense control of gene expression in mycobacteria	30
2.5 External guide sequence mediated attenuation of gene expression	32
2.6 Knockout mutagenesis	33
2.6.1 Allelic exchange mutagenesis	34
2.6.2 Transposon mutagenesis	35
2.7 Aims of this study	36

Chapter 2

Natural and artificial control of gene expression

2.1 Post-transcriptional control of gene expression

Post-transcriptional or translational control of gene expression in prokaryotes has been extensively studied (reviewed by Gold, 1988; Mc Carthy and Gualerzi, 1990; Mc Carthy and Brimacombe, 1994). Translational control is usually negative and the negative control elements can be divided into three main categories: trans-acting proteins; trans-acting RNAs (called antisense RNAs); and, *cis*-acting elements in the mRNAs. Typically, protein repressors bind and sequester an essential part of the translation initiation region. Antisense RNAs inhibit target RNA function by a number of different mechanisms, including direct sequestration of a translation initiation region; accelerated mRNA decay; target RNA refolding; and transcription attenuation (reviewed by Wagner and Simons, 1994; Eguchi *et al.*, 1991). However, the most common theme of negative translational control is the formation of RNA secondary structures that sequester an essential part of the translation initiation region. Usually, such sequestration involves local folding of the SD regions into stable stem-loop structures (de Smit and van Duin, 1990b), but in some cases long-range RNA-RNA interactions are implicated (Saito and Richardson, 1981; Petersen 1989).

2.2 Naturally occurring antisense RNAs

Naturally occurring antisense RNAs in prokaryotes are usually small, untranslated and highly structured. Most of these RNAs have been shown to consist of stem-loop structures, often connected by single-stranded regions or flanked by single stranded tails. Extensive research has been carried out on plasmids R1 and ColE1, where antisense RNAs control plasmid copy number, and on transposon IS10, where translation of transposase mRNA is inhibited by an antisense RNA. These systems are useful sources of information regarding binding rate, specificity and mechanisms of target RNA inactivation.

2.2.1 Plasmid ColE1

The copy number of *E. coli* plasmid ColE1 is regulated by interaction between two plasmid-encoded RNA molecules (reviewed by Eguchi *et al.*, 1991; Figure 2.1). One of the RNAs (RNA II) of 555 nucleotides serves as a primer for initiation of DNA replication by forming a RNA-DNA hybrid that is cleaved by RNase H. The primer formation is negatively regulated by interaction of RNA II with antisense RNA I of 108 Nucleotides. The kinetics of interaction play a key role in copy number regulatory circuits of ColE1 family plasmids and other antisense RNA-regulated plasmids (Wagner and Simons, 1994). The interplay between the rates of various processes involved, like bimolecular interactions, conformational changes and RNA degradation, results in a system that maintains a steady state of plasmid copy number. In particular, the inhibitory effect of ColE1 RNA I on primer formation is mainly dependent on the kinetics of reaction with RNA II (Tomizawa, 1990). Complex formation (and inhibition of replication) is enhanced by the plasmid-encoded Rom protein. Rom protein promotes conversion of the unstable RNA I-RNA II complex to a stable complex. This serves to decrease copy number but its presence or absence does not seem to qualitatively alter the copy number control mechanism (Brendel and Perelson, 1993).

RNA I is complementary to 108 nucleotides at the 5'-end of RNA II and the reaction results in RNA-RNA duplex formation preventing the formation of a functional RNA II structure. However, the initial steps of binding are determined by intramolecular secondary structures of both molecules. The reaction starts from relatively unstable loop-loop interactions ('kissing complex') that finally lead through a series of steps to an RNA-RNA hybrid along the entire length of RNA I (Figure 2.1). The kinetic parameters of some steps were studied experimentally (Tomizawa, 1990) and allowed to develop a quantitative model of RNA I-RNA II binding kinetics (Perelson and Brendel, 1993).

The intramolecular conformational changes in RNA II are thought to play some role in RNA I-RNA II interaction kinetics. During its transcription, RNA II undergoes structural modifications involving the first 200 nucleotides, replacing one of the three

stem-loop structures (stem III) by the alternative structure IV (Fitzwater *et al.*, 1988). This structural transition changes the binding mode of RNA II to RNA I and, consequently, the second-order rate constant of RNA II-RNA I interaction (Wong and Polisky, 1985). There are indications that the rate of the changes is of considerable functional importance. First, it is known that RNA I is an effective inhibitor only if it interacts with RNA II during a particular time window, while, although RNA I-RNA II interaction after this window occurs, it does not inhibit RNA II-DNA hybrid formation (Tomizawa, 1986). The second line of evidence for the importance of the RNA folding kinetics comes from the data on some copy-number mutations in the RNA II sequence. These mutations lead to an increased plasmid copy number due to the defects in the inhibition by the RNA I (Fitzwater *et al.*, 1988). However, the nucleotide changes in some of the mutants influence neither the complementary region of RNA I and RNA II nor any of the proposed alternative intermediate RNA II structures and the role of the RNA II folding pathway in the kinetics of RNA I-RNA II binding (Polisky *et al.*, 1990).

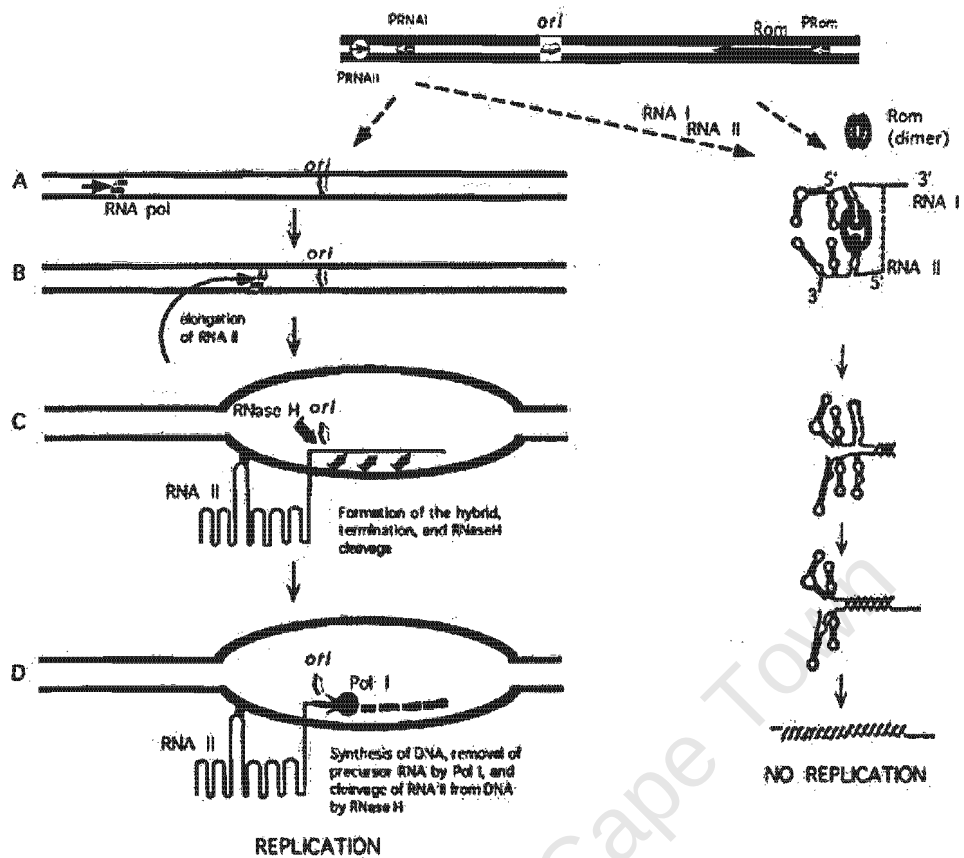


Figure 2.1 Mechanism of initiation of replication of ColE1 and its regulation (Perelson and Brendel 1993). The diagram in the top shows a genetic map of the initiation of replication region of ColE1. The left portion of the diagram depicts the mechanism of initiation of replication. The right portion shows the action of RNA I and the Rom protein. The arrows inside black or white circles indicate initiation of transcription locations for RNA II and RNA I. The origin of replication (*ori*) site (nucleotide 555 of RNA II) and the site of action of RNase H are indicated. A through D represents different steps in the initiation of replication. The proposed mechanism of the succession of events involved in the regulation of the initiation of replication of ColE1 by RNA I and Rom are shown on the right portion. The interaction between RNA I and RNA II leads to the inhibition of DNA synthesis is shown. The first interaction between RNA I and RNA II ("kissing") is reversible and stabilized by the Rom protein. A more detailed scheme of the possible interactions between the RNA and Rom species has been reported elsewhere (Kues and Stahl, 1989).

2.2.2 Plasmid R1

In plasmid R1, an antisense RNA (CopA) is the main copy number control element (Nordstrom *et al.*, 1984). CopA is constitutively transcribed and turned over rapidly *in vivo*. This implies that the intracellular concentration of CopA RNA represents a measure of the copy number of the plasmid (reviewed by Nordstrom and Wagner, 1994). CopA, by binding to its target, the leader region of the RepA mRNA (CopT), inhibits the synthesis of the initiator protein, RepA. This inhibition is direct and involves translational coupling between leader peptide reading frame (*tap*) and *repA* (Blomberg *et al.*, 1994).

Once again, the relationship of structure and function of antisense RNAs is a topic of interest, since features that govern performance of the inhibitor RNA are amenable to study, and lessons learned from the well defined systems may guide the design of artificial antisense RNAs. Previous studies showed that nucleotides in loop regions are involved in primary recognition. For plasmid R1, this conclusion was supported by the copy number phenotypes of naturally occurring point mutations as well as those created by site-specific mutagenesis (Nordstrom *et al.*, 1984, Blomberg *et al.*, 1994, Hjalt and Wagner, 1992). An optimal loop size of the recognition loop II for high binding rates was shown to be 5-7 bases (Hjalt and Wagner, 1992). For complete antisense/target RNA pairing, a single-stranded region also has to be present (Persson *et al.*, 1990).

2.2.3 IS10

Insertion sequence IS10 encodes a single protein, the transposase, which acts at the ends of the element to promote transposition (Kleckner, 1988). IS10 also encodes an antisense RNA whose promoter is located within the transposase gene. The antisense RNA template sequence extends across and beyond the template sequence for the 5'-end of the transposase mRNA, and antisense pairing inhibits transposase gene expression (Case *et al.*, 1988; Lee and Schmidt, 1985; Simons and Kleckner, 1983).

IS10s antisense control is part of a regulatory process called multicopy inhibition which decreases the rate of transposition per IS10 copy with increasing efficiency as the

number of IS10 elements in the cell increases (Kleckner, 1988). As more and more copies of IS10 accumulate, the concentration of the antisense inhibitor increases with a subsequent reduction in the rate of transposase expression per element decreases.

Any such decrease results in a directly proportional reduction in the rate of transposase expression (and transposition) per element, because transposase acts preferentially in *cis* (Kleckner, 1988). If transposase acts freely *in trans*, the effects of antisense control on transposase expression would be counteracted by the increase in effective transposase concentration due to the presence of multiple IS10 copies. Multicopy inhibition is observed experimentally as a reduction in the transposition frequency of a single marked copy of IS10 (or its parent transposon, Tn10) when additional IS10 copies are present, usually on a multicopy plasmid.

Biologically, multicopy inhibition should slow down the rate of accumulation of IS10 copies in a bacterial population. Consistent with this role, antisense inhibition has little effect when IS10 is present in a single copy, because the antisense RNA concentration is too low (Simons and Kleckner, 1988).

2.2.4 Chromosomal antisense RNAs

As outlined above, the majority of antisense RNAs were identified in plasmids, transposons, and bacteriophages. In all of these cases, the antisense RNA-encoding gene overlaps the target gene (*cis*-encoded), but is transcribed off the complementary strand in the opposite direction. In this class, complete complementarity between the RNAs is therefore ensured, and only one target gene is regulated. By contrast, only a small number of antisense genes have been identified in bacterial chromosomes. Most of these are *trans*-encoded, at genetic loci other than their target genes. Therefore, antisense-target complementarity is incomplete, and regulation must be achieved by formation of partial and imperfect duplexes (Altuvia and Wagner, 2000). In *E. coli*, four small RNAs with these characteristics, DicF, MicF, OxyS, and DsrA, have been suggested to regulate gene expression (Wassarmen *et al.*, 1999). Because these RNAs are *trans*-encoded and contain only short regions of target complementarity, this could facilitate recognition of multiple targets. The example of two riboregulators, OxyS and DsrA, may serve to illustrate this property. OxyS is a

small RNA that is induced in response to oxidative stress. It acts as a global regulator by affecting the expression of multiple genes and protecting against DNA damage (Altuvia *et al.*, 1997). As schematically indicated in Figure 2.2, both OxyS and DsrA carry separate determinants for regulation of each of their two different target RNAs. OxyS inhibits the translation of *fhlA*, a transcriptional activator for formate metabolism by an antisense mechanism (Altuvia *et al.*, 1998), and *rpoS* translation is repressed indirectly by titration of Hfq, a host factor required for melting out of an inhibitory mRNA structure (Zhang *et al.*, 1998). As indicated in Figure 2.2 (left, light and dark green lines), different regions of OxyS are involved in these activities. DsrA uses an antisense mechanism for both up-regulation of *rpoS* and down regulation of *hns*. Mutational analyses have defined different segments of DsrA as the modules involved in target interactions (Majdalani *et al.*, 1998; Lease *et al.*, 1998; Figure 2.2 Right).

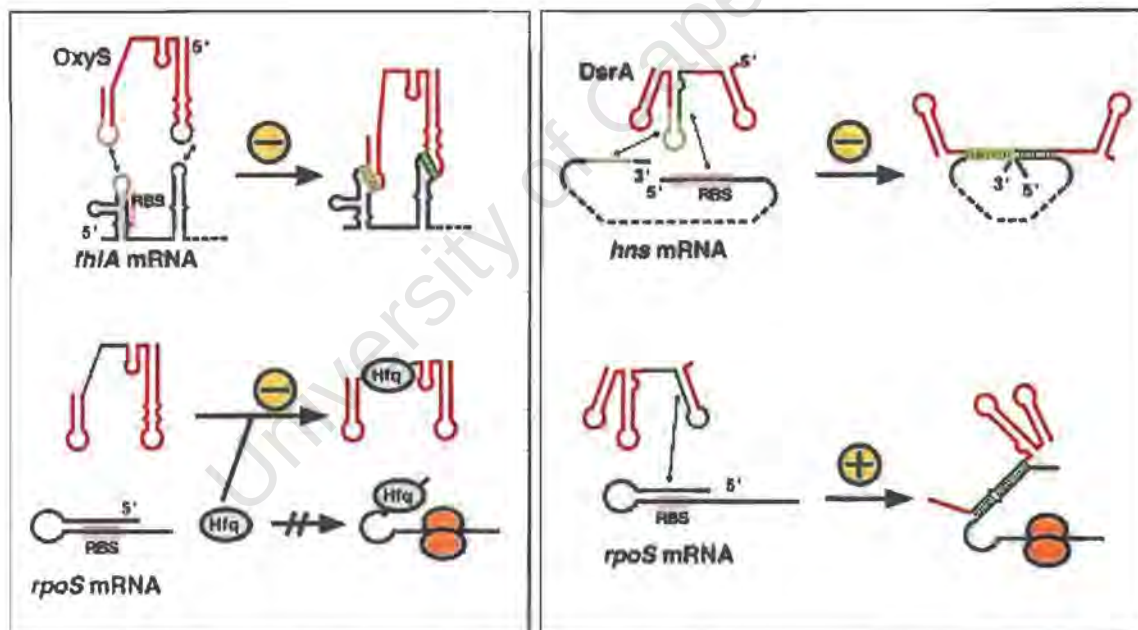


Figure 2.2. Regulation of target RNAs by OxyS and DsrA. The left box shows a schematic of inhibition mechanisms of OxyS and the right box inhibition by DsrA. Target RNAs are shown in blue and antisense RNAs in red. Interacting regions are indicated in green. For clarity, different shades of green are used in the bipartite interactions in the upper two drawings. “+” and “-” indicate activation and inhibition, respectively. RBS indicates the approximate position of the ribosome-binding site. Hfq is host factor Q (see text), and orange mushrooms represent translating ribosomes (Figure from Altuvia and Wagner, 2000).

OxyS and DsrA carry out their functions by forming a relatively modest number of base pairs as determined by the region of complementarity [8+11(for both helicies; see Figure 2.2), DsrA-*hns*; 7+9, OxyS-*fhlA*]. The stereotypical copy number regulators of, e.g., plasmids ColE1, R1, pIP501 and others (Wagner and Simons, 1994), are complementary to their targets over 90-140 nucleotides. At first glance, one might expect the more the better. However, association rates, not the thermodynamic stability (ΔG) of a complex between antisense and target RNA, constitute the most important parameters for inhibitory efficiency (Nordstrom and Wagner, 1994). OxyS forms complexes with *fhlA* mRNA characterized by a dissociation constant of 25 nM (Argaman and Altuvia, 2000). The association rate constant was determined to be $\approx 6 \times 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$, i.e., similar to that of the plasmid antisense systems. This implies that formation of the presumed inhibitory kissing complex is very rapid, dissociation occurs at a half life of $>1 \text{ min}$, and the K_d suggests that as few as 100 molecules per cell will inhibit the majority of the RNA targets (Altuvia and Wagner, 2000). Thus relatively short sequences can suffice for efficient binding and regulation, provided they can interact rapidly with exposed target sequences. Furthermore, results obtained in recent years have shown that even antisense RNAs that were expected to hybridize fully with targets fail to do so. For both plasmid R1 (Kolb *et al.*, 2000) and ColIb-P9 (Asano and Mizobuchi, 2000), the inhibitory complex is a peculiar cruciform-like structure stabilized by an additional intermolecular helix (Altuvia and Wagner, 2000). Thus, the topological difficulties of unwinding stable stem-loop structure may have selected solutions in which more or less extensive kissing complexes are used for regulation (Wagner and Brantl, 1998).

2.3 Mechanism of inhibition of gene expression by antisense control

2.3.1 Direct occlusion of the ribosome binding site

It is highly likely that antisense inhibition of gene expression results in some cases from steric interference by fully or partially paired antisense RNA with binding of ribosomes to the target mRNA. In several antisense systems (IS10, P22 *sar*, *micF*, and R1162), the region of complementarity between the antisense RNA and the

target mRNA includes the start codon for the affected gene and/or its ribosome binding site. Since ribosomes probably only bind efficiently to single stranded RNA and not to duplex RNA (Stormo, 1986), the antisense interaction should be sufficient to block gene expression. This appears to be the case for IS10: 30S ribosomal subunits are unable to bind efficiently to the transposase ribosome binding site when RNA-IN is paired to RNA-OUT (Ma and Simons, 1990).

2.3.2 RNase digestion

An alternative model for antisense inhibition is that pairing renders the target RNA susceptible to digestion by double-strand specific RNase III or other cellular RNases. This mechanism has been excluded in the case of IS10, where RNase III cleavage occurs but is not required for inhibition of transposase expression. The antisense/target RNA duplex (RNA-IN + RNA-OUT) is rapidly degraded at identified sites *in vivo* by RNase III (Case *et al.*, 1990). However, in a strain deficient for RNase III activity (*rnc*-), antisense control occurs at 90 to 95% the normal level (Simons and Kleckner, 1988).

2.3.3 Alternative Mechanisms

In other antisense systems, the antisense RNA is not complementary to the translation initiation region of the target mRNA. Diverse mechanisms have been proposed for inhibition of gene expression at a step subsequent to transcription initiation.

In two cases, the antisense RNA is proposed to cause premature termination of target RNA transcription. For *tic* RNA (Okamoto and Freundlich, 1986), binding of the antisense species to *crp* message is proposed to create a duplex region downstream of the transcription complex that mimics the role of GC stem-loops in normal transcription termination. For pT181, a more complex model is proposed in which antisense RNA binding causes conformational changes in the target mRNA (Novick *et al.*, 1985). In the absence of the antisense RNA, termination of target mRNA transcription at an internal RHO-independent termination site is prevented by pairing between a proportion of the terminator stem-loop and an upstream sequence, the pre-emptor. The pT181 antisense RNA is proposed to bind to the pre-emptor segment,

thus allowing formation of the terminator structure and then termination of transcription.

Antisense controls IncFII *repA* expression by binding to upstream sequences in a polycistronic transcript. Antisense pairing is proposed to alter the secondary structure of distal messenger sequences in such a way that the otherwise accessible *repA* ribosome binding site is now sequestered in a stable double-stranded stem (Womble *et al.*, 1987). Translation of the central gene in the target message is proposed to modulate the antisense interaction owing to effects of ribosomes on target RNA structure (Wagner *et al.*, 1987). Transcriptional pausing in this region may also play a role (Dong *et al.*, 1987).

2.4 Artificial antisense control of gene expression in mycobacteria

Synthesized antisense oligoribonucleotides, which base pair with gene transcripts, constitute a relatively new technology for the control of gene expression in prokaryotes and eukaryotes, including mammalian cells (Zamecnik and Stephenson, 1978). This technique is particularly attractive in situations where it is lethal to inactivate a gene but reduction of the level of protein may be tolerated. Used in conjunction with available sequence data (Cole *et al.*, 1998), antisense oligoribonucleotides are a powerful tool in studying genes involved in mycobacterial pathogenesis.

Antisense oligoribonucleotides have been used with varying degrees of success in mycobacteria (Rapaport *et al.*, 1996; Parish and Stoker, 1997; McDermot *et al.*, 1998; Wilson *et al.*, 1998; Harth *et al.*, 2000). Parish and Stoker (1997) showed that the introduction into wild-type *M. smegmatis* of a construct expressing the antisense strand of *hisD* from the amidase promoter could produce a phenotype, which is similar to a genuine histidine auxotroph, under induced conditions. Auxotrophy was not observed under non-induced conditions. For this they made use of the inducible acetamidase promoter. Briefly, the amidase gene of *M. smegmatis* enables the organisms to grow on simple amides as the sole carbon source, and is highly inducible in the presence of acetamide (Draper, 1967). The conditional system described (Parish and Stoker, 1997) can be used in two ways: (i) specific genes can be studied in

the same way as *hisD*, (ii) essential genes can be identified using a shotgun approach and screening for clones which grow in the absence, but not the presence of acetamide.

In another example of antisense attenuation of gene expression, this time in the slow growing *M. bovis*, Wilson *et al.* (1998) used antisense RNA to *ahpC*, an oxidative stress defence gene involved in isoniazid resistance, to demonstrate an important role of *ahpC* in the virulence of both wild-type and INH-resistant strains. Furthermore, they showed that overexpression of *ahpC* can compensate for the loss of catalase peroxidase activity that occurs in some INH-resistant strains. For this work the entire *ahpC* gene was reversed and placed under the control of the *ahpC* promoter. AhpC levels were shown to be significantly reduced in bacteria electroporated with antisense expressing constructs.

In yet another study, Rapaport *et al.* (1996) assessed the potential efficacy of modified antisense oligonucleotides against drug resistant *M. smegmatis*. Under normal conditions, oligonucleotide phosphorothioates do not enter mycobacteria, and several strategies were successfully utilized to afford entry of oligonucleotides into the mycobacterial cells. One involved the presence of very low levels of ethambutol, which enables the entry of oligonucleotides into mycobacteria because of its induced alterations in the cell wall. Another involved the utilization of oligonucleotides covalently attached to a D-cycloserine molecule, whereby entry into the mycobacterial cell is achieved by a receptor mediated process. Another low molecular weight, covalently attached ligand that enabled the entry and subsequent antimycobacterial activities of the oligonucleotide phosphorothioates in the absence of a cell wall modifying reagent was biotin. Significant sequence-specific growth inhibition of wild-type, as well as of drug-resistant, *M. smegmatis* was obtained by modified oligonucleotides complementary in sequence to a specific region of the mycobacterium aspartokinase (*ask*) gene when utilized in combination with ethambutol (as compared to ethambutol alone) or as D-cycloserine or biotin covalent adducts without the presence of any other cytotoxic or cytostatic agent.

Controlled gene expression systems, which allow selective genes to be regulated (i.e., titrated up or down) and thereby functionally analyzed, have the potential to provide more quantitative data on the gene product.

2.5 External guide sequence mediated attenuation of gene expression

Recently it has become possible to utilize the features of a ubiquitous cellular enzyme, ribonuclease P (RNase P), to cleave target mRNA (Yuan *et al.*, 1992). External guide sequences (EGSs), pioneered by Altman *et al.* (Forster and Altman, 1990; Altman, 1993), are short oligoribonucleotides that, upon hybridization with a RNA target via complementary Watson-Crick base pairing, can form a tRNA-like complex and induce sequence-specific cleavage of the RNA target by endogenous RNase P. In all cells, the ribonucleoprotein RNase P processes tRNA precursors by cleaving the 5' leader sequence, thereby generating mature tRNA. It was demonstrated that RNase P from *E. coli* as well as human RNase P can recognize bimolecular constructs that resemble tRNA precursor and can hydrolyze one of the strands at a specific position. The RNA strand that is cleaved is defined as the target strand, and the other strand, the external guide sequence (EGS) (Forster and Altman, 1990; Yuan *et al.*, 1992). *In vitro*, the RNA subunit of RNase P alone is sufficient to catalyze his reaction and it is therefore a ribozyme. Unlike most ribozymes however, RNase P recognizes its substrate through tertiary interactions rather than through extensive Watson-Crick base pairing (Pace and Brown, 1995).

Results published by researchers using this methodology in unrelated organisms indicate a higher efficiency of attenuation using the EGS approach as compared to the "antisense-alone" method. In *E. coli*, under conditions that assure a high copy number of EGSs and/or multiple target sites in a designated mRNA, expression of the target genes is virtually abolished (Guerrier-Takada *et al.*, 1997). A reduction of about 80% in the expression of both thymidine kinase (*TK*) mRNA and protein has been observed in human cells in which EGSs directed to the *TK* of herpes simplex virus 1 are expressed (Kawa *et al.*, 1998). Effective inhibition of influenza virus

production in cultured cells by EGS has also been demonstrated (Plehn-Dujowich and Altman, 1998).

Ribonucleic acid and their synthetic oligonucleotides, by virtue of their ability to bind to a nucleic acid target in a sequence specific manner, are being developed as potential human therapeutic agents. The main advantage of this approach, in addition to its anticipated high specificity, lies in the fact that oligonucleotide drug candidates can be developed rapidly once the sequence of the target is known. These reagents could also have great value in identifying the function of various gene products; and with the rapid explosion in the acquisition of sequences of human genes aided by the Human Genome Project, the utility of such reagents is gaining importance (Ma *et al.*, 1998). Several antisense phosphorothioate oligoribonucleotides are currently in clinical trials for a variety of disease indications, and chimeric molecules with higher orders of complexity in terms of chemical composition and mode of action are likely to follow (Crooke and Bennett, 1996; Matteucci and Wagner, 1996).

2.6 Knockout mutagenesis

Gene disruption/inactivation technology remains an important tool for identifying essential genes for bacterial growth and pathogenesis. Various strategies have been used in bacterial systems to achieve functional inactivation of gene products. Most of these involve gene knockout methods that use insertional techniques, deletions (e.g., allelic replacement), and point mutations (Cheung *et al.*, 1995; Hensel *et al.*, 1995; Akerley *et al.*, 1995; Arigoni *et al.*, 1998). Strains carrying null mutations in genes essential for growth are generally not recoverable for further analysis; thus, conditional lethal mutants have been used successfully to identify and characterize genes essential for viability. The genetic analysis of mycobacteria was initially hampered by a lack of efficient tools for generating defined mutants. However, recent advances in both allelic exchange and transposition mutagenesis have enabled the identification of virulence factors following Koch's molecular postulates (Falkow, 1988; Jacobs, 1992).

2.6.1 Allelic exchange mutagenesis

The “reverse genetics” approach to understanding gene function is to disrupt the gene of interest by exploiting the homologous recombination properties of the cell to replace a functional allele with an inactivated copy (Ruvkun and Ausubel, 1981). Usually, the gene is inactivated by the insertion of an antibiotic-resistance marker, such that the recombination event is easily detected on a selective medium. This methodology was shown to be effective in *M. smegmatis* using the *pyrF* gene (Husson *et al.*, 1990). However, performing allelic exchange has been relatively cumbersome due to the rarity of double-cross-over events, requiring extensive screening to isolate a gene exchange mutant. This method has proven particularly inefficient in slow-growing mycobacteria where homologous recombination is less frequent than illegitimate recombination. Several workers have been unable to detect any gene replacement (Aldovini *et al.*, 1993; Kalpana *et al.*, 1991). However, several recent studies have succeeded in identifying exchange, albeit at low frequency, in slow-growing mycobacteria (Marklund *et al.*, 1995; Norman *et al.*, 1995; Reyrat *et al.*, 1995; Balasubramanian *et al.*, 1996; Parish *et al.*, 1999; Armitage *et al.*, 2000; De Voss *et al.*, 2000; Wards *et al.*, 2000; Springer *et al.*, 2001).

More recently, the double-selection strategy using the particular properties of a counter-selectable marker has been developed for allelic exchange mutagenesis in TB. Screening for mutants is facilitated if transformants retaining the delivery vector are eliminated on plates containing a counter-selective compound. Several counter-selectable markers (*katG*, *rpsL*, *sacB*) have been described and used to facilitate allelic exchange mutagenesis in mycobacteria (McFadden 1996). In addition to the *sacB* gene, Pelicic *et al.* (1996a) utilized a thermosensitive mycobacterial origin of replication in the second-generation ts-*sacB* vectors. They demonstrated that ts-*sacB* vectors could be used for allelic exchange mutagenesis in *M. tuberculosis* of *purC*, a gene of the purine biosynthesis pathway (Jackson *et al.*, 1996) that was refractory to mutagenesis using previously available methods.

In many cases an antibiotic resistance marker may not be desirable as a disruption insertion (Pavelka and Jacobs, 1999). It may not be known whether a gene is essential, and targeted disruption does not let one ascertain essentiality. The failure

to obtain a mutant might be due to the failure of the methodology and not the essentiality of the gene. Furthermore the possibility of polar effects from an inserted antibiotic resistance marker can prevent the disruption of a nonessential gene if that gene is located in an operon upstream of an essential gene. Also, there are a limited number of antibiotic resistance genes available for use in mycobacteria and making marker mutation excludes one antibiotic from further consideration. In addition, mutants that are potential vaccine candidates should not contain antibiotic resistance determinants (Pavelka and Jacobs, 1999). With this in mind Pavelka and Jacobs (1999) utilized a system of whereby in-frame deletions were created negating the concerns with using a targeted disruption method. Such mutants are antibiotic sensitive and cannot revert, and the mutations should not have polar effects on the expression of downstream genes.

2.6.2 Transposon mutagenesis

Much of our current understanding of bacterial pathogenesis is based on data generated by insertional mutagenesis using mobile elements that, upon transposition, randomly disrupt chromosomal genes. Insertional mutagenesis involves using a mobile element to disrupt genes randomly in the chromosome upon transposition. Using a conventional strategy, i.e., delivery of the transposon on a vector unable to replicate in mycobacterial cells, two groups simultaneously reported efficient transposon mutagenesis in *M. tuberculosis*. IS1096 derivatives engineered to contain kanamycin resistance cassettes were used to create transposon mutagenesis libraries containing more than 10^6 mutants of *M. tuberculosis* (Pelicic *et al.*, 1997). Bardarov *et al.* (1997) followed a considerably different route to achieve transposon mutagenesis in *M. tuberculosis*. They used a conditionally replicating phage system for delivering the transposon into the *M. tuberculosis* chromosome. The greatest advantage of using a transposon-containing phage is that the natural process of phage infection is extremely efficient: virtually every cell of a bacterial population can be infected, thus optimizing the number of mutants that can be generated. In brief: thermosensitive phasmids derived from mycobacteriophages D29 and Tm4 were isolated after chemical mutagenesis. Although still able to infect mycobacteria, the phasmids selected were not able to replicate or integrate at a non-permissive temperature. IS1096 derivatives were inserted into these phasmids and the resulting

constructs were used to infect TB and BCG at a non-permissive temperature. Transposition, however still occurs and several thousand mutants were generated, all of them resulting from the transposition of the mobile element into the chromosome. The insertions mapped throughout the genome, indicating that the mutant libraries were indeed representative.

Both of these methods have been used to good effect especially in the construction of attenuated strains (Hensel and Holden, 1996, Parish and Stoker, 2000; Gordhan *et al.*, 2002; Hondalus *et al.*, 2000)

Signature-tagged mutagenesis (STM) allows the screening of large pools of tagged mutants for those containing insertions in genes required for growth within the infected host (Hensel *et al.*, 1995). Pooled mutants are used to infect the host, permitted to multiply, and hybridization used to identify those strains present in the initial inoculum but absent from bacteria recovered after infection. While this technique does not provide information concerning the differential expression of genes in vivo with respect to ex vivo conditions, in vivo expression is assumed due to the requirement of gene expression for survival within the host. Signature-tagged mutagenesis requires an efficient insertional mutagenesis system to generate tagged mutations representing the genome of the target organism. Camacho *et al.* (1999) described an efficient transposon mutagenesis system based on a plasmid carrying a thermosensitive replicon and the counter selectable marker *sacB*. By using this mutagenesis system they have been able to successfully adapt STM to *M. tb*.

Rubin *et al.* (1999) developed simple transposition systems by using minitransposons, made up of short inverted repeats flanking antibiotic resistance markers. These elements efficiently transpose after expression of transposase from an appropriate bacterial promoter. They found that transposition of *mariner*-based elements in *Escherichia coli* produces diverse insertion mutations in either a targeted plasmid or a chromosomal gene. With *HimarI*-derived transposons they were able to isolate phage-resistant mutants of both *E. coli* and *Mycobacterium smegmatis*. *mariner*-based transposons provide valuable tools for mutagenesis and genetic manipulation of bacteria that currently lack well developed genetic systems.

2.7 Aims of this study

The aim of part A is the attenuation of gene expression in mycobacteria. Chapter 3 is a promoter study identifying a promoter suitable for driving expression of antisense molecules. This chapter also investigates the half-life of luciferase protein to illuminate promoter activity analysis. Chapter 4 looks at attenuation of the MPB83 lipoprotein using short segments of *mpb83* antisense RNA. We have a monoclonal anti-MPB83 antibody that was used to assess the levels of MPB83 attenuation using this method.

Chapter 5 and 6 is an extension of the antisense theme of attenuation. It involves the attenuation of specific gene expression using external guide sequences (EGSs) based on the work performed by Dr Sidney Altman who described RNase P, a ribozyme responsible for the maturation of pre-tRNA to tRNA. This technology has been adapted for use in mycobacteria partly because of the general difficulty in obtaining targeted gene replacements but also with a view of later use as a possible therapeutic agent. This chapter contains evidence of an exciting new method of attenuation that has made it possible to design EGSs to any of the 3924 annotated genes of the entire 4,411,529 bp *M. tuberculosis* genome. Any gene can therefore be targeted. EGSs are short oligoribonucleotides designed to bind to a given target mRNA to form a precursor tRNA-like complex. This complex can be recognised by endogenous RNase P, resulting in specific cleavage of the target mRNA. The result is attenuation of gene expression by irreversible cleavage of the target mRNA.

Chapter 3

Construction of a transcription vector for Antisense RNA expression in mycobacteria

Contents	page
3.1 Summary	40
3.2 Introduction	40
3.2.1 Promoter choice	41
3.2.2 Reporter choice	43
3.2.3 Reporter construct design	45
3.2.4 Aims	46
3.3 Materials and Methods	47
3.3.1 Bacterial strains and plasmids	47
3.3.2 Electroporation	48
3.3.3 Luciferase assays	49
3.3 Results	49
3.4.1 Amplification and cloning of the <i>M. tuberculosis</i> <i>katG</i> promoter	49
3.4.2 Deletion of <i>tssA</i> of pJCKatG creating pJCKatG2	50
3.4.3 Cloning the <i>hsp60</i> promoter into pJCLuc	51
3.4.4 Deletion of 5'-UTR of <i>hsp60</i> promoter creating pJCHsp1	52
3.4.5 Cloning the <i>mpb83</i> promoter into pJCLuc	54
3.4.6 Promoter expression analysis	54
3.4.6.1 <i>mpb83</i> Expression analysis	54
3.4.6.2 <i>katG</i> and <i>hsp60</i> expression analysis	55
3.4.7 Stability of luciferase in <i>M. smegmatis</i> .	58
3.4.8 Development of transcription expression cassette	59
3.4.9 PCR amplification of <i>mpb83</i> and cloning into pJCKatG	63

3.5	Conclusions	66
3.5.1	Effect of secondary structure at the site of ribosome attachment	66
3.5.2	Effect of secondary structure on mRNA decay	69
3.5.2.1	mRNA length	70
3.5.2.2	5' terminus	70
3.5.2.3	Higher-order RNA structure	71

University of Cape Town

Chapter 3

Promoter expression studies for use in a transcription vector

3.1 Summary

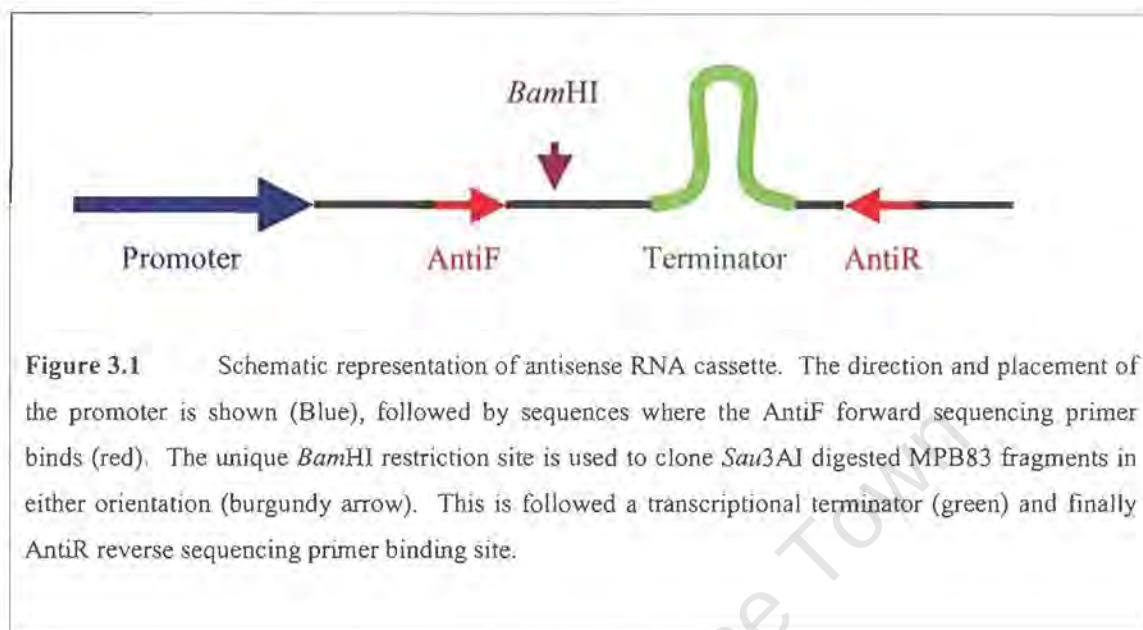
Promoters suitable for over-expressing antisense mRNA were screened using the firefly luciferase reporter gene. These included the *M. tuberculosis katG* promoter (Mulder *et al.*, 1999), and the *hsp60* promoter from *M. bovis* BCG (Stover *et al.*, 1991). Chimeric promoter constructs, in which the 5'UTRs were deleted, were assayed. The *M. tuberculosis katG* promoter was identified as the most suitable promoter for the expression of antisense mRNA. The expression profile of MPB83 was determined using the *M. bovis* BCG (Tokyo) *mpb83* promoter cloned upstream of the firefly luciferase gene. Antisense *mpb83* fragments approximately 200 bps in length were cloned downstream of the *M. tb katG* promoter and upstream of a transcription terminator.

3.2 Introduction

In this chapter a transcription vector capable of over-expressing antisense mRNA is designed. For optimized antisense mRNA expression, a highly active promoter capable of driving expression in exponential phase of growth is required. For this, a translation fusion vector was used using the firefly luciferase gene as a reporter for promoter activity.

The transcription vector contains the following basic features: A promoter is located upstream of antisense sequences; antisense sequences are cloned into a unique *BamHI* site, which has compatible ends with *Sau3AI* and allows for cloning in either orientation; a transcription terminator is located downstream of the *BamHI* restriction site. Confirmation of sequence orientation was performed using AntiF and AntiR

sequencing primers which bind to regions upstream of the *Bam*HI restriction site and downstream of the transcription terminator respectively (Figure 3.1).



3.2.1 Promoter choice

The first step in designing the antisense RNA expression cassette is selecting a suitable promoter. For this, promoters that had previously been described, with known transcription start sites were used. These included the *M. tuberculosis katG* promoter (Mulder *et al.*, 1999), and the *hsp60* promoter from *M. bovis* BCG (Stover *et al.*, 1991). Both of these promoters utilize two transcriptional start sites shown in figure 3.2.

The *M. tuberculosis katG* promoter encodes a dual-function catalase-peroxidase enzyme, which protects the cell against excess hydrogen peroxide and therefore contributes to its survival in macrophages (Middlebrook and Cohn, 1953; Jackett *et al.*, 1978; Wilson *et al.*, 1995; Heym *et al.*, 1997). Two transcriptional start sites were identified for the *M. tuberculosis katG* promoter in both *E. coli* and *M. smegmatis* (Mulder *et al.*, 1999)(Figure 3.2 A). This suggests that the gene is transcribed from two promoters and that both hosts recognize the same promoter sequences. The presence of two promoters upstream of mycobacterial genes is not uncommon and has

been reported previously (Stover *et al.*, 1991; Suzuki *et al.*, 1991; Dhandayuthapani *et al.*, 1997; Movahedzadah *et al.*, 1997). It is possible that the promoters are recognized by different RNA polymerase holoenzymes, and that they are utilized to different extents during growth.

The *hsp60* gene also utilizes two transcription start sites at positions 183 and 216 bp (tssA and tssB) in *M. bovis* BCG (Stover *et al.*, 1991), although only tssA is recognized in *M. smegmatis* (Levin and Hatfull, 1993). This promoter is constitutively expressed but is induced under conditions of stress (Figure 3.2 B).

Both the *katG* and *hsp60* promoters have extended 5' untranslated regions (5'-UTRs). Since the mechanism of antisense attenuation is primarily dependant on secondary structure of antisense and target mRNA, it is important that transcription from the promoter be followed as closely as possible by antisense sequences. For this reason, the promoters were fused as close to the transcription start sites as possible to a reporter gene for promoter activity assessment.

A

-35_B -35_A -10_B -10_A

TCATCTACTGGGGTCTATGTCCTGATTGTTCTGATATCCGACACTTCGCGATCACATCCG

↑ *tts_B* ↑ *tts_A*

TGATCACAGCCCGATAACACCAACTCCT**GGAAGGAAT**GCT GTG CCC GAG

RBS M P E

B

-35_B -10_B -35_A

CAAGGTCGAACGAGGGGCATGACCCGGTGCGGGGCTTCTTGCACTCCCCATAGGCGAGT

↑ *tts_B*

GCTAAGAATAACGTTGGCACTCGCGACCGGTGAGTCGTAGGTCGGCACCCCTCAGGCCAG

↑↑ *tts_A*

GCCCGTCGTCGCAGCGAGTGGCAGCGAGGACAAGTAGAGCCCTCCCTCCCGGGGCACTGC

GCCCGGCCAGCGTAAGTAGCGGGGTTGCCGTCACCAGGTGACCCCCCTTTTCATCCCCG

ATCC**GGAGGAAT**CACTTCGCA ATG GCC AAG

RBS M A K

Figure 3.2 Nucleotide sequence of: **A** *M. tuberculosis katG* promoter region including the first three codons. **B** *M. bovis* BCG *hsp60* promoter region including the first three codons. The transcription start sites (*tss*) are indicated with arrows. The putative -10 and -35 regions associated with each transcription start site are underlined, and the ribosome binding site (RBS) is shown in bold.

3.2.2 Reporter choice

The activity of one promoter or a group of promoters using different reporter systems (Kremer *et al.*, 1995; Dhandayuthapani *et al.*, 1995; Dellagostin *et al.*, 1995; Curcic *et al.*, 1994; Sarkis *et al.*, 1995) has previously been reported. A host of different reporter systems were used in these studies including firefly luciferase and green

fluorescent protein (GFP) from the cnidarian *Aequorea victoria*. In an attempt to characterize the expression of a number of different promoters, relative to each other, we chose to use the firefly luciferase reporter system previously described in our lab by Mulder *et al.* (1999).

The luciferase genes from *Vibrio fischeri*, *Vibrio harvei*, and the firefly, *Photinus pyralis*, have been used for this purpose (Engelbrecht *et al.*, 1985, Legocki *et al.*, 1986; Schmetterer *et al.*, 1986; Carmi *et al.*, 1987). The bacterial system requires up to 7 genes in the *luc* operon, however luminescence can be obtained from just 2 subunits of the luciferase enzyme, encoded by *luxA* and *luxB* (Meighen, 1988).

In luminous bacteria, the luciferase enzyme catalyzes the following reaction:



In which R is an aliphatic moiety containing at least 7 carbon atoms, FMN is a flavin mononucleotide, and FMNH₂ is its reduced form. One luciferase molecule gives rise to approximately one photon of light, therefore very small amounts of luciferase can be detected using a luminometer. The reaction, however, requires the addition of exogenous substrates such as decanal. The substrate is generally added in excess to ensure that it is not limiting, however it becomes strongly inhibiting at high concentrations, therefore complicating the interpretation of results. Other problems with decanal are that it is highly volatile, poorly soluble in aqueous systems, lipophilic and is cytotoxic at higher concentrations. The bacterial luciferase has been used as a reporter in mycobacteria (Gordon *et al.*, 1994).

The firefly luciferase enzyme, which catalyzes the oxidation of luciferin in the presence of ATP, has also been used as a reporter in mycobacteria (Hickey *et al.*, 1996, Arain *et al.*, 1996; Sarkis *et al.*, 1995, Mulder *et al.*, 1999). The primary advantage of the firefly luciferase as a reporter is the extreme sensitivity of the assay. Less than 10-20 moles (0,001 pg) of luciferase have been detected under ideal conditions (Wood *et al.*, 1991a), which enables the study of low level expression, for example from weak promoters (Williams *et al.*, 1989). Sensitivity is approximately 100-fold greater than the chloramphenicol acetyl-transferase assay (Alam and Cook,

1990). Other advantages are that the assay is rapid, easy to perform and relatively inexpensive. Endogenous luminescence is absent from most biological samples; therefore there is no interference with the signal generated by luciferase, and no background. The luciferase gene encodes a 62-kDa protein, which requires no post-transcriptional modifications for activity, and can function immediately after translation (Wood *et al.*, 1984; de Wet *et al.*, 1985). It is also structurally very simple and contains no prosthetic groups, bound metals or disulphide bonds, and can therefore be expressed in virtually any host or cell type (Wood, 1990). The protein is stable at room temperature and loses only 5% of its activity after 10 hours. Firefly luciferase catalyzes the oxidation of beetle luciferin with the concomitant production of a photon of light. The oxidation normally occurs from an enzyme intermediate, luciferyl-AMP, however in the presence of co-enzyme A, it occurs through luciferyl-CoA, resulting in light production without the characteristic self-inhibition of luciferase (Wood, 1991b). In conventional luciferase assays (Wood, 1991a), a flash of light is generated, which rapidly decays after the enzyme and substrate are mixed. The Promega luciferase assay system, on the other hand, allows for greater enzymatic turnover of luciferase (Wood, 1991b), resulting in greater light intensity, which remains constant for several minutes. Thus this system does not require rapid mixing protocols or automated injection devices.

3.2.3 Reporter construct design

Possibly the most important point to remember when designing and analyzing reporter constructs or results, is that the activity obtained is a measure of transcription initiation, promoter strength and the efficiency of the ribosome binding site. The use of translation fusion vectors reduces the scale of interpretation error by supplying the sequences required for translation. If these sequences are constant throughout, problems such as the influence of strong or weak Shine Dalgarno sequences, spacing between SD and initiation codon, and use of either ATG or GTG initiation codons can be reduced when assessing relative activities of different promoters. These translation fusion vectors are however not without problems for relative promoter activity assessment.

Due to the complexity of mycobacterial promoter sequences and the high frequency of multiple promoter sequences and transcription start sites (Reviewed in Mulder *et al.*, 1997), deletions of the often large 5'- untranslated region (5'-UTR) during cloning can lead to altered expression. For example, the complex secondary structures found in the extended 5'-UTR of promoters such as *hsp60* (Stover *et al.*, 1991), could play a role in the post-transcriptional regulation of this promoter. The secondary structure of the entire *hsp60* transcript would be different to that of the 5'-UTR region fused to the reporter mRNA. This could lead to differences both in post-transcriptional gene regulation and also translation efficiency due to changes in the accessibility of the RBS to 16S rRNA. Secondly, little is known about the possible structural events leading to altered mRNA degradation. If a transcript has an altered half-life due to cloning procedures, this would affect translation. Yet another possibility is the impact of large deletions upstream or downstream of the promoter itself due to cloning. Deletions of regions of the upstream enhancer region of the *katG* promoter of *M. tuberculosis* resulted in transcription from a previously uncharacterized transcription start site (Mulder *et al.*, 1999). Finally, one also has to consider the impact of multiple copies of the reporter fusion construct in terms of regulation by chromosomally encoded regulators when the promoters are episomally transcribed. With these points in mind it is however still possible to obtain an estimate of relative promoter activity using translation fusion vectors.

3.2.4 Aims

The aim of this chapter is to construct a transcription vector capable of expressing antisense RNA in mycobacteria. Luciferase assays were used as an indicator of relative promoter strength.

3.3 Materials and Methods

3.3.1 Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are summarized in table 3.1. For sub-cloning purposes Two *E. coli* strains were used. For *dam* methylation sensitive restriction enzymes such as *BclI*, JM110 (Yanisch-Perron *et al.*, 1985) was used. For all other cloning steps, LKIII (Zabeau and Stanley, 1982) was used. *M. smegmatis* LR222 (Beggs *et al.*, 1995) was used for promoter studies. Details of the commonly used techniques and solutions are given in Appendices B and C, respectively.

Table 3.1
Constructs prepared or used in this study

Bacterial strains and plasmids used in this study		
Bacterial strains	Genotype	Reference
<i>E. coli</i> LKIII	<i>recA, lacI, lacZ, lacY⁺</i>	Zabeau and Stanley, 1982
<i>E. coli</i> JM110	<i>F', traD36, lac^gΔ(lacZ)M15, proA⁺B⁺/rpsL (Str^r), thr, leu, thi, lacy, galK, galT, ara, fhuA, dam, dcm, supE44Δ(lac-proAB)</i>	Yanisch-Perron <i>et al.</i> , 1985
<i>M. smegmatis</i> LR222	High frequency transformation strain	Beggs <i>et al.</i> , 1995
Plasmid	Key Features	Technical Notes
pJCLuc	Promoterless luciferase gene in pJC86.	Michelle Mulder, PhD thesis, University of Cape Town.
pJCKatG	Luciferase gene under the control of <i>M. tuberculosis</i> <i>katG</i> promoter in pJCLuc.	Michelle Mulder, PhD thesis, University of Cape Town.
pUCBM21	M13-based cloning vector	Roche

pAT100B	PUC19 containing 1.8 kb region of <i>M. bovis</i> BCG <i>mpb83</i> gene	P Tippoo, University of Cape Town
pUC21Oligo17	trp terminator cloned into the <i>HindIII/SphI</i> restriction sites of pUCBM21	A plasmid used to subclone MPB83 <i>Sau3AI</i> fragments upstream of the trp terminator, creating pUC21Oligo17-5, 36, 24, 25
pJCKatG-MPB83-5	PJCKatG harbouring region 1 of <i>mpb83</i> in the sense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-36	PJCKatG harbouring region 1 of <i>mpb83</i> in the antisense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-24	PJCKatG harbouring region 2 of <i>mpb83</i> in the sense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-25	PJCKatG harbouring region 2 of <i>mpb83</i> in the antisense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>

3.3.2 Electroporation

M. smegmatis LR222 cells were prepared for electroporation by the method of Jacobs *et al.* (1991). A stationary phase culture of *M. smegmatis* LR222, in Middlebrook-ADC-Tween, was diluted 100-fold into fresh medium, and grown to exponential phase at 37°C with shaking. The cells were harvested by centrifugation, washed 3-4 times in ice-cold 10% (v/v) glycerol, and resuspended in a final volume of 2ml/100ml original culture volume in 10% (v/v) glycerol. For electroporation, 60µl cells were combined with 140µl of 10% (v/v) glycerol and 1µg of DNA, in a 5µl volume, and incubated on ice for 5 minutes. Electroporations were performed in a 0.1 cm gap Gene Pulser® cuvettes (Bio-Rad Laboratories, CA U.S.A) at 1.8kV, 25µF and 1000 Ω, using a Gene Pulser™ apparatus (Bio-Rad Laboratories, CA U.S.A.), which resulted in a field strength of 18.0 kV/cm. After electroporation, 900µl of Middlebrook-ADC was added, and the cells were incubated at 37°C for 3-4 hours to facilitate expression of the kanamycin resistance gene. Transformants containing the recombinant plasmids were selected on Middlebrook-ADC agar plates containing kanamycin at a concentration of 50 µg.ml. A control electroporation with no added DNA was included to monitor the frequency of spontaneous mutation to kanamycin

resistance. After incubation at 37°C for ten days, transformants were cultured in liquid medium for plasmid preparations.

3.3.3 Luciferase assays

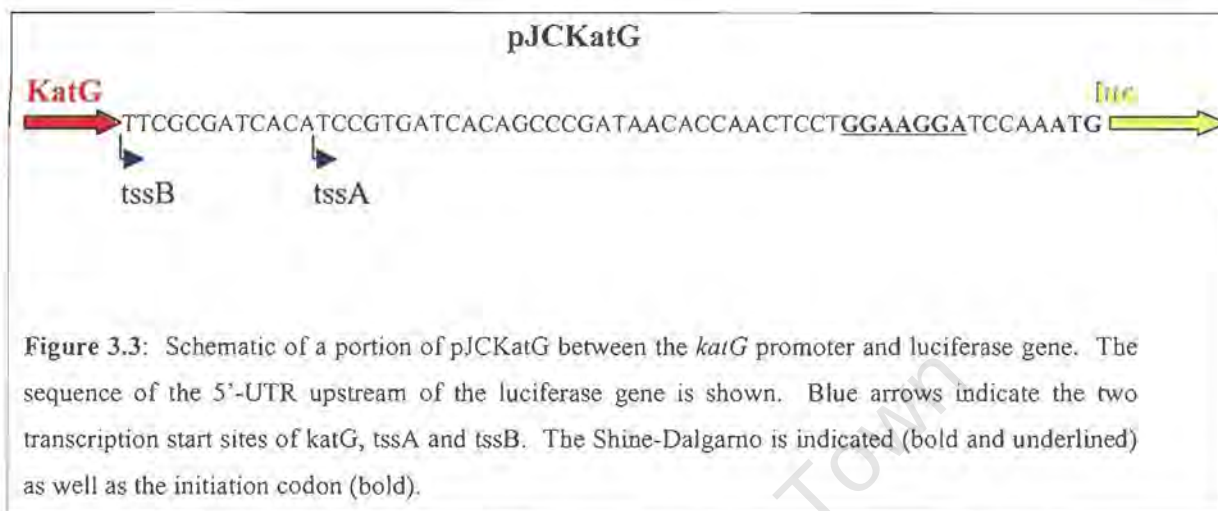
For the luciferase activity assays, 5-ml cultures were grown to late exponential phase (25 hr p.i.) at 37 °C with aeration (Orbital shaker, Hågar Designs, at 150 r.p.m.). *M. smegmatis* cell extracts were prepared by resuspension of cell pellets in Cell Culture Lysis Reagent, followed by processing in the FastPrep FP120 instrument for two, 40s cycles at speed setting 6. Unlysed cells and cell debris were removed by centrifugation. The supernatants were assayed for luciferase activity as follows. For each assay, 100 µl of reconstituted luciferase reagent was added to 20 µl of cell extract, vortexed for 5 s, and the light emission measured in a Bio-Orbit 1253 luminometer. The protein concentrations of the cell extracts were measured using the D_c Protein Assay kit (Bio-Rad), and the luciferase activities are quoted as relative light units (RLU) per mg protein in each assay.

3.4 Results

3.4.1 Amplification and cloning of the *M. tuberculosis katG* promoter

The *E. coli*-mycobacterial promoter probe vector, pJCLuc (Mulder *et al.*, 1999) was used to assess promoter strength in *M. smegmatis*. The *M. tuberculosis katG* promoter was amplified as a 559 bp fragment by PCR (Mulder *et al.*, 1999). This fragment was cloned upstream of the firefly luciferase gene in pJCLuc such that the distance between the *Bam*HI site and the GTG start codon was equal to that between the *Bam*HI and the ATG start codon of the luciferase gene in pJCLuc, ie., 2 bp (Mulder *et al.*, 1999), thus creating pJCKatG. This was done to eliminate any differences in translation efficiency that may occur due to alterations in the distance between the translation initiation codon and the ribosome binding site.

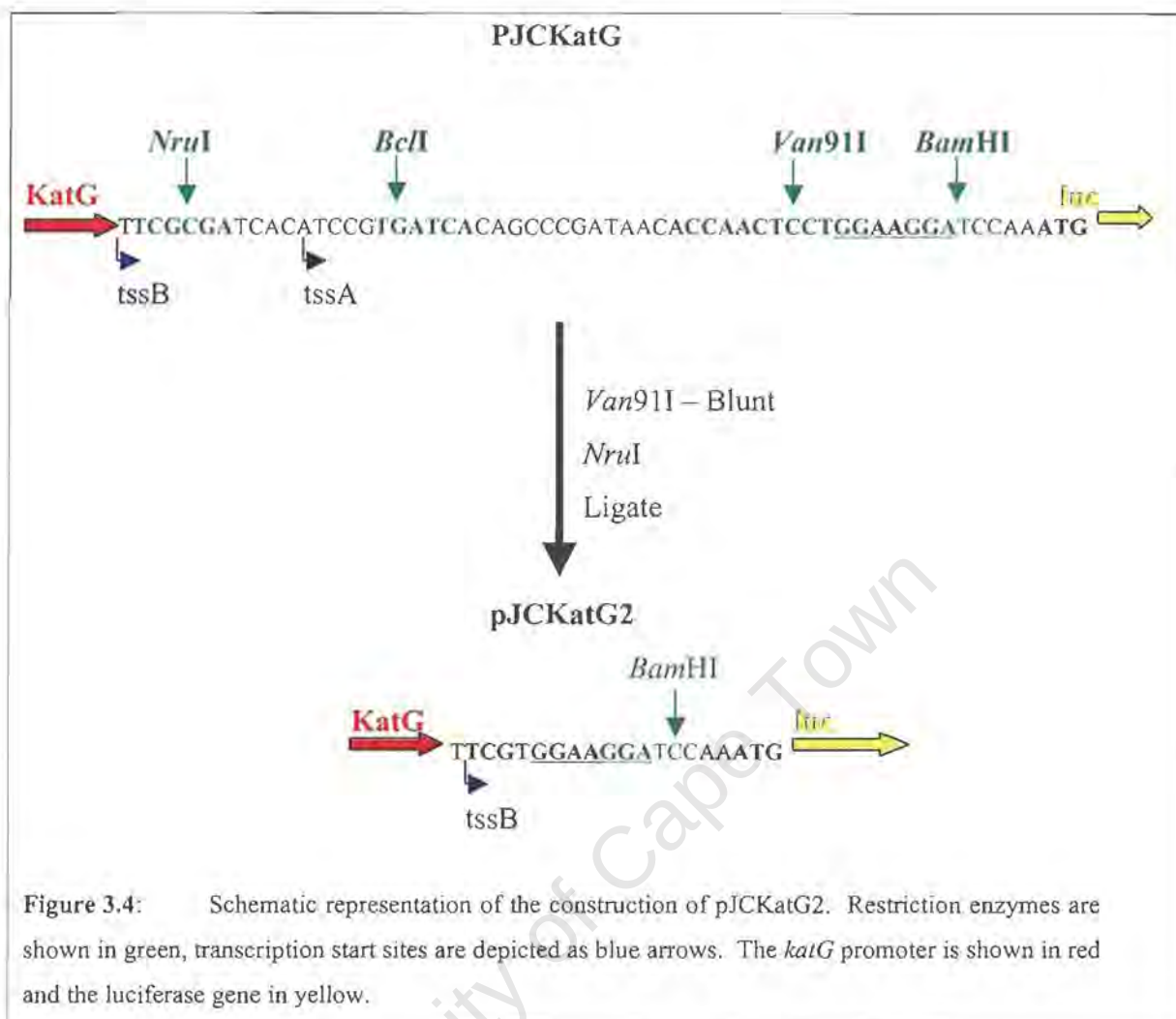
Primer extension analysis performed on *M. smegmatis* and *E. coli* cells harboring pJCKatG (Mulder *et al.*, 1999) indicate that two sites located 47 (tssA) and 56 (tssB) bp upstream of the initiation codon (Figure 3.3)



3.4.2 Deletion of tssA of pJCKatG creating pJCKatG2

To create a promoter construct that enables the use of the *Sau3A*I compatible *Bam*HI site, in close proximity to the transcription start site, part of the 5'-UTR of pJCKatG, including tssA was deleted. To do this, restriction endonucleases, *Van91*I and *Nru*I were used (Figure 3.4)

To form compatible ends with the blunt-end generating restriction endonuclease *Nru*I, the *Van91*I restriction site was digested with mung bean nuclease prior to ligation. Clones were screened for the loss of a *Bcl*II restriction site. This generated pJCKatG2, which was confirmed by sequence analysis using the sequencing primer luc2 (5'-GCAGTTGCTCTCCAGCGG-3'). It has a reduced 5' region upstream of the *Bam*HI restriction endonuclease site, from 38 (tssA) and 49 (tssB) bp in pJCKatG to 10 bp (tssB) in pJCKatG2.



3.4.3 Cloning the *hsp60* promoter into pJCLuc

The *M. bovis* BCG *hsp60* promoter was cloned from pUS967 (Dellagostin *et al.*, 1995) into the *E. coli*-mycobacterial promoter probe vector, pJCLuc. The *hsp60* promoter was excised from pUS967 as a 436 bp fragment by restriction enzyme digestion, first with *EcoRI* followed by blunt-ending with T4 DNA polymerase in the Amersham DNA Blunting Kit (Amersham Life Science, UK), the 5' protruding ends were "filled in". The linear vector was then cleaved with the restriction endonuclease *HindIII*. (Figure 3.5). pJCLuc was prepared for ligation with this fragment by digestion with *ApaI*. The cohesive ends resulting from restriction endonuclease digestion were blunt-ended using the Amersham DNA Blunting kit (Amersham Life Science, UK). The T4 DNA polymerase removes the 3'-protruding ends with its 3'-

5' exonuclease activity. This was followed by digestion with *Hind*III. Ligation of these two fragments resulted in the plasmid pJCHsp60 (Figure 3.5).

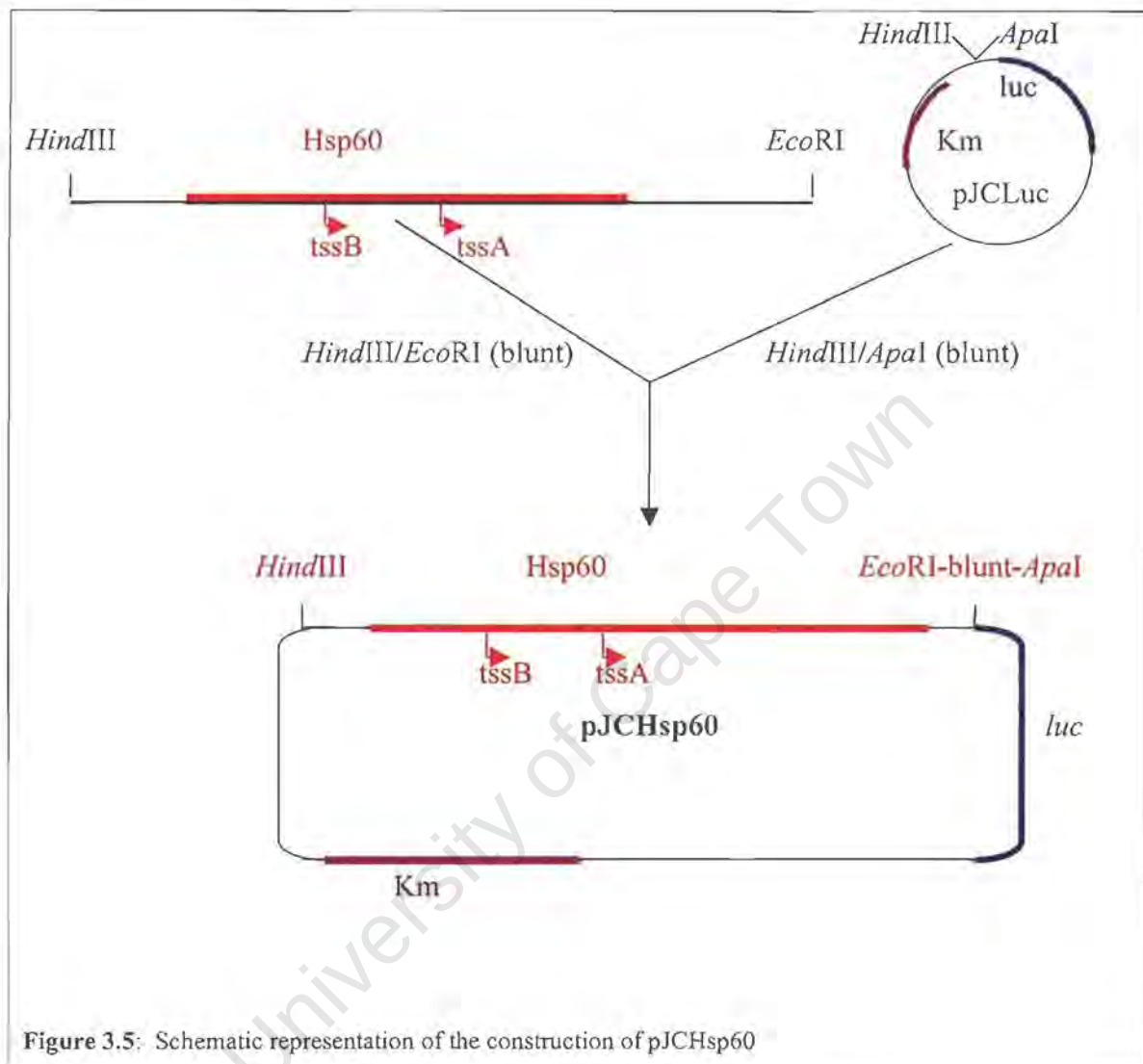


Figure 3.5: Schematic representation of the construction of pJCHsp60

3.4.4 Deletion of 5'-UTR of *hsp60* promoter creating pJCHsp1

To assess the feasibility of using a chimeric, shortened form of the *hsp60* promoter for the expression of antisense constructs, most of the 5'-UTR of the *M. bovis hsp60* promoter was deleted (Figure 3.6). As is the case with the *katG* promoter, the *hsp60* promoter utilizes 2 transcription start sites, *tssA* and *tssB*. However, the 5'-UTR in this case is much longer than that of the *katG* promoter, 183 and 286 bps respectively. The blunt-end restriction endonuclease *Nru*I, cleaves pUS967 6 bases downstream of

tssA. Utilized in conjunction with *Hind*III, it releases a fragment of 269 bps containing the *hsp60* promoter without much of the 5'-UTR. This fragment was cloned into pJCKatG digested with the restriction endonuclease *Van*91I, blunt-ended with T4 DNA polymerase from the Amersham DNA blunting kit (Amersham Life Science, UK), prior to digestion with *Hind*III and removal of the *katG* promoter from pJCKatG. Ligation of the *hsp60* promoter with the pJCLuc backbone created the vector pJCHsp1, which was confirmed by sequence analysis.

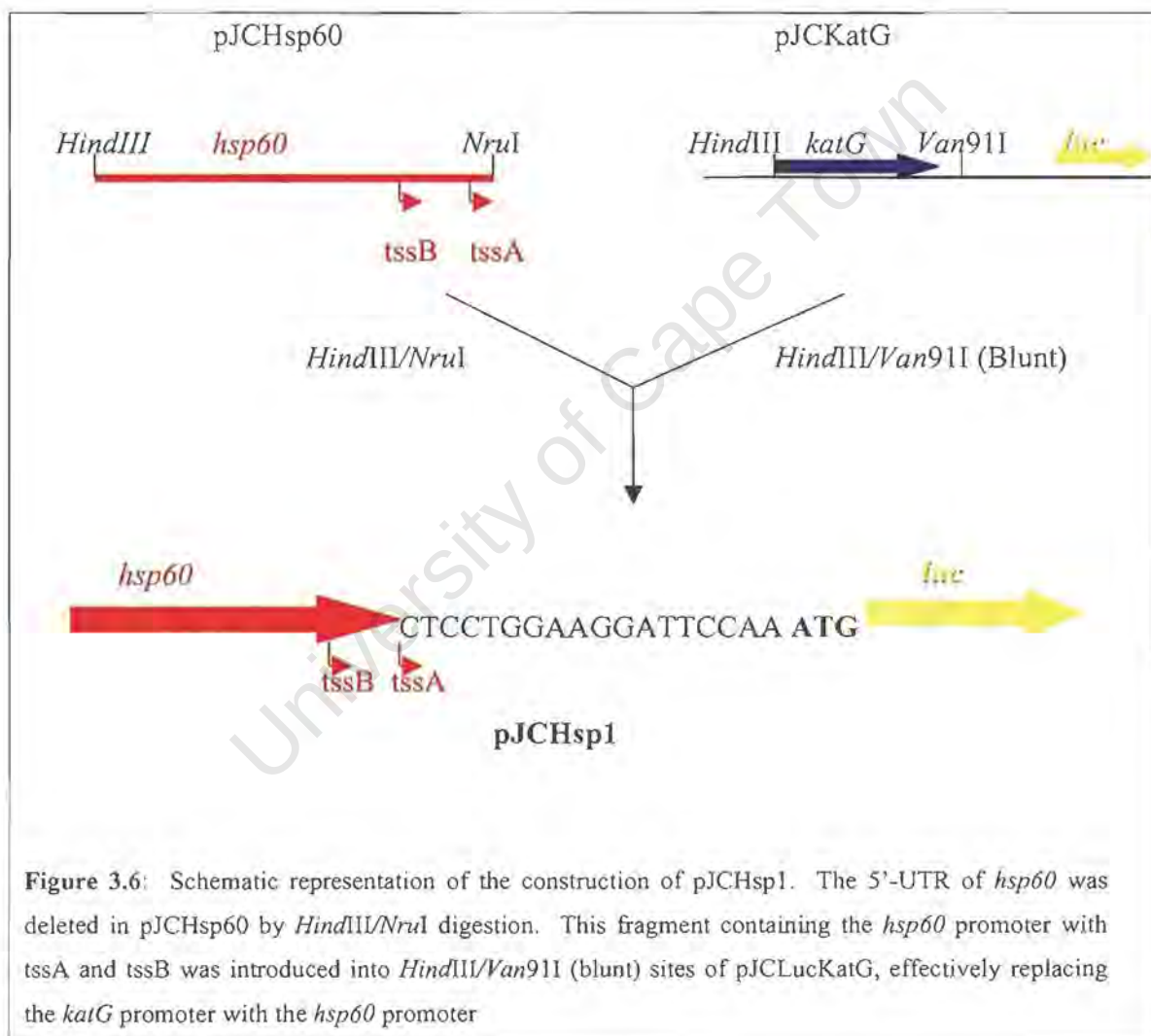


Figure 3.6: Schematic representation of the construction of pJCHsp1. The 5'-UTR of *hsp60* was deleted in pJCHsp60 by *Hind*III/*Nru*I digestion. This fragment containing the *hsp60* promoter with *tssA* and *tssB* was introduced into *Hind*III/*Van*91I (blunt) sites of pJCLucKatG, effectively replacing the *katG* promoter with the *hsp60* promoter

3.4.5 Cloning the *mpb83* promoter into pJCLuc

The MPB83 promoter was amplified by PCR as a 773 bp fragment using p18a as template and MPB83P3 (5'-CTAAGATGCCATGCGTG-3') and MPB83Pup (5'-TTCCTTTGCTTCTGCG-3') as primers (Vosloo PhD. Thesis). This fragment was gel-purified and cloned upstream of the promoterless luciferase gene in pJCLuc (Vosloo PhD. Thesis) creating pJCMPB83. This construct was used in this study to obtain a growth-phase expression profile of MPB83 expression.

3.4.6 Promoter expression analysis

3.4.6.1 *mpb83* expression analysis

The expression profile of *mpb83* was determined. For this pJCMPB83 was introduced into *M. smegmatis* by electroporation, single colonies isolated on solid middlebrook-7H9 medium supplemented with ADC and Kanamycin. Starter cultures were grown to stationary phase in liquid Middlebrook-7H9 medium supplemented with Tween 80, ADC and Kanamycin, which were used to inoculate 200ml liquid cultures for experimental purposes. Luciferase activity was measured at various time points from 16 to 64 hours representing lag, exponential and stationary phases of growth. This data was plotted as luciferase activity (RLU/mg total protein) against time (Figure 3.7). pJCKatG was used as a reference to assess relative promoter expression. From this data it is clear that although the expression profiles are similar, pJCKatG is more than 2 fold more efficient at transcription and translation of the luciferase reporter gene than pJCMPB83. The profile of *katG* expression was similar to that reported by Mulder *et al.* (1999). Maximal expression of both of these promoters occurs at late exponential – early stationary phase. Thereafter there is a rapid loss in luciferase expression indicating little or no further luciferase expression. The exponential loss of luciferase over time from the maximal point would correlate with a cessation in luciferase production and protein degradation half-life studies at 37 °C (section 3.4.7).

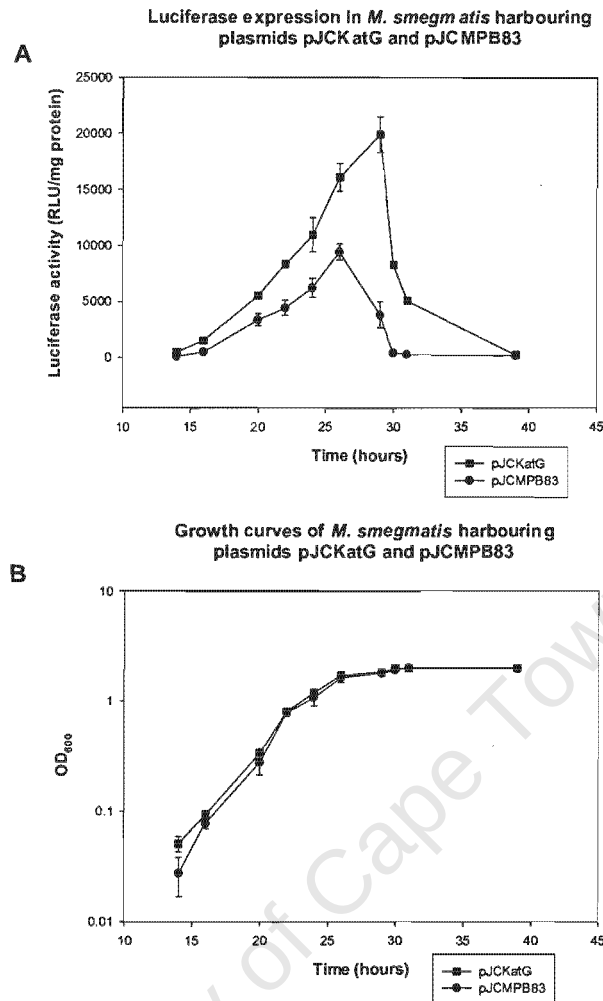


Figure 3.7: A: Luciferase expression in *M. smegmatis* harbouring pJCMPB83 and pJCKatG. B: OD₆₀₀ Growth curves of *M. smegmatis* harbouring pJCMPB83 and pJCKatG

3.4.6.2 *katG* and *hsp60* promoter expression analysis

To maximize the production of antisense molecules and hence the efficiency of such a system of attenuation, it is advantageous to use a powerful promoter. For this reason the expression profiles of 4 promoters were analysed. pJCKatG, pJCKatG2, pJCHsp60, and pJCHsp1 were introduced into *M. smegmatis* as described above. Both *katG* and *hsp60* are complex promoters with 2 transcription start sites each. In the case of *katG*, two different constructs were designed (Figure 3.8).

A pJCKatG

TTCGCGATCACATCCGTGATCACAGCCCGATAACACCAACTCCTGGAAGGATCCAAATG

B pJCKatG2

TTCGTGGAAGGATCCAAATG

Figure 3.8: Sequence of the region from the transcription start site (underlined red) to translation start site (underlined blue) of; A: pJCKatG and, B: pJCKatG2. Ribosome binding sites are underlined in green.

pJCKatG contains the *katG* promoter fused to the luciferase gene such that both transcriptional start sites are active. The downstream transcription start site has been deleted in pJCKatG2. This deletion results in a 50% reduction in luciferase expression (Figure 3.9 A).

Luciferase expression in *M. smegmatis* harbouring plasmids pJCKatG and pJCKatG2 Luciferase expression in *M. smegmatis* harbouring plasmids pJCHsp60 and pJCHsp1

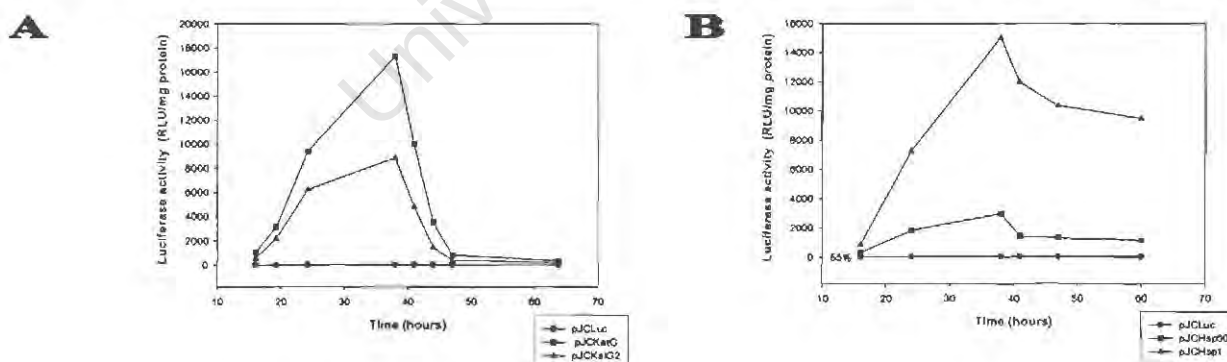


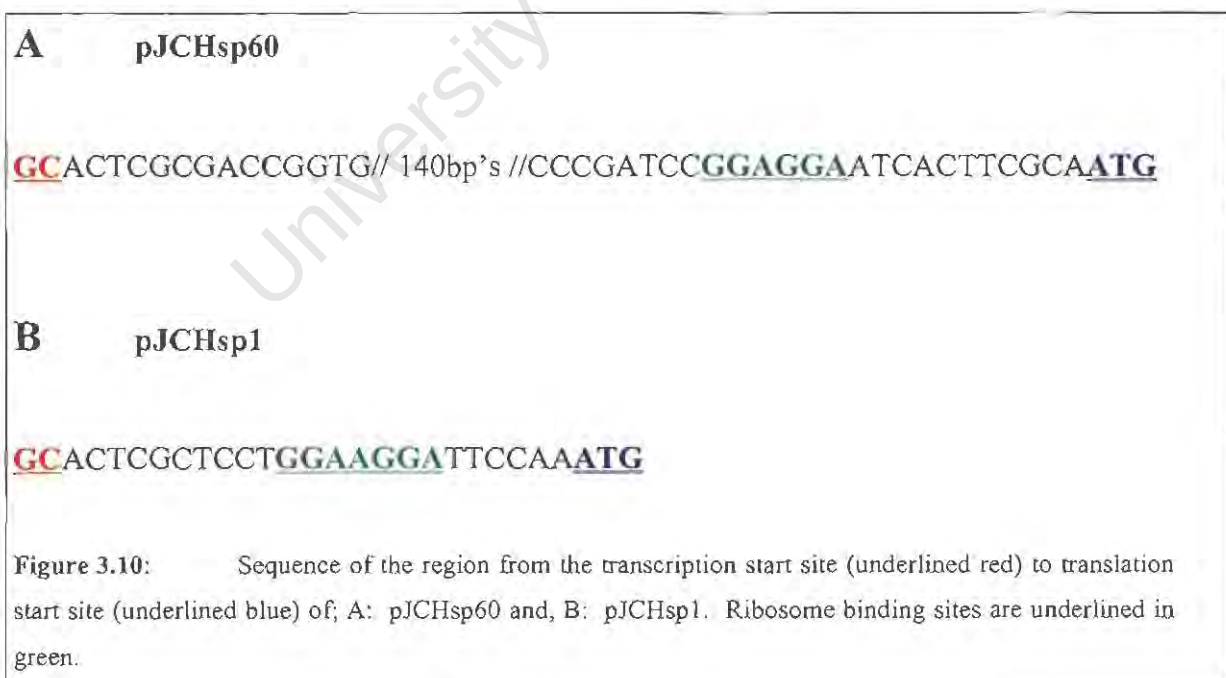
Figure 3.9: Luciferase expression in *M. smegmatis* harbouring; A: pJCKatG and pJCKatG2, B: pJCHsp60 and pJCHsp1.

The advantage of using pJCKatG2 is that it allows for insertion of antisense sequences into the *Bam*HI site, 3bp's downstream of the transcription start site. This

minimizes the possibility of misfolding of the antisense molecules due to long stretches of non-specific 5' mRNA.

In the case of the *hsp60* promoter, two constructs were also designed. pJChsp60 is based on the full length *M. bovis* BCG *hsp60* promoter cloned to the luciferase gene (Figure 3.10). This construct contains the extremely long 5'-UTR from transcription start site A (tssB is inactive in *M. smegmatis*) to the translation start site (183 bp's). This long 5'-UTR has been deleted in pJChsp1. In this construct the 5'-UTR has been reduced to 23 bp's and the distance between the Shine-Dalgarno and the start codon also reduced from 11 bp's in pJChsp60 to 6 bp's (pJChsp1).

Expression profiles of these two constructs produced the following results (Figure 3.9 B). pJChsp60 is active from early exponential phase right through to late stationary phase. The expression levels are relatively low in comparison to *katG* and *mpb83*. When the 5'-UTR is deleted and the distance between RBS and start codon reduced to 6 bp's, expression increases 6.5 fold. The expression profile is however still similar to pJChsp60.



3.4.7 Stability of luciferase in *M. smegmatis*

Interpretation of growth-phase dependent expression profiles using translation fusion vectors is complicated because the final luciferase expression is determined by real time transcription/translation of the *luc* gene in conjunction with previous luciferase degradation. To validate the growth curve luciferase results it was decided to perform functional protein half-life studies on luciferase. This was performed in *M. smegmatis* using rifampicin.

Rifampicin is a semi-synthetic derivative of the naturally occurring rifamycin B, in which the aromatic ring system of the parent antibiotic has been substituted with a 4-methylpiperazinyloxyethyl side-chain resulting in markedly enhanced antibacterial activity. Rifampicin inhibits the growth of Gram-positive bacteria and *M. tuberculosis* at extremely low concentrations. In intact bacteria it selectively inhibits RNA synthesis (Hartmann *et al.*, 1967), and this effect is well correlated with its bactericidal action (Lancini *et al.*, 1969). *In vitro* it is one of the most potent inhibitors of bacterial DNA-dependent RNA polymerase known; the enzyme from *E. coli* is inhibited 50 % by 2×10^{-8} M rifampicin. In complete contrast, the DNA-dependent DNA polymerase I reaction is not affected by rifampicin at concentrations even 4 orders of magnitude higher (Hartmann *et al.*, 1967). Rifampicin inhibits the RNA polymerase reaction by interfering specifically with the initiation process: it has no effect if added after the polymerization process has begun (Sippel and Hartmann, 1968; Umezawa *et al.*, 1968; Mizuno *et al.*, 1968). It does not prevent the binding of the polymerase to its template (Umezawa *et al.*, 1968).

Rifampicin (50 µg/ml) was added to growing cultures of cells, and light output measured. After an initial steady increase in luciferase activity (up to 60 min after addition of rifampicin), a gradual decline in activity was observed (Figure 3.11). Luciferase activity declined with a half-life of 4 hours from time point 60 min to 6 hr 21 min, which is consistent with findings by Sarkis *et al.* (1995). The steady increase in luciferase activity over the first 60 min after addition of the drug could be attributed to either slow or impaired uptake of rifampicin or a highly stable message due to the presence of a highly structured and elongated 5'-UTR transcribed from the *hsp60*

promoter fused to the luciferase gene. The latter of these two possibilities is another factor that has to be taken into account when interpreting results of chimeric promoter/reporter fusion constructs.

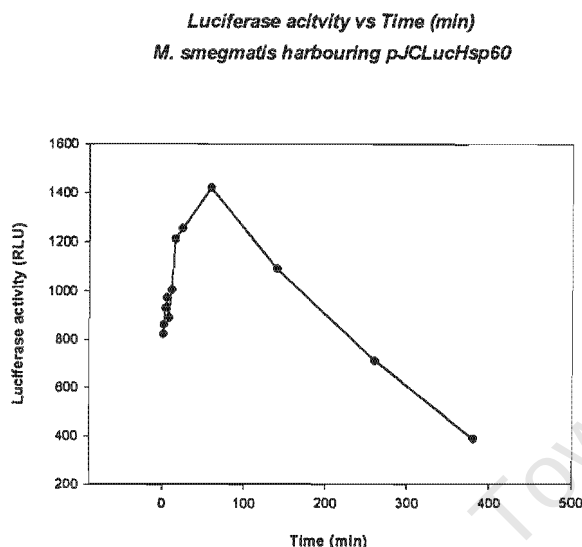


Figure 3.11: Stability of luciferase activity in *M. smegmatis* harbouring pJCLuc60. A mid-log culture was diluted and rifampicin (50 µg/ml) was added to the culture. Luciferase activities were determined immediately (time zero) or at various times after drug addition.

3.4.8 Development of transcription expression cassette

For efficient expression of MPB83 antisense mRNA in mycobacteria, an *E. coli*/mycobacterium shuttle vector was constructed as shown in figure 3.12. This vector is based on pJCKatG (Michelle Mulder, PhD thesis UCT). It contains both *E. coli* and mycobacterial origins of replication, and the *M. tuberculosis* *katG* promoter.

Annealed complementary oligonucleotides coding for the rho-independent *trp* transcription terminator, were cloned into pUCBM21 (Roche) digested with *HindIII/SphI*, creating pUC21Oligo17. This construct was used to subclone *Sau3AI* digested DNA into the unique *BclI* restriction site.

The gene encoding MPB83 was isolated from a λgt11 BCG genomic library by screening recombinant plaques with Mab3C4. A positive clone with an insert of 1.8 kb was identified and was sub-cloned as a *KpnI* fragment in pUC19 to generate the

the recombinant plasmid pAT100B (prepared and obtained by P.Tippoo, our laboratory). This fragment consists of 69 bp of upstream region, the MPB83 coding sequence and 1090 bp of downstream region (Figure 3.13). The coding region of MPB83 as well as upstream (69 bp) and downstream (130 bp) was released from pAT100B by *KpnI*/*SphI* digestion.

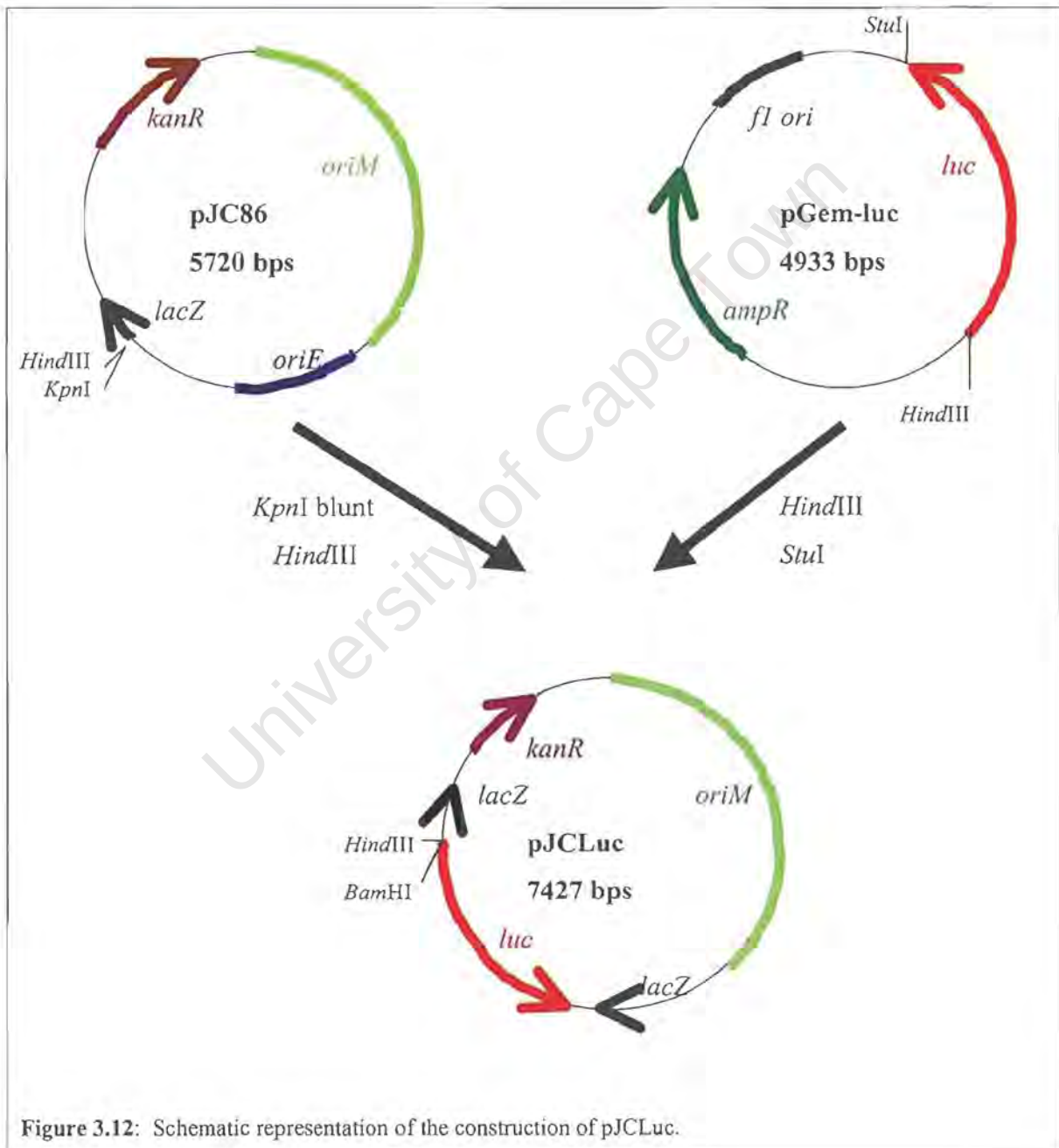


Figure 3.12: Schematic representation of the construction of pJCLuc.

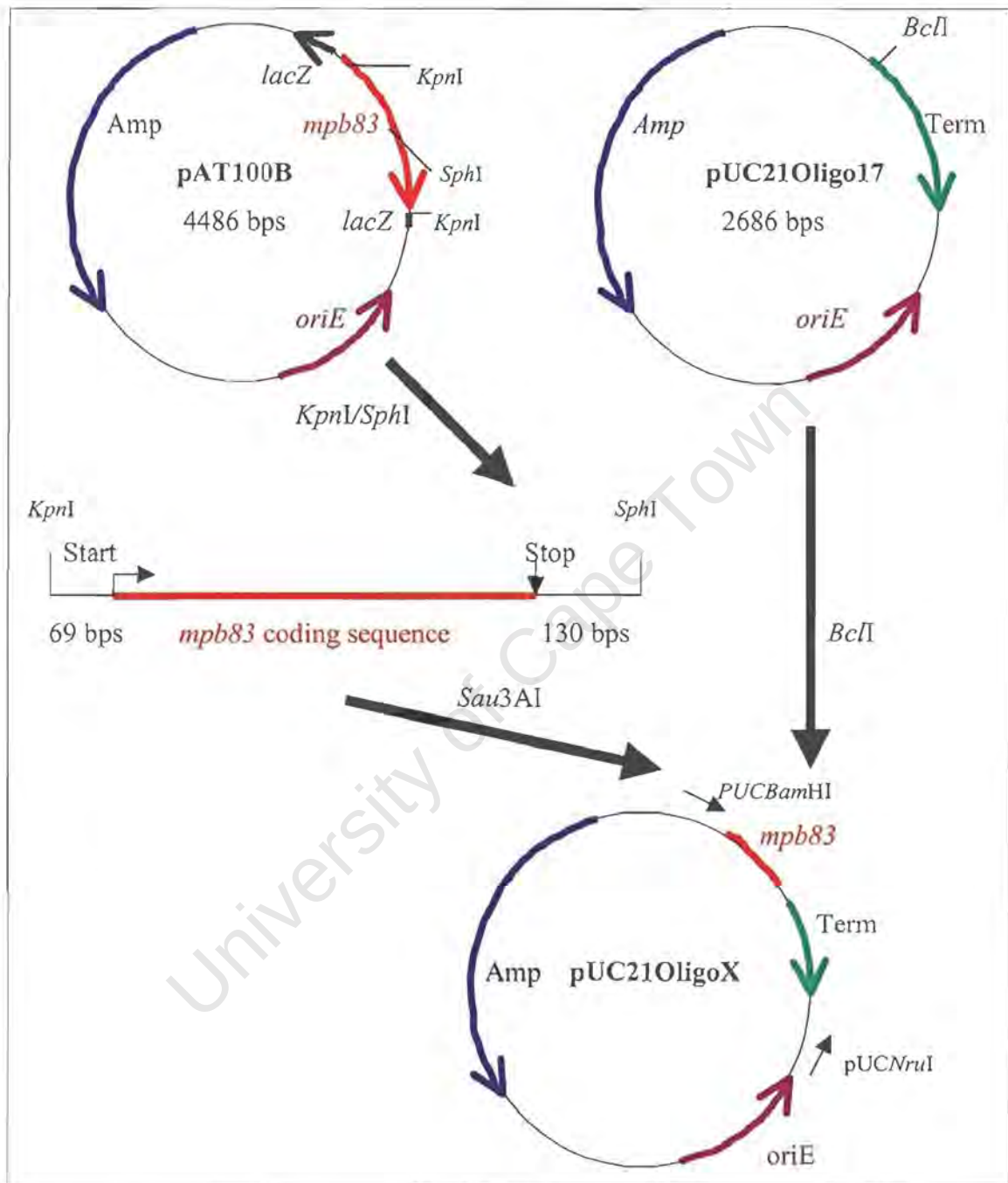


Figure 3.13: Schematic representation of the construction of pUC21Oligo17-5, pUC21Oligo17-36, pUC21Oligo17-24, pUC21Oligo17-25. These constructs differ in the region and orientation of *mpb83* *Sau3AI* insert. The *KpnI/SphI* *mpb83* fragment was removed from pAT100B and digested with *Sau3AI*. DNA corresponding to different regions of *mpb83* was ligated upstream of the transcriptional terminator (*Term*) in pUC21Oligo17, previously digested with *BclI*.



This fragment was purified by phenol extraction after separation by electrophoresis through a 1% TAE agarose gel. This was followed by digestion with *Sau3AI* for 2 hrs at 37° C, followed by heat inactivation of the restriction enzyme at 65° C for 30 min. pUC21Oligo17 was prepared for ligation with MPB83 *Sau3AI* fragments by digestion with the restriction enzyme *BclI*. For this pUC21Oligo17 had to be transformed into the *dam*⁻ *E. coli* strain, JM110. This renders the plasmid pUC21Oligo17 sensitive to *BclI* digestion at 50° C. Recombinant plasmids were verified by restriction analysis and sequence confirmation.

Constructs with sequences corresponding to both sense and anti-sense orientation of two different regions of the *mpb83* gene were selected for further analysis. Clones pUC21Oligo17-5 and pUC21Oligo17-36 correspond to region 1 of the MPB83 coding sequence (Figure 3.14) in the sense and anti-sense orientation respectively. Clones pUC21Oligo17-24 and pUC21Oligo17-25 correspond to region 2 of MPB83 (Figure 3.14) in the sense and anti-sense orientation respectively.

3.4.9 PCR amplification of *mpb83* and cloning into pJCKatG

MPB83 constructs were subcloned from pUC21OligoX (Figure 3.13) into the transcription vector pJCKatG for expression in *M. bovis*. For this primers pUC*Bam*HI (5'-gctggatCCATGATTACG-3') and pUC*Nru*I (5'-gttgatCGCGACGG-3') were used to amplify *mpb83* and transcription terminator sequences from pUC21Oligo17-5;36;24;25 clones. The lower case nucleotides indicate the mismatches, and the underlined nucleotides represent the restriction endonuclease sites for *Bam*HI and *Nru*I, respectively. PCR amplification was performed as described in Appendix B, using plasmid DNA (1 pg) as the template, and an annealing temperature of 53 °C. Control reactions included those with no template DNA or primers, or those with single primers only, to ensure that the PCR products were not the result of non-specific binding of primers, or due to contaminating DNA. PCR products were separated by agarose gel electrophoresis, stained with ethidium bromide, and viewed under 310 nm UV transillumination. The PCR products were excised from the agarose gel, purified by the phenol squeeze method described in Appendix B, and

digested to completion with the restriction endonucleases *Bam*HI and *Eco*CRI. The resulting fragments were cloned into the compatible restriction endonuclease sites of pJCKatG, to yield: pJCKatG-MPB83-5, pJCKatG-MPB83-36, pJCKatG-MPB83-24, and pJCKatG-MPB83-25. The sequences of the inserts were verified by sequencing (Figure 3.15). Cloning the fragments using the *Nru*I/*Bam*HI restriction sites results in the expression of antisense molecules with the same expression profile as pJCKatG2. This results in higher expression of luciferase than pJCMPB83. It also has the advantage of minimal 5' sequences that could interfere with secondary folding of the *mpb83* transcripts after transcription.

pJCKatG-MPB83-5,24,25,36 sequence

A

katG promoter sequences -35 -10 TSS pUCNruI MPB83 S/AS

GGACGAGGCGGAGGTCATCTACTGGGGTCTATGTCCTGATTGTTCGATATCCGACACTTCGATGTGGCGATGAACGCTCTGGATC-----
 -----GATCAGCCCGCCTAATGAGCGTTTTTTTTTAAAGCTTGGCGTAATCATGGATCCAAATGGAAGACGCCAAAAACATAAAAG

Terminator sequences pUCBamHI LucI

B

pJCKatG-mpb83-5
Region 1 sense strand

GATCAACGTTTCAAGCCAAACCGGCCGAGCAGCGAGCCTCGCAGCCATCGCGATTGCGTTCTTAGCGGGTTGTTTCGAGCACCAAACCGGTGTCGCAAGACACCAGCCCGAAACCGGCGACAGCCCGCGCGCCCGTTACACGGCGGAATGGCTGACCCCGCAGCGGACCTGATTGGTCGTGGGTGCGCGCAATACGCGCGCAAAATCCCAACCGTCCCGGATC

pJCKatG-mpb83-36
Region 1 antisense strand

GATCCGGGACCGGTGGGATTTTGCGCCCGTATTGCGCGCACCCACGACCAATCAGGTCCGCTGCGGGGTCAAGCAATTGCCGCCGTGGTAACGGGCGCCCGCGGGCTGGTCGCCGGTTTCGGGCTGGTGCTTGCGACACGGTTTGGTGCTCGAACAACCCGCTAAGAACGCAATCGCGATGGCTGCGAGGCTCGCTGCTGCGGCCGTTTGGCCTGAACGTTGATC

pJCKatG-mpb83-24
Region 2 sense strand

GATCGGTGGCCGAATGGCGCAAGACCCGGTCGCTACCGCGGCTTCCAACAACCCGATGCTCAGTACCCTGACCTCGGCTCTGTCCGGCAAGCTGAACCCGGATGTGAATCTGGTCGACACCCTCAACGGCGCGAGTACACCGTTTTCGCCCCACCAACGCCGATTGACAAAGCTGCCGCGGGCCACTATCGATC

pJCKatG-mpb83-25
Region 2 antisense strand

GATCGATAGTGGCCGCCGGCAGCTTGTGCAATGCGGCGTTGGTGGGGGCGAAAACGGTGTAAGTCTCGCCGCCGTTGAGGGTGTGACCAAGATTACATCCGGGTTACAGTTGCCCGACAGAGCCGAGGTCAGGGTACTGACATCGGGTTGTTGAAGCCCGGTAGCGACCGGGTCTTGCGCCATTCCGGCCACCGATC

Figure 3.15: **A:** Single stranded DNA sequence of the transcription vector, pJCKatG, containing *mpb83* sequences. The -35 and -10 regions of the *katG* promoter are shown in red. The transcription start site is indicated (TSS). PCR cloning primers pUCNruI (purple) and pUCBamHI (green) were also used for sequencing purposes. Transcription termination sequences are shown in brown. **B** *mpb83* sequences corresponding to region 1 (figure 5.10) of *mpb83* in sense, pJCKatG-mpb83-5, and antisense, pJCKatG-mpb83-36 orientations. Region 2 of *mpb83* (Figure 3.14) was also cloned in the sense, pJCKatG-mpb83-24, and antisense orientations, pJCKatG-mpb83-25.

3.5 Conclusions

In this chapter I have described the construction of a vector capable of expressing both sense and antisense *mpb83* RNA. The vector utilizes the *M. tuberculosis katG* promoter that has been cloned in such a way that transcription starts 10 bases upstream of the antisense sequences. Transcription is terminated immediately after antisense sequence by the *trp* transcription terminator. The *M. tuberculosis katG* promoter was chosen to transcribe the antisense sequences. Transcription start site A of this promoter was deleted to minimize 5'-UTR sequences that might interfere with secondary structure folding of antisense RNA. To do this, the promoter was truncated and fused to the firefly luciferase gene. As expected, levels of translation dropped by approximately 50% when compared with a promoter construct containing both transcription start sites, pJCKatG. Translation levels from pJCKatG2 indicated higher levels of translation than that obtained with the *mpb83* promoter. The expression profiles of both of these promoters were similar in that they are both active in exponential phase of growth.

Direct comparison of promoters using fusion vectors is impossible to achieve. I used a system whereby the translation machinery was identical for all promoters tested. This system still requires knowledge of how translation can be affected by a host of factors often introduced during cloning and overlooked in subsequent interpretation of data. The next section is a brief review of some of these factors.

3.5.1 Effect of secondary structure at the site of ribosome attachment

The efficiency of translation in *E. coli* is determined primarily at the stage of initiation, which includes binding of the 30S ribosome to 5' segments (from approximately nt -20 to +15) of mRNA spanning the SD sequence and the initiation codon (Draper 1996). Unlike transcription where promoters are usually unaffected by the surrounding nucleotide content (Harley and Reynolds, 1987; Hawley and McClure, 1983; Kobayashi *et al.*, 1990; Lissner and Margalit, 1993), it is now clear

that the wide range of efficiencies in the translation of different mRNAs is predominantly due to the unique structural features at the 5' end of each mRNA species. Thus, in contrast to the portable promoters, no universal sequence for the efficient initiation of translation has been devised.

The secondary structure at the translation initiation region of mRNA plays a crucial role in the efficiency of gene expression (de Smit and van Duin 1990b; Hall *et al.*, 1982; Hartz *et al.*, 1991; Iserentant and Fiers, 1980; Kastelein *et al.*, 1983). It is believed that the occlusion of the SD region and/or the start codon by stem-loop structure prevents accessibility of the 30S ribosomal subunits and inhibits translation (de smit and van Duin, 1994; McCarthy and Brimacombe, 1994; Nagai *et al.*, 1991; Yuzawa *et al.*, 1993).

De Smit and Van Duin (1994) presented evidence that translational initiation on a structured messenger is the result of a direct competition between the intramolecular and intermolecular base pairing (secondary structure and SD interaction). Using this competition, they obtained quantitative data on the way in which SD mutations affect ribosomal affinity and expression (Figure 3.16). Mutational studies with the ribosome-binding site of the bacteriophage MS2 coat protein indicated that a strong SD interaction might compensate for the unfavorable influences of intramolecular mRNA secondary structures (de Smit and van Duin, 1994).

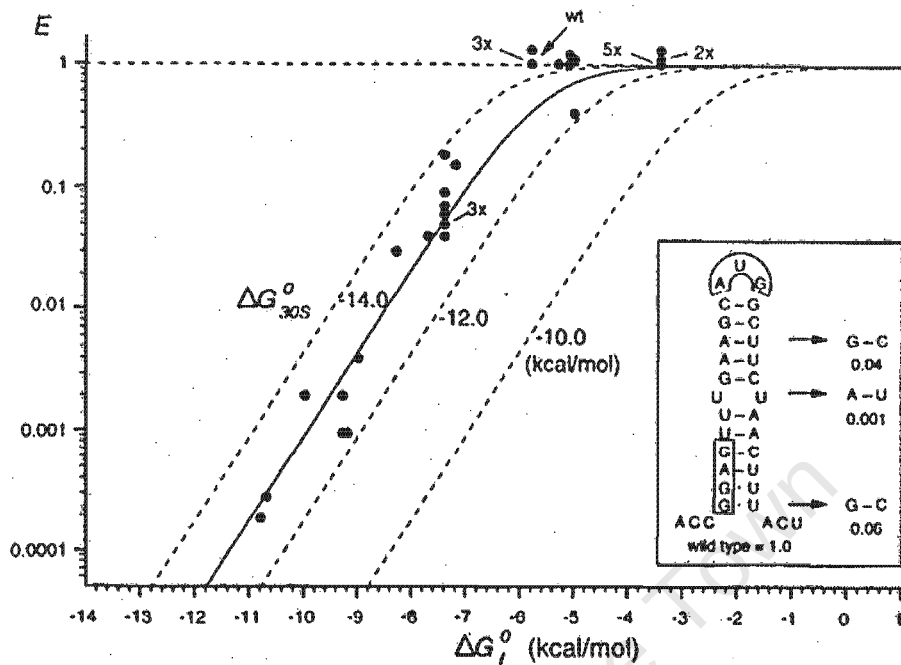


Figure 3.16. Graphic representation of the quantitative relationship between expression (E) and the stability of the secondary structure of the ribosome binding site (ΔG°_I). Filled circles indicate experimental data, obtained by site-directed mutagenesis of the coat-protein gene from bacteriophage MS2; where relevant the number of coinciding data points is given. Broken curves show the theoretical relationship for three ribosome-mRNA affinities (ΔG°_{30S}). A similar, continuous curve was fitted to the experimental data. ΔG°_{30S} values are for a concentration of free 30 S subunits of $8.5\mu\text{M}$. The insert shows the wild-type structure at the coat-protein start, as well as several base-pair changes and the corresponding expression level relative to the wild-type. The initiation codon and the SD sequence are boxed (de Smit and van Duin, 1994).

The naturally occurring phenomenon of translational coupling in bacteria (Schumperli *et al.*, 1982) is an example of this. Translational coupling is able to coordinate expression of different proteins from polysistronic mRNAs. It was shown that the moderately strong *gal* promoter could direct the synthesis of galactokinase at very high levels when *galK* was translationally coupled to an upstream gene, suggesting that even a weak RBS may be highly efficient if it is accessible to ribosomes (Schumperli *et al.*, 1982).

In the case of *rpoH*, part of the ribosome binding site (nt +1 to 20) including the AUG codon and downstream box was thought to be masked through the formation of base pairs with the internal region. This idea was initially suggested by computer prediction (Nagai *et al.*, 1991) and subsequently supported by mutational analyses (Nagai *et al.*, 1991; Yuzawa *et al.*, 1993) and structural conservation among the *rpoH* homologs (Nakahigashi *et al.*, 1995). Such a structure seemed most likely to restrict translation by preventing ribosome entry under non-stress conditions and was supported by further deletion analyses (Morita *et al.*, 1999).

3.5.2 Effect of secondary structure on mRNA decay

If a particular mRNA has a very short half-life, it can only be translated a limited number of times before it will be functionally inactivated. This result allows for another level of gene regulation beyond transcription initiation and translation initiation. Several different RNases participate in mRNA degradation in *E. coli*, including endonucleases (RNase E, RNase K, and RNase III) and 3' exonucleases (RNase II and polynucleotide phosphorylase [PNPase]); no 5' exonuclease has been identified to date in prokaryotes (Belasco 1996). In *E. coli*, it has been shown that the half-lives of various mRNA transcripts can vary between 40s to 20 min (Kushner 1996). Clearly there are many different factors that can influence the stability of a transcript and these can be *cis*- or *trans*-acting.

- | | |
|------------------------------|---|
| <i>cis</i> factors: | 1: Secondary structures that retard or increase decay; |
| <i>trans</i> factors: | 1: Exo- and endonucleases that lead to phosphodiester bond cleavages; |
| | 2: Proteins that lead to slower decay |

Here is a brief overview of a few of the factors that could influence the half-life of the transcripts in translation fusion vectors:

3.5.2.1 mRNA length

As mentioned above some promoters have extended 5'-UTR eg., *hsp60* (Stover *et al.*, 1991) whilst others have short 5'-UTR eg., *katG* (Mulder *et al.*, 1999). This could lead to large differences in reporter construct length (if the 5'-UTR is included in the cloning). Transcript length is however not a good indicator of mRNA stability. For example:

<i>lpp</i> mRNA 325 nt	Half life = 15-17 min (Hirashima <i>et al.</i> , 1973)
<i>trxA</i> mRNA 293 nt	Half life = 2-3 min (Arraiano <i>et al.</i> , 1988)
<i>ompA</i> mRNA 1215 nt	Half life = 15 min (Von Gabain <i>et al.</i> , 1983)

It is clear that the rate determining step in mRNA decay is not transcript length and indeed very unpredictable.

3.5.2.2 5' Terminus

Since there is no evidence of 5' – 3' exonucleases or 5' capping in *E. coli*, one would assume that the 5'-end of an mRNA would not be important for stability. However, analysis of the *ompA* gene clearly indicates that the presence of a short 5'-UTR can dramatically affect mRNA stability. In this case the *ompA* gene has a half-life of approximately 15-17 min if the 5'-UTR is left intact (Von Gabain *et al.*, 1983; Emory *et al.*, 1992) and 3.6 to 4.4 min if it is removed (Emory *et al.*, 1992). Moreover, if the 5'-UTR is transferred to the 5' end of a short-lived message such as *bla*, the half life of the chimeric mRNA increases from 3-4 min to >13 min (Belasco *et al.*, 1986). It was proposed that *E. coli* mRNAs beginning with a long single-stranded segment are preferentially targeted by an RNase that interacts with the 5' terminus before cleaving the mRNA at internal sites (Emory *et al.*, 1992). It appears, therefore, that the addition of *ompA* 5' stabilizer to heterologous genes might enhance gene expression in *E. coli*. However, it is possible that mRNAs, which contain internal RNA-processing sites, are not protected by the presence of the 5' stabilizer (Di Mari and Bechhofer 1993).

The 5'-UTR of the bacteriophage T4 gene 32 increases the half-life of unstable mRNAs in *E. coli* but only in T4-infected cells (Duvoisin *et al.*, 1986). The erythromycin resistance genes (*erm*) of gram-positive bacteria such as *Staphylococcus aureus* and *Bacillus subtilis* encode mRNAs that contain stabilizing elements in their 5'-UTRs. However stabilization of mRNAs by the *ermC* and *ermA* 5'-UTR is induced by an antibiotic that inhibits translation and causes ribosome stalling (Bechhofer and Dubanau 1987; Sandler and Weisblum, 1988; Sandler and Weisblum, 1989)

There is also evidence suggesting that stem-loop present in the 3'-UTR can provide protection against exonucleolytic degradation (Wong and Chang, 1986).

3.5.2.3 Higher-order RNA structure

It is not clear whether specific nucleotide sequences, comparable to either type I or type II DNA restriction sites, for example, play a significant role in the decay of mRNAs (Kushner 1996). Thus cleavage site specificity for endonucleases involved in mRNA decay most likely is very dependent on secondary and tertiary structures. The apparent absence of traditional 5'- to 3' exonucleases – enzymes that degrade RNA 1 nucleotide at a time – in *E. coli* makes the importance of 5' termini as discussed in the case of *ompA* mRNA very intriguing. Possibly there is an exo-ribonuclease that requires a 5' terminus to bind but cleaves the RNA at a distance from the binding site. The Rec BCD enzyme is an example of a DNase that functions in this way (Dixon and Kowalczykowski, 1993).

Chapter 4

Attenuation of MPB83 expression in *M. bovis*

Contents	page
4.1 Summary	73
4.2 Introduction	73
4.2.1 Aims	74
4.3 Materials and Methods	74
4.3.1 Bacterial strains and plasmids	74
4.3.2 Monoclonal antibody	75
4.3.3 Electroporation	75
4.3.4 SDS-PAGE and immunoblot analysis of protein	76
4.3.5 Band intensity quantitation	77
4.4 Results	77
4.4.1 Attenuation of MPB83 expression by antisense constructs	77
4.5 Conclusions	79

Chapter 4

Attenuation of MPB83 expression in *M. bovis*

4.1 Summary

Antisense constructs from the previous chapter were electroporated into *M. bovis* BCG (Tokyo) and MPB83 levels monitored by Western blotting using an MPB83-specific monoclonal antibody, Mab3C4. Attenuation levels of 21% and 32% were achieved with antisense fragments directed against region 1 and region 2 of *mpb83* respectively. The same constructs in the sense orientation resulted in no significant attenuation in MPB83 levels.

4.2 Introduction

A brief review to natural and artificial control of gene expression by antisense methods was given in chapter 2. In this chapter the objective is to artificially attenuate MPB83 expression by using antisense molecules directed to *mpb83* mRNA. The most common method of achieving antisense control has been by using the entire gene in the opposite orientation (Parish and Stoker, 1997). Here we use approximately 200 bp portions of the *mpb83* gene rather than the entire gene in an attempt to find conserved structures on the *mpb83* mRNA that can be preferentially targeted.

Most naturally occurring antisense RNAs display a high degree of secondary structure with stem-loops as their most predominant feature. In *copA/copT* and RNA I/RNA II, binding of the antisense molecule occurs in at least two experimentally distinguishable steps: formation of a transient “kissing complex” formed between complementary loop nucleotides, followed by a subsequent nucleation step that initiates complete RNA duplex formation (Hjalt and Wagner, 1992). It is clear that for efficient antisense binding to occur, primary binding must take place at loop

regions between target and antisense RNA. The loops involved often consist of 5-7 unpaired nucleotides with internal loops and bulged-out nucleotides often found in close proximity.

MPB83 DNA was digested with the 4 bp restriction enzyme, *Sau3AI*, resulting in fragments of approximately 200 bps. These fragments were used to target *mpb83* mRNA. Attenuation efficiency was determined by Western blotting using an anti-MPB83 monoclonal antibody, Mab3C4 (Vosloo *et al.*, 1997). For this experiment we used the antisense RNA cassette constructed in chapter 3.

4.2.1 Aims

This chapter aims to attenuate MPB83 expression in *M. bovis* BCG (Tokyo) using ± 200 bp antisense fragments of *mpb83*. Antisense constructs were introduced into *M. bovis* BCG (Tokyo) as an episomal plasmid. MPB83 levels were assayed at various time intervals during growth by Western blot analysis using an MPB83 specific monoclonal antibody.

4.3 Materials and Methods

4.3.1 Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are summarized in table 4.1. For all cloning steps, LKIII (Zabeau and Stanley, 1982) was used. *M. bovis* BCG (Tokyo) strain 172 (State Vaccine Institute, Pinelands, Cape Town, South Africa) was used for MPB83 attenuation experiments. Details of the commonly used techniques and solutions are given in Appendices B and C, respectively.

Table 4.1
Constructs prepared for this study

Bacterial strains and plasmids used in this study		
Bacterial strains	Genotype	Reference
<i>E. coli</i> LKIII	<i>recA, lacI, lacZ, lacY⁺</i>	Zabeau and Stanley, 1982
<i>M. bovis</i> BCG (Tokyo) strain 172	Attenuated vaccine strain 172	State Vaccine Institute, Pinelands, S.A.
Plasmid	Key Features	Technical Notes
pJCKatG-MPB83-5	PJCKatG harbouring region 1 of <i>mpb83</i> in the sense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-36	PJCKatG harbouring region 1 of <i>mpb83</i> in the antisense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-24	PJCKatG harbouring region 2 of <i>mpb83</i> in the sense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-25	PJCKatG harbouring region 2 of <i>mpb83</i> in the antisense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>

4.3.2 Monoclonal antibody

Mab3C4 previously raised against *M. bovis* BCG Tokyo sonicates (Vosloo *et al.*, 1997), shown to belong to isotype IgG_{2a} and specific to MPB83 was used for immunoblot experiments.

4.3.3 Electroporation

M. bovis BCG Tokyo strain 172 cells were prepared for electroporation by the method of Jacobs *et al.* (1991). A stationary phase culture of *M. bovis* BCG Tokyo, in Middlebrook-ADC-Tween, was diluted 100-fold into fresh medium, and grown to

exponential phase at 37°C with shaking. The cells were harvested by centrifugation, washed 3-4 times in ice-cold 10% (v/v) glycerol, and resuspended in a final volume of 2ml/100ml original culture volume in 10% (v/v) glycerol. For electroporation, 60µl cells were combined with 140µl of 10% (v/v) glycerol and 1µg of DNA, in a 5µl volume, and incubated on ice for 5 minutes. Electroporations were performed in a 0.1 cm gap Gene Pulser® cuvettes (Bio-Rad Laboratories, CA, U.S.A) at 1.8kV, 25µF and 1000 Ω, using a Gene Pulser™ apparatus (Bio-Rad Laboratories, CA U.S.A.), which resulted in a field strength of 18.0 kV/cm. After electroporation, 900µl of Middlebrook-ADC was added, and the cells were incubated at 37°C for 3-4 hours to facilitate expression of the kanamycin resistance gene. Transformants containing the recombinant plasmids were selected on Middlebrook-ADC agar plates containing kanamycin at a concentration of 50 µg.ml. A control electroporation with no added DNA was included to monitor the frequency of spontaneous mutation to kanamycin resistance. After incubation at 37°C for ten days, transformants were cultured in liquid medium for plasmid preparations.

4.3.4 SDS-PAGE and immunoblot analysis of protein

Stationary phase cultures of *M. bovis* BCG (Tokyo) strain 172 transformants harbouring pJCKatG-MPB83-5, pJCKatG-MPB83-36, pJCKatG-MPB83-24, pJCKatG-MPB83-25, and pJCLuc were diluted 100-fold into middlebrooks 7H9 (Difco Laboratories) supplemented with ADC (see Appendix C), 0.05% Tween 80 and kanamycin, and incubated at 37 °C with shaking. After 13, 20, and 30 days post inoculation, corresponding to early exponential, late exponential and stationary phases respectively, 10 ml sample was removed, and the cells harvested by centrifugation (13000 rpm, 5 min), washed with 10 ml PBS (pH7.2, see Appendix C) and then resuspended in 1 ml PBS (pH7.2). The bacterial cells were lysed using a Fast-Prep FR120 APPARATUS (Bio101, Savant Instruments) (2X40 second cycles at speed setting 6, cooling on ice 5 min between cycles). Cell debris was removed by centrifugation (13000 rpm, 10 min). Protein concentrations were determined using the Lowery colorimetric assay (Bio-Rad DC protein assay kit), using dilutions of bovin serum albumin (1.5 mg/ml – 0.180mg/ml) as standards. All samples were heated to boiling for 5 min, loaded on a 10% (w/v) SDS-polyacrylamide gel, and

separated by electrophoresis at a constant current of 40 mA (see Appendix B). Proteins were visualized by staining with coomassie® brilliant blue R250 (Bio-Rad Laboratories). For immunoblots, the proteins in unstained gels were transferred onto an Amersham C-super nitrocellulose membrane (see Appendix B). Immunoblots were performed using the chemo-luminescence blotting substrate (POD) (Roche), according to the manufactures instructions. The membrane was blocked for 60 min in blocking solution (supplied in the kit) and diluted to 1% in TBS (see Appendix C). All incubations and washes were performed at room temperature with gentle agitation. Mab3C4 was diluted 1:1000 in 0.5% blocking solution (supplied in the kit) and incubated with the membrane for 60 min. The membrane was washed for 5 min with 0.5% blocking solution. The secondary antibody, POD-labeled anti-mouse IgG (Roche) was diluted 1:5000 in 0.5% blocking solution and incubated with the membrane for 30 min. Following four washes of 10 min each with TBST (see Appendix C), the protein bands were visualized using the detection solution supplied in the kit. Excess fluid was drained from the membrane and the detection solution was added to the membrane for 60 sec. The membrane was sandwiched between X-ray film (Agfa CP-BU), and after a 60 sec exposure, the film was developed. Longer or shorter exposures were performed as necessary.

4.3.5 Band intensity quantitation

Results of Western Blot experiments were analysed by photographing X-ray negatives with a Kodak DC 120 camera. The images were analysed with 1D Image Analysis Software (Kodak Digital Science™).

4.4 Results

4.4.1 Attenuation of MPB83 expression by antisense constructs

Transcription vectors; pJCKatG-mpb83-5, pJCKatG-mpb83-36, pJCKatG-mpb83-24, pJCKatG-mpb83-25, containing portions of *mpb83* in either the sense or antisense orientation, were electroporated into *M. bovis* BCG Tokyo. After growth on solid

middlebrook-7H10 medium supplemented with ADC in the presence of kanamycin, single colonies were used inoculate liquid medium containing Middlebrook-7H9, ADC, Tween 80 and kanamycin. Bacterial cell were extracted at various time intervals for protein extraction and purification. Protein samples were separated on polyacrylamide gels and immunoblotted according to section 4.3.4.

Results indicate a reduction in levels of MPB83 in bacteria harbouring plasmids expressing *mpb83* in antisense orientation compared with those expressing *mpb83* in either sense orientation or *M. bovis* BCG (Tokyo) without plasmid (figure 4.1).

This experiment was performed in *M. bovis* BCG (Tokyo) and not *M. smegmatis* because the latter does not express MPB83. The three time intervals chosen for sampling; day 13, day 20 and day 30 correspond to early exponential, mid exponential, and stationary phases respectively. At the first time point (day 13), no significant attenuation of MPB83 had taken place between cultures harbouring the various control and antisense plasmids. This is expected as the antisense molecules are under the control of the chimeric promoter contained in pJCKatG2, and the expression profile of this construct indicates that expression only occurs well into exponential phase.

At the second time point (day 20), the effect of antisense constructs: pJCKatG-MPB83-36 and pJCKatG-MPB83-25 becomes evident. This effect is further enhanced at day 30 where the attenuation of pJCKatG-MPB83-36 is 21.3 ± 5.9 % and pJCKatG-MPB83-25 is 32.3 ± 7.3 %. The sense control constructs; pJCKatG-MPB83-5 and pJCKatG-MPB83-24 did not have an effect on the levels of MPB83 expression.

These results taken in conjunction with the expression profiles of both MPB83 and pJCKatG2 indicate that attenuation of MPB83 is possible using ± 200 bp antisense fragments of *mpb83*. The reason why antisense fragments corresponding to region 2 of the *mpb83* gene are more effective than those from region 1 in attenuating MPB83 expression could be as a result of more favorable/stable secondary structure of both the antisense mRNA and the target mRNA in that region.

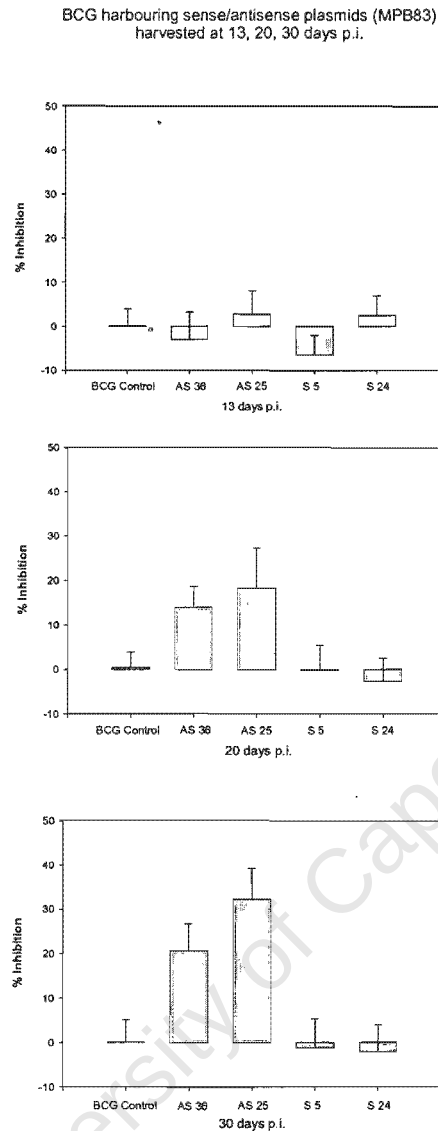


Figure 4.1: Attenuation effect of sense (S) and antisense (AS) molecules on MPB83 expression in *M. bovis* BCG Tokyo as measured by Western Blotting and subsequent signal quantification. Results are presented as % inhibition of MPB83 normalised to BCG control.

4.5 Conclusions

Attenuation of gene expression in Mycobacterial systems is a vital tool for our understanding of gene function. This is commonly performed by gene knock out via targeted gene replacement. This is very useful for the study of non-essential genes but not for genes that are essential to the survival of the organism. One method of overcoming this problem is to down-regulate gene expression without mutating the gene by expression of the antisense strand. Two studies have been reported in

mycobacteria where the entire gene in antisense orientation has been used for this purpose. In the first, phenotypic inactivation of the *M. smegmatis* *hisD* gene was reported by Parish and Stoker (1997). They used the entire *hisD* gene cloned downstream of the inducible acetimidase gene of *M. smegmatis*. When this construct was induced *in vivo*, the bacteria became phenotypically histidine auxotrophs. This auxotrophy was restored by histidine supplementation. Auxotrophy was not observed under non-induced conditions. In the second study, Wilson *et al.* (1998) demonstrated a down-regulation of AphC, an oxidative-stress defence protein involved in isoniazid resistance, by using antisense RNA constructs of *ahpC* under the control of different strength promoters.

The work performed in this chapter has concentrated on using portions of *mpb83* in the antisense orientation. Results reveal attenuation of 21% and 32 % in regions corresponding to region 1 and region 2 of *mpb83* respectively. The same constructs in the sense orientation resulted in no significant attenuation of MPB83 levels. These results indicate that by utilizing portions of *mpb83*, it is possible to attenuate expressing of MPB83 and perhaps even differentially regulate expressing depending on which region is targeted. Understanding why region 2 in the antisense orientation is more efficient at blocking MPB83 translation than region 1 could help us to achieve a system of predictable differential attenuation.

One possibility could lie in the secondary folding characteristics of the target and antisense molecules. Secondary structure prediction is an extremely complicated area of study. Structures are the result of a series of thermodynamic calculations (Zuker 1994), the most stable of which is the most probable to predominate. These structures are however putative due to the numerous factors involved in folding dynamics not included in the calculation such as interaction with alternate RNA/protein molecules.

In an attempt to explain the differences attenuation efficiencies between MPB32-36 and MPB83-25, the most energetically favorable structures for the target transcript, *mpb83* (Figure 4.2), as well as the two antisense transcripts was determined (Zuker 1994)(Figure 4.3). The target mRNA transcript is particularly difficult to determine since the transcription termination site is unknown. We do however know the transcription start site (Vosloo *et al.*, 1998). The antisense transcripts have pre-

determined transcription start and stop sites. Conserved stem-loop structures and single stranded regions were compared with the hypothesis that; in order for attenuation to take place, be it by ribosomal binding site occlusion, ribosome stalling or induced mRNA degradation, hydrogen bonding between target and antisense transcript would have to occur. Analysis of these secondary structure however reveals few conserved regions where local “unzipping” of the target transcript could take place, as is the case with *copA/copT* and RNA I/RNA II (Hjalt and Wagner,1992). A simple calculation of single stranded bases reveals that 24 out of a possible 80 single stranded bases are available for hydrogen bonding of antisense transcript MPB83-36 (region 1) to target *mpb83* mRNA. This represents 30% of the available single stranded bases in the *mpb83* transcript in region 1. Similarly, if we perform the same calculation on MPB83-25 (region 2), 21 out of a possible 64 single stranded bases are available, representing 34% of the single stranded bases in region 2 of *mpb83* mRNA. This difference seems insufficient to explain the observed differences in attenuation between the constructs.

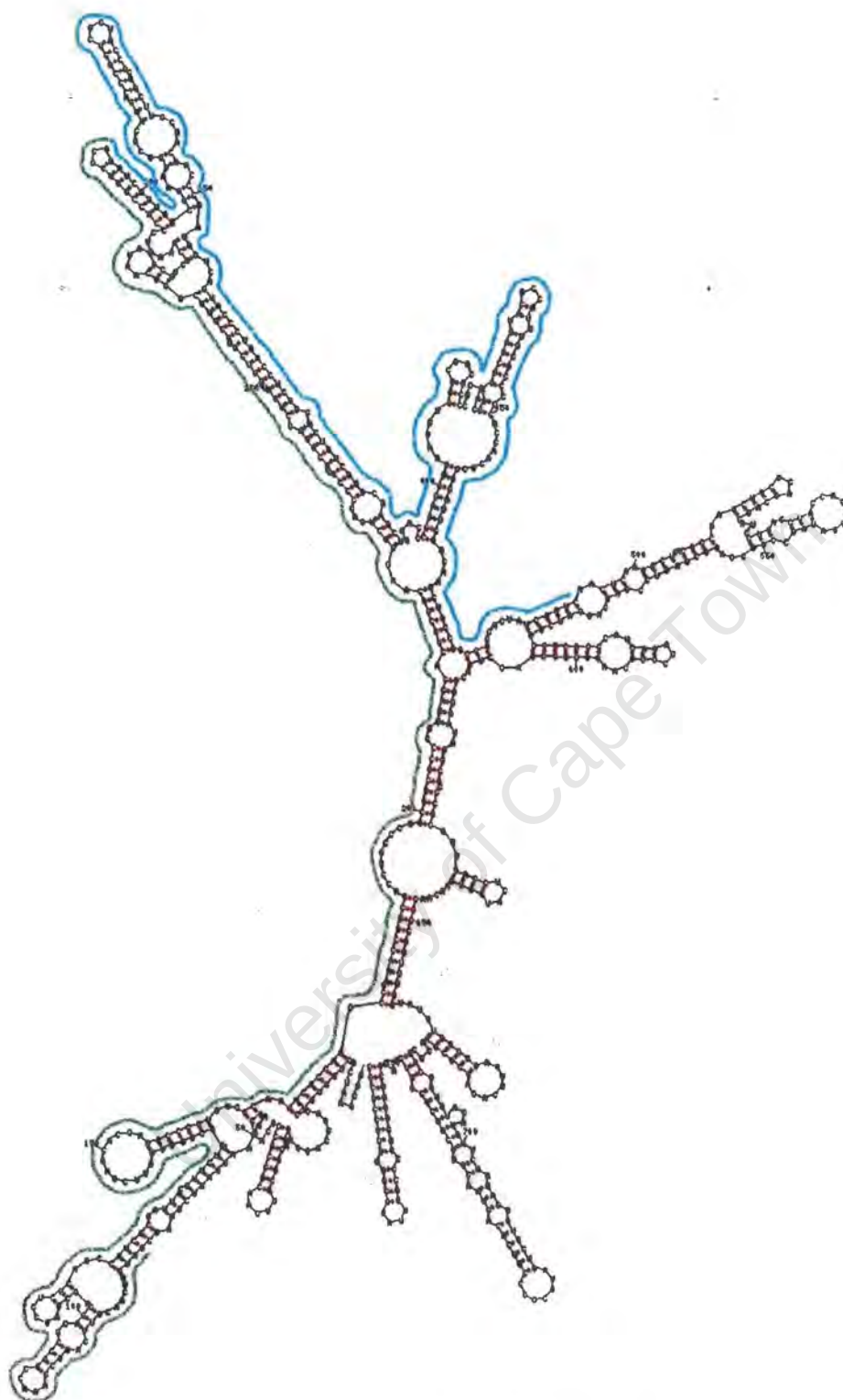


Figure 4.2: Putative secondary structure of *mpb83* mRNA. Area outlined in green represents region 1 and blue represents region 2 as targeted by MPB83-36 and MPB83-5 respectively. This structure represents the most thermodynamically stable structure calculated by mfold (Zuker 1994).

A



B



Figure 4.3: Putative secondary structures of: A) MPB83-36 antisense mRNA, B) MPB83-5 antisense mRNA. Regions in red represent single stranded regions in the antisense molecule that correspond to single stranded regions in the target *mpb83* transcript (Figure 5.18). Regions in blue represent double stranded regions now unavailable for binding to single stranded regions in target *mpb83* mRNA. These structures represent the most thermodynamically stable structures calculated by mfold (Zuker 1994).

Chapter 5

External Guide Sequences

Contents	page
5.1 Summary	85
5.2 Introduction	85
5.3 Materials and Methods	89
5.3.1 Bacterial strains and plasmids	89
5.3.2 Design of EGSs	91
5.3.3 Plasmids for <i>in vitro</i> transcription	93
5.3.3.1 Construction of pGemTMPB83	94
5.3.3.2 Construction of pGem3Zf(+)-MPB83EGS1/2	95
5.3.4 Construction of pJCImpCEGS1/2	96
5.3.5 Synthesis of <i>in vitro</i> EGS and target RNA	100
5.3.6 <i>In vitro</i> cleavage assays	100
5.3.7 Northern Blots	100
5.4 Results	101
5.4.1 <i>In vitro cis</i> cleavage of MPB83 EGS sequences	101
5.4.2 Specific binding of EGS to <i>mpb83</i>	103
5.5 Conclusions	104

Chapter 5

External Guide Sequences

5.1 Summary

The objective of this portion of the study is to introduce and test a novel, more efficient method of achieving attenuation in mycobacteria. This method involves using external guide sequences. External guide sequences (EGSs) are short oligoribonucleotides designed to bind to a given target mRNA to form a precursor tRNA-like complex. External guide sequences are short oligoribonucleotides that, upon hybridisation with a RNA target via complementary Watson-Crick base pairing, can form a tRNA-like complex and induce sequence-specific cleavage of the RNA target by endogenous RNase P. A high copy number transcription vector has been designed for the purpose of overexpressing EGSs in mycobacteria. It consists of an efficient promoter followed by a hammerhead ribozyme, a multiple cloning site, followed by a second hammerhead ribozyme, and finally a transcription terminator. *In vitro* cleavage analysis showed that efficient expression and *in cis* release of EGS sequences from full length transcribed RNA.

5.2 Introduction

The construction of defined mutants of *Mycobacterium tuberculosis* is an important tool for investigating the virulence mechanisms of this pathogen and for attenuating potential vaccine strains. It is possible to generate mutants of *M. tuberculosis* either by transposon mutagenesis (McAdam *et al.*, 1995; Pelicic *et al.*, 1998) or by targeted gene replacement (Balasubramanian *et al.*, 1996; Parish *et al.*, 1999). These techniques are however not without problems. Firstly, the frequency of recombination is not high and more importantly they are limited to the study of non-essential genes. In this chapter a transcription vector capable of expressing EGS's

was designed. *In vitro* cleavage assays were performed to demonstrate correct 5' and 3' processing of Ess as well as specific binding of EGSs to target mRNA.

Antisense approaches to controlling gene expression rely on single-stranded oligonucleotides and include antisense and EGS methods. Both of these methods involve Watson-Crick base pairing of a complementary oligonucleotide to nascent mRNA. The mechanism of inhibition of antisense methods involves either competitive binding of antisense oligonucleotides with the ribosome at the site of ribosome attachment to the nascent mRNA; or a steric block to the ribosome elongation complex by the formation of double stranded RNA-DNA structures; or even induction of RNase's (Reviewed in Chapter 2). Both antisense and EGS methods result in a reduction in translation. The EGS method however is very specific in its mechanism of action as apposed to the antisense method, which can be unpredictable in the site, and mechanism of attenuation.

RNase P is the enzyme responsible for maturation of the 5' termini of tRNA precursor molecules *in vivo* (Altman *et al.*, 1993). In all organisms investigated so far, it has been found that the enzyme is composed of two kinds of subunits, one is RNA and the other protein, and both are essential for the activity of RNase P *in vivo* (Kole *et al.*, 1980). The RNA component (M1 RNA) of RNase P from *E. coli* and from other bacteria (Guerrier-Takada *et al.*, 1983; Guerrier-Takada and Altman 1984; Baer and Altman, 1985) can, by itself, carry out the catalytic function *in vitro*. The function of the protein component *in vivo* is unknown but, *in vitro*, presence of the protein stimulates the rate of the cleavage reaction (Guerrier-Takada and Altman, 1984) and affects the substrate specificity (Guerrier-Takada *et al.*, 1984; McClain *et al.*, 1987).

During studies of the recognition of its substrates by RNase P from *E. coli*, it was found that the enzyme could cleave substrates that lack specific domains (the D and anticodon stems and loops Fig. 5.1) of the usual tRNA structure (McClain *et al.*, 1987). Studies of substrate recognition by M1 RNA and RNase P (Altman *et al.*, 1993; Foster and Altman, 1990; Li *et al.*, 1992; McClain *et al.*, 1987; Perreault and Altman, 1992; Oh and Pace, 1994; Kirsebom and Svard, 1994) have lead to the development of a general strategy of gene targeting in which M1 RNA and RNase P can be used as tools to cleave any specific mRNA sequence (Figure 5.1). The small

model substrate designed to contain a structure equivalent to the acceptor stem, the T-stem, the 3' CCA sequence and the 5' leader sequence of a ptRNA molecule (McClain *et al.*, 1987) can be recognized and cleaved by RNase P.

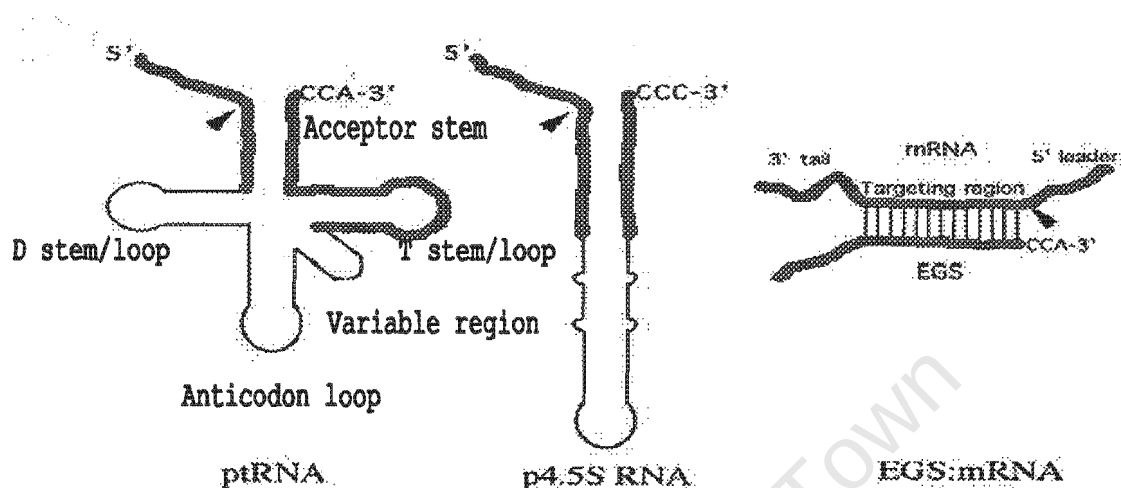


Figure 5.1: Schematic representation of natural substrates (ptRNA and p4.5S RNA), and a small model substrate (EGS:mRNA) for ribonuclease P and M1 RNA from *E. coli*. The structural components common to both ptRNA and p4.5 RNA substrates are highlighted. The site of cleavage by RNase P or M1 RNA is marked with a filled arrow (Adapted from Hsu *et al.*, 2000).

Results published by researchers using similar methodology in unrelated organisms indicate a higher efficiency of attenuation using the EGS approach compared with the “antisense-alone” method. Theoretically, it is possible for one molecule of ribozyme to cleave multiple target mRNAs. For high-level activity, an antisense RNA needs a cellular factor, such as an RNase III-type nuclease to destroy the target RNA, as is the case with the “antisense alone” method. As a consequence, a ribozyme-bases system, such as EGS/RNaseP, tends to show a higher inhibition effect than does an “antisense-alone” RNA (Inokuchi *et al.*, 1994).

A reduction of 60% in the levels of fully induced alkaline phosphatase and β -galactosidase expression was observed when plasmid encoded EGSs were introduced into *E. coli* (Guerrier-Takada *et al.*, 1995). The inability to achieve complete inhibition of gene expression in *E. coli* may be due to several factors, among which are a limit to the steady-state copy number per cell of EGSs in the constructs and the

inaccessibility of a fraction of every population of mRNA due to the tight coupling of transcription and translation in *E. coli* (Guerrier-Takada *et al.*, 1995).

To further demonstrate the application of the EGS method, Guerrier-Takada *et al.* (1997) targeted genes responsible for drug resistance in *E. coli*. Both the chloramphenicol (*cat*) and β -lactamase (*bla*) genes were recognized as substrates after hybridizing with specific EGSs. These substrates were recognized by RNase P, and the target mRNA cleaved and inactivated, thereby rendering the host cell drug-sensitive. Complete phenotypic conversion in a population of drug-resistant *E. coli* was reported.

Inhibition of influenza virus reproduction in cells in tissue culture by the simultaneous targeting of two influenza virus mRNAs using EGSs was demonstrated (Plein-Dujowich and Altman, 1998). Here, the success in targeting two genes was emphasized by the fact that inhibition by a single EGS yielded an inhibition in protein synthesis of about 50% compared with an average of 90% using a double target EGS at a lower MOI. A reduction of about 80% in the expression of both thymidine kinase (*TK*) mRNA and protein was observed in human cells in which EGSs directed to the *TK* of herpes simplex virus 1 was expressed (Kawa *et al.*, 1998).

The gyrase A gene of *E. coli*, which catalyses chromosomal DNA supercoiling during replication (Levine *et al.*, 1998) was targeted by McKinney *et al.* (2001). Gyrase A is the molecular target of quinolone antimicrobials (Hooper and Wolfson, 1993), and can mediate quinolone resistance. The EGSs were based on species-specific gyrase A sequences that are not present in *Salmonella typhimurium*, but which encode identical gyrase A protein sequences in both species, and the C5 protein subunit of RNase P. Using the EGS technology, the levels of microbial viability was reduced to less than 10% of the wild-type strain. The EGS effect was additive with roughly double the effect on bacterial viability observed when a second distinct EGS directed against a second portion of a single mRNA target was added. Compared with a single EGS, quadruple effects were observed when four distinct EGSs targeting two different mRNAs: two EGSs against gyrase A and two EGSs against RNase P. In addition to this, the effect of the EGSs was dependent on the number of nucleotides paired when

attacking gyrase A mRNA. In this case, three nucleotides unpaired out of a 15-mer EGS still favour complete inhibition by the EGS but 5 unpaired nucleotides do not. This suggests mechanisms for a degree of exquisitely narrow spectrum antimicrobial activity, discriminating between and selectively inhibiting even closely related species of bacteria based on fine differences in bacterial coding sequences (McKinney *et al.*, 2001).

Neither antisense nor EGS methods result in complete inhibition of translation, but rather translation attenuation. This makes them suitable for attenuating genes essential for the viability of the organism.

The motivation behind using the EGS method includes the fact that it: i) complements available genome data, ii) has a highly specific mechanism of action, iii) enables us to study gene expression in real-time using an inducible promoter, iv) has the advantage of being able to study genes essential to the survival of the organism.

5.3 Materials and Methods

5.3.1 Bacterial strains and plasmids

The bacterial strains, and plasmids used in this study are summarized in table 5.1. All cloning steps were performed in *E. coli* LKIII (Zabeau and Stanley, 1982) grown in 2YT medium (Miller, 1992). Details of the commonly used techniques and solutions are given in Appendices B and C, respectively.

Table 5.1
Constructs prepared for this study

Bacterial strains and plasmids used in this study		
Bacterial strains	Genotype	Reference
<i>E. coli</i> LKIII	<i>recA, lacI, lacZ, lacY⁺</i>	Zabeau and Stanley, 1982
<i>M. smegmatis</i> LR222	High frequency transformation mutant	Beggs <i>et al.</i> , 1995
Plasmid	Key Features	Technical Notes
pJCLuc	Promoterless luciferase gene in pJC86.	<u>Michelle Mulder, University of Cape Town.</u>
pJCKatG	Luciferase gene under the control of <i>M. tuberculosis katG</i> promoter in pJCLuc	<u>Michelle Mulder, University of Cape Town.</u>
p283HH	Transcription vector containing hammerhead sequences and transcription terminator	<u>Dr Celia Guerrier-Takada, Yale University</u>
pJCImpCEGS1/2	Mycobacterium transcription vector expressing <i>M. smegmatis</i> ImpC EGS 1 or 2	Transcription controlled by <i>katG</i> promoter. EGSs cleaved from transcripts <i>in cis</i> by flanking hammerhead sequences.
pMtbImpCEGS1	External guide sequence complementary to nucleotides 22-36 of <i>M.tb impC</i> mRNA.	EGS synthesised <i>in vitro</i> using T7 polymerase. Complementary to a region of <i>impC</i> mRNA that is unlikely to form secondary structures.

pMtbImpCEGS2	External guide sequence complementary to nucleotides 136-150 of M.tb <i>impC</i> mRNA.	EGS synthesised <i>in vitro</i> using T7 polymerase. Complementary to a region of <i>impC</i> mRNA that is unlikely to form secondary structures.
pGemTMPB83	<i>mpb83</i> lipoprotein gene cloned in pGemT.	<i>mpb83</i> synthesised <i>in vitro</i> from pGemTMPB83 using T7 polymerase (819 base transcript).
pGem3Zf(+)-MPB83EGS1	External guide sequence complementary to nucleotides 10-25 of <i>mpb83</i> cloned in pGemT.	EGS synthesised <i>in vitro</i> using T7 polymerase. Complementary to a region of <i>mpb83</i> mRNA that is unlikely to form secondary structures (200 base full-length transcript; 74 b 5', 42 b EGS, 84 b 3').
pGem3Zf(+)-MPB83EGS2	External guide sequence complementary to nucleotides 138-153 of <i>mpb83</i> cloned in pGemT.	EGS synthesised <i>in vitro</i> using T7 polymerase. Complementary to a region of <i>mpb83</i> mRNA that is unlikely to form secondary structures (200 base full-length transcript; 74 b 5', 42 b EGS, 84 b 3').
pUC β -lactamase	<i>bla</i> gene cloned in pUC18.	<i>bla</i> gene synthesised <i>in vitro</i> with authentic 5' and 3' ends (1121 base transcript).

5.3.2 Design of EGSs

EGSs are designed to bind specifically to target mRNA sequences. For optimal binding to occur, the secondary structure of the entire mRNA sequence (where possible) was modeled with the MFOLD energy minimization program of Zuker (1994). A target site was chosen on the basis of the prediction that it would be single-stranded and that the first nucleotide downstream of the intended site of cleavage by RNase P would be a G, a factor important in determining the efficiency of cleavage by RNase P (Perreault and Altman, 1992; Kufel and Kirsebom, 1996). The minimal structural motif recognized by RNase P is a short RNA double-stranded helix (7-19 base-pairs) between target mRNA and the EGS (Liu and Altman, 1996). The EGSs

must terminate in the sequence CCA, which is common to all tRNA. All of these features were incorporated in the design of EGSs as shown in figure 5.2. The 25-30 base-pair EGSs were made by automated DNA synthesis. For cloning purposes each complementary oligonucleotide was designed to contain the required unique restriction site at 5' and 3' ends after annealing. A list of the EGSs and their specific features is shown in Table 5.2.

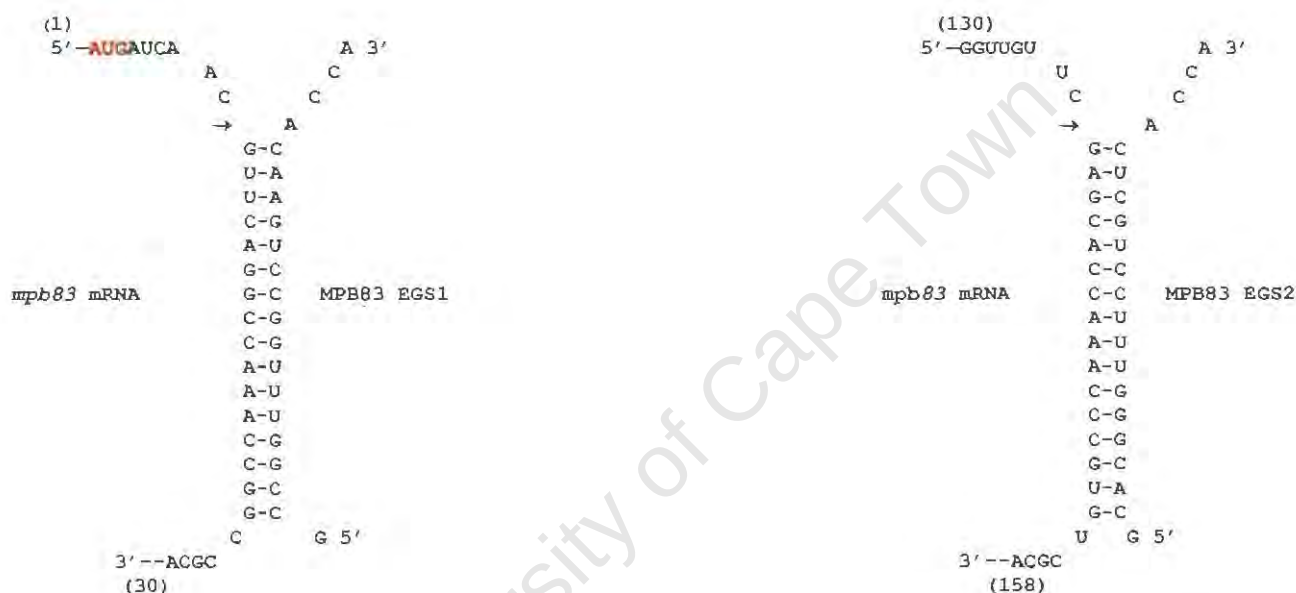


Figure 5.2: Structure of the external guide sequences directed against mRNAs encoded by the *mpb83* gene. Only the segment of each mRNA, to which the EGS hybridizes is shown. The arrows (→) indicate the sites of cleavage by RNase P. Numbers in parentheses indicate the distance in nucleotides from the start codon.

Table 5.2
EGSs used in this study

Name	Annealed oligosequence and restriction sites *↓ represent cleavage sites on the target mRNA
<i>M. bovis</i> MPB83 EGS1	10 ↓ 5' - GCCGGTTTGGCCTGAACACCAG 3' - ACGTCGGCCAAACCGGACTTGTGGTCCAGTG -3' -5' PstI BstEII
<i>M. bovis</i> MPB83 EGS2	78 ↓ 5' - GCACGGGTTTGGTGCTCACCAG 3' - ACGTCGTGCCCAAACCACGAGTGGTCCAGTG -3' -5' PstI BstEII
<i>M. smegmatis</i> ImpC EGS1	22 ↓ 5' - TGCATCAGCCGTCGCAGCTCACCCTA 3' - ACGTACGTAGTCGGCAGCGTCGAGTGGTGATCAGTG -3' -5' PstI NsiI SpeI BstEII
<i>M. smegmatis</i> ImpC EGS2	136 ↓ 5' - TGCATCGCACGGTCGACCGCACCCTA 3' - ACGTACGTAGCGTGCCAGCTGGCGTGGTGATCAGTG -3' -5' PstI NsiI SpeI BstEII

5.3.3 Plasmids for *in vitro* transcription

In vitro expression of target mRNA and EGSs was performed for a number of reasons.

- 1: To determine *in cis* cleavage, by hammerhead ribozymes, and subsequent release of EGSs from full length RNA.
- 2: To demonstrate specific binding of EGSs to target RNA.

For this, the following constructs were designed and placed under the control of the T7 promoter (Table 5.1).

5.3.3.1 Construction of pGemTMPB83

For *in vitro* expression of *mpb83* mRNA target mRNA, pGemTMPB83 was constructed (Figure 5.3). This plasmid contains the *mpb83* gene (from the transcription start site to the translation stop codon) cloned immediately downstream of the T7 promoter. The *mpb83* gene was amplified from p18a by PCR using primers; MPB83 P1 (5'-CGGGATCCTCCGAATCACTGGTGTC-3') and MPB83 P_T (5'-GGGGTACCGCGCCGAACGTTACTG-3'), containing *Bam*HI and *Kpn*I restriction sites (underlined) respectively. The 746 bp amplicon was ligated into pGemT (Promega). pGemTMPB83 was confirmed by sequencing.

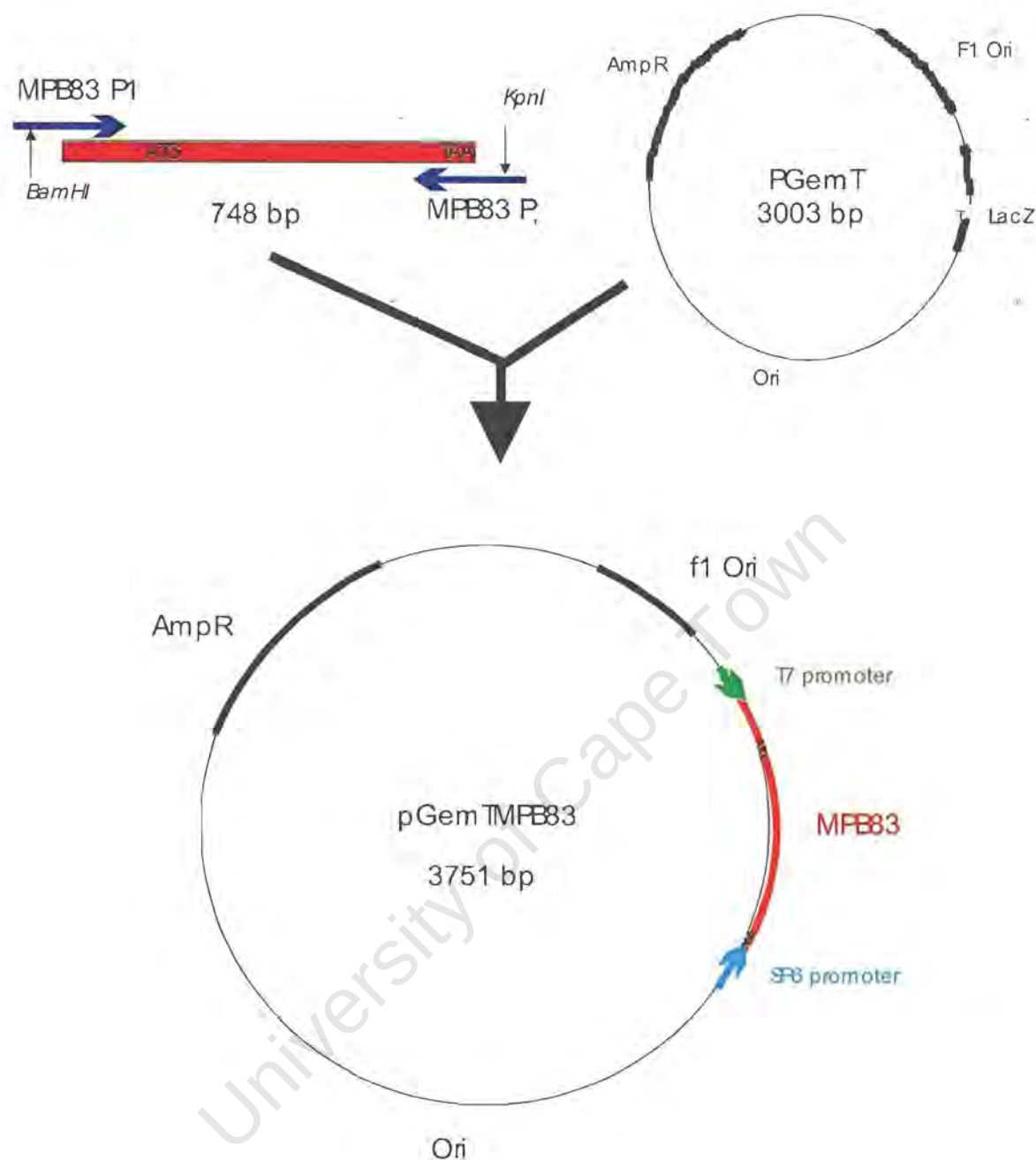


Figure 5.3: Schematic representation of the construction of pGemTMPB83 (not to scale)

5.3.3.2 Construction of pGem3Zf(+)-MPB83EGS1/2

For *in vitro* expression of MPB83 EGS1 and EGS2, pGem3Zf(+)-MPB83EGS1 and pGem3Zf(+)-MPB83EGS2 were constructed (Figure 5.4). MPB83 EGS1 and MPB83 EGS2 positioned between hammerhead sequences and upstream of the M1 terminator were amplified from p283HHMPB83EGS1/2 by PCR. The hammerhead, terminator,

and EGS sequences were amplified by PCR using the universal pUC18 forward 17-mer sequencing primer (5'- GTTTCCAGTCACGAC-3' Promega) and pUCBamHI (a pUC18 based primer containing a *Bam*HI restriction sequence at the 5' end of the primer 5'-GCTGGATCCATGATTACG-3') primers. The 488 base-pair product was digested with *Bam*HI to release two fragments of 300 and 188 base-pairs respectively. The 188 bp fragment containing hammerhead, terminator and EGS sequence was gel purified and quantitated before being cloned into pGem3Zf(+) (Promega) previously digested with *Bam*HI.

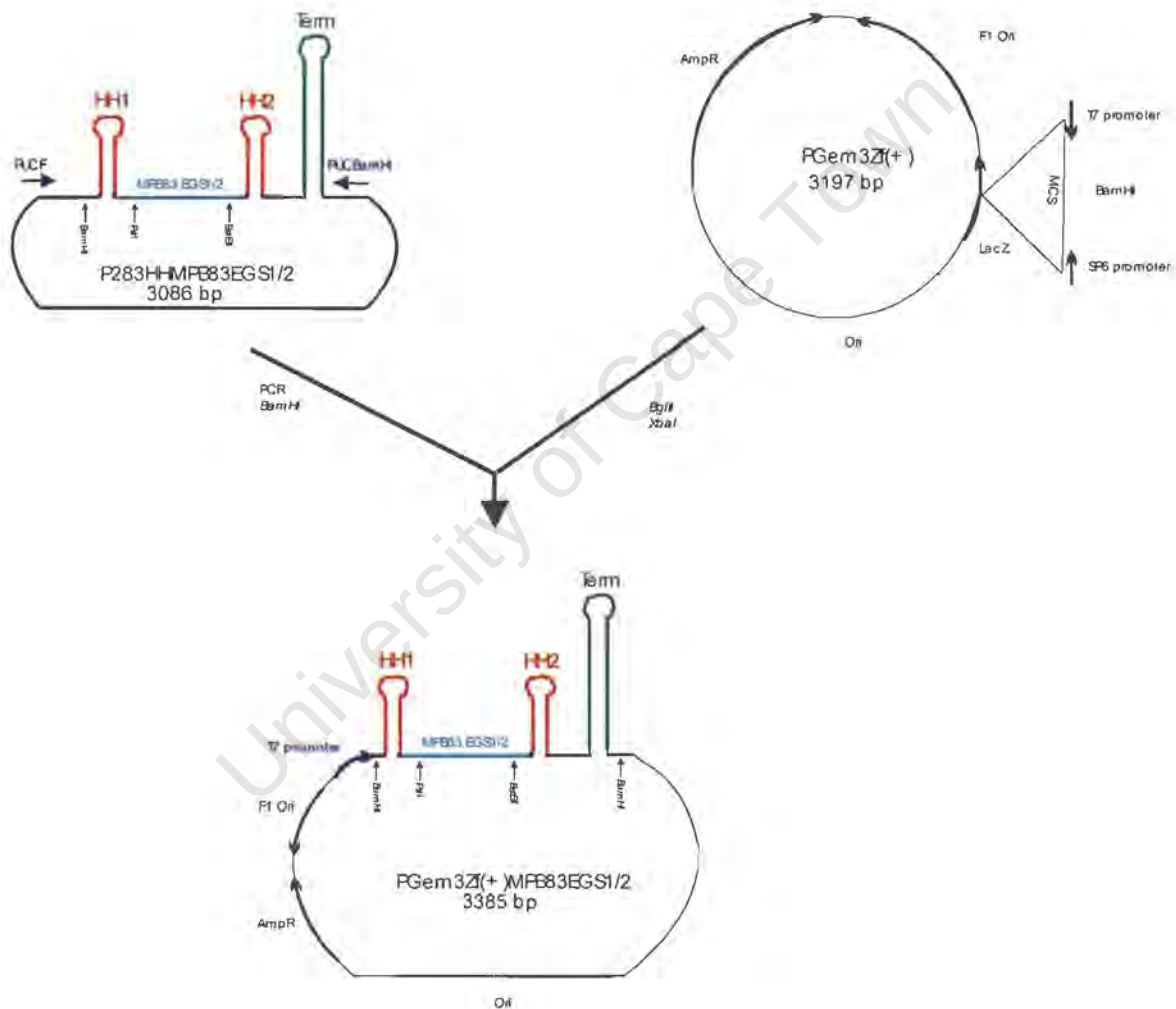


Figure 5.4: Schematic representation of construction of pGem3Zf(+)MPB83EGS1 and pGem3Zf(+)MPB83EGS2 (not to scale)

5.3.4 Construction of pJCImpCEGS1/2

For the purpose of efficient expression of EGSs in mycobacteria, an *E. coli*/Mycobacterium shuttle vector was constructed as shown in figure 5.5. This vector is based on the vector pJCKatG (Michelle Mulder, PhD thesis UCT). It contains both *E. coli* and mycobacterial origins of replication, and the *M. tuberculosis* *katG* promoter upstream of a 1748 bp promoterless firefly luciferase gene (from pGEM-*luc*, promega). The EGSs are cloned into unique sites within hammerhead sequences and upstream of the M1 transcription terminator. The hammerhead sequences 5' and 3' of the EGS ensures efficient *in cis* processing of the EGS enabling subsequent release of the EGS from upstream and downstream sequences that could otherwise inhibit efficient binding of the EGS to specific target mRNA. The hairpin structure of the M1 terminator sequence protects the EGS RNA from 3' to 5' exonuclear degradation prior to *in cis* cleavage. The hammerhead and M1 transcription terminator sequences were obtained from Dr. Celilia Guerrier-Takada (Yale University), in the form of the plasmid, p283HH. For the purpose of expressing MPB83 EGSs, MPB83 EGS1 and MPB83 EGS2 were cloned between the hammerhead sequences.

The vector, p283HH (Dr. Celia Guerrier-Takada, Yale University), was prepared by digestion with *BstEII* overnight at 60°C, followed by ethanol precipitation (See Appendix B) and digestion with *PstI*. The stuffer fragment was removed by ethanol precipitation and the concentration of digested p283HH determined by quantitative gel electrophoresis.

EGSs directed against *M. smegmatis* *impC*, ImpC EGS1 and ImpC EGS2 were designed to hydrogen-bond to 15 base-pairs (positions 22-36 and 136-150 in the coding region of the *impC* gene respectively) in the *impC* target mRNA to generate a structure that resembles the aminoacyl stem of the tRNA. Annealed oligonucleotides corresponding to ImpC EGS1 and ImpC EGS2 (with *PstI* and *BstEII* site at their 5' and 3' ends, respectively) were cloned between the *PstI* and *BstEII* sites of p283HH. The recombinant plasmids, p283HHImpCEGS1 and p283HHImpCEGS2, were verified by PCR analysis using the specific EGS as a forward primer and the universal

pUC19 reverse 17-mer sequencing primer (Promega). These recombinant plasmids were sequenced using the pUC19 reverse 17-mer sequencing primer (Promega) to confirm the sequence of the inserts.

To make the plasmids pJCImpCEGS1 and pJCImpCEGS2, the hammerhead, terminator, and EGS sequences were amplified by PCR using the universal pUC18 forward 17-mer sequencing primer (Promega) and pUCBamHI (a pUC18 based primer containing a *Bam*HI restriction sequence at the 5' end of the primer) primers. The 488 base-pair product was digested with *Bam*HI to release two fragments of 300 and 188 base-pairs respectively. The 188 bp fragment containing hammerhead, terminator and EGS sequence was gel purified and quantitated. The shuttle vector, pJCKatG (Michelle Mulder, UCT), was digested with *Bam*HI, the enzyme heat inactivated at 65°C for 15 minutes, and dephosphorylated using Shrimp alkaline dephosphorylase (Roche). Constructs pJCImpCEGS1 and pJCImpCEGS2 were made by insertion of 188 bp hammerhead, terminator, and ImpC EGS1/2 into pJCKatG digested with *Bam*HI and dephosphorylated as described above. The inserts could ligate in either orientation. The correct orientation was verified by restriction digestion with *Hind*III (correct orientation yielded a DNA band corresponding to 732 bp as apposed to 584 bp in the opposite orientation).

In this strategy, two unique restriction sites, namely *Nsi*I and *Spe*I, were introduced into pJCImpCEGS1/2 (Figure 6.3). This vector can now be used as a backbone for the introduction of further EGSs in a one step cloning procedure. Annealed EGS oligonucleotides designed with terminal *Nsi*I/*Spe*I overhangs can be cloned directly into pJCImpCEGS1/2 (digested with *Nsi*I/*Spe*I). For selection purposes, the annealed oligos are designed to contain a *Hind*III restriction site.

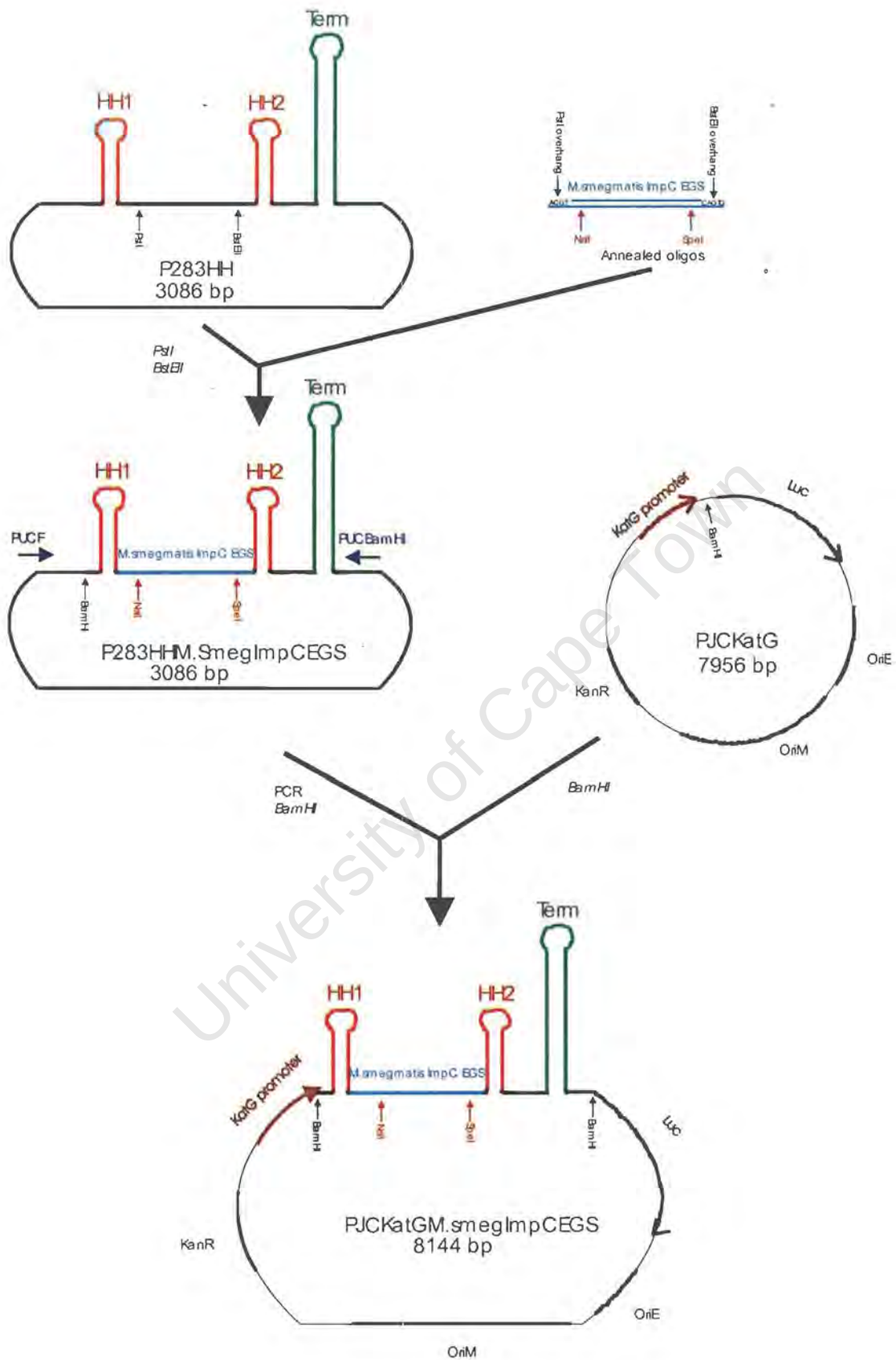


Figure 5.5: Schematic representation of construction of pJCImpCEGS (not to scale)

5.3.5 Synthesis of *in vitro* EGS target RNA

All RNA species required for cleavage assays were synthesized from the DNA templates by T7 polymerase. The synthesized RNA was subjected to gel purification in polyacrylamide gels which contain 7M urea. Subsequently, RNA was resuspended into buffer C (50mM Tris, pH 7.5, 100 mM NH₄Cl, 10mM MgCl₂). Prior to the assays for RNA enzymatic activity, these RNA enzymes were heated to 75°C for 3 min, and then allowed to renature by cooling slowly to room temperature.

5.3.6 *In vitro* cleavage assays

Self-cleavage reactions during transcription reactions were carried out with 1µg linearised template DNA. These fragments are cleaved at the 3' end of hammerhead sequences. Reaction mixtures (20µl) were in TSC buffer (40mM Tris-HCl, pH7.9, 6mM MgCl₂, 2mM Spermidine, 10mM NaCl), 0.5mM dNTPs, 20µCi [α³²P]-GTP, with 40 U of ribonuclease inhibitor (Promega) and 20 U of T7 RNA polymerase (Promega). After overnight incubation at 37°C, the reaction mixtures were analysed on 8% (w/v) sequencing gels.

5.3.6 Northern Blots

Specific *in vitro* RNA generated from T7 run-off transcription as previously described was gel purified. [α³²P]-GTP labeled EGS RNA was added to target mRNA in the presence of annealing buffer and hybridized at 37°C for 2 hours. These complexes were then separated in a 3% non-denaturing agarose gel in 1X TBE. The acrylamide gel was "pre-run" in a vertical apparatus at 20 watts (W) in 1X TBE buffer for 20 minutes. Loading buffer was added to the RNA samples prior to loading. The gel was run at 40 W. After electrophoresis, the gel was transferred onto blotting paper and dried under a vacuum on a Hoefer slab gel drier Model SE1160 for one hour. The dried gel was then exposed to an Agfa Curix X-ray film for autoradiography.

5.4 Results

5.4.1 *In vitro cis* cleavage of MPB83EGS sequences

RNA catalysis was first described by Altman (1990) and Cech (1990) with the discovery of the RNA component of RNase P and the group I intron, respectively. This was the first time that biochemical catalysis had been observed without the presence of protein-based enzymes, and such ribonucleotides were termed ribozymes. One of the smallest is the hammerhead ribozyme, which is composed of only approximately 30 nucleotides and is able to induce the site-specific cleavage of a phosphodiester bond. The hammerhead ribozymes used in this study were obtained from Dr. Celia Guerrier-Takada, the sequence is shown in figure 5.6.

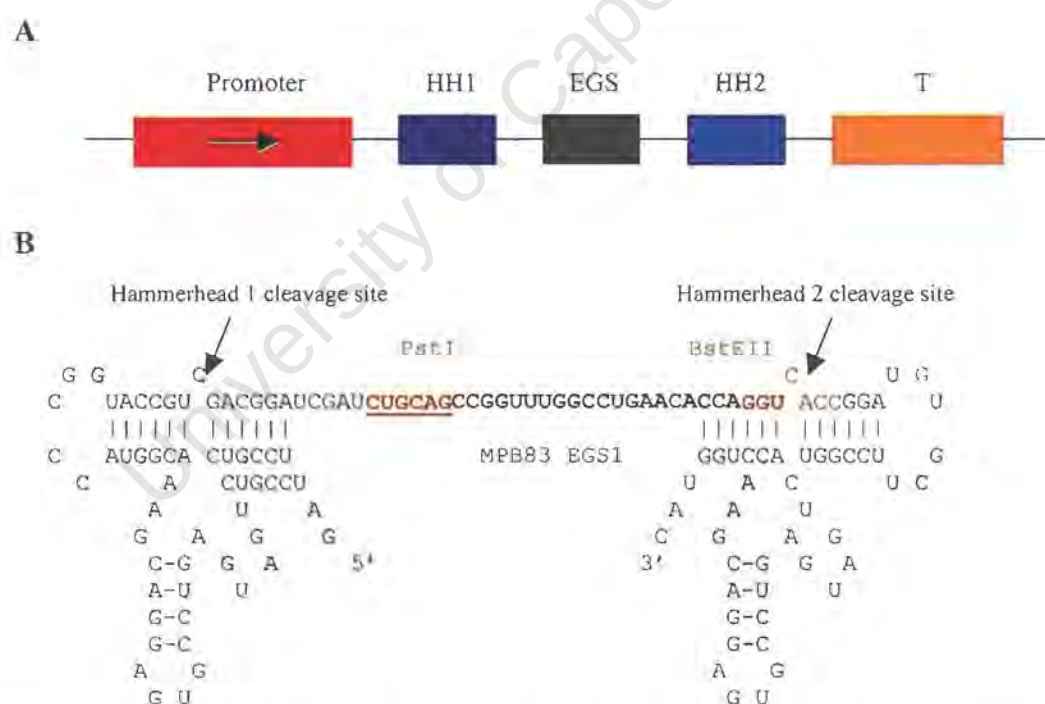


Figure 5.6: A: Schematic representation of a synthetic gene that codes for an EGS. EGS, external guide sequence; HH, hammerhead ribozyme; T, terminator of transcription. B: The sequence for hammerhead ribozymes 1 and 2 (Blue), and the MPB83 EGS (Black). The arrows indicate the cleavage sites for the two ribozymes. Restriction sites are underlined.

The hammerhead sequences (HH1 and HH2) flank the EGS. During transcription the hammerhead sequences fold into a Y-shaped conformation. This triggers cleavage *in*

cis at the sites shown in figure 5.6 (Dr. Celia Guerier-Takada personal communication). Cleavage at these sites release the EGS from the hammerhead and terminator sequences.

In order to verify correct 5' and 3' processing of EGS mRNA by hammerhead sequences, radiolabelled *in vitro* run-off transcripts of MPB83EGS mRNA were incubated at 37°C for 1 hour after transcription in Buffer A (50mM Tris pH 7.5, 100mM NH₄Cl, 100mM MgCl₂). These transcripts were then separated on a 10% denaturing polyacrylamide gel (Figure 5.7).

Results presented in figure 5.7 indicate that cleavage in *cis* is occurring and that the EGSs are being released from the full length transcripts.

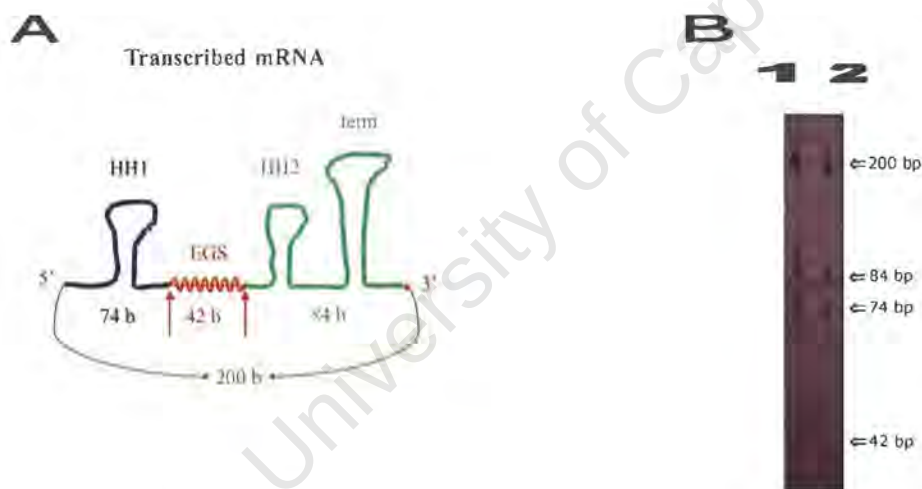


Figure 5.7: A) Schematic representation of *in vitro* transcribed MPB83EGS mRNA (full length 200 bases). Cleavage sites by hammerhead 1 (HH1) and hammerhead 2 (HH2) sequences are shown (↑), yielding a 5' HH1 fragment (74 bases), 3' HH2 and terminator fragment (84 bases), and EGS (42 bases). B) PAGE of *in vitro* transcribed [α -³³P] dCTP labelled MPB83EGS mRNA. Lane 1) MPB83EGSa; lane 2) MPB83EGSb

5.4.2 Specific binding of EGS to *mpb83*

To demonstrate that MPB83EGS1 and MPB83EGS2 bind to *mpb83* target mRNA, EGSs that had previously been released from HH sequences (see 5.4.1), were radiolabelled and used to probe *mpb83* by northern blotting (Figure 5.8).



Figure 5.8: Northern blot of *in vitro* transcribed mRNA using either MPB83EGS1 or MPB83EGS2 as radiolabeled probe prior to non-denaturing PAGE. Lane 1) *bla* mRNA:MPB83EGS1, lane 2) MPB83EGS1, lane 3) MPB83EGS2, lane 4) *mpb83* mRNA:MPB83EGS1, lane 5) *mpb83* mRNA:MPB83EGS2.

For this experiment, *bla* mRNA transcribed from pUC β -lac was used as a negative control and hybridized with labeled MPB83 EGS1 (lane 1). Labeled MPB83 EGS1 and MPB83 EGS2 were placed in lane 2 and 3 respectively; due to the small size of these transcripts they migrated off the polyacrylamide gel. Lanes 4 and 5 contain *mpb83* mRNA hybridized to specific MPB83 EGS1 and MPB83 EGS2 respectively.

Specific binding between both MPB83EGS1 and MPB83EGS2 and *mpb83* has taken place whilst no hybridization has occurred between MPB83EGS1 and *bla* mRNA. This suggests that the two MPB83EGSs are able to bind to target *mpb83* mRNA specifically.

5.5 Conclusions

External guide sequences have previously been shown to be effective at attenuating target gene expression by reducing the amount of translatable mRNA (Inokuchi *et al.*, 1994; Guerrier-Takada *et al.*, 1995; Guerrier-Takada *et al.*, 1997; Plein-Dujowich and Altman 1998; Kawa *et al.*, 1998; Levine *et al.*, 1998; McKinney *et al.*, 2001). In this study it was shown *in vitro*, that EGSs flanked by hammerhead sequences are processed 5' and 3', releasing EGSs for target binding and subsequent presentation to RNase P for cleavage (Altman *et al.*, 1993).

Chapter 6

Attenuation of gene expression by external guide sequences

Contents	page
6.1 Summary	106
6.2 Introduction	106
6.3 Materials and Methods	108
6.3.1 Bacterial strains and plasmids	108
6.3.2 Electroporation	109
6.3.3 Growth rates in liquid culture	110
6.3.4 Total RNA isolation	110
6.3.5 Quantification of specific transcripts with LightCycler RT-PCR	111
6.4 Results	112
6.4.1 ImpC EGS expression affects growth of <i>M. smegmatis</i>	112
6.4.2 Real-time quantitative RT-PCR analysis of <i>impC</i>	115
6.5 Conclusions	120

Chapter 6

Attenuation of gene expression by External Guide Sequences

6.1 Summary

External guide sequences targeting *M. smegmatis* *impC* were designed, cloned into transcription vectors and electroporated into *M. smegmatis*. Differences in both growth rate and culture turbidity were observed in *M. smegmatis* harbouring ImpC EGS expressing plasmids. The growth phase at which these phenotypes are observed is consistent with maximal EGS expression from the *katG* promoter. Real-time quantitative RT-PCR analysis of *impC* indicated a reduction in full-length *impC* RNA isolated from *M. smegmatis* harbouring ImpC EGS constructs. *M. smegmatis* harbouring ImpC EGS1 resulted in the highest level of *impC* cleavage.

6.2 Introduction

In chapter 4, I showed that attenuation of a non-essential gene (*mpb83*) could be achieved in a slow growing mycobacterium. The external guide sequence-based approach to attenuation of gene expression was adapted to increase the efficiency of attenuation. In this chapter I looked for a gene that resulted in a more overt phenotype when partially knocked out, therefore, targeted a putative essential gene. For this, *impC* was chosen.

One of the defining characteristics of mycobacteria is their unique cell wall. It is composed of a complex structure containing long-chain fatty acid molecules, the mycolic acids that are covalently attached to the peptidoglycan by arabinogalactan polymer. Species-specific glycolipids and peptidoglycolipids surround this polymer, forming the outer leaflet of a lipid bilayer (Besra and Catterjee 1994). In addition, lipoarabinomannan (LAM) molecules are present in the envelope. The cell envelope is a highly effective permeability barrier (Trias and Benz, 1994), which may be important for survival of pathogenic species inside the host environment.

The synthesis of phosphatidylinositol in *M. smegmatis* is essential for growth and myo-inositol is a key metabolite for mycobacteria (Parish *et al.*, 1997; Jackson *et al.*, 2000). Little is known about the biosynthesis of inositol in mycobacteria and the only known *de novo* pathway for myo-inositol biosynthesis involves a two-step process (Figure 6.1). First, cyclization of glucose 6-phosphate to afford myo-inositol 1-phosphate via inositol-1-phosphate synthase and, secondly, dephosphorylation of myo-inositol 1-phosphate by inositol monophosphatase (IMP) to afford myo-inositol (Parish *et al.*, 1997; Nigou and Besra 2002).

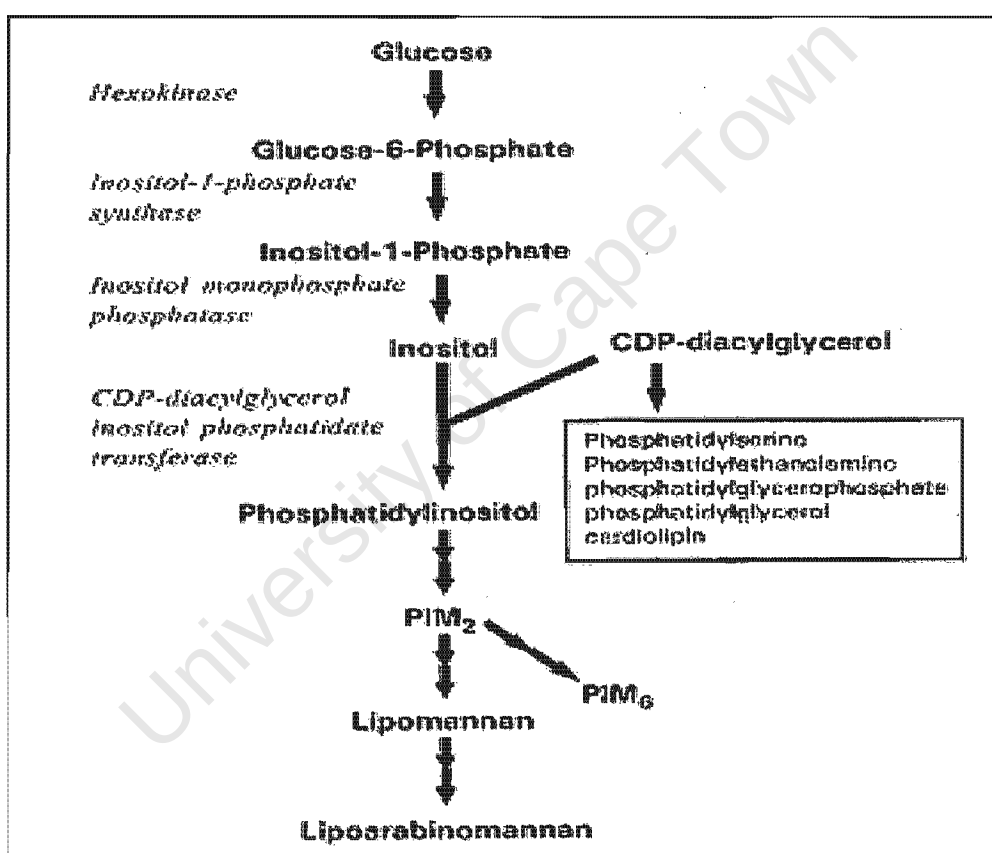


Figure 6.1: Proposed pathway for inositol metabolism in mycobacteria (Parish *et al.*, 1997)

Both *M. smegmatis* and *M. tuberculosis* contain three genes (*impA-C*), which encode proteins with inositol monophosphate phosphatase motifs. One or more of these are responsible for synthesizing inositol present in LAM and mycothiol. Recently, using complementation studies, it has been shown that an *impC* knockout in *M. tuberculosis* is lethal (Dr Farah Movahedzadeh, personal communication). Also, mutants of *M.*

smegmatis that lack mycothiol are viable, but grow poorly *in vitro* (Newton *et al.*, 1999).

Two EGS were designed to target different regions of *M. smegmatis impC*. They were cloned between hammerhead ribozymes downstream of the highly expressed *M. tuberculosis* H37Rv *katG* promoter in the mycobacterium transcription vector pJCKatG. These constructs, pJCImpCEGS1 and pJCImpCEGS2 were electroporated into *M. smegmatis* MC² along with a control plasmid, pJCLuc.

The objective of this portion of the study is the specific attenuation of *impC* expression in *Mycobacterium smegmatis* using *impC*-specific external guide sequences.

6.3 Materials and Methods

6.3.1 Bacterial strains and plasmids

The bacterial strains, and plasmids used in this study are summarized in table 6.1. Details of the commonly used techniques and solutions are given in Appendices B and C, respectively.

Table 6.1
Constructs prepared for this study

Bacterial strains and plasmids used in this study		
Bacterial strains	Genotype	Reference
<i>M. smegmatis</i> LR222	High frequency transformation mutant	Beggs <i>et al.</i> , 1995
Plasmid	Key Features	Technical Notes
pJCLuc	Promoterless luciferase gene in pJC86.	<i>Michelle Mulder, University of Cape Town.</i>
pJCImpCEGS1/2	Mycobacterium transcription vector expressing <i>M. smegmatis</i> ImpC EGS 1 or 2	Transcription controlled by <i>katG</i> promoter. EGSs cleaved from transcripts <i>in cis</i> by flanking hammerhead sequences.

6.3.2 Electroporation

M. smegmatis cells were prepared for electroporation by the method of Jacobs *et al.* (1991). A stationary phase culture of *M. smegmatis* LR222 (Beggs *et al.*, 1995), in Middlebrook-ADC-Tween, was diluted 100-fold into fresh medium, and grown for 48 hours at 37°C with shaking. The cells were harvested by centrifugation, washed 3-4 times in ice-cold 10% (v/v) glycerol, and resuspended in a final volume of 2ml/100ml original culture volume in 10% (v/v) glycerol. For electroporation, 60µl cells were combined with 140µl of 10% (v/v) glycerol and 1µg of DNA, in a 5µl volume, and incubated on ice for 5 minutes. Electroporations were performed in a 0.1 cm gap Gene Pulser® cuvettes (Bio-Rad Laboratories, CA U.S.A) at 1.8kV, 25µF and 1000

Ω , using a Gene Pulser™ apparatus (Bio-Rad Laboratories, CA U.S.A.), which resulted in a field strength of 18.0 kV/cm. After electroporation, 900 μ l of Middlebrook-ADC was added, and the cells were incubated at 37°C for 3-4 hours to facilitate expression of the kanamycin resistance gene. Transformants containing the recombinant plasmids were selected on Middlebrook-ADC agar plates containing kanamycin at a concentration of 50 μ g/ml. A control electroporation with no added DNA was included to monitor the frequency of spontaneous mutation to kanamycin resistance. After incubation at 37°C for three days, transformants were cultured in liquid medium for plasmid preparations.

6.3.3 Growth rates in Liquid Culture

Studies of the growth of *M. smegmatis* cells transformed with various plasmids were carried out as follows: Overnight cultures of *M. smegmatis* cells harbouring EGS expressing plasmids in Mb-ADC-Tw-kan (Middlebrook-ADC-Tween containing kanamycin at a concentration of 50 μ g/ml) were diluted 1/100 in Mb-ADC-Tw-kan, and the cultures were incubated at 37°C. At predetermined time points 1ml of culture was extracted and culture turbidity determined by OD₆₀₀.

6.3.4 Total RNA isolation

Samples were extracted at various time points for RNA isolation. At each time point 35-45ml of sample was harvested by centrifugation at 3000rpm for 15 min at 4 °C. The cell pellets were frozen at -70 °C until all samples had been collected. After rapid thawing, the cell pellets were lysed in 1 ml of TRIzol (In Vitrogen, Carlsbad, California). This suspension was transferred into 2ml screw capped tubes approximately one-third filled with 0.1 mm diameter zirconia-silica beads. Cells were disrupted by three 20s pulses at speed 6, cooling on ice between cycles in a FastPrep FP120 instrument (Savant Instruments). After centrifugation at 13, 000 rpm for 1 min, the aqueous top layer was transferred to a fresh 1.5 ml tube containing 300 μ l chloroform:isoamyl alcohol (24:1) and Phase Lock gel (Eppendorf) and separation performed according to the manufactures instructions. RNA was precipitated from

the aqueous layer and resuspended in 70 µl water. Contaminating DNA was degraded by digesting RNA samples with amplification grade DNase I (Sigma).

6.3.5 Quantification of specific transcripts with LightCycler RT-PCR

LightCycler RT-PCR was performed on total *M. smegmatis* RNA. For each sample approximately 4µg of total RNA was used. A two-step RT-PCR methodology was used. cDNA synthesis was performed using the 1st strand cDNA synthesis kit for RT-PCR (AMV) (Roche Biochemicals). PCR reactions were performed using the LightCycler FastStart DNA master SYBR green I kit (Roche Biochemicals). Master mixes were prepared following the manufacturers instructions, using the primers listed in Table 6.2. For the first step, cDNA synthesis was performed at +25 °C for 10 minutes then at 42 °C for 60 minutes. 2 µl from each of these reactions was used as template for PCR reactions. For this the following temperature profile was used for amplification: denaturation for 1 cycle at 95°C for 10 minutes and 50 cycles at 95°C for 1 s (temperature transition, 20°C/s), 68°C (for 16s rRNA primers) or 56°C (for ImpCSmeg primers) for 5 s, and 72 °C extension for 15 s in single mode. Melting-curve analysis was done between 56°C and 95 °C (temperature transition, 0.2 °C/s) with stepwise fluorescence acquisition. Sequence-specific standard curves were generated using serial dilutions for the specific RNA standards. The relative concentration of each sample transcript was then determined with the aid of the LightCycler software. The specificity of the PCR reactions was verified by ethidium bromide staining on 3% agarose gels.

Table 6.2

Oligonucleotide primers used for quantitative RT-PCR using the LightCycler

Target gene	Primer	Sequence	Source
<i>impC</i>	ImpCSmeg F1	5'- GTGAGCGAATTGATGGACG -3'	This study
<i>impC</i>	ImpCSmeg R1	5'- TCGCGAAGTACGGAATGC -3'	This study
<i>rrsB</i>	16S rRNA F	5'- GGTGTGGGTTTCCTTCCTTGG -3'	This study
<i>rrsB</i>	16S rRNA R	5'- ACATCTCACGACACCAGCTG -3'	This study

6.4 Results

6.4.1 ImpC EGS expression affects growth of *M. smegmatis*

Two EGS were designed to target different regions of *M. smegmatis impC* (Figure 6.2). They were cloned between hammerhead ribozymes downstream of the highly expressed *M. tuberculosis* H37Rv *katG* promoter in the mycobacterium transcription vector pJCKatG. These constructs, pJCImpCEGS1 and pJCImpCEGS2 were electroporated into *M. smegmatis* MC² along with a control plasmid, pJCLuc.

The effects of ImpC EGS expression on the growth of *M. smegmatis* were monitored over time (Figure 6.3). Results indicate that ImpCEGS1 is altering the growth dynamics of *M. smegmatis*. The growth rate as depicted by log₁₀ CFU vs Time, indicate that there is little difference between control, pJCLuc, and ImpCEGS1 during early exponential phase. Differences in both growth rate and culture turbidity however are observed coinciding with expression of the *katG* promoter (Figure 6.4). Since ImpA activity is optimal at the end of the exponential growth phase (Nigou and Besra, 2002), it is encouraging to observe a difference in growth rates occurring concurrently with maximal EGS expression.

<u><i>M. smegmatis</i> ImpC EGS1</u>	
1	GTGACCCACCCGGCCGACACC <u>GAGCTGCGACGGCTG</u> GCGGTGAAGGCCGCCGAACGCGGC
1	M T H P A D T E L R R L A V K A A E R G
61	GCCGACGTGTGCCTGCGTCACCTCGGTGAACTGGTCTCCGTCGCGACCAAGAGCGCGGCA
21	A D V C L R H L G E L V S V A T K S A A
<u><i>M. smegmatis</i> ImpC EGS2</u>	
121	GGCGACGTGCGTGACC <u>GCGGTGCGACCGTGCC</u> GCCGAAGAGGCCGTGCGGGCCGTGCTGCGC
41	G D V V T A V D R A A E E A V R A V L R
181	GAGGCCCGCCCGGATGACTCGATGCTCGGCGAGGAGATCGCCGAGCACACCGGATCCGGT
61	E A R P D D S M L G E E I A E H T G S G
241	CCGCTGCAGTGGGTGATCGACCCGCTCGACGGCACCACCAACTACACGCGGCGCATTCGG
81	P L Q W V I D P L D G T T N Y T R R I P
301	TACTTCGCGACCTCGGTGCGGTGCGGCGGGTCTCCGACGGCACCTGGCTGGCCGGGGCC
101	Y F A T S V A V R R V S D G T W L A G A
361	GTGTGCGCGCCGGCCCTGGGCAGCACCTGGAGCGCGGCCAGAGCCGAGGGCGCGCGACTG
121	V C A P A L G S T W S A A R G E G A R L
421	GATTCGATCACCCAGTCGGCGCAGCTGCCGCTGCAGCTGATCGATTGACGGCCCGGCTG
141	D S I T Q S A Q L P L Q L I D S T A R L
481	GTGGGCACGGGCCTGTCTACGATCCGGGCCTGCGCCGCCGCCAGTTGCGGGACCTGGGT
161	V G T G L S Y D P G L R R R Q L R D L G
541	GCACTGGTCTGCGAGTACACCGACATGCGCCGGTTGCGCGCGGCCGCGGTGGACCTGTGC
181	A L V C E Y T D M R R F G A A A V D L C
601	CTGGTGGCGCAGGGCAGCCTGGACGCGTTTCGTCGAGGCCGATCTCGCGGTGCACGACTGG
201	L V A Q G S L D A F V E A D L A V H D W
661	GCGGCTGGCGCACTGATCGCCGAGGAGGCCGCGCCGAGGTGACGCGGCCGTGCGCCGAG
221	A A G A L I A E E A G A E V T R P C A E
721	GGGGATCCGGTGTGACGCGCTGGCGCTGACCTGGCGTTGACGCACACCACCTGTACGCC
241	G D P V S A R W R * P G V D A H H L Y A
781	CGCGGCCTCCAGCGCCTTGA
261	R G L Q R L

Figure 6.2: Sequence of *M. smegmatis* *impC* gene (<ftp://ftp.tigr.org>). Sites where *M. smegmatis* ImpC EGS1/2 bind to target sequence are shown in red.

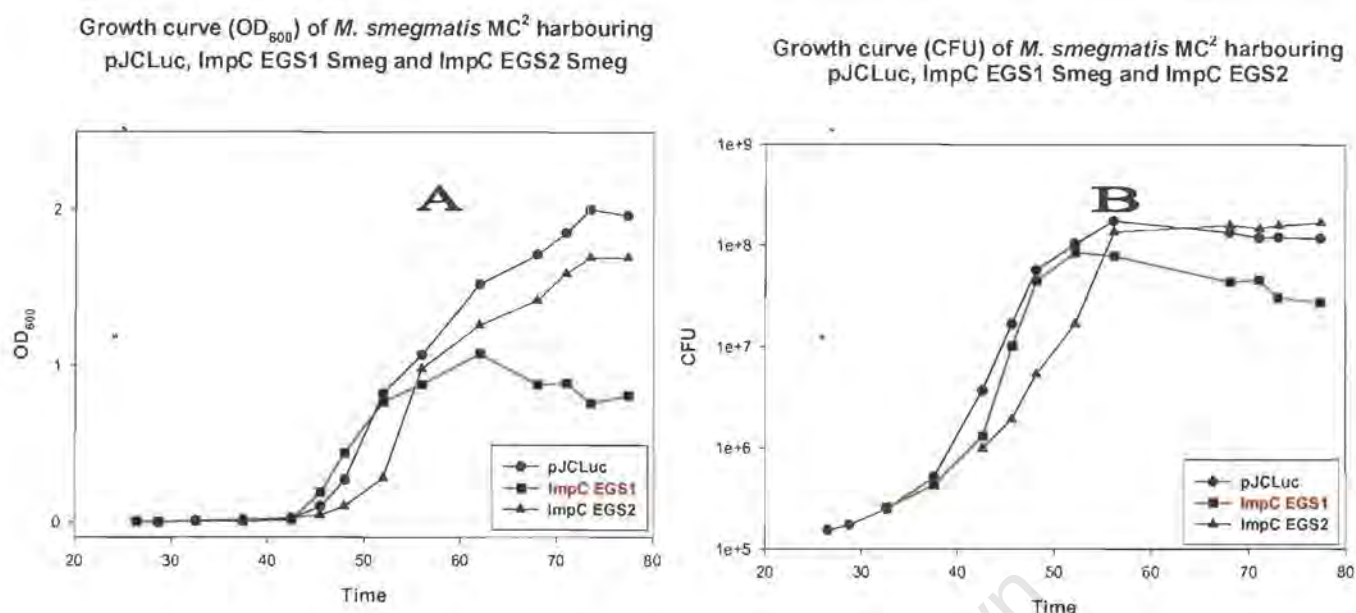


Figure 6.3: Growth curve of *M. smegmatis* harbouring ImpC EGS expressing plasmids and pJCLuc control plasmids. Plotted as a measure of; **A)** Absorbance at 600 nm (OD₆₀₀) and; **B)** colony forming units (CFU).

At this point (ie., late exponential phase) *M. smegmatis* expressing ImpCEGS1 appears to alter its growth relative to the control. Culture turbidity peaks at 1.08 in *M. smegmatis* expressing ImpC EGS1 whilst the control peaks at 2.01. The 2-fold reduction in culture turbidity in ImpC EGS1 expressing cells could indicate changes in cell morphology and resultant cell clumping or cessation of cell growth or both. The CFU data indicates that there is more than a 3-fold reduction in viable cells between 52-77 hrs in *M. smegmatis* expressing ImpCEGS1 compared to a 1.5 fold reduction in control cells. Although there seems to be some loss of viability determined by CFU, this alone would not be sufficient to account for the dramatic differences in culture turbidity between control and ImpC EGS1 expressing *M. smegmatis*. *M. smegmatis* harboring pJCImpCEGS2 is not affected in the same way as pPCImpCEGS1. Although the growth rate is slower than the control, the CFU peak at 1.69×10^8 CFU/ml, similar to the control 1.75×10^8 CFU/ml. The culture turbidity is 1.81 as compared with 2.01 in the control. These results suggest that ImpC EGS1 is either more efficient in binding to target *impC* than ImpC EGS2 or alternatively it is able to mimic a tRNA more effectively than ImpC EGS2 once bound to target *impC* and is hence cleaved by RNase P.

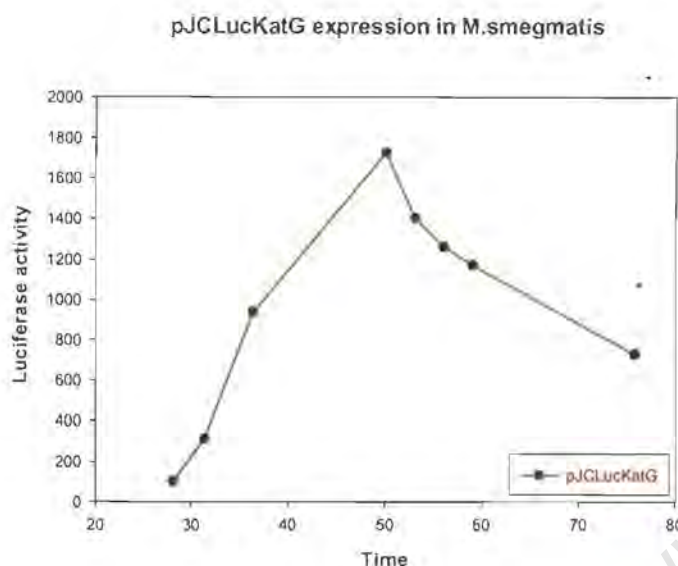


Figure 6.4: Luciferase activity of the *M. tuberculosis* H37Rv *katG* promoter monitored over a 75 hour growth period in *M. smegmatis*.

6.4.2 Real-time quantitative RT-PCR analysis of *impC*

To measure the transcript levels of *impC* in *M. smegmatis*, a quantitative RT-PCR assay was used. In this method, total RNA is used as a template to synthesise cDNA. The amount of cDNA produced, which is proportional to the amount of specific transcript present in the original RNA sample, is measured by quantitative PCR. An internal standard is included to normalize RNA in each sample.

To estimate the levels of *impC* mRNA present in *M. smegmatis* when challenged with different EGS expressing plasmids, total RNA was extracted from cultures of *M. smegmatis* harboring two EGS expressing plasmids, pJCImpCEGS1 and pJCImpCEGS2, as well as a negative control, pJCLuc, at a variety of time points. Three aliquots of total RNA from independent preparations were subjected to reverse transcription with random hexamer primers, and the resulting cDNAs were amplified by PCR. Samples containing known amounts of target cDNA were also subjected to PCR in order to provide standard curves (Figure 6.5) from which mRNA levels were determined. RNA samples not treated with reverse transcriptase were also subjected

to PCR in order to measure the level of contamination with chromosomal DNA, all showing no amplification.

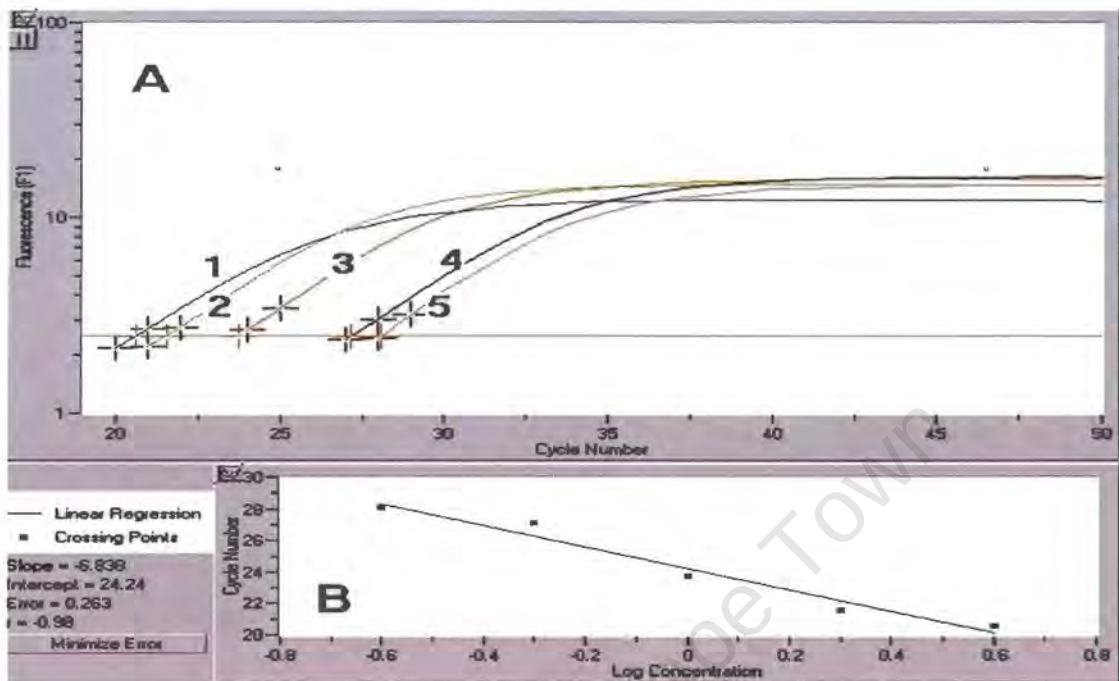


Figure 6.5: Detection of cDNA standards for *impC* using the LightCycler technique. (A) Amplification of 4 ng (1), 2 ng (2), 1 ng (3), 0.5 ng (4), 0.25 ng (5) of cDNA. (B) The generated standard curve.

To study the levels of mRNA from different cultures at different time points requires the use of an internal standard. The 16S rRNA gene was used to quantitate the levels of RNA in each sample (Fig 6.6). The *impC* RNA levels were then normalized by comparison with the levels of transcription of 16S rRNA.

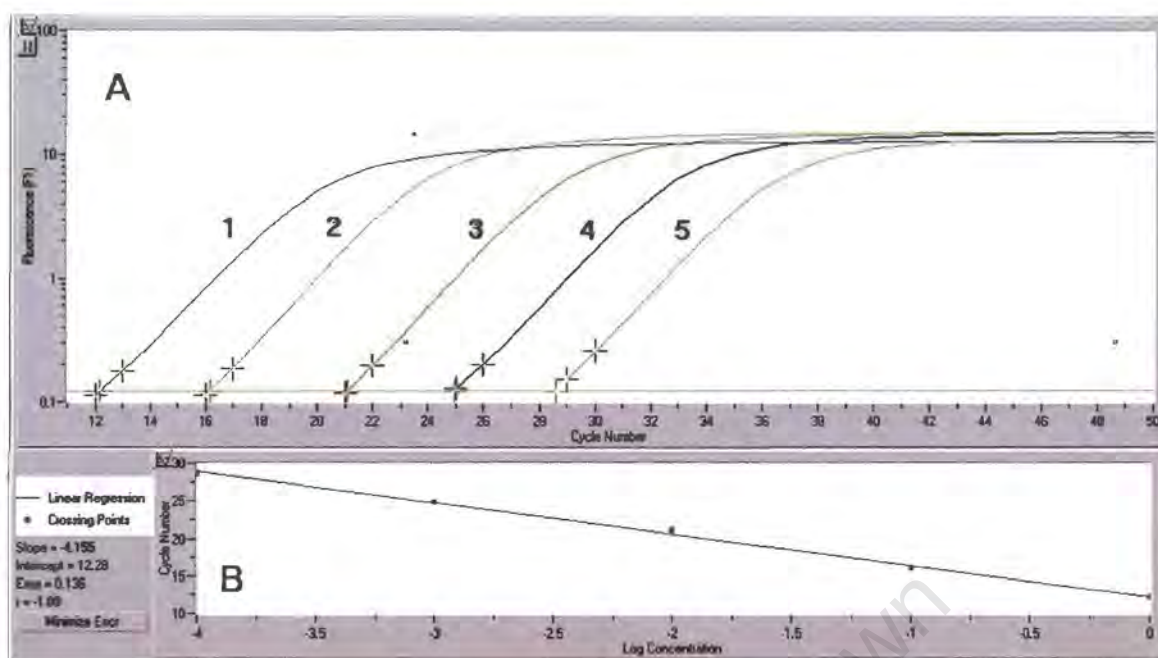


Figure 6.6: Detection of cDNA standards for 16S rRNA using the LightCycler technique. (A) Amplification of 2 ng (1), 0.2 ng (2), 20 pg (3), 2 pg (4), 0.2 pg (5) of cDNA. (B) The generated standard curve.

Standard curves of both *impC* and 16S rRNA revealed good linearity over a range of cDNA concentrations (Figure 6.5 and 6.6). Specificity was determined by melting curve analysis (Figure 6.7), indicating a single product with a melting temperature of 90°C. Analysis of the PCR amplicons on a 2% agarose gel was also performed (Figure 6.8). The size of the amplicons correlated to the predicted sizes in each case; 251 bps for 16S rRNA primers and 354 bps for *impC* primers.

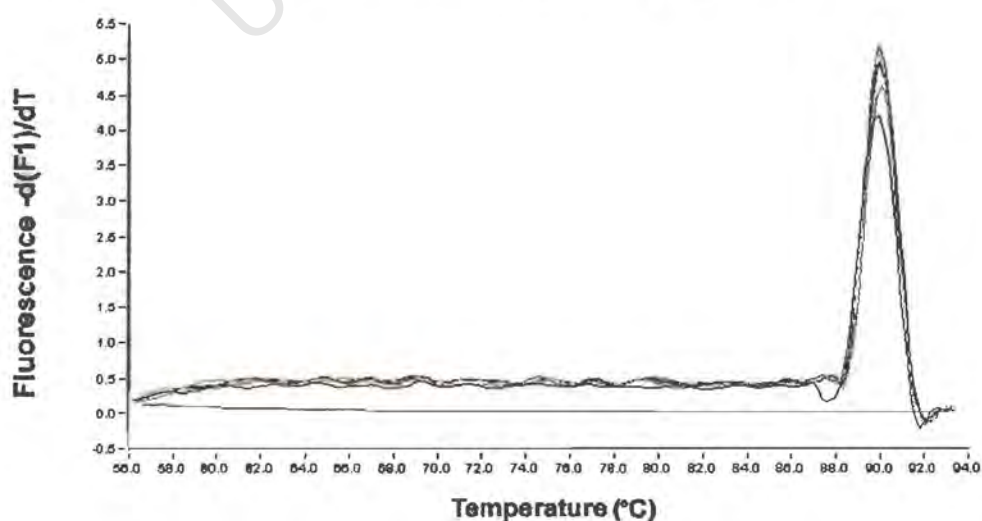


Figure 6.7: Melting temperature curve of 16S rRNA standards showing specific product

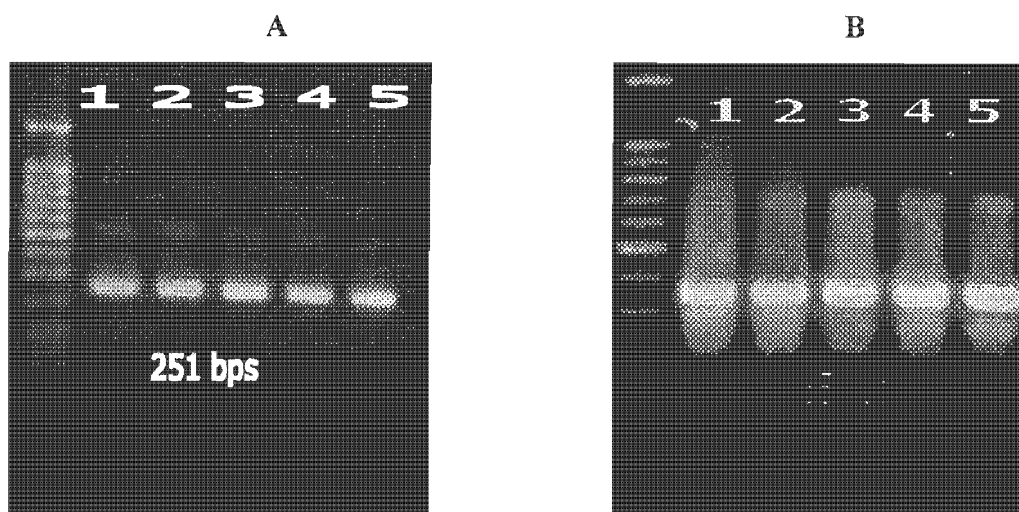


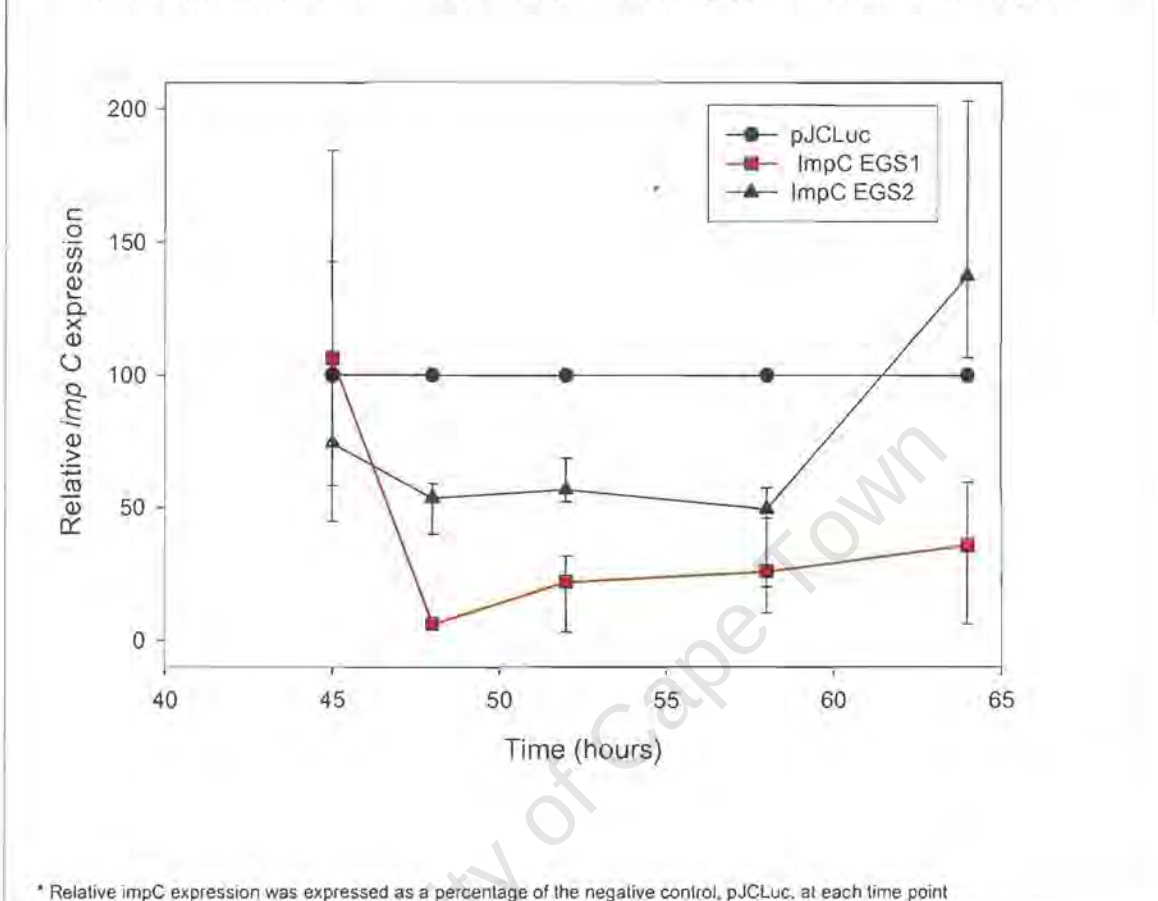
Figure 6.8: Detection of the specific cDNA standards for 16S rRNA (A), lanes 1-5 containing: 2 ng, 0.2 ng, 20 pg, 2 pg, and 0.2 pg respectively; and *impC* (B), lanes 1-5 containing 4 ng, 2 ng, 1 ng, 0.5 ng, and 0.25 ng of cDNA respectively, using the LightCycler technique. The molecular weight marker is a 100bp ladder (Promega).

The degree of *impC* cleavage by ImpC EGS was measured by quantitative RT-PCR. Results shown in Figure 6.9, indicate relative levels of intact *impC* mRNA. Results are presented as a percentage of the negative control, pJCLuc, at each time point. The degree of *impC* cleavage in the clones that contained genes coding for ImpC EGS was variable. *M. smegmatis* harboring ImpC EGS1 resulted in the highest level of *impC* cleavage. The percentage of intact *impC* dropped to a minimum of 5.83% at 48 hrs post inoculation. Cleavage then stabilizes to 17.15% at 52 hrs, 22.13% at 58 hrs, and 28.20% at 64 hrs post inoculation. Expression of EGSs under the control of the *katG* promoter is maximal at 50 hrs (Figure 6.9). The cleavage data presented here correlates accurately with the presence of ImpC EGS1 at the 48 hr time point.

M. smegmatis harboring the second ImpC EGS construct, ImpC EGS2, follows a similar pattern of *impC* cleavage over time. The level of cleavage is however less, 49.45% at 48 hrs, and may indicate a lower binding affinity if ImpC EGS2 to *impC* mRNA at the intended site of hydrogen binding.

Figure 6.9

Relative *impC* expression of *M. smegmatis* MC² harbouring pJCLuc, ImpC EGS1 and ImpC EGS2 constructs. Determined by quantitative RT analysis.



The effect of ImpC EGS expression is clearly shown in figure 6.10. Samples of *M. smegmatis* harboring control plasmid, pJCLuc, and two ImpC EGS expressing plasmids, pJCImpCEGS1 and pJCImpCEGS2, were isolated at OD_{600nm} of 0.4 and 0.9, respectively. Total RNA was isolated, normalized by 16S rRNA RT-PCR then subjected to RT-PCR using *impC* primers (Table 6.2). Results indicate *impC* levels below detectable limits at OD_{600nm} of 0.4 for all three samples. At OD_{600nm} of 0.9 however, the control sample of *M. smegmatis* harbouring pJCLuc indicates high levels of *impC* expression. ImpC EGS1 expression results in cleavage of full-length *impC* to levels below detectable limits. ImpC EGS2 expression results in partial cleavage of *impC* and an attenuation of *impC* can clearly be seen in this sample.

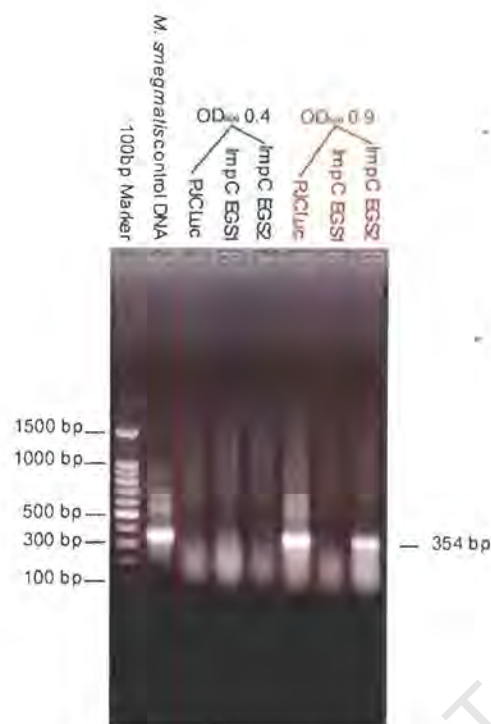


Figure 6.10: Agarose gel of quantitative *impC* RT-PCR showing attenuation effect of ImpC EGS constructs in *M. smegmatis* relative to negative control, harboring the plasmid pJCLuc.

When analysed in conjunction with the phenotypic growth curve data presented in figure 6.3, the *impC* cleavage data is striking. Not only do the time points of *impC* cleavage correspond, but also *M. smegmatis* harboring ImpC EGS1 is affected to a greater extent than those harboring ImpC EGS2.

6.5 Conclusions

External guide sequences have previously been shown to be effective at attenuating target gene expression by reducing the amount of translatable mRNA (Inokuchi *et al.*, 1994; Guerrier-Takada *et al.*, 1995; Guerrier-Takada *et al.*, 1997; Plein-Dujowich and Altman 1998; Kawa *et al.*, 1998; Levine *et al.*, 1998; McKinney *et al.*, 2001).

ImpC EGS aimed at binding to two different regions of *M. smegmatis impC* were designed. They were cloned into a mycobacterial transcription vector between hammerhead ribozyme sequences. The *katG* promoter drove transcription, which is active during exponential growth. After electroporation into *M. smegmatis*, a

reduction in turbidity as well as colony forming units was observed when compared with *M. smegmatis* harboring the negative control plasmid, pJCLuc.

Analysis of *impC* mRNA revealed that there is a reduction in full-length *impC* in *M. smegmatis* harboring either pJCImpCEGS1 or pJCImpCEGS2 when compared with bacteria harboring the negative control, pJCLuc. This reduction is greatest in cells expressing ImpC EGS1 and correlates with the phenotypic observations both in time to onset of effect as well as ImpC EGS1 being more effective ImpC EGS2 at altering the growth dynamics of *M. smegmatis*. These results are suggestive of specific reduction in full-length *impC* being responsible for the phenotypic observations.

An increased tendency to clump in a liquid culture and slower growth relative to the wild type of a *M. smegmatis* ImpA transposon mutant was previously observed (Parish *et al.*, 1997). They detected weak cross-hybridizing bands in *M. smegmatis* and found three related sequences in a search of approximately two thirds of the *M. tuberculosis* genome. They hypothesized that there is at least one other IMPase in *M. smegmatis*, which partially compensates for the mutation, allowing the cells to survive. This data could also explain the results obtained using EGSs. There are 4 possibilities:

- 1) The EGSs do not result in complete cessation to ImpC protein production and there is an attenuation effect that results in a reduced growth rate.
- 2) There is more than one gene involved in the essential inositol metabolism pathway at the point where myo-inositol-1-phosphate is converted into inositol. Knocking out one of these enzymes results in a reduced growth rate.
- 3) A combination of the two points above results in the reduced growth rate.
- 4) The reduced growth rate is due to a recessive mutation effect whereby ImpC protein is being diluted in progeny cells resulting in the phenotype observed.

The current hypothesis is that the three IMPases present in both *M. smegmatis* and *M. tuberculosis*; ImpA, ImpB and ImpC, have different roles:

- 1) ImpA: Produces inositol for the synthesis of phosphatidylinositol, PIMS and LAM.

- 2) ImpB: Orthologue of *E. coli subB*. In *E. coli* the function is unknown and since *E. coli* does not produce inositol it may not be involve in this pathway.
- 3) ImpC: Produces inositol for the synthesis of mycothiol (Stoker personal communication).

Furthermore the ImpC EGS1 results are consistent with results published by Newton *et al.*, 1999 in which mutants of *M. smegmatis* that lack mycothiol were viable, but grew poorly *in vitro*.

PART B

Introduction to Translation Initiation

University of Cape Town

Chapter 7

Introduction

7.1 Translation Initiation Overview

Translation efficiency in *Escherichia coli* is generally determined at the stage of initiation (Lang *et al.*, 1989; de Smit and van Duin, 1990a; Hartz *et al.*, 1991). Initiation is a major determinant of the overall expression level of a gene and is also a target for a variety of translational control mechanisms. Although the outlines of translational initiation have been evident for 30 years, a quantitative understanding of the factors determining the overall initiation rate has developed slowly. Initiation is the sequence of events leading up to formation of the first peptide bond. The components of the initiation complex have been identified for many years and include: the ribosome, mRNA, fMet-tRNA^{Met}, and three initiation factors (IF1, IF2, and IF3) (Noll and Noll, 1972).

Initiation of translation is a complex process and can roughly be subdivided into the following steps (de Smit and van Duin, 1994). (1) Sequence-dependent binding of a 30S subunit to the RNA. (2) Recognition of the initiation region through specific interactions. (3) Codon-anticodon interaction between the initiation codon and fMet-tRNA. (4) Association of the 50S subunit, followed by codon-anticodon interaction at the second codon and start of elongation (Gualerzi *et al.*, 1977; Ringquist *et al.*, 1992).

A kinetic scheme to determine translation efficiency was first proposed by Gualerzi and collaborators (1990, Figure 7.1). A version of that scheme is as follows:

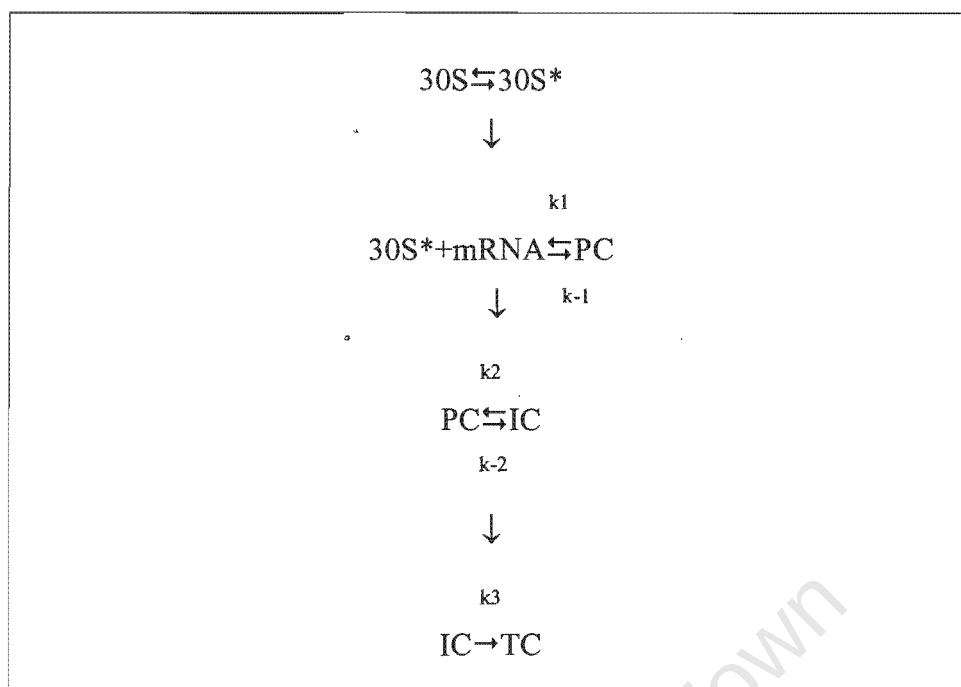


Figure 7.1: Kinetic scheme of translational efficiency (Gualerzi and Pon, 1990)

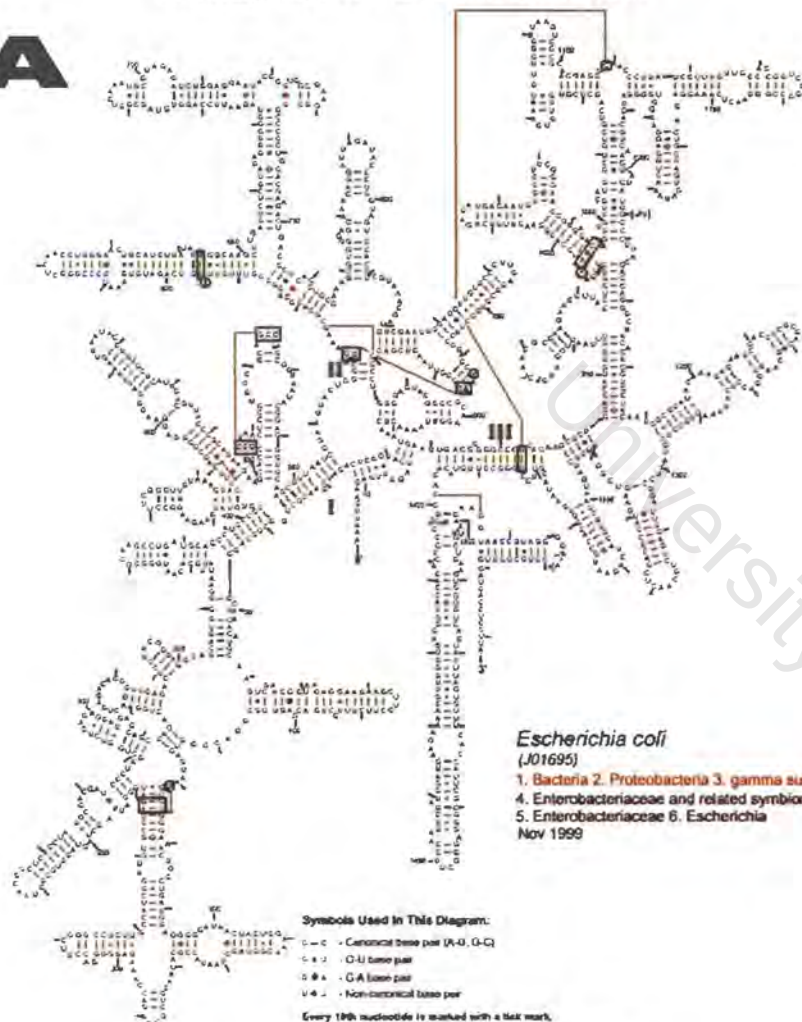
In this scheme, 30S is the small ribosomal subunit; 30S*, a complex of the small subunit with initiation factors IF1, IF2, IF3, and fMet-tRNA^{Met}; PC, a pre-initiation complex between 30S* and a specific messenger RNA in which the Shine Dalgarno (SD) region of the mRNA is base-paired to the 3' end of the 16S rRNA (Shine and Dalgarno, 1974), but no base pairing exists between the anticodon of fMet-tRNA^{Met} and the initiation codon; IC, an initiation complex in which, in addition to the SD interaction, codon-anticodon pairing exists (i.e., resulting from a first-order rearrangement of PC); TC, translation complex, in which full 70S particles are actively engaged in peptide bond formation. In this scheme 30S subunits fused with initiation factors (30S*) bind to the mRNA in an initial step stabilized in part by the SD complementarity forming a pre-initiation complex (PC). The kinetics of this reaction k_1 and k_{-1} are determined by the degree of complementarity of the SD and presence or absence of intramolecular structures in the SD region. The next step involves unimolecular isomerization of PC in which the tRNA^{fMet} anticodon pairs with the codon resulting in the formation of the initiation complex (IC). This reaction is controlled by the rate constants k_2 and k_{-2} , which are determined by factors such as spacing between the SD and the initiation codon and presence or absence of intramolecular structure in the region of the initiation codon.

7.2 The Ribosome

E. coli ribosomes are composed of 38% protein, comprising 52 different proteins (Wittman, 1982), and 62% RNA (Noller and Nomura, 1996). Three species of rRNA are found as integral structural components of the ribosomes of *E. coli* and other bacteria. The 30S subunit contains 16S rRNA, whereas 5S and 23S rRNA are found in the 50S subunit. The 5S rRNA was the first ribosomal RNA molecule to be sequenced (Brownlee *et al.*, 1967). It is composed of 120 nucleotides and contains no post-transcriptionally modified nucleotides (Noller and Nomura, 1996). The 16S rRNA contains 1542 nucleotides (in *E. coli*), 10 of which (positions 527, 966, 967, 1207, 1402, 1407, 1498, 1516, 1517, and 1518) are methylated and one of which is a pseudouridine (Bakin *et al.*, 1994). Analysis of the phylogenetic conservation of sequence in 16S rRNA has provided insight into the relative biological significance of different regions of the molecule (Gutell *et al.*, 1985; Woese *et al.*, 1983). The occurrence of post-transcriptional modifications in regions of high sequence conservation suggests that they are functionally important (Noller and Nomura, 1996). The 16S rRNA secondary structure is subdivided into 3 major structural domains and one minor domain by 3 sets of long-range base-paired interactions (Figure 7.2). The 5' domain (residues ca. 26-557) is defined by the 27-37/547-556 helix; the central domain (residues ca. 564-912) is defined by the 567-569/881-883 helix; and the 3' major (residues ca. 926-1391) and 3' minor (residues ca. 1392-1542) domains are defined by the 923-933/1384-1393 helix (Noller and Nomura, 1996). The ability of certain ribosomal proteins to remain bound to these fragments or even to rebind to RNA fragments from which proteins have been removed (Ehresman *et al.*, 1981) supports the suggestion that these domains are true structural entities. Structurally the 16S rRNA of *E. coli* (Figure 7.2a) and *M. tuberculosis* (Figure 6.7b) are similar although there is only a 74% sequence homology between them. The 23S rRNA consists of 2904 nucleotides in *E. coli* (Brosius *et al.*, 1980) and, like the 16S rRNA contains numerous pseudouridines as well as methylations. Little or no significant sequence homology has been detected between the three different rRNAs; nor has any been found between different regions of the individual rRNAs, beyond that which would be anticipated to occur by chance, arguing for independent evolutionary origins for the different rRNAs (Noller, 1993).

In *E. coli*, rRNA synthesis is a rate-limiting step in ribosome production since r-protein expression is regulated by feedback mechanisms sensitive to the rRNA concentration (Nomura et al., 1984). In several bacteria, including *E. coli*, the number of ribosomes varies linearly with the growth rate over a range of conditions. This phenomenon has been termed growth rate-dependent control (GRDC) of ribosome synthesis, and it serves to maintain the cellular pool of ribosomes at a level commensurate with the requirement of the cell for protein synthesis at all times (Gausling, 1977). Slow growing members of the Genus *Mycobacterium* such as *M. tuberculosis* depend on a single *rrn* operon driven by two promoters for their total ribosome pool, while fast growing members, including *M. smegmatis*, typically have two *rrn* operons per genome (Bercovier et al., 1986) and possess multiple promoters to increase their capacity for rRNA synthesis (Gonzalez-y-Merchand et al., 1997).

A



Citation and related information available at <http://www.rna.tomb.vt.edu>

B

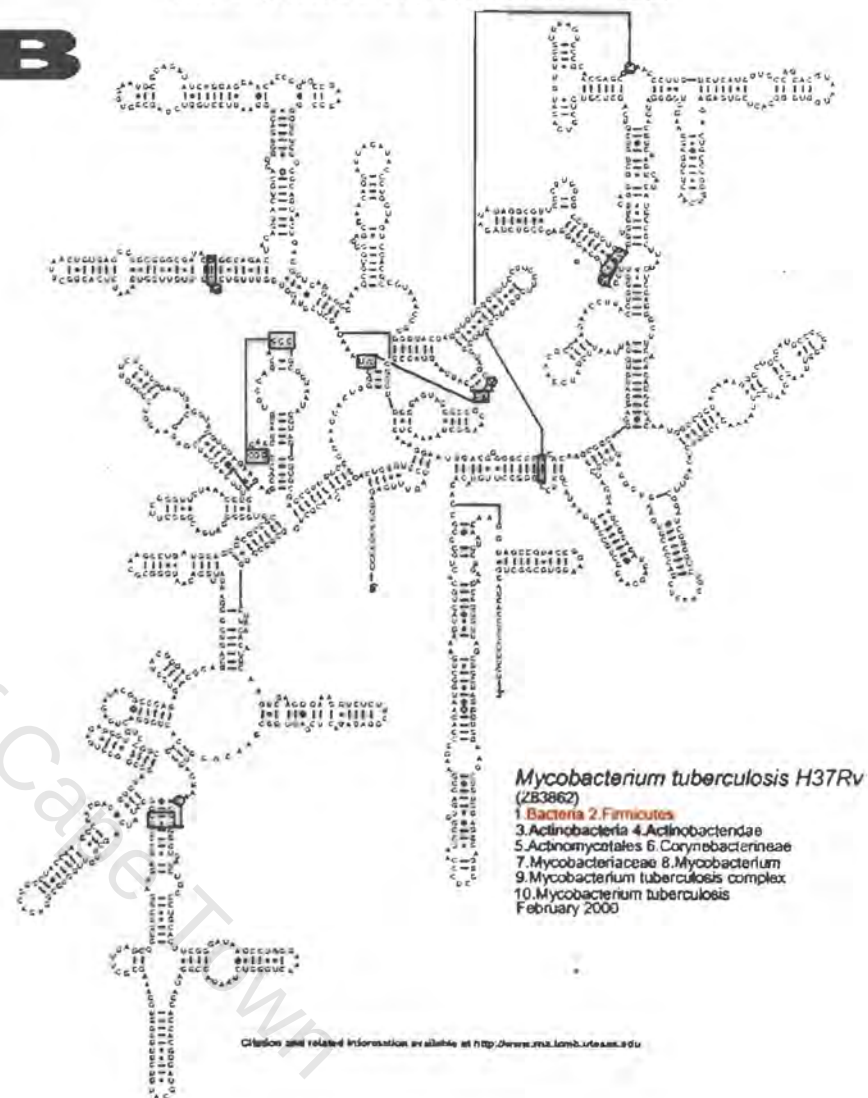


Figure 7.2: 16S rRNA (Gutell 1994). A) *E. coli*, B) *M. tuberculosis*.

7.3 Features of mRNA that affect translation efficiency

Several features of the mRNA that contribute to the various rate constraints of translational efficiency (Dreyfus, 1988; Gold, 1988; Mc Carthy and Brimacombe, 1994; deBoer and Hui, 1990; Gold and Stormo, 1990; Gualerzi and Pon, 1990; Mc Carthy and Gualerzi, 1990; Ringquest *et al.*, 1992) are:

(1) The accessibility of the mRNA. The ribosome binding site has been defined as that portion of mRNA protected from RNase digestion by ribosomes that have bound to mRNA but have not initiated translation (Steitz, 1969; Hindley and Staples, 1969; Gupta *et al.*, 1970). This site can be hidden from ribosomes by intramolecular base-pairing (secondary structure) (Gold, 1988; Stormo, 1986; de Smit and Van Duin, 1990a), or by bound proteins or by antisense RNA (translational repression) (Deckman and Draper, 1987). In this case k_1 is reduced (Figure 7.1). Conversely, k_1 is increased by mRNA sequences that have no intramolecular structure involving the Shine and Dalgarno sequence (SD).

(2) In the SD interaction, the 3' end of the 16S rRNA base-pairs to a complementary sequence usually present just downstream of the initiation start site (Shine and Dalgarno, 1974; Hui and de Boer, 1987; Jacob *et al.*, 1987). The forward rate constant is increased if the SD is long. k_1 is fast if the SD is short or does not include the central core sequence of the SD (Stormo *et al.*, 1982a; Stormo *et al.*, 1982b). In the absence of SD complementarity, as with leaderless genes (Winzeler and Shapiro, 1997), ribosomes still bind to RNA, presumably through protein-RNA interactions. Ribosomal protein S1, which appears to be the main contributor, has a preference for U-stretches of RNA that are found upstream of various well translated genes (Boni *et al.*, 1991). Additional ribosome-mRNA contacts have been proposed (Mc Carthy and Gualerzi, 1990). One such contact is the downstream box (DB) sequence which was originally described as a translational enhancer of several *E. coli* and bacteriophage mRNA's (Sprengart *et al.*, 1990). Another is the upstream box (UB), a conserved 13-base sequence present in the 5'-UTR of the cold-shock-inducible genes, *cspA*, *cspB*, *cspG*, and *cspI* of *E. coli* (Yamanaka *et al.*, 1999).

(3) k_2 is a function of the spacing between the SD and the initiation codon; the maximal rate is not markedly affected if the spacing is longer than the optimum, but

can reduce sharply if the spacing is too short ; k_2 will be large if the initiation codon is not involved in an intramolecular structure, and smaller if a structure is present.

(4) k_2 is made larger if the initiation codon is other than AUG, both because of the intrinsic weakness of the codon-anticodon pair and the action of IF3 (Berkhout *et al.*, 1986). k_2 increases in the order AUG>GUG>UUG (Adhin and van Duin, 1989; Ringquest *et al.*, 1992).

(5) k_3 is related to second codon choice and codon usage. AAA and GCU are frequently second codons in *E. coli*, and would increase k_3 . While a stable codon-anticodon interaction at the second codon may benefit initiation directly, elongation at the subsequent few codons will influence translational efficiency by determining the rate at which the start site is cleared for binding of the next ribosome (Liljenstrom and von Heijne, 1987; Chen and Inouye, 1990; Rosenberg *et al.*, 1993).

Based on the features of the mRNA required to form a ribosome binding site and to direct ribosomes to the correct translational start site (Table 7.1), translational efficiency is difficult to predict. For example, it is accepted that a strong SD interaction enhances ribosome binding and translational efficiency (Chapon, 1982; Busby and Dreyfus, 1983; Gillam *et al.*, 1984; Munson *et al.*, 1984; Hui and de Boer, 1987; Jacob *et al.*, 1987; Calogero *et al.*, 1988; Dalboge *et al.*, 1988; Hartz *et al.*, 1991; Ivey-Hoyle and Steege, 1992; Ringquest *et al.*, 1992). Nevertheless, it has been notoriously difficult to find a quantitative relationship between the degree of complementarity and extent of its effect. A more detailed description of each of the properties listed in Table 7.1 follow:

Table 7.1

Properties of mRNA that influence Translation efficiency			
		Positive	Negative
1	Secondary structure at RBS	No intramolecular structure	Translational repression e.g. Antisense or protein occlusion
2	SD/DB/UB sequence	High homology	Low homology or missing core sequences
3	SD/Initiation codon spacing	Optimal	Shorter than optimum
4	Initiation codon	AUG>GUG>UUG	
5	Second codon and codon usage	Abundant codons	Rare codons

Table 7.1: Summary of various properties of mRNA that influence translation efficiency.

7.3.1 Accessibility of RBS

Because translation initiation is governed by base pair interactions between the ribosome and the mRNA, it is not surprising that the efficiency of ribosome binding is primarily determined by the secondary structure of the messenger (de Smit and van Duin, 1990a). For example, a single nucleotide substitution can reduce the expression of an efficiently translated gene by 3 orders of magnitude, when it turns a mismatched base-pair in the native secondary structure into a match (de Smit and van Duin, 1990 b, 1993).

de Smit and van Duin (1994) examined the quantitative relationship between protein production and the stability of the secondary structure in the initiation region. The graphic representation of the quantitative relationship between relative expression (E)

and the stability of the secondary structure of the ribosome binding site (ΔG°_f) (Figure 7.3), demonstrates the importance secondary structure in the ribosome binding site of mRNA in determining translation efficiency. They concluded that the efficiency of translation is determined by the overall stability of the structure at the ribosome binding site, whether the initiation codon itself is base paired or not. Structures weaker than -6 kcal/mol usually do not reduce translational efficiency. Below this threshold, all systems show a tenfold decrease in expression for every -1.4 kcal/mol, as predicted from theory (de smit and van Duin, 1994). Based on the effects that intramolecular folding has on translational efficiency, it has not been possible to devise a universal sequence for the expression of heterologous genes.

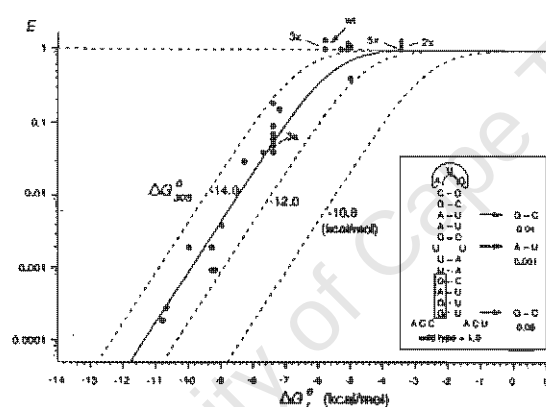


Figure 7.3. Graphic representation of the quantitative relationship between expression (E) and the stability of the secondary structure of the ribosome binding site (ΔG°_f). Filled circles indicate experimental data, obtained by site-directed mutagenesis of the coat-protein gene from bacteriophage MS2; where relevant the number of coinciding data points is given. Broken curves show the theoretical relationship for three ribosome-mRNA affinities (ΔG°_{30S}). A similar, continuous curve was fitted to the experimental data. ΔG°_{30S} values are for a concentration of free 30 S subunits of $8.5\mu\text{M}$. The insert shows the wild-type structure at the coat-protein start, as well as several base -pair changes and the corresponding expression level relative to the wild-type. The initiation codon and the SD sequence are boxed (de Smit and van Duin, 1994).

An example of this is the cold-shock response in *E. coli*. Here, a number of proteins, called the cold shock proteins, are transiently induced when a culture of *E. coli* is shifted from 37 to 15 or 10°C . Nine genes encoding CspA-like proteins, *cspA* to *cspI* have been identified (Yamanaka *et al.*, 1998). Among these cold-shock-inducible genes, *cspA* has been extensively investigated for the mechanism of its cold shock induction. It has been demonstrated that the *cspA* mRNA is extremely unstable at

37°C but becomes stable upon cold shock, indicating that the stability of mRNA places a crucial role in cold shock induction of *cspA* (Brandi *et al.*, 1996; Fang *et al.*, 1997; Goldenberg *et al.*, 1996). *cspA* expression is regulated in a complex manner at the level of transcription, mRNA stability, and translation (Yamanaka *et al.*, 1999). An important feature of the *cspA* mRNA is its unusually long 5' untranslated region (5'-UTR) consisting of 159 bases (Tanabe *et al.*, 1992). This feature is also shared with several other cold shock genes, which are dramatically induced after temperature downshift, such as *cspB* (Etchegaray *et al.*, 1996), *cspG* (Nakashima *et al.*, 1996), and *cspI* (Wang *et al.*, 1999). Mutation analysis of the 5'-UTR of the *cspA* mRNA in *E. coli* suggests that the secondary structure of the 5'-UTR at low temperatures, in particular surrounding the SD and UB sequences, might differ from that at 37°C (Yamanaka *et al.*, 1999). This structural change might make *cspA* mRNAs more accessible to ribosomes. It is thus possible that the 5'-UTR of the *cspA* mRNA might act as an RNA thermometer, as was recently proposed for the *rpoH* mRNA (Morita *et al.*, 1999).

Numerous studies, for example, those with *lamB* (Hall *et al.*, 1982), the *gal* operon (Queen and Rosenberg, 1981), the *trp* operon (Das *et al.*, 1983), λ *cro* (Iserentant and Fiers, 1980), and the MS2 lysis gene (Kastelein *et al.*, 1983), also suggest that mRNA translation is less efficient when the SD sequence or the AUG initiation codon is sequestered in a double-stranded region of the mRNA, thereby blocking the accessibility of these sequences to the ribosomes. Thus sequences outside the ribosome binding site can alter translational efficiency if they base pair with the SD sequence or the initiation codon.

7.3.2 16S rRNA interactions

7.3.2.1 SD Interactions

In most ribosome binding sites, the start codon is preceded by a purine-rich region at a distance of 5-12 bases. This so-called Shine-Dalgarno (SD) sequence (Shine and Dalgarno, 1974) shows a variable degree of complementarity with a region close to the 3' end of 16S rRNA; the anti-Shine-Dalgarno (ASD) sequence. The SD and ASD-containing sequences can be co-isolated from initiation complexes as an RNA

duplex (Shine and Dalgarno, 1974). The SD region is thought to assist the 30S particle in positioning itself at the proper place with respect to the start codon on the mRNA (Gold 1988). The presence of a SD, which is complementary to the ASD sequence, enhances the efficiency of initiation greatly (de Boer *et al.*, 1985). Lack of complementarity still allows for protein initiation to occur, but does so at reduced rates (Jacob *et al.*, 1987).

The endogenous mRNAs use a variety of SD sequences ranging from sequences like 5'-GGAGG, with perfect 5-base complementarity, to barely recognizable SD sequences or no SD sequence at all (Gren, 1984). Whether SD sequences that are classified as poor, i.e., with little or no complementarity with the ASD sequence, are still able to establish base pairing with the 3' end of the 16S rRNA molecule, is not known. An example of a mutation that maps within the SD sequence and which enhances SD-ASD complementarity has been reported by Chapon (1982). Here, it concerned a base change from 5'-GAAG to 5'-GGAG of the *malT* operon. This mutation increased *malT* mRNA translation more than 10-fold. Using a different approach, they varied the length of the SD region from 4 to 13 bases. For that purpose a series of DNA segments were synthesized which were introduced in front of the start codon (de Boer *et al.*, 1983). It appeared that the translatability of the mRNA did not increase as one might naively have expected, but decreased by about 50% with increasing SD length from 4 to 13 base pairs (Gold and Stormo, 1987). This observation suggested that base interaction must be sufficient for mRNA-rRNA binding to occur but should not exceed certain strength or else the ribosome may bind too tightly to the mRNA. As at the onset of elongation the SD-ASD interaction must be disrupted, a "too-tight" binding might impede the transition from initiation to elongation phase of protein synthesis.

7.3.2.2 Initiation codon interactions

It is widely accepted that translation of canonical prokaryotic mRNAs is achieved mainly through ribosomal recruitment signals present in the 5'-untranslated leader region. In contrast, the mechanism(s) leading to translation of leaderless mRNA, in which the start codon is either preceded by only a few nucleotides or which starts directly with a 5' terminal AU, has remained elusive.

In the ribosome-initiator-tRNA model of translation initiation, the 5'-terminal AUG is the only necessary and constant element of a leaderless mRNA that is recognized by the translational machinery (Moll *et al.*, 2002). Support for this premise comes from the studies which have demonstrated that the translational efficiency of mRNAs deprived of their 5'-UTR (including the SD sequence) depended exclusively on the presence of a 5'-AUG start codon (van Etten and Janssen, 1998). In fact, substitution of the 5'-AUG with the otherwise canonical initiation triplets GUG or UUG resulted in a complete loss or at least in a dramatic loss in translation (Moll *et al.*, 2002). As the ribosome-initiator-tRNA model relies entirely on base pairing between the anticodon of initiator-tRNA and the initiation triplet of leaderless mRNA, the above observation could be explained by an intrinsic instability of ternary initiation complexes built in the absence of ribosome-mRNA contacts at the 5'-side of a wobbly codon-anticodon interaction (Moll *et al.*, 2002). This explanation agrees with a recently published study wherein the translational efficiency of a λ cI-lacZ leaderless reporter constructs beginning with AUG, CUG, GUG, or UUG was determined (O' Donnell and Janssen, 2001). Only the construct with the AUG start codon supported a high level expression, whereas a low level of expression was obtained with GUG, and no expression at all with the constructs beginning with UUG or CUG. Nevertheless, the finding that after conversion of its AUG start codon to an amber codon, a leaderless λ cI-lacZ construct was not expressed even though an amber suppressor initiator-tRNA was co-expressed, has been interpreted as showing that it is not the codon-anticodon complementarity but, instead, the AUG codon *per se* which is required for translation initiation of leaderless mRNAs (van Etten and Janssen, 1998).

7.3.2.3 Alternative mRNA sequences involved in 16S rRNA recognition

The non-random distribution of nucleotides located upstream of the SD region and downstream of the initiation codon (Stormo *et al.*, 1982 b; Schneider *et al.*, 1986) indicates that other primary mRNA sequences might also be involved directly in the recognition process. Based on theoretical studies, additional recognition signals have been proposed after sequence comparison of translation initiation regions (Petersen *et al.*, 1988; Thanaraj and Pandit, 1989), but experimental support for these postulated interactions is still lacking.

7.3.2.3.1 The Upstream Box

The Upstream Box (UB) is a conserved 13-base sequence present in the 5'-UTR of the cold-shock-inducible genes, *cspA*, *cspB*, *cspG*, and *cspI* of *E. coli* (from base +123 to +135 of *E. coli cspA*, Figure 7.4). Interestingly, these 13-base sequences are located upstream of the SD sequence and contain a palindromic sequence to form a stable secondary structure ($\Delta G = -9.5$ kcal/mol) (Yamanaka *et al.*, 1999). The UB is also complementary to the region from base 1023 to 1035 of *E. coli* 16S rRNA (Figure 7.2a). Yamanaka and colleagues (1999) speculated that the UB sequence might function to punctuate the formation of a stable secondary structure immediately upstream of the SD sequence, allowing it to be highly accessible to ribosomes. Alternatively, it is possible that the UB sequence may be another *cis* element, which may enhance translation efficiency forming a duplex with the 16S rRNA in addition to the SD and the DB (Mitta *et al.*, 1997).

16S rRNA 991-1037	3'-CAAGGGCUUCCGUGUAAGAGUAGAGACUUUUGAAGGCACCUACAGUU-5'
<i>cspA</i> +123-+162	5'-GCCG AAAGGCACAC UUAAUUAAUUAAGGUAUUACACUAUG- 3'
<i>cspB</i> +133-+163	5'-GCCG AAAGGCUCA -----AGUU AAGGAA -UGUAGAAUG- 3'
<i>cspG</i> +125-+158	5'-GCCG AAAGGCCCA -----AAAUG AAGGAAGU AAAAU AUG - 3'
<i>cspI</i> +124-+148	5'-GCCG AAAGGCUAA UUUCUCAA-- AAGGUAUA -AUUAUG- 3'

Figure 7.4. Sequence similarities of *cspA*, *cspB*, *cspG*, and *cspI* mRNAs around the SD sequence and potential base pairing between *cspA* mRNA and 16S rRNA. Nucleotide numbers of *cspA*, *cspB*, *cspG*, and *cspI* mRNA are given starting from the major transcription initiation site as +1. Nucleotides identical in the four *csp* mRNAs are shown in boldface. The 13 base homologous sequences in *cspA*, *cspB*, *cspG*, and *cspI* are in red (UB). Positions of the SD sequence and the initiation codon are underlined. Potential base pairings between *cspA* mRNA and 16S rRNA are indicated by vertical lines (Yamanaka *et al.*, 1999).

7.3.2.3.2 The Downstream Box

The importance of a specific region located just downstream of the initiation codon was determined by expression studies with the highly active RBS of the bacteriophage

T7 gene 0.3 (Sprengart *et al.*, 1990). The efficiency of this RBS was strongly dependent on nucleotides +15 to +26 ('downstream box'; DB) of the coding region, which are complementary to a region in 16S rRNA (Sprengart *et al.*, 1990). Similar sequences complementary to 16S rRNA nucleotides 1469-1483 ('anti-downstream box'; ADB) are located in many other *E. coli* and bacteriophage mRNA downstream of the start codon, and therefore interactions between these regions were proposed (Sprengart *et al.*, 1990). Subsequently, the presence of a DB has been reported for an increasing number of *E. coli* and bacteriophage genes (Faxen *et al.*, 1991; Nagai *et al.*, 1991; Shean and Gottesman, 1992; Helke *et al.*, 1993; Ito *et al.*, 1993; Pohlner *et al.*, 1993; Lange and Hengge-Aronis, 1994), but up until now it has been regarded as a translational enhancer element of minor importance compared with the SD sequence (McCarthy and Brimacombe, 1994).

7.3.3 Spacing between the SD sequence and the initiation codon

Spacing between the SD sequence and the initiation codon would not be expected to affect binding of the message to the 30S subunit but would affect the rate of the first-order rearrangement to an initiation complex with the start codon base paired to the anticodon of the tRNA. Thus, longer spacings would require some looping out of intervening sequences, while shorter spacings might even involve melting of the SD-16S rRNA duplex. Gold (1988) has described a message with a spacing of 2 bases (counting from the most 3' base that can base pair to the 16S rRNA, to the first position of the initiation codon) that is cold sensitive for translation in *E. coli*, suggesting such a scenario (Gold and Stormo, 1987). Spacing effects have been tested experimentally in *E. coli* (Gheysen and Fiers, 1982; Thomas *et al.*, 1982), but an interpretation of the results in many cases is complicated by differences in sequence context and potential secondary structures. Hartz *et al.* (1991) constructed a series of plasmids containing the SD sequence at various positions from the initiation codon separated by all A's to minimize secondary-structure effects. They found that in *E. coli*, spacing in the range of 6 to 10 nucleotides had a minor effect on translational yield. Using the same constructs Vellanoweth and Rabinowitz (1992) found similar results for *Bacillus subtilis*. The optimum spacing in both organisms was around 7 to 9 bases, which is consistent with the average spacing found in sequenced genes (Hager and Rabinowitz, 1985).

7.3.4 Initiation codon

A major difference between *E. coli* and mycobacteria lies in the choice of initiation codon. While 87% of 5056 sequenced *E. coli* genes start with AUG (Table 7.2), 36% of 3231 *M. tuberculosis*, 30% of 77 *M. bovis*, and 41% of 95 *M. smegmatis* genes sequenced start with GUG. Similarly, nearly 30% of *B. subtilis* and other clostridial branch genes start with UUG or GUG (Hager and Rabinowitz, 1985). Mutations of an AUG initiation codon to GUG to UUG have generally resulted in decreased expression in *B. subtilis*, as has been found in *E. coli* (Reddy *et al.*, 1985). Ambulos *et al.* (1986) found the order AUG>UUG>GUG in *B. subtilis* which contrasts with the relative order AUG>GUG>UUG in *E. coli*. These orders of preference correlate with the relative codons in both organisms (Gold *et al.*, 1981; Hager and Rabinowitz, 1985), but the significance of their differential use is unclear.

7.3.5 2nd and subsequent codons

Most amino acids can be coded for by more than one codon, and the relative frequency with which these different synonymous codons occur is referred to as the codon usage. In an organism like *E. coli*, many genes seem to have very little preference for particular codons (Table 7.3), i.e., there is little codon bias. However, highly expressed genes in *E. coli* show a higher degree of codon bias, which reflects the different levels of the relevant tRNA species that are specific for the same amino acid but with different codon specificity (Dale and Patki, 1990). With mycobacteria, all genes sequenced so far show a relatively high degree of codon bias, which is reflected in a preponderance of those codons with G or C at position 3. Many authors in respect to individual genes have commented on this effect on, and it is apparent from the codon usage table for *M. smegmatis* (Table 7.4). The effect is much more pronounced for *M. tuberculosis* than for *M. leprae*, which is consistent with the lower G + C content of the DNA of the latter. For example, of the four available alanine codons, the two "G/C" codons, GCG and GCC account for 82% of the alanine residues in *M. tuberculosis* genes, while in *M. leprae* the distribution of the same alanine codons makes up 66% of alanine residues. In *M. smegmatis*, which has a similar G + C content to *M. tuberculosis*, "G/C" alanine codons accounts for 88% and in *M. bovis* 81% of the total alanine residues. In *E. coli* this figure drops to 55%.

Table 7.2: Average nucleotide frequency at the translation initiation region

	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13
E. coli																																	
%G	20	18	21	21	19	22	24	23	29	34	37	42	35	29	26	23	21	24	19	19	13	0	100	21	17	20	23	14	22	21	15	19	23
%A	33	32	28	32	32	32	33	33	31	33	32	27	32	33	30	31	32	34	26	29	87	0	0	47	35	31	41	39	31	36	30	32	35
%T	25	28	29	25	27	27	24	24	22	16	17	18	19	22	26	26	25	22	31	29	0	100	0	14	22	30	16	27	28	21	33	29	19
%C	22	22	22	21	23	20	19	19	18	17	14	12	15	16	17	20	22	20	23	23	0	0	0	18	26	19	20	19	19	23	21	20	22
Total	5056																																
M. tb																																	
%G	33	29	33	33	30	36	35	31	35	40	40	41	39	35	37	32	32	35	25	33	36	0	100	27	22	30	38	20	30	33	20	33	36
%A	21	22	18	20	21	18	22	24	23	26	22	22	21	20	20	20	20	18	18	17	64	0	0	40	16	14	24	26	15	25	21	15	23
%T	16	18	18	17	19	18	16	18	17	13	15	15	16	20	17	17	17	15	19	15	0	100	0	12	16	17	14	25	18	13	26	16	14
%C	31	31	31	30	30	29	27	26	25	22	23	22	23	26	26	30	31	32	37	35	0	0	0	21	45	38	24	30	37	29	33	35	26
Total	3231																																
M. bovis																																	
%G	36	27	21	29	30	33	39	35	44	49	45	40	30	40	34	24	19	39	18	36	30	0	100	36	17	27	38	21	43	30	14	37	28
%A	23	28	26	29	28	28	28	33	28	22	26	23	23	17	28	25	31	22	20	13	70	0	0	34	17	23	30	30	20	31	27	17	31
%T	18	13	14	14	15	20	11	14	6	11	17	16	24	20	18	20	13	6	24	13	0	100	0	7	16	19	10	23	18	9	32	21	16
%C	22	32	40	28	27	20	23	19	21	17	12	21	23	23	20	30	37	33	39	38	0	0	0	23	51	31	22	26	19	30	27	26	26
Total	77																																
M. smeg																																	
%G	35	23	27	35	33	41	37	27	45	41	42	52	36	30	35	23	36	35	21	39	41	0	100	25	21	30	47	17	35	31	15	42	30
%A	24	25	19	19	17	16	27	26	21	28	23	20	19	25	23	30	21	15	22	17	59	0	0	39	10	10	21	33	14	26	24	9	28
%T	19	14	17	22	17	18	14	13	15	12	13	9	20	21	20	18	14	13	17	8	0	100	0	10	13	21	12	18	15	15	21	13	8
%C	22	38	37	24	34	25	22	34	19	19	23	20	26	25	23	30	29	36	39	36	0	0	0	25	55	38	21	33	36	28	40	36	34
Total	95																																

These tables also reveal some more specific features, such as the almost complete absence of the AGA/AGG codons for arginine, the former of which makes up only 1.3% of arginine codons in *M. tuberculosis*, 0.7% in *M. smegmatis*, 2% in *M. bovis* and 2.8% in *M. leprae*. In all three species, the favoured arginine codon is CGC by a considerable margin, accounting for 39% in *M. tuberculosis*, 51% in *M. smegmatis*, 42% in *M. bovis*, and 31% in *M. leprae*.

Codons used most frequently include those with high G/C content and include: Ala (GCC) > Gly (GGC) > Leu (CTG) > Ala (GCG) in *M. tuberculosis*, Ala (GCC) > Leu (CTG) > Asp (GAC) > Ala (GCG) in *M. smegmatis*, Ala (GCC) > Gly (GGC) > Leu (CTG) > Asp (GAC) in *M. bovis*, and Ala (GCC) > Val (GTG) > Leu (CTG) > Asp (GAC) in *M. leprae*. Rare codons follow a similar pattern in that they are predominantly consist of high A/T codons: Arg (AGA) < Leu (TTA) < Ile (ATA) < Ser (TCT) in *M. tuberculosis*, Leu (TTA) < Ile (ATA) < Arg (AGA) < Phe (TTT) in *M. smegmatis*, Arg (AGA) < Leu (TTA) < Ser (TCT) < Ile (ATA) in *M. bovis*, and Arg (AGA) < Arg (AGG) < Cys (TGT) < Ile (ATA) in *M. leprae*. *M. leprae*, which has a lower G + C content, has the AGG arginine codon listed as the second most rare codon whilst in *M. tuberculosis* it is the sixth least used codon and in *M. smegmatis* it is only the fifteenth most rare codon.

E. coli codon usage				
AmAcid	Codon	Number	/1000	Fraction
Gly	GGG	20861.00	11.52	0.16
Gly	GGA	16779.00	9.27	0.13
Gly	GGT	43952.00	24.27	0.33
Gly	GGC	50278.00	27.77	0.38
Glu	GAG	33616.00	18.56	0.33
Glu	GAA	69547.00	38.41	0.67
Asp	GAT	58837.00	32.49	0.63
Asp	GAC	34351.00	18.97	0.37
Val	GTG	45263.00	25.00	0.36
Val	GTA	20355.00	11.24	0.16
Val	GTT	34077.00	18.82	0.27
Val	GTC	26618.00	14.70	0.21
Ala	GCG	56120.00	30.99	0.34
Ala	GCA	36551.00	20.19	0.22
Ala	GCT	28592.00	15.79	0.17
Ala	GCC	44697.00	24.68	0.27
Arg	AGG	3699.00	2.04	0.04
Arg	AGA	5906.00	3.26	0.06
Ser	AGT	17547.00	9.69	0.16
Ser	AGC	28397.00	15.68	0.26
Lys	AAG	21254.00	11.74	0.26
Lys	AAA	61896.00	34.18	0.74
Asn	AAT	36143.00	19.96	0.48
Asn	AAC	38512.00	21.27	0.52
Met	ATG	48845.00	26.97	1.00
Ile	ATA	11428.00	6.31	0.10
Ile	ATT	54201.00	29.93	0.50
Ile	ATC	43396.00	23.97	0.40
Thr	ACG	25662.00	14.17	0.26
Thr	ACA	15522.00	8.57	0.16
Thr	ACT	17438.00	9.63	0.18
Thr	ACC	39857.00	22.01	0.40
Trp	TGG	26978.00	14.90	1.00
End	TGA	1916.00	1.06	0.33
Cys	TGT	9709.00	5.36	0.46
Cys	TGC	11597.00	6.40	0.54
End	TAG	519.00	0.29	0.09
End	TAA	3452.00	1.91	0.59
Tyr	TAT	31392.00	17.34	0.58
Tyr	TAC	22273.00	12.30	0.42
Leu	TTG	24407.00	13.48	0.13
Leu	TTA	26387.00	14.57	0.14
Phe	TTT	41488.00	22.91	0.58
Phe	TTC	29468.00	16.27	0.42
Ser	TCG	15875.00	8.77	0.14
Ser	TCA	15356.00	8.48	0.14
Ser	TCT	17040.00	9.41	0.15
Ser	TCC	15820.00	8.74	0.14
Arg	CGG	11238.00	6.21	0.11
Arg	CGA	7227.00	3.99	0.07
Arg	CGT	35692.00	19.71	0.35
Arg	CGC	37411.00	20.66	0.37
Gln	CAG	51354.00	28.36	0.65
Gln	CAA	27184.00	15.01	0.35
His	CAT	23467.00	12.96	0.58
His	CAC	17147.00	9.47	0.42
Leu	CTG	88767.00	49.02	0.47
Leu	CTA	7831.00	4.32	0.04
Leu	CTT	21339.00	11.78	0.11
Leu	CTC	19335.00	10.68	0.10
Pro	CCG	39251.00	21.68	0.50
Pro	CCA	15766.00	8.71	0.20
Pro	CCT	13673.00	7.55	0.17
Pro	CCC	10227.00	5.65	0.13

Table 7.3: Codon usage in *E. coli*. These data were produced at the Sanger Centre and can be obtained from <ftp://sanger.otago.ac.nz/pub/Transtern/Data/codons/bct/Esccoli.cod>

<i>M. smegmatis</i> codon usage				
AmAcid	Codon	Number	/1000	Fraction
Gly	GGG	551.00	11.09	0.13
Gly	GGA	334.00	6.72	0.08
Gly	GGT	926.00	18.63	0.22
Gly	GGC	2479.00	49.88	0.58
Glu	GAG	2074.00	41.73	0.76
Glu	GAA	659.00	13.26	0.24
Asp	GAT	628.00	12.64	0.19
Asp	GAC	2657.00	53.46	0.81
Val	GTG	2205.00	44.37	0.49
Val	GTA	92.00	1.85	0.02
Val	GTT	208.00	4.19	0.05
Val	GTC	1973.00	39.70	0.44
Ala	GCG	2550.00	51.31	0.42
Ala	GCA	474.00	9.54	0.08
Ala	GCT	265.00	5.33	0.04
Ala	GCC	2796.00	56.26	0.46
Arg	AGG	122.00	2.45	0.03
Arg	AGA	35.00	0.70	0.01
Ser	AGT	104.00	2.09	0.04
Ser	AGC	607.00	12.21	0.24
Lys	AAG	1097.00	22.07	0.88
Lys	AAA	155.00	3.12	0.12
Asn	AAT	81.00	1.63	0.07
Asn	AAC	1004.00	20.20	0.93
Met	ATG	880.00	17.71	1.00
Ile	ATA	17.00	0.34	0.01
Ile	ATT	61.00	1.23	0.03
Ile	ATC	2071.00	41.67	0.96
Thr	ACG	992.00	19.96	0.31
Thr	ACA	176.00	3.54	0.06
Thr	ACT	91.00	1.83	0.03
Thr	ACC	1895.00	38.13	0.60
Trp	TGG	676.00	13.60	1.00
End	TGA	79.00	1.59	0.69
Cys	TGT	66.00	1.33	0.19
Cys	TGC	274.00	5.51	0.81
End	TAG	24.00	0.48	0.21
End	TAA	11.00	0.22	0.10
Tyr	TAT	136.00	2.74	0.13
Tyr	TAC	929.00	18.69	0.87
Leu	TTG	367.00	7.38	0.08
Leu	TTA	8.00	0.16	0.00
Phe	TTT	44.00	0.89	0.03
Phe	TTC	1409.00	28.35	0.97
Ser	TCG	1104.00	22.21	0.43
Ser	TCA	99.00	1.99	0.04
Ser	TCT	53.00	1.07	0.02
Ser	TCC	577.00	11.61	0.23
Arg	CGG	935.00	18.81	0.24
Arg	CGA	273.00	5.49	0.07
Arg	CGT	537.00	10.80	0.14
Arg	CGC	2001.00	40.26	0.51
Gln	CAG	1339.00	26.94	0.89
Gln	CAA	161.00	3.24	0.11
His	CAT	195.00	3.92	0.17
His	CAC	933.00	18.77	0.83
Leu	CTG	2734.00	55.01	0.57
Leu	CTA	47.00	0.95	0.01
Leu	CTT	162.00	3.26	0.03
Leu	CTC	1503.00	30.24	0.31
Pro	CCG	1576.00	31.71	0.57
Pro	CCA	114.00	2.29	0.04
Pro	CCT	137.00	2.76	0.05
Pro	CCC	938.00	18.87	0.34

Table 7.4: Codon usage in *Mycobacterium smegmatis*. These data were produced at the Sanger Centre and can be obtained from <ftp://sanger.otago.ac.nz/pub/Tranterm/Data/codons/bct/Mycsme.cod>

7.4 Role of Initiation Factors

Translation initiation factor IF3 acts as a fidelity factor by promoting the dissociation of aminoacyl-tRNAs other than initiator fMet tRNA^{Met} from the ribosomal P-site, and by discriminating against non-canonical start codons, such as AUU (Gualerzi *et al.*, 2000). Recent data (Petrelli *et al.*, 2001) clearly indicate that this discrimination is not based on a direct “inspection” of the codon-anticodon interaction by IF3 (Hartz *et al.*, 1990), but is indirectly caused by the effect of this factor on the conformational dynamics of the 30S ribosomal subunit. Previous studies have shown that IF3 is a negative effector of translation initiation complex formation on 5' terminal start codons and that an artificial increase of the intracellular IF3 concentration inversely correlates with the translational efficiency of a leaderless reporter construct (Tedin *et al.*, 1999).

IF2 stimulates fMet tRNA^{Met} (initiator tRNA) binding to the ribosomal P-site (Gualerzi *et al.*, 2000). The stimulation of leaderless mRNA translation seems to be exerted by IF2 not only through the promotion of 30S-initiator-tRNA complex formation, but also by conferring an increased stability to the initiation complex (Grill *et al.*, 2001). Most probably, this effect of IF2 is mediated through interactions of the factor with the ribosome and/or with fMet tRNA^{Met} on the ribosome (Gualerzi *et al.*, 2000).

A role for IF1 becomes apparent only when 70S ribosomes are used instead of 30S subunits. Neither IF2 and IF3 together nor IF1 alone favours fMet- tRNA^{Met} with 70S subunits, but the combination of all three factors promotes fMet- tRNA^{Met}-containing complexes (Hartz *et al.*, 1991). A number of other studies indicate that the initiation factors do not effect mRNA-30S subunit binding affinity, but do alter the kinetics of complex formation and dissociation, the positioning of the mRNA within the subunit, and selectivity for initiator tRNA (Gualerzi and Pon 1990).

7.5 Conclusion

Much of our current knowledge concerning the process of translation in prokaryotes comes from extensive studies of one organism, *E. coli*. These investigations have provided detailed insights into many aspects of protein synthesis, including the mechanism and kinetics of initiation, the cyclical reactions involving transfer RNA, protein factors, GTP, and ribosomal proteins in elongation; and the fine structure and relationships of proteins and RNA in the ribosomal machinery itself (Noller and Nomura, 1996). Results obtained through the use of *E. coli* have provided the basis of the genetic code, the recognition of specific mRNA signals directing ribosomes to initiation points, and insights into subtle translational regulatory mechanisms. Although this wealth of knowledge has unravelled many of the mysteries concerning the transfer of the one-dimensional information inherent in DNA to three-dimensional active molecules, biochemical evidence suggests that some of the specific conclusions cannot be extrapolated to all other bacteria. For example, not all bacteria contain the same number of ribosomal proteins. One ribosomal protein (S1) missing in many bacteria is known to play a critical role in initiation in *E. coli*, suggesting that fundamental processes in translation may differ among the bacteria. Another indication of such differences is that mRNAs cannot be freely exchanged among all bacteria and still result in efficient recognition and translation. Lodish (1969) was the first to observe that ribosomes from *E. coli* and *Bacillus stearothermophilus* do not translate f2 coliphage RNA into three products equally well. Not only was overall incorporation by *B. stearothermophilus* only 5% that of *E. coli*, but the products were made in vastly different relative amounts by the two different ribosomes. It was further demonstrated that the inability of *B. stearothermophilus* ribosomes to efficiently translate the f2 message was a function of the small ribosomal subunit; neither the source of the crude initiation factors nor that of large ribosomal subunit had any effect on the observed translational selection (Lodish, 1970).

7.6.1 Aims of this study

During homology searches of the initiation regions of bacteria in general and mycobacteria in particular, a conserved region within the coding region was identified. This region was identified as the downstream box, a controversial region thought to be involved in ribosome recognition and subsequent translation initiation. This region corresponds with the anti-downstream box in the 16S rRNA, which is identical in all mycobacteria. This however represents the region of highest sequence variation between *E. coli* and mycobacteria. The aim of this study is to determine the significance of this and to ask the question: Is translation in the absence of a Shine-Dalgarno region possible in the presence of a downstream box?

Chapter 8

Translation augmentation by downstream box sequences in *E. coli* and *M. smegmatis* in the absence of a Shine Dalgarno sequence

Contents	page
8.1 Summary	149
8.2 Introduction	149
8.3 Materials and Methods	153
8.3.1 Bacterial strains and plasmids	153
8.3.2 DNA manipulations	153
8.3.3 Plasmid constructs	155
8.3.4 Electroporation	156
8.3.5 <i>In vitro</i> coupled transcription/translation	156
8.3.6 Luciferase assays	156
8.3.7 Secondary structure prediction	157
8.4 Results and Discussion	160
8.4.1 16S rRNA analysis	160
8.4.1.1 DB design	161
8.4.2 Expression studies in <i>E. coli</i> and <i>M. smegmatis</i>	162
8.4.3 <i>In vitro</i> transcription/translation analysis	165
8.5 Conclusion	165

Chapter 8

Translation augmentation by Downstream Box sequences in *Escherichia coli* and *Mycobacterium smegmatis* in the absence of a Shine Dalgarno sequence

8.1 Summary

The downstream box is a putative translational enhancer of several *Escherichia coli* and bacteriophage mRNA's. It is located within the coding region of a gene, downstream of the initiation codon. Here, we investigate the effect on translation of downstream box sequences optimised for binding to *Mycobacterium smegmatis* and *Escherichia coli* 16S rRNA in the absence of a Shine-Dalgarno region. The results are presented as relative translational efficiency of each construct in either *Mycobacterium smegmatis* or *Escherichia coli*. The influence of altered mRNA secondary structure due to downstream box optimisation on ribosome accessibility to the initiation codon between constructs is therefore constant relative to each host. Eradication of the Shine-Dalgarno region in the absence of a downstream box abolished the translation activity. In contrast, optimised downstream box constructs promoted efficient protein synthesis in the absence of a Shine-Dalgarno region.

8.2 Introduction

Initiation of protein synthesis in *E. coli* requires the mRNA RBS to direct the ribosome to the appropriate translational start site (Dreyfus, 1988; Gold, 1988; McCarthy and Brimacombe, 1994). Two well-characterized elements of the *E. coli* RBS are the initiation codon and the SD region, a purine-rich sequence 3-10 bp upstream of the coding region that is complementary to the 3' end of the 16S rRNA (Shine and Dalgarno, 1974; Gold, 1988). However, the identification of leaderless genes (Winzeler and Shapiro, 1997), and the non-random distribution of nucleotides

located upstream of the SD region and downstream of the initiation codon (Stormo *et al.*, 1982b; Schneider *et al.*, 1986) may indicate that other primary mRNA sequences form part of the translation initiation region (TIR). These regions have been shown to be A-rich (Rocha *et al.*, 1999) and therefore may not form stable mRNA secondary structures that influence translation initiation (de Smit and van Duin, 1994).

The downstream box (DB) is a 12-15 base sequence located downstream of the mRNA initiation codon and is complementary to bases 1469-1483 (anti-downstream box – aDB) of the *Escherichia coli* 16S rRNA (Sprengart *et al.*, 1990; Sprengart *et al.*, 1996, Figure 8.1). Although direct biochemical evidence for a DB-aDB interaction is lacking, there are numerous studies implicating this interaction in the initiation of translation: i) The efficiency of expression of the bacteriophage T7 gene 0.3 was strongly dependent on nucleotides +15 to +26 of the coding region (Sprengart *et al.*, 1990); ii) as the DB-aDB match was increased from 10 to 12 - 15 bases, translation increased concurrently (3-8 fold) following induction of the gene (Etchegaray and Inouye, 1999); iii) affinity enrichment of DB sequences in an mRNA pool occurred when the 30S ribosomal subunit was used as the ligand (Ringquist *et al.*, 1995)

In *E. coli*, the DB has been proposed to enhance translation in the absence of a Shine-Dalgarno (SD) region (Sprengart *et al.*, 1990; Sprengart *et al.*, 1996). Furthermore, leaderless mRNAs are translationally active in the absence of a SD region (Winzeler and Shapiro, 1997; Janssen *et al.*, 1989; Klock and Hillen, 1986; Shean and Gottesman, 1992) suggesting that there are additional translational signals downstream of the initiation codon.

It has been suggested that enhancement of translation by the DB does not involve mRNA-rRNA base pairing (O'Connor *et al.*, 1999), yet enrichment of SD and DB by SELEX, an *in vitro* selection-technique in which large random pools of nucleic acids can be screened for a particular functionality, using S1-depleted 30S ribosomal subunits, demonstrated an affinity for sequences containing DB (Ringquist *et al.*, 1995; Etchegaray *et al.*, 1998). Also, SELEX targeting ribosomal protein S1, known to be required for translation initiation, enriched DB-like sequences (Ringquist *et al.*, 1995) suggesting that S1 protein facilitates the interaction between mRNA and

ribosomes, presumably by increasing the local concentration of the AUG codon in the decoding region of the 16S rRNA (Tedin *et al.*, 1999). The affinity of the ribosome for the DB-containing mRNA enhances translation (Sprengart *et al.*, 1996).

It has been shown that shifting *E. coli* cells to a low temperature results in a global inhibition of cellular protein synthesis (Jones and Inouye, 1994). However, under these conditions, maximal levels of cold shock proteins are observed. Therefore, it has been speculated that the cold shock mRNAs, unlike the majority of cellular mRNAs are capable of bypassing the protein synthesis inhibitory effect triggered by cold shock (Jones and Inouye, 1996). In this regard, Jones and Inouye (1996) have proposed that cold shock mRNAs may be efficiently translated due to a higher affinity of these mRNAs for the ribosome at low temperature. Recently, it has also been demonstrated that artificial DB sequences inserted in the *lacZ* gene cause a significant translational enhancement (Etchegaray and Inouye, 1999).

However, evidence in support of the DB-aDB interaction from biochemistry, mRNA-rRNA co-variation and rRNA mutagenesis is lacking. Chemical probing studies have failed to show protection of the putative DB of the leaderless λ cI mRNA bound in 70S initiation complexes (Resch *et al.*, 1996). Moreover, the 3D crystal structure of the *Thermus thermophilus* 30S subunit (Wimberley *et al.*, 2000) revealed that the whole shoulder of the body of the ribosomal particle is situated between the putative DB of the mRNA and that part of helix 44 of 16S rRNA comprising the proposed aDB (Moll *et al.*, 2001). This would render any model in which the start codon is placed in the ribosomal P-site when the adjacent DB interacts with the aDB (Sprengart and Porter, 1997) or in which the DB base pairs synergistically with the SD—anti-SD interaction (Sprengart *et al.*, 1996) untenable. Also, neither natural λ cI mRNA nor a derivative of this leaderless mRNA with an optimized basepairing potential with the aDB formed a binary complex with ribosomes (Moll *et al.*, 2001). This result was consistent with previous kinetic toeprint experiments (Resch *et al.*, 1996), as well as the findings that the aDB is not accessible to a complementary DNA oligonucleotide, and that the aDB *in situ* is in a double-stranded conformation not amenable to basepairing (La Teana *et al.*, 2000).

This study evaluates the expression of luciferase by constructs with their DBs optimized for two unrelated bacterial hosts, *Mycobacterium smegmatis* and *Escherichia coli*. Expression of the constructs, which lacked functional SD sequences, was most efficient in the host for which the DB had been optimized.

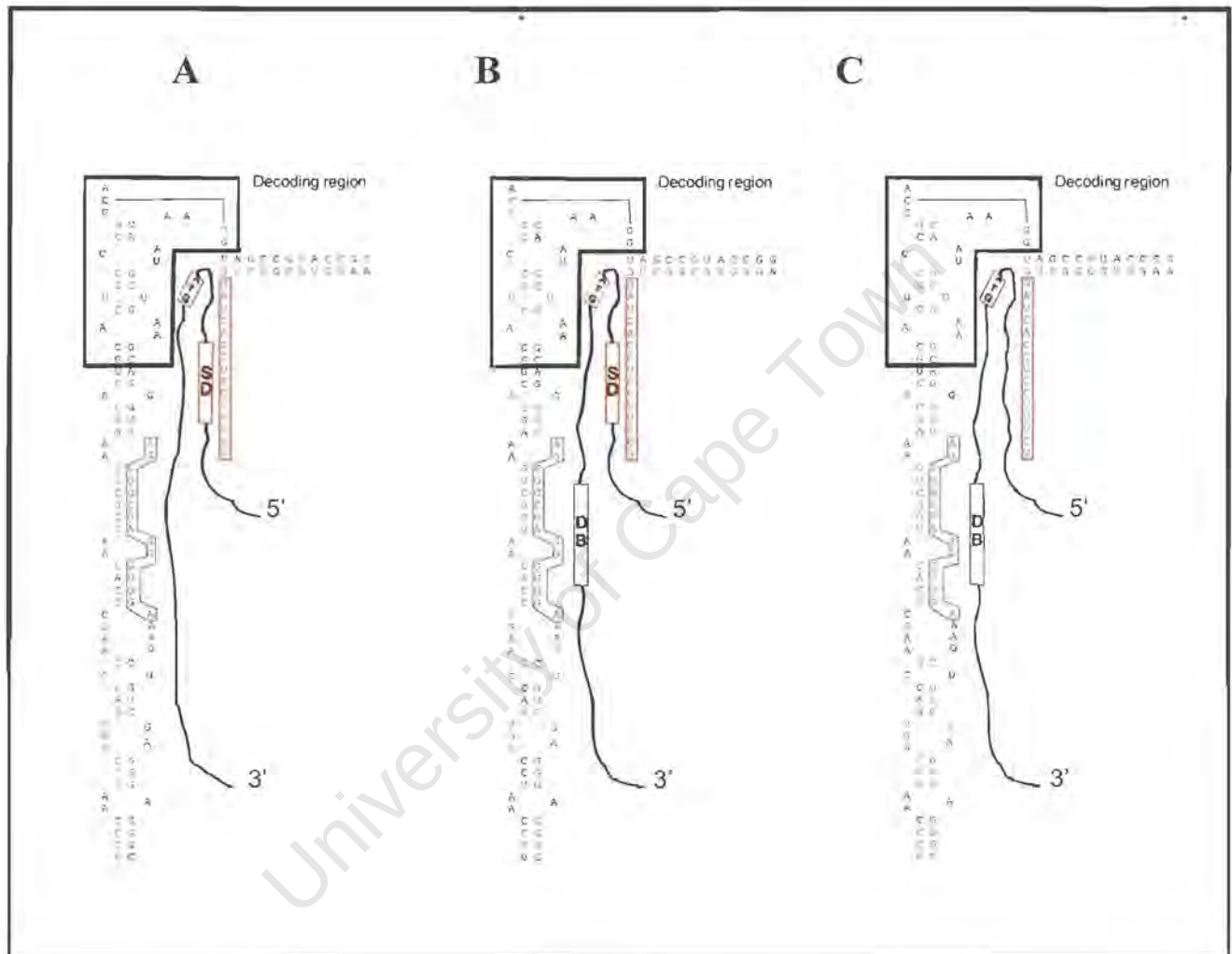


Figure 8.1

Schematic representation of mRNA and 16S rRNA during translation initiation at the penultimate stem-and-loop of 16S rRNA. The Decoding region, where interactions between initiator tRNA and start codons occur, is boxed (Black). The initiation codon is highlighted in brown. A) SD interaction of mRNA with the 16S rRNA anti-SD region (Red box). B) Downstream box (DB) interaction of mRNA with the 16S rRNA anti-downstream box (Green box) region in the presence of a SD. C) SD-independent, DB-mediated interaction of mRNA with the 16S rRNA aDB region (Sprengart *et al.*, 1996). The 16S rRNA sequence shown here is that of *Mycobacterium bovis* BCG.

8.3 Materials and Methods

8.3.1 Bacterial strains and plasmids

The bacterial strains, plasmids and PCR templates used in this study are summarized in Table 8.1. All cloning steps were performed in *E. coli* LKIII (Zabeau and Stanley, 1982) grown in 2YT medium (Miller, 1992). *In vivo* expression studies were performed in *E. coli* LKIII (Zabeau and Stanley, 1982), and *M. smegmatis* LR222, a high-frequency transforming strain (Beggs *et al.*, 1995). Mycobacterial strains were cultivated in Middlebrook 7H9 medium (Difco) supplemented with catalase-free ADC (albumin and dextrose) and 0.05% (v/v) Tween 80, or on Middlebrook-ADC agar plates. pJCLucKatG was described previously (Mulder *et al.*, 1999a).

8.3.2 DNA Manipulations

Standard recombinant DNA methods were performed by previously described protocols (Ausubel *et al.*, 1987; Sambrook *et al.*, 1989). DNA fragments were sequenced by the dideoxy sequencing protocol (Sanger *et al.*, 1977) using a T7 Polymerase Sequencing kit (Pharmacia), and a synthetic oligonucleotide (5'-GCGGAGGTCATCTACTGG-3') complementary to the *katG* promoter.

Table 8.1

Bacterial strains and constructs used in this study			
Bacterial strains	Genotype		Reference
<i>E. coli</i> LK111	<i>recA, lacI, lacZ, lacY⁺</i>		Zabeau and Stanley, 1982
<i>M. smegmatis</i> LR222	High frequency transformation mutant		Beggs <i>et al.</i> , 1995
Plasmid constructs	Size (bp)	Marker	Description
pJCLuc	7420	Kan ^R	<i>E. coli</i> -mycobacterial shuttle vector containing luciferase gene (pGem®-luc)
pJCLucKatG	7956	Kan ^R	pJCLuc with <i>M. tb katG</i> promoter upstream of luciferase gene
pJCLucKatGΔSD	7945	Kan ^R	pJCLucKatG with SD sequence upstream of <i>luc</i> gene removed
pEcLuc	7984	Kan ^R	SD upstream of <i>luc</i> removed and DB optimized for binding to <i>E. coli</i> 16S rRNA
pMsLuc	7984	Kan ^R	SD upstream of <i>luc</i> removed and DB optimized for binding to <i>M. smegmatis</i> 16S rRNA
PCR templates	Size (bp)	Description	
T7LucΔSD	1803	pJCLucKatGΔSD template with <i>katG</i> promoter replaced with T7 promoter.	
T7Luc	1814	pJCLucKatG template with <i>katG</i> promoter replaced with T7 promoter	
T7EcLuc	1842	pEcLuc template with <i>katG</i> promoter replaced with T7 promoter	
T7MsLuc	1842	pMsLuc template with <i>katG</i> promoter replaced with T7 promoter	

8.3.3 Plasmid constructs

Plasmids containing optimised DB sequences under the control of the highly expressed *katG* promoter were made by insertion of double-stranded (annealed) oligonucleotides at the *Van95I/XbaI* sites of pJCLucKatG (Mulder *et al.*, 1999a). These oligonucleotides correspond to the optimized DB for *M. smegmatis* and *E. coli* respectively. Annealed oligonucleotides MycDB-1 (5'-CTGGATGAGCTCATCGCCGATCCCACCTTCT-3') with MycDB-2 (5'-CTAGAGAAGGTGGGATCGGCGATGAGCTCATCCAGGAG-3') and EcDB-1 (5'-CTGGATGAGCTCATCATGGATCACAAAGTGT-3') with EcDB-2 (5'-CTAGACACTTTGTGATCCATGATGAGCTCATCCAGGAG-3') were used to create pMsΔluc and pEcΔluc fusion constructs, respectively. To repair the deleted 16 N-terminal amino acids of the luciferase protein in the fusion constructs, the *luc* gene was amplified from pJCLucKatG using the primers: Luc-1 (5'-TATACTAGTCGAAGACGAA-3') and Luc-2 (5'-GCTATGACCATGATTACG-3'). The amplicon was digested with *SpeI* and *ClaI* and ligated into the fusion constructs digested with *XbaI* and *ClaI*, creating the KatG-DB-luc constructs; pEcLuc (*E. coli* optimised DB) and pMsLuc (*M. smegmatis* optimised DB), respectively. These constructs encode active luciferase protein. Plasmid, pJCLucKatGΔSD, in which the SD sequence has been deleted after digestion of pJCLucKatG with *Van91I/BamHI*, blunt ended with *Klenow* enzyme and subsequent religation was made as a negative control for expression studies.

For *in vitro* expression T7-DB-luc constructs were created. PCR primer T7-DB (5'-ATATCAGCTGAGATCTTAATACGACTCACTATAGTTCGCGATCACATCCG-3'), containing a portable T7 promoter sequence was designed to replace the *M. tuberculosis katG* promoter. Fifteen bases at the 3' terminus of this primer are homologous to pJCKatG-DB-Luc template. This primer was used in conjunction with Luc 2 (5'-GCTATGACCATGATTACG-3'), a reverse primer that anneals downstream of the luciferase gene in pJCKatG-DB-Luc. PCR using these primers resulted in an amplicon that was used as a linear template for *in vitro* transcription/translation experiments.

8.3.4 Electroporation

M. smegmatis cells were prepared for electroporation as described by Jacobs *et al.*, 1991. For electroporation, 60 μ l of cells were combined with 140 μ l of 10% (v/v) glycerol and 1 μ g of plasmid DNA (5 μ l), and placed on ice for 5 min. Electroporation was performed using Gene Pulser™ cuvettes (0.1 cm gap, Bio-Rad) at 1.8 kV, 25 μ F and 1000 Ω , using a Gene Pulser™ apparatus (Bio-Rad). After electroporation, 900 μ l of Middlebrook-ADC medium was added, and the cells were incubated at 37 °C for 3-4 hr. The cells were plated on Middlebrook-ADC agar plates containing 30 μ g kanamycin ml⁻¹ and incubated at 37 °C for 4-5 days.

8.3.5 *In vitro* coupled transcription/translation

In vitro synthesis of protein from linear plasmid DNA templates was performed using the *E. coli* 30S Extract system system for linear templates (Promega Corp., Madison, USA). Template DNA was isolated after digestion with *Hind*III and *Bst*EII and quantitated on a 2% agarose gel using Kodak Digital Science™ 1D image analysis software. Reactions were performed according to the manufacturer's specifications. Five microlitre aliquots were removed and used in subsequent luciferase assays.

8.3.6 Luciferase assays

For the luciferase activity assays, 5-ml cultures were grown to late exponential phase at 37 °C with aeration (Orbital shaker, Hägar Designs, at 150 r.p.m.). *E. coli* cell extracts were prepared by harvesting cells from 1 ml of culture and resuspension of the cells in 100 μ l of ice-cold 1 x Cell Culture Lysis Reagent (Luciferase Assay System, Promega). The cells were lysed in a FastPrep FP120 instrument (Savant Instruments) for two 20-s cycles at a speed setting of 6. Cell debris was removed by centrifugation, and the supernatant equilibrated to room temperature before being assayed for luciferase activity. Luciferase activity assays were performed using the Promega Luciferase Assay System, according to the manufacturer's instructions. Light output was measured in a Bio-Orbit 1253 luminometer. *M. smegmatis* cell extracts were prepared by resuspension of cell pellets in Cell Culture Lysis Reagent,

followed by processing in the FastPrep FP120 instrument for two, 40s cycles at speed setting 6. Unlysed cells and cell debris were removed by centrifugation. The supernatants were assayed for luciferase activity as follows. For each assay, 100 μ l of reconstituted luciferase reagent was added to 20 μ l of cell extract, vortexed for 5 s, and the light emission measured in a Bio-Orbit 1253 luminometer. The protein concentrations of the cell extracts were measured using the D_c Protein Assay kit (Bio-Rad), and the luciferase activities are quoted as relative light units (RLU) per mg protein in each assay.

8.3.7 Secondary structure prediction

The stabilities of putative sense:antisense RNA interactions were estimated as intramolecular interactions using the energy minimization program, Mfold (Zuker, 1994). For this purpose, the sense and antisense sequences were appended to the 3' and 5' ends of a stable RNA hairpin (Table 8.2).

Table 8.2

Predicted secondary structures involving the 16s rRNA anti-Downstream Box				
	<i>E. coli</i> 16S rRNA		<i>M. smegmatis</i> 16S rRNA	
	Predicted RNA:RNA interaction ^a	ΔG_{37}° (kcal/mol) ^a	Predicted RNA:RNA interaction ^a	ΔG_{37}° (kcal/mol) ^a
Penultimate stem loop of 16S rRNA	CAC GA UUUGUGAUUCAU AAACGUUGGGUG AGA AG	-10.3	GAA GG GA GGUG AUCGGC CCAC UGGCUG AGC AA AA	-12.6
Downstream Box optimised for <i>E. coli</i> 16S rRNA	CACUUUGUGAUUCAUGA GUGAAACACUAGGUACU	-21.9	GAAG GA GC GUGG UCG GA CACU GGU CU GAAA -A -A	-5.5
Downstream Box optimised for <i>M.</i> <i>smegmatis</i> 16S rRNA	CACUU U CA UG GAUU UGA AC CUAG GCU CUUCC C CC	-2.7	GAAGGUGGGAUCGGCGA CUUCCACCCUAGCCGCU	-31.2

Calculated free energy changes for putative RNA-RNA stem-loop interactions in *E. coli* and *M. smegmatis* 16S rRNA. Canonical base pairing is highlighted in bold. ^a Predicted free energy changes and predicted interactions for RNA-RNA duplex formation at 37°C using the Mfold energy minimization program of Zuker (1994).

A

<i>M. tuberculosis</i>	:	TCGGTAACACCCGAAGCCAGTGGCCTAACCCCTCGGGAGGGAGCTGTGCAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGC
<i>E. coli</i>	:	TGGGTTGCAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAAC
Consensus	:	T GGT CA GAAG GT GC TAACC TCGGGAGGG GCT C A <u>TG GAT</u> CA TGGG GAAGTCGTAACAAGGTA C

B

<i>M. tuberculosis</i>	:	TCGGTAACACCCGAAGCCAGTGGCCTAACCCCTC--GGGAGGGAGCTGTGCAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCC
<i>M. bovis</i>	:	TCGGTAACACCCGAAGCCAGTGGCCTAACCCCT---GGGAGGGAGCTGTGCAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCC
<i>M. avium</i>	:	TCGGTAACACCCGAAGCCAGTGGCCTAACCCCTTTGGGAGGGAGCTGTGCAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCC
<i>M. kansasii</i>	:	TCGGTAACACCCGAAGCCAGTGGCCTAACCCCTC--GGGAGGGAGCTGTGCAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCC
<i>M. microti</i>	:	TCGGT AACACCCGAAGCCAGTGGCCTAACCCCTC--GGGAGGGAGCTGTGCAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCC
<i>M. smegmatis</i>	:	TCGGTAACACCCGAAGCCGGTGGCCTAACCCCTT-GTGGAGGGAGCCGTGCAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCC
Consensus	:	TCGGTAACACCCGAAGCC GTGGCCTAACCCCT GGAGGGAGC GTCGAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCC

Figure 8.2

Sequence alignments of the 3' end of 16S rRNA. Anti-Downstream Box sequences are boxed and consensus sequences are underlined. A) The sequence homology between *E. coli* 16S rRNA and *M. tuberculosis* 16S rRNA is 74%. This homology drops to 47% over the aDB region (bases 1467-1481 in *E. coli* 16S rRNA). B) All published Mycobacteriaceae mRNA sequences are identical in the aDB region (6 sequences are depicted).

8.4 Results and Discussion

8.4.1 16S rRNA Analysis

In prokaryote 16S rRNA secondary structure, the penultimate stem between bases 1409-1491, although not conserved at the sequence level, is a conserved structure (Gutell *et al.*, 1985; Raue *et al.*, 1988). The sequence homology between *E. coli* 16S rRNA and *M. tuberculosis* 16S rRNA is 74%. This homology drops to 47% over the aDB region. Although the sequence variation between *E. coli* and *M. tuberculosis* increases to 53% in the aDB region, the structure of the 16S rRNA remains conserved (Gutell *et al.*, 1985 and Raue *et al.*, 1988) (Figure 8.3). Although there is a degree of sequence variation in the 16S rRNA in members of the Mycobacteriaceae, all are identical in the aDB region (Figure 8.2b).

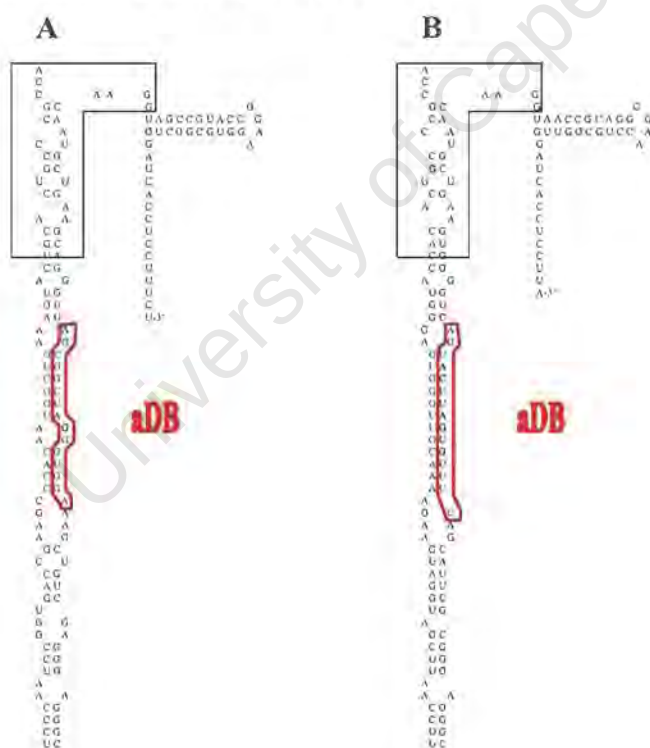


Figure 8.3

Schematic representation of the aDB in helix 44 of: A) *Mycobacterium smegmatis*, and B) *E. coli* 16SrRNA. The decoding region, where interactions between initiator tRNA and start codons occur, is boxed (Black). The aDB region is boxed in red.

8.4.2 DB design

Downstream boxes, optimized for binding to either *M. smegmatis* or *E. coli* 16S rRNA aDB regions, were designed. Hypothetical RNA:RNA interactions were predicted for the interaction of each DB to each of the two host 16S rRNA. The relative stability of each interaction was determined using the Mfold energy minimization program of Zuker, 1994 (Table 8.2).

The design of optimized DB constructs was based on the following hypothesis: for translation to occur, 16S rRNA has to unfold in the aDB region of the penultimate stem and bind to the DB region of the mRNA. The DB region of the mRNA must also undergo unfolding. The driving force behind these interactions is the calculated minimum folding energy of each structure. Binding may be viewed as a competitive event between displacement of 16S rRNA aDB in the penultimate stem and subsequent binding to optimized DB mRNA (if this is more energetically favourable) or not. The minimum folding energy of *E. coli* 16S rRNA in the aDB region is -10.3 kcal/mol at 37 °C. The minimum folding energy of the optimized *E. coli* DB/*E. coli* 16S rRNA aDB interaction is -21.9 kcal/mol at 37 °C. If the hypothesis holds true, one would expect the 16S rRNA aDB to bind to the optimized *E. coli* DB mRNA in much the same way as a SD/aSD interaction occurs. The optimized *M. smegmatis* DB/*E. coli* 16S rRNA aDB interaction however would not favour such an event, as the minimum folding energy for this interaction is only -2.7 kcal/mol at 37 °C, and one would predict little or no association between these molecules and subsequently little or no translation. In *M. smegmatis*, a similar pattern would apply for the optimized DB mRNA. In this host, the optimized *M. smegmatis* DB/*M. smegmatis* 16S rRNA aDB interaction would be favored over that of the *E. coli* DB/*M. smegmatis* aDB interaction (Table 8.2).

The DB constructs optimized for expression in *E. coli* or *M. smegmatis* were cloned downstream of the *M. tuberculosis* H37Rv *katG* promoter (Mulder *et al.*, 1999a) (Table 8.1). A promoterless firefly luciferase gene, lacking a SD, was cloned downstream of the DB sequences, resulting in two constructs; pEcluc (*E. coli* optimized DB) and pMsluc (*M. smegmatis* optimized DB). These constructs were

used in conjunction with a positive control plasmid, pJClucKatG (Mulder *et al.*, 1999a), that contains a functional SD that results in high levels of luciferase expression and a second plasmid, pJCLucKatG Δ SD, that lacks a functional SD sequence for ribosome attachment.

8.4.3 Expression studies in *E. coli* and *M. smegmatis*

Expression of the DB constructs in *E. coli* and *M. smegmatis* results in luciferase expression in the absence of a SD in both organisms (Table 8.3). In *E. coli*, the construct with DB sequences optimized for expression in *E. coli*, pEcLuc, results in a 13.5x higher expression than the mycobacterium optimized DB construct, pMsLuc. When these constructs are electroporated into *M. smegmatis* LR222, the pattern is reversed and the *M. smegmatis* optimised DB construct results in 18.5x higher expression than the *E. coli* optimised DB construct, indicating a DB-specific effect on translation. The efficiency of optimized DB constructs in both hosts is noticeably lower than the control, pJCLucKatG, which has a SD sequence. Translational efficiencies measured by RLU/mg total protein for each of the constructs relative to pJCLucKatG Δ SD (a construct lacking a functional SD sequence) indicate a DB-dependant relationship. In *E. coli* LK111, pEcLuc results in 110x higher luciferase expression than the negative control, pJCLucKatG Δ SD, whilst pMsLuc results in 8x higher luciferase expression. When the host bacteria are changed to *M. smegmatis*, the relative translational efficiencies are 260x and 14x for pMsLuc and pEcLuc respectively.

Table 8.3

Relative Translation Efficiency (RLU/mg protein)			
Construct	<i>E. coli</i> LK111	<i>M. smegmatis</i> LR222	<i>In vitro</i> (<i>E. coli</i> 30 S Extract Promega)
pJCLucKatGΔSD	1	1	-
pJCLucKatG	1327 ± 132	18753 ± 674	-
pEcLuc	110 ± 19	14 ± 1.8	-
pMsLuc	8 ± 1.2	260 ± 27	-
T7LucΔSD	-	-	1
T7Luc	-	-	660 ± 43
T7EcLuc	-	-	85 ± 9.8
T7MsLuc	-	-	15 ± 2.3

Translation efficiency as measured by RLU/mg total protein over a limited growth period corresponding to late exponential phase. The DB constructs, pEcLuc and pMsLuc were transformed into *E. coli* LK111 and *M. smegmatis* LR222 along with the positive control, pJCLucKatG (this construct has a functional SD for translation initiation), and negative control, pJCLucKatGΔSD. *In vitro* transcription/translation was performed on PCR generated templates, T7LucΔSD, T7Luc, T7EcLuc and T7MsLuc.

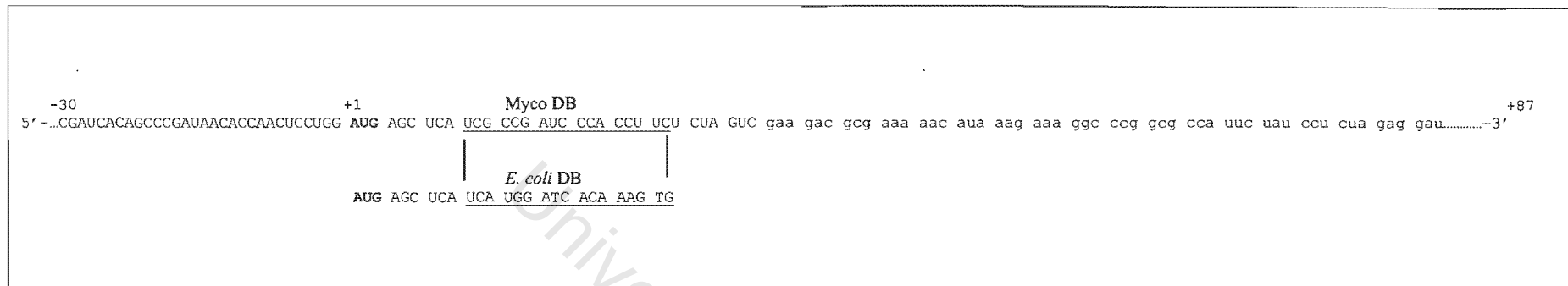


Figure 8.4

5' mRNA sequence of pMSluc and pEcluc (From -30 to +87 with regard to the AUG. The AUG translation start site is in bold. The Mycobacterium and *E. coli* optimized DB sequences are underlined. The 5' portion of the firefly luciferase sequence is in small caps starting at codon 2. Apart from the underlined DB portion, the sequences are identical.

8.4.4 *In vitro* transcription/translation analysis

In vitro synthesis of luciferase protein from PCR templates was performed using the *E. coli* 30S Extract System for linear templates (Promega Corp., Madison, USA). For this the *M. tuberculosis* H37Rv *katG* promoter was replaced with a T7 promoter cassette. Gel purified template DNA was quantified using Kodak Digital Science 1D image analysis software. Results of the *in vitro* transcription/translation experiments using constructs under the control of the T7 promoter mirror those obtained from bacterial hosts shown in Table 8.3. Expression from the T7 promoter results in 5.6x higher RLU from T7Ecluc as compared with T7Msluc. These results reflect the fact that *E. coli* 30S extract was used in the transcription/translation experiments. Higher luciferase expression was therefore observed in the construct with an *E. coli* optimized DB sequence. Translational efficiencies for each of the constructs relative to pJCLucKatGΔSD are 85x and 15x for pEcLuc and pMsLuc respectively.

8.5 Conclusion

Along with the SD, the DB has been implicated in translation enhancement in *E. coli* and bacteriophage gene expression (Sprengart *et al.*, 1996). The data presented here support the idea that the DB plays a role in the efficiency of translation in mycobacteria and that the DB is a conserved feature across species at a functional level. The mechanism by which the DB enhances translation has yet to be identified as direct interaction between DB/ADB proposed earlier is not supported by *in vitro* filter binding assays (Moll *et al.*, 2001).

Chapter 9

General conclusions

Trends in the way molecular biology research is performed have changed in dramatic fashion over the last decade. As outlined in chapter one, automated DNA sequencing has resulted in a massive source of information that requires an *in silico* approach to analysis. The work presented in this thesis was based for the most part on trying to tap the vast amount of information provided by DNA sequencing labs worldwide.

Part A of the thesis deals with the issue of antisense control or attenuation of gene expression. In chapter 4, I looked at the possibility of using segments of genes for attenuating gene expression. This work was performed before the sequence of *M. tuberculosis* became available and was performed with the idea of “shotgun” cloning entire *Sau3AI* fragments of *M. tuberculosis* in antisense orientations and subsequent screening for growth deficient phenotypes. Levels of attenuation achieved were 21% and 32% for antisense constructs directed against region 1 and 2 of the *mpb83* transcript respectively. In 1998, when Cole *et al.* (1998) published the *M. tuberculosis* sequence such an approach became less useful. Further study into the mechanism of attenuation could however lead to the development of a system of differential attenuation of expression of genes when required. Just such a system could be envisaged for genes that are required for the survival of the organism.

Chapter 5 and 6 includes work performed in the post *M. tuberculosis* genome sequence era. The concept of targeted gene attenuation by the use of short external guide sequences designed directly from available genome sequence sounded extremely attractive. Proof of concept work was performed in *M. smegmatis* using *impC* as target. Both *M. smegmatis* and *M. tuberculosis* contain three genes (*impA-C*), which encode proteins with inositol monophosphate motifs. ImpC is thought to be involved in the production of inositol required for the synthesis of mycothiol (Stoker personal communication). Using complementation studies, it was shown that

an *impC* knockout in *M. tuberculosis* was lethal (Dr Farah Movahedzadeh, personal communication). Work performed in *M. smegmatis* revealed altered growth dynamics when EGSs directed against *ImpC* were introduced. *ImpC* EGS1 was more efficient than *ImpC* EGS2 and resulted in a 2-fold reduction in culture turbidity when compared with control pJCLuc. The results suggest that *ImpC* EGS1 is either more efficient in binding to target *impC* than *ImpC* EGS2 or alternatively it is able to mimic a tRNA more effectively than *ImpC* EGS2 once bound to target *impC* and is hence cleaved by RNase P. Analysis of *impC* mRNA levels reveals that there is a reduction in full-length *impC* mRNA in *M. smegmatis* harbouring both *ImpC* EGS constructs, pJCImpCEGS1 and pJCImpCEGS2. The reduction in full-length *impC* correlates to both the phenotypic growth analysis time points and to maximal expression of the EGSs by the *katG* promoter. These results are indicative of specific *ImpC* attenuation and point to several exciting future prospects.

In chapter 8, regions of difference in the conserved 16S rRNA were identified. These sequences were shown to interact with a region within the coding region of mRNA transcripts (downstream box). It was demonstrated that species-specific expression from mRNA transcripts lacking traditional ribosome binding sequences but harboring optimized downstream box sequences was possible. The levels of translation were low in comparison to transcripts harboring Shine-Dalgarno sequences but were nevertheless elevated in optimized transcripts when compared to non-optimized controls.

Future work

Selection of more efficient EGS

Yuan and Altman (1994) showed that an EGS custom-designed for mRNA for chloramphenicol transacetylase (*cat*) can direct the specific cleavage of the mRNA by human RNase P *in vitro* or in cells in tissue culture. However, the cleavage reaction was inefficient compared to cleavage of natural tRNA precursor substrates. The

precise details of the relation between the tertiary structure of a tRNA precursor molecule and its ability to be recognized and cleaved efficiently by RNase P are not completely understood, so there are no comprehensive guidelines for engineering EGSs with improved efficiency.

The EGSs used to direct human RNase P for targeted cleavage resemble three-quarters of a tRNA molecule and consist of two sequence elements: a targeting sequence, complementary to the mRNA sequence; which is a portion of the natural tRNA sequence and is required for RNase P recognition. While little is known about the rate-limiting step of EGS-targeting reactions in cultured cells, recent studies on ribosomes and antisense phosphorothioate molecules suggest that the binding of these molecules to their target RNA appear to be rate-limiting *in vivo* (Lee *et al.*, 1999; Pal *et al.*, 1998; Sullenger and Cech 1993; Putlitz *et al.*, 1999; Yu *et al.*, 1998). In an attempt to increase the efficiency of EGS/substrate recognition by RNase P, Liu *et al.* (2002) demonstrated the feasibility of developing effective EGSs for gene-targeting applications by employing an *in vitro* selection procedure to isolate highly active EGSs from an EGS pool that contained random mutations. It is believed that the acceptor stem and D stem and loop sequences within a natural tRNA molecule (figure 9.1) are involved in tertiary interactions with different parts of the tRNA (eg. Variable and T stem and loop regions) and are important for folding the tRNA and interactions with RNase P (Altman and Kirsebom 1999; Frank and Pace 1998).

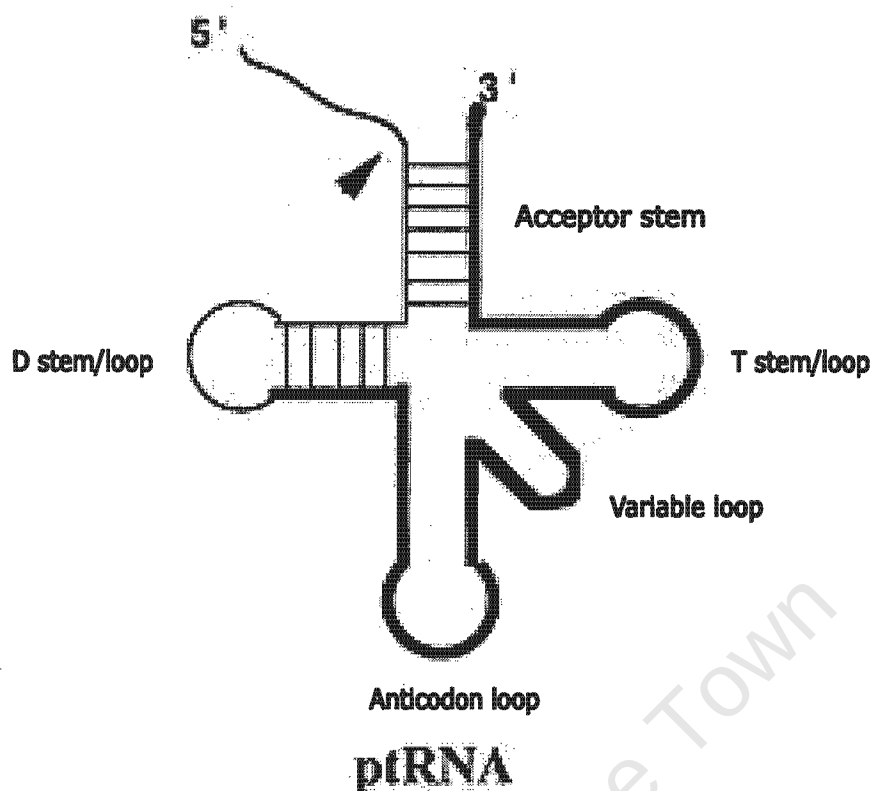


Figure 9.1: Schematic representation of a precursor tRNA. The site of RNase P cleavage is indicated by the arrow to generate mature tRNA

Previously they showed that EGSs efficiently direct human RNase P to cleave the mRNA sequence encoding the thymidine kinase (TK) of herpes simplex virus 1 (HSV-1) *in vitro* (Kawa *et al.*, 1998). A reduction of 50-70% in the TK mRNA and protein expression was observed in HSSV-1 infected cells expressing EGSs. However the cleavage reaction was inefficient compared to cleavage of a natural tRNA substrate (ie. $\text{ptRNA}^{\text{Ser}}$). The reduced susceptibility of the TK 46-EGS complex to be cleaved by RNase P, compared to a natural tRNA, was probably the result of replacing the natural tRNA sequence in the complex of TK 46 and the EGS (Zhou *et al.*, 2002). It was suggested that the substitutions might disrupt some of the tertiary interactions that are potentially important in maintaining the proper tRNA-like conformation required for RNase P recognition. The variable region, and the T stem loop were randomly mutated to develop a more efficient construct for RNase P recognition.

One of the EGSs, TK1, was at least 35 times more effective at directing RNase P in cleaving TK mRNA *in vitro* than the EGS derived from a natural tRNA sequence.

These results demonstrate the potential for using the selection procedure as a general approach for the generation of highly effective EGSs for gene-targeting applications.

Synthetic oligonucleotides as therapeutic agents

Synthetic oligonucleotides and their analogs, by virtue of their ability to bind to nucleic acid target in a sequence-specific manner, are being developed as potential human therapeutic agents. Several antisense phosphorothioate oligodeoxyribonucleotides are currently in clinical trials for a variety of disease indications, and chimeric molecules with higher orders of complexity in terms of chemical competition and mode of action are likely to follow (Crooke and Bennett, 1996; Matteucci and Wagner 1996).

Unmodified oligoribonucleotides are extremely sensitive to various nucleases present in all biologic fluids, the half-life being in the order of a few seconds in human serum (Beeiglelman *et al.*, 1995). RNAs are also much more difficult to prepare in large quantities despite the remarkable progress realized in the RNA synthesis field (Wincott *et al.*, 1995). Simple 2'-O-methyl modifications of the RNA backbone were shown to confer nuclease resistance (Ma *et al.*, 1998). They developed a relatively short nuclease-resistant EGS that can cleave an RNA target in a sequence-specific manner in the presence of human RNase P. The resulting lead prototypes retained full reactivity in inducing RNase P-mediated target cleavage when compared with an all-RNA control. Furthermore, the serum half-lives of these compounds were extended from a few seconds for the all-RNA EGS to more than 24 hours (Ma *et al.*, 1998). These results provide useful insights into the design and application of oligonucleotide-based EGSs as potential regulators of gene expression and tools for the identification and validation of drug discovery targets.

EGS Delivery

There are two basic strategies for EGS delivery into cells:

- 1) Endogenous delivery, where a gene for the EGS is delivered. For this the gene encoding the EGS is inserted into a vector, which can be delivered into target cells. This can provide transient or stable expression of the EGS.
- 2) Exogenous delivery, where a pre-synthesized EGS is applied to the cells. This can be considered as a form of chemotherapy. Since oligonucleotides have low cytotoxicity and high sequence specificity they lack the drawbacks of chemotherapeutic agents making the strategy very attractive (Reviewed in Birikh *et al.*, 1997). The greatest hurdle for the exogenous delivery is cellular uptake of EGSs.

Cellular uptake

Numerous strategies have been employed to introduce oligonucleotides into bacterial and human cells. Rapaport *et al.* (1996) introduced oligonucleotide phosphorothioates into *M. smegmatis* using a variety of strategies. One involved the presence of a very low level of ethambutol, which enables the entry of oligonucleotides into mycobacteria because of its induced alterations in the cell wall, and another involved the utilization of oligonucleotides covalently attached to a D-cycloserine molecule, whereby entry into the mycobacterial cell is achieved by a receptor-mediated process. Another low molecular weight, covalently attached ligand that enabled the entry and subsequent antimycobacterial activities of oligonucleotide phosphorothioates in the absence of a cell wall modifying reagent is biotin.

The delivery of antisense oligonucleotides and ribozymes to HIV-1 infected or potentially infected cells by antibody targeted liposomes, certain cationic lipid formulations and pH-sensitive liposome results in significant anti-HIV-1 activity. These carriers not only facilitate cytoplasmic delivery but also protect the drugs from nuclease digestion (Duzgunes *et al.*, 2001).

REFERENCES

- Adhin, M.R., van Duin, J. (1989) Translational regulation of the lysis gene in RNA bacteriophage ϕ requires a UUG initiation codon. *Mol. Gen. Genet.* 218(1):137-42.
- Akerley, B.J., E. J. Rubin, A. Camilli, D. J. Lampe, H. M. Robertson, and J. J. Mekalanos (1995) Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* 95(15):8927-32.
- Alam, J., Cook, J.L. (1990) Reporter genes: application to the study of mammalian gene transcription. *Anal. Biochem.* 188(2):245-54.
- Aldovini, A., R. N. Husson, R. A. Young (1993) The *uraA* locus and homologous recombination in *Mycobacterium bovis* BCG. *J. Bacteriol.* 175: 7282-7289.
- Alland, D., Kalkut, G.E., Moss, A.R., McAdam, R.A., Hahn, J.A., Bosworth, W., Drucker, E., Bloom, B.R. (1994) Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med.* 330(24):1710-6.
- Altman, S., and Kirsebom, L.A. (1999) In: *The RNA world* (Gesteland, R.F., Ceck, T.R., and Atkins, J.F., eds.), 2nd Ed., pp. 351-380. Cold Spring Harbour Laboratory Press. CSH.
- Altman, S. (1990) Ribonuclease P. *J Biol Chem.* 265(33):20053-6.
- Altman, S. (1993) RNA enzyme-directed gene therapy. *Proc. Natl. Acad. Sci U S A.* 90(23):10898-900.
- Altman, S., Kirsebom, L., Talbot, S. (1993) Recent studies of ribonuclease P. *FASEB J.* 7(1):7-14. Review.
- Altuvia, S., A. Zhang, L. Argaman, A. Tiwari, G. Storz (1998) The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J.* 17:6069-75.
- Altuvia, S., D. Weinstein-Fischer, A. Zhang, L. Postow, G. Storz (1997) A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell.* 90:43-53.
- Altuvia, S., and E. G. H. Wagner (2000) Switching on and off with RNA. *Proc. Natl. Acad. Sci. U. S. A.* 97: 9824-9826.
- Ambulos, N. P. Jr, E. J. Duvall, P. S. Lovett (1986) Analysis of the regulatory sequences needed for induction of the chloramphenicol acetyltransferase gene *cat-86* by chloramphenicol and ampicillin. *J Bacteriol.* 167(3): 842-849.
- Arain, T.M., Resconi, A.E., Singh, D.C., Stover, C.K. (1996) Reporter gene technology to assess activity of antimycobacterial agents in macrophages. *Antimicrob Agents Chemother.* 40(6):1542-4.
- Argaman, L., and S. Altuvia (2000) *fhlA* repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. *J. Mol. Biol.* 300:1101-12.
- Arigoni, F., F. Talabot,, M. Peitsch, M. D. Edgerton, E. Meldrum, E. Allet, R. Fish, T. Jamotte, M. L. Curchod, and H. Lofrerer (1998) A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* 16(9): 851-856.
- Armitige, L. Y., C. Jagannath, A. R. Wanger, S. J. Norris (2000) Disruption of the genes encoding antigen 85A and antigen 85B of *Mycobacterium tuberculosis* H37Rv: effect on growth in culture and in macrophages. *Infect. Immun.* 68: 767-778.

- Arraiano, C.M., Yancey, S.D., Kushner, S.R. (1988) Stabilization of discrete mRNA breakdown products in *ams pnp rnb* multiple mutants of *Escherichia coli* K-12. *J Bacteriol.* 170(10):4625-33.
- Asano, K., and K. Mizobuchi (2000) Structural analysis of late intermediate complex formed between plasmid ColIb-P9 Inc RNA and its target RNA. How does a single antisense RNA repress translation of two genes at different rates? *J. Biol. Chem.* 275:1269-74.
- Ausubel, F. M., Brent, R. Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Strhl, K. (1987) *Current Protocols in Molecular Biology*. Cambridge, MA: Wiley Interscience.
- Azad, A.K., Sirakova, T.D., Rogers, L.M., Kolattukudy, P.E. (1996) Targeted replacement of the mycocerosic acid synthase gene in *Mycobacterium bovis* BCG produces a mutant that lacks mycosides. *Proc Natl Acad Sci U S A.* 14;93(10):4787-92.
- Baer, M.F., Reilly, R.M., McCorkle, G.M., Hai, T.Y., Altman, S., RajBhandary, U.L. (1988) The recognition by RNase P of precursor tRNAs. *J Biol Chem.* 263(5):2344-51.
- Bakin, A., Konalok, J. A., McCloskey, J. A., and Ofengand, J. (1994) The single pseudouridine residue in *E. coli* 16S rRNA is located at position 516. *Nucleic Acids Res.* 22:3681-3684.
- Balasubramanian, V., Pavelka, M.S. Jr., Bardarov, S.S., Martin, J., Weisbrod, T.R., McAdam, R.A., Bloom, B.R., Jacobs, W.R. Jr. (1996) Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. *J Bacteriol* 178(1):273-9
- Bardarov, S., J. Kriakov, C. Carriere, S. Yu, C. Vaamonde, R. A. McAdam, B.R. Bloom, G. F. Hatfull, W. R. Jacobs (1997) Conditionally replicating mycobacteriophages: a system for transposon delivery to *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 94:10961-10966.
- Bechhofer, D. H., Dubnau, D. (1987) Induced mRNA stability in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U S A.* 84(2):498-502.
- Beggs, M. L., Crawford, J. T. and Eisenach, K. D. (1995) Isolation and sequencing of the replication region of *Mycobacterium avium* plasmid pLR7. *J. Bacteriol.* 177, 4836-4840.
- Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S., Small, P.M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science.* 284(5419):1520-3.
- Beigelman, L., McSwiggen, J.A., Draper, K.G., Gonzalez, C., Jensen, K., Karpeisky, A.M., Modak, A.S., Matulic-Adamic, J., DiRenzo, A.B., Haeberli, P., et al. (1995) Chemical modification of hammerhead ribozymes. Catalytic activity and nuclease resistance. *J Biol Chem.* 270(43):25702-8.
- Belasco, J.G., Nilsson, G., von Gabain, A., Cohen, S.N. (1986) The stability of *E. coli* gene transcripts is dependent on determinants localized to specific mRNA segments. *Cell.* 46(2):245-51.
- Bercovier, H., Kafri, O., Sela, S. (1986) Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochem. Biophys. Res. Commun.* 136(3):1136-41.
- Berkhout, B., van der Laken, C. J. and Knippenberg, P. H. (1986) *Biochim. Biophys. Acta.* 866. 144.
- Berthet, F.X., Lagranderie, M., Gounon, P., Laurent-Winter, C., Ensergueix, D., Chavarot, P., Thouron, F., Maranghi, E., Pelicic, V., Portnoi, D., Marchal, G., Gicquel, B. (1998) Attenuation of virulence by disruption of the *Mycobacterium tuberculosis* *erp* gene. *Science.* 282(5389):759-62.
- Besra, G. S., and Chatterjee, D. (1994) Lipids and carbohydrates of *Mycobacterium tuberculosis*, pp. 283-306. In B. R. Bloom (ed.), *Tuberculosis: pathogenesis, protection, and control*, A. S. M., Washington D. C.

Birikh, K.R., Heaton, P.A., Eckstein, F. (1997) The structure, function and application of the hammerhead ribozyme. *Eur J Biochem.* 245(1):1-16. Review.

Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453-74.

Blomberg, P., H. M. Engdahl, C. Malmgren, P. Romby, E. G. Wagner (1994) Replication control of plasmid R1: disruption of an inhibitory RNA structure that sequesters the repA ribosome-binding site permits tap-independent RepA synthesis. *Mol Microbiol.* 12:49-60.

Boni, I.V., Isaeva, D.M., Musychenko, M.L., Tzareva, N.V. (1991) Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res.* 19(1):155-62.

Brandi, A., Pietroni, M. L., Gualerzi, C. O., and Pon, C. L. (1996) Post-transcriptional regulation of CspA expression in *E. coli*. *Mol. Microbiol.* 19:231-240.

Brendel, V. and Perelson, A.S. (1993) Quantitative model of ColE1 plasmid copy number control. *J Mol Biol.* 20;229 (4):860-72.

Brennan, P. J. and P. Drapper (1994) In "Tuberculosis: Pathogenesis, Protection and Control" (ed. Bloom, B. R.) p. 271-284 (A.S.M., Washington DC).

Brosch, R., Gordon, S.V., Eiglmeier, K., Garnier, T., Tekaia, F., Yeramian, E., Cole, S. T. (1999) In *Molecular Genetics of mycobacteria*, eds. Hatful, G. F. and Jacobs, W. R. Jr. (A.S.M., Washington, D. C.), pp. 19-36.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., Cole, S.T. (2002) A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A.* 99(6):3684-9.

Brosius, J., Dull, T. J., and Noller, H. F. (1980) Complete nucleotide sequence of a 23S rRNA gene from *E. coli*. *Proc. Natl. Acad. Sci. USA* 77:201-204.

Brownlee, G. G., Sanger, F., and Barrell, B. G. (1967) Nucleotide sequence of 5S rRNA from *E. coli*. *Nature (London)* 215:735-736.

Busby, S., Dreyfus, M. (1983) Segment-specific mutagenesis of the regulatory region in the *Escherichia coli* galactose operon: isolation of mutations reducing the initiation of transcription and translation. *Gene* 21(1-2):121-31.

Calogero, R.A., Pon, C.L., Canonaco, M.A., Gualerzi, C.O. (1988) Selection of the mRNA translation initiation region by *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* 85(17):6427-31.

Carmi, O.A., Stewart, G.S., Ulitzur, S., Kuhn, J. (1987) Use of bacterial luciferase to establish a promoter probe vehicle capable of nondestructive real-time analysis of gene expression in *Bacillus* spp. *J. Bacteriol.* 169(5):2165-70.

Case, C. C., S. M. Roels, J. E. Gonzalez, E. L. Simons, and R. W. Simons (1988) Analysis of the promoters and transcripts involved in IS10 anti-sense RNA control. *Gene* 72: 219-236.

Case, C. C., E. L. Simons, and R. W. Simons (1990) The IS10 transposase mRNA is destabilized during antisense RNA control. *EMBO J.* 9:1259-66.

CDC, Morb. Mortal. Wkly. Rep 41 (RR20), 1 (1998)

Cech, T.R. (1990) Self-splicing of group I introns. *Annu Rev Biochem.* 59:543-68. Review.

Chapon, C. (1982) Expression of *malT*, the regulator gene of the maltose region in *Escherichia coli*, is limited both at transcription and translation. *EMBO J.* 1(3):369-74.

Chen, G.F., Inouye, M. (1990) Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res.* 18(6):1465-73.

Cheung, A. L., C. Wolz, M. R. Yeaman, and A. S. Bayer (1995) Insertional inactivation of a chromosomal locus that modulates expression of potential virulence determinants in *Staphylococcus aureus*. *J. Bacteriol.* 177(11):3220-6.

Chung, C.T. and Millar, R.H. (1988) A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Research.* 16: 3580.

Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG. (2001) Massive gene decay in the leprosy bacillus. *Nature.* 22;409(6823):1007-11.

Cole S.T., Brosch R, Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E.III., Tekaia F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver K., Osborne J., Quail M.A., Rajandream, M-A., Rogers J., Rutter S., Seeger K., Skelton J., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S. and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544

Crooke, S. T., C. F. Bennett (1996) Progress in antisense oligonucleotide therapeutics. *Annu. Rev. Pharmacol. Toxicol.* 36: 107-29. Review.

Curcic, R., Dhandayuthapani, S., Deretic, V. (1994) Gene expression in mycobacteria: transcriptional fusions based on *xylE* and analysis of the promoter region of the response regulator *mirA* from *Mycobacterium tuberculosis*. *Mol. Microbiol.* 13(6):1057-64.

Dalboge, H., Carlsen, S., Jensen, E.B., Christensen, T., Dahl, H.H. (1988) Expression of recombinant growth hormone in *Escherichia coli*: effect of the region between the Shine-Dalgarno sequence and the ATG initiation codon. *DNA.* 7(6):399-405.

Dale, J. and J. Patki. In "Molecular biology of the Mycobacteria". (J. Mc Fadden, ed.) P180-192. Academic Press Inc. San Diego, CA 92101 (1990).

D' Alessio, J. M. (1982) RNA sequencing. In: Gel electrophoresis of nucleic acids: a practical approach. D. Rickwood and B. D. Hames eds. (IRL, Oxford), pp. 173-197.

Das, A., J. Urbanowski, H. Weissbach, J. Nestor, C. Yanofsky (1983) In vitro synthesis of the tryptophan operon leader peptides of *Escherichia coli*, *Serratia marcescens*, and *Salmonella typhimurium*. *Proc. Natl. Acad. Sci. U. S. A.* 80(10):2879-83.

Davis, J. (1996) Bacteria on the rampage. *Nature.* 19;383(6597):219-20.

de Boer, H. A., L. J. Comstock, A. S. Hui, E. Wong, and M. Vasser. (1983) A hybrid promoter and portable Shine-Dalgarno regions of *Escherichia coli*. *Biochem Soc Symp.* 48:233-44.

de Boer, H. A., P. Ng, and A. S. Hui. (1985) In "Sequence specificity in transcription and translation," p. 419. Alan R. Liss, New York

- de Boer, H.A., Hui, A.S. (1990) Sequences within ribosome binding site affecting messenger RNA translatability and method to direct ribosomes to single messenger RNA species. *Methods Enzymol.* 185:103-114.
- Deckman, I.C., Draper, D.E. (1987) S4- α mRNA translation regulation complex. II. Secondary structures of the RNA regulatory site in the presence and absence of S4. *J Mol Biol.* 196(2):323-32.
- De Smit, M. H. and van Duin, J. (1990a) Control of prokaryotic translation initiation by mRNA secondary structure. *Progr. Nucl. Acid Res. Mol. Biol.* 38. 1-35.
- De Smit, M. H. and van Duin, J. (1990b) Secondary structure of the ribosome binding site determines translational efficiency; A quantitative analysis. *Proc. Natl. Acad. Sci. U.S.A.* 87, 7668-7672.
- De Smit, M. H. and van Duin, J. (1993) Translational initiation of the coat-protein gene of phage MS2; Native upstream RNA relieves inhibition by mRNA secondary structure. *Mol. Microbiol.* 9. 1079-1088.
- De Smit, M. H. and van Duin, J. (1994) Control of translation by secondary structure in *E. coli*. *J. Mol. Biol.* 244. 144-150.
- De Voss, J.J., K. Rutter, B. G. Schroeder, H. Su, Y. Zhu, D. E. Barry (2000) The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages. *Proc. Natl. Acad. Sci. U. S. A.* 97: 1252-1257.
- de Wet, J.R., Wood, K.V., Helinski, D.R., DeLuca, M. (1985) Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. *Proc. Natl. Acad. Sci. U S A.* 82(23):7870-3.
- 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):1785-92.
- Dhandayuthapani, S., Mudd, M., Deretic, V. (1997) Interactions of OxyR with the promoter region of the oxyR and ahpC genes from *Mycobacterium leprae* and *Mycobacterium tuberculosis*. *J. Bacteriol.* 179(7):2401-9.
- DiMari, J., Bechhofer, D.H. (1993) Initiation of mRNA decay in *Bacillus subtilis*. *Mol Microbiol.* 7(5):705-17.
- Dixon, D.A., Kowalczykowski, S.C. (1993) The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. *Cell.* 73(1):87-96.
- Dong, X., D. D. Womble, R. H. Rownd (1987) Transcriptional pausing in a region important for plasmid NR1 replication control. *J. Bacteriol.* 169: 5353-5363.
- Draper, P. (1967) The aliphatic acylamide amidohydrolase of *Mycobacterium smegmatis*: its inducible nature and relation to acyl-transfer to hydroxylamine. *J. Gen. Microbiol.* 46 (1):111-123.
- Draper P, Kandler O, Darbre A. (1987) Peptidoglycan and arabinogalactan of *Mycobacterium leprae*. *J. Gen. Microbiol.* 133 (Pt 5):1187-94.
- Draper, D.E. (1996) Strategies for RNA folding. *Trends Biochem Sci.* 21(4):145-9. Review.
- Dreyfus, M. (1988) What constitutes the signal for the initiation of protein synthesis on *Escherichia coli* mRNAs? *J Mol Biol.* 204(1):79-94.
- Duvoisin, R.M., Belin, D., Krisch, H.M. (1986) A plasmid expression vector that permits stabilization of both mRNAs and proteins encoded by the cloned genes. *Gene.* 45(2):193-201.

- Duzgunes, N., Simoes, S., Konopka, K., Rossi, J.J., Pedroso, de Lima, M.C. (2001) Delivery of novel macromolecular drugs against HIV-1. *Expert Opin Biol Ther.* 1(6):949-70. Review.
- Dye, C., Williams, B.G., Espinal, M.A., Raviglione, M.C. (2002) Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science.* 15;295(5562):2042-6. Review.
- Eguchi, Y., T. Itoh, and J. Tomizawa (1991) Antisense RNA. *Annu. Rev. Biochem.* 60:631-52.
- Ehresmann, C., Stieger, P., Carbon, P., Ungewickell, E., and Garrett, R. A. (1981) The topography and the 5' end of the 16S rRNA in the presence and absence of ribosomal proteins, S4 and S20. *Eur. J. Biochem.* 103: 439-446.
- Emory, S.A., Bouvet, P., Belasco, J.G. (1992) A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev.* 6(1):135-48.
- Engbrecht, J., Simon, M., Silverman, M. (1985) Measuring gene expression with light. *Science.* 227(4692):1345-7.
- Etchegaray, J. P., Jones, P. G., and Inouye, M. (1996) Differential thermoregulation of two highly homologous cold-shock genes, *cspA* and *cspB* of *E. coli*. *Genes Cells* 1:171-178.
- Etchegaray, J. P., Xia, B., Jiang, W. and Inouye, M. (1998) Downstream box: a hidden translational enhancer. *Mol Microbiol.* 27(4):873-4.
- Etchegaray, J. P., and Inouye, M. (1999) Translational enhancement by an element downstream of the initiation codon of *E. coli*. *J. Biol. Chem.* 274:10079-10085.
- Falkow, S. (1988) Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.* 10: S274-6.
- Fang, L., Jiang, W., Bae, W., and Inouye, M. (1997) Promoter-independent cold-shock induction of *cspA* and its derepression at 37 C by mRNA stabilization. *Mol. Microbiol.* 23:355-364.
- Farmer, P., Kim, J.Y. (1998) Community based approaches to the control of multidrug resistant tuberculosis: introducing "DOTS-plus". *BMJ.* 5;317(7159):671-4. Review.
- Faxen, M., J. Plumbridge, and L. A. Isaksson. (1991) Codon choice and potential complementarity between mRNA downstream of the initiation codon and bases 1471-1480 in 16S ribosomal RNA affects expression of *glnS*. *Nucleic Acids Res.* 19(19):5247-51.
- Feinberg, M.B. (1996) Lessons from and about AIDS....*Lancet.* 348(9036):1188.
- Fitzgerald, J. R., J. M. Musser (2001) Evolutionary genomics of pathogenic bacteria. *Trends Microbiol.* 9(11):547-53. Review.
- Fitzwater, T., X. Y. Zhang, R. Elble, B. Polisky (1988) Conditional high copy number ColE1 mutants: resistance to RNA1 inhibition *in vivo* and *in vitro*. *EMBO J.* 7:3289-97.
- Forster, A. C., S. Altman (1990) External guide sequences for an RNA enzyme. *Science* 249:783-786.
- Frank, D.N., Pace, N.R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu Rev Biochem.* 67:153-80. Review.
- Gausing, K. (1977) Regulation of ribosome production in *Escherichia coli*: synthesis and stability of ribosomal RNA and of ribosomal protein messenger RNA at different growth rates. *J Mol Biol.* 115(3):335-54.

Gheysen, D., D. Iserentant, C. Derom, W. Fiers. (1982) Systematic alteration of the nucleotide sequence preceding the translation initiation codon and the effects on bacterial expression of the cloned SV40 small-t antigen gene. *Gene*. 17(1):55-63.

Gheysen, D., Fiers, W. (1982) Expression and excretion of human fibroblast beta 1 interferon in monkey cells after transfection with a recombinant SV40 plasmid vector. *J. Mol. Appl. Genet.* 1(5):385-94.

Gillam, S., Astell, C.R., Jahnke, P., Hutchison, C.A. 3rd, Smith, M. (1984) Construction and properties of a ribosome-binding site mutation in gene E of phi X174 bacteriophage. *J. Virol.* 52(3):892-6.

Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo. (1981) Translational Initiation in Prokaryotes. *Annu. Rev. Microbiol.* 35: 365-403.

Gold, L., and G. Stormo. (1987) Translation initiation, p. 1302-1307. In F. C. Neidhardt, J. L. Ingraham, K. B. Lon, B. Magasanik, M. Schaechter, and H. E. Umbarger (eds.), *E. coli* and *S. typhimurium*: cellular and Molecular biology, vol. 2. ASM Washington, D. C.

Gold, L. (1988) Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu Rev Biochem.* 57:199-233.

Gold, L., Stormo, G.D. (1990) High-level translation initiation. *Methods Enzymol.* 185:89-93.

Goldenberg, D., Azar, L., and Oppenheim, A. B. (1996) Differential mRNA stability of the *cspA* gene in the cold-shock response of *E. coli*. *Mol. Microbiol.* 19:241-248.

Gonzalez-y-Merchand, J.A., Garcia, M.J., Gonzalez-Rico, S., Colston, M.J., Cox, R.A. (1997) Strategies used by pathogenic and nonpathogenic mycobacteria to synthesize rRNA. *J Bacteriol.* 179(22):6949-58.

Gordon, S., Parish, T., Roberts, L.S., Andrew, P.W. (1994) The application of luciferase as a reporter of environmental regulation of gene expression in mycobacteria. *Lett Appl Microbiol.* 19(5):336-40.

Gordhan, B.G., Smith, D.A., Alderton, H., McAdam, R.A., Bancroft, G.J., Mizrahi, V. (2002) Construction and phenotypic characterization of an auxotrophic mutant of *Mycobacterium tuberculosis* defective in L-arginine biosynthesis. *Infect Immun.* 70(6):3080-4.

Gren, E. J. (1984) Recognition of messenger RNA during translational initiation in *Escherichia coli*. *Biochimie.* 66(1):1-29.

Grill, S., Moll, I., Hasenohrl, D., Gualerzi, C.O., Blasi, U. (2001) Modulation of ribosomal recruitment to 5'-terminal start codons by translation initiation factors IF2 and IF3. *FEBS Lett.* 495(3):167-71.

Gualerzi, C. O., Pon, C. L., Parvlik, R. T., Canonaco, M. A., Paci, M. and Wintermeyer, M. In "Structure, function and genetics of ribosomes" (B. Hardesty and G. Kramer, eds), p. 621. Springer-Verlag, Berlin and New York 1986.

Gualerzi, C., and Pon, L. S. (1990) Initiation of mRNA translation in Prokaryotes. *Biochemistry* 29: 5881-5889.

Gualerzi, C. O., L. Brandi, E. Caserta, A. La Teana, R. Spurio, J. Tomsic, and C. Pon (2000) Translation initiation in bacteria. In the Ribosome: Structure, Function, Antibiotics and Cellular Interactions. Garrett, R. A., Douthwaite, S. R., Liljas, A., Matheson, A. T., Moore, P. B., and Noller, H. F. (eds). Washington, DC: ASM Press, pp 477-494.

Gualerzi, C., Risuleo, G., Pon, C.L. (1977) Initial rate kinetic analysis of the mechanism of initiation complex formation and the role of initiation factor IF-3. *Biochemistry.* 16(8):1684-9.

- Gupta, S. L., J. Chen., L. Schaefer, P. Lengyel, and S. M. Weissman.** (1970) Nucleotide sequence of a ribosome attachment site of bacteriophage f2 RNA. *Biochem Biophys Res Commun.* **39**(5):883-8.
- Gutell, R. R.** (1994) Collection of Small Subunit (16S- and 16S-like) ribosomal RNA structures: 1994. *Nucleic Acids Research* **22**: 3502-3507.
- Gutell, R. R., Weiser, B., Woese, C. R. and Noller, H. F.** (1985) Comparative anatomy of 16S-like ribosomal RNA. *Prog Nucleic Acid Res Mol Biol.* **32**:155-216.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S.** (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell.* **35**(3 Pt 2):849-57.
- Guerrier-Takada, C., Altman, S.** (1984) Catalytic activity of an RNA molecule prepared by transcription *in vitro*. *Science.* **20**;223(4633):285-6.
- Guerrier-Takada, C., McClain, W.H., Altman, S.** (1984) Cleavage of tRNA precursors by the RNA subunit of *E. coli* ribonuclease P (M1 RNA) is influenced by 3'-proximal CCA in the substrates. *Cell.* **38**(1):219-24.
- Guerrier-Takada, C., Li, Y., Altman, S.** (1995) Artificial regulation of gene expression in *Escherichia coli* by RNase P. *Proc Natl Acad Sci U S A.* Harth, G., P. C. Zamecnik, J. Y. Tang, D. **21**;92(24):11115-9.
- Guerrier-Takada, C., Salavati, R., Altman, S.** (1997) Phenotypic conversion of drug-resistant bacteria to drug sensitivity. *Proc. Natl. Acad. Sci. USA.* **94**: 8468-8472.
- Hager, P. W., and J. C. Rabinowitz** (1985) Translational specificity in *B. subtilis*, p. 1-29. In D. Dubnau (ed.). "The molecular biology of the Bacilli." Academic press. Inc., N.Y.
- Hall, M.N., J. Gabay, M. Debarbouille, M. Schwartz.** (1982) A role for mRNA secondary structure in the control of translation initiation. *Nature.* **295**(5850):616-8.
- Harley, C.B., Reynolds, R.P.** (1987) Analysis of *E. coli* promoter sequences. *Nucleic Acids Res.* **15**(5):2343-61.
- Harth, G., Zamecnik, P.C., Tang, J.Y., Tabatadze, D., Horwitz, M.A.** (2000) Treatment of *Mycobacterium tuberculosis* with antisense oligonucleotides to glutamine synthetase mRNA inhibits glutamine synthetase activity, formation of the poly-L-glutamate/glutamine cell wall structure, and bacterial replication. *Proc Natl. Acad. Sci. U. S. A.* **97**(1): 418-23.
- Hartmann, G., Honikel, K.O., Knusel, F., Nuesch, J.** (1967) The specific inhibition of the DNA-directed RNA synthesis by rifamycin. *Biochim Biophys Acta.* **145**(3):843-4.
- Hartz, D., McPheeters, D. S., and Gold, L.** (1989) Selection of the initiator tRNA by *E. coli* initiation factors. *Genes Dev.* **3**:1899-1912.
- Hartz, D., J. Binkley, T. Hollingsworth, and L. Gold** (1990) Domains of initiator tRNA and initiation codon crucial for tRNA selection by *E. coli* IF3. *Genes Dev.* **4**: 1790-1800.
- Hartz, D., McPheeters, D. S., and Gold, L.** (1991) Influence of mRNA determinants on translation initiation in *E. coli*. *J. Mol. Biol.* **218**: 83-97.
- Hawley, D.K., McClure, W.R.** (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* **11**(8):2237-55. Review.
- Helke, A., R. M. Geisen, M. Vollmer, M. L. Sprengart, E. Fuchs.** (1993) An unstructured mRNA region and a 5' hairpin represent important elements of the *E. coli* translation initiation signal determined by using the bacteriophage T7 gene 1 translation start site. *Nucleic Acids Res.* **21**(24):5705-11.

Hensel, M., J. E. Shea, C. Gleeson, M. D. Jones, E. Dalton, and D. W. Holden (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science* 269(5222): 400-403.

Hensel, M., D. W. Holden (1996) Molecular genetic approaches for the study of virulence in both pathogenic bacteria and fungi. *Microbiology* 142:1049-58.

Heym, B., Honore, N., Truffot-Pernot, C., Banerjee, A., Schurra, C., Jacobs, W.R. Jr., van Embden, J.D., Grosset, J.H., Cole, S.T. (1994) Implications of multidrug resistance for the future of short-course chemotherapy of tuberculosis: a molecular study. *Lancet*. 344(8918):293-8.

Heym B, Stavropoulos E, Honore N, Domenech P, Saint-Joanis B, Wilson TM, Collins DM, Colston MJ, Cole ST. (1997) Effects of overexpression of the alkyl hydroperoxide reductase AhpC on the virulence and isoniazid resistance of *Mycobacterium tuberculosis*. *Infect Immun*. 65(4):1395-401.

Hickey, M.J., Arain, T.M., Shawar, R.M., Humble, D.J., Langhorne, M.H., Morgenroth, J.N., Stover, C.K. (1996) Luciferase *in vivo* expression technology: use of recombinant mycobacterial reporter strains to evaluate antimycobacterial activity in mice. *Antimicrob Agents Chemother*. 40(2):400-7.

Hindley, J., and D. H. Staples. (1969) Sequence of a ribosome binding site in bacteriophage Q-beta-RNA. *Nature*. 224(223):964-7.

Hirashima, A., Childs, G., Inouye, M. (1973) Differential inhibitory effects of antibiotics on the biosynthesis of envelope proteins of *Escherichia coli*. *J Mol Biol*. 79(2):373-89.

Hjalt, T., Wagner, E.G. (1992) The effect of loop size in antisense and target RNAs on the efficiency of antisense RNA control. *Nucleic Acids Res*. 20(24):6723-32.

Hjalt, T.A., and E. Wagner (1995) Bulged-out nucleotides protect an antisense RNA from RNase III cleavage. *Nucleic Acids Res*. 23:571-579.

Hondalus, M.K., Bardarov, S., Russell, R., Chan, J., Jacobs, W.R., Jr, Bloom, B.R. (2000) Attenuation of and protection induced by a leucine auxotroph of *Mycobacterium tuberculosis*. *Infect Immun*. 68(5):2888-98.

Hooper, D.C., Wolfson, J.S. (1993) In: *Quinolone Antimicrobial Agents*, eds. Hooper, D.C. and Wolfson, J.S. (Am. Soc. Microbiol, Washington, D.C), pp. 53-75.

Hsu, A.W., Kilani, A.F., Liou, K., Lee, J., Liu, F. (2000) Differential effects of the protein cofactor on the interactions between an RNase P ribozyme and its target mRNA substrate. *Nucleic Acids Res*. 15;28(16):3105-16.

Hui, A., de Boer, H.A. (1987) Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A*. 84(14):4762-6.

Husson, R.N., B. E. James, R. A. Young (1990) Gene replacement and expression of foreign DNA in mycobacteria. *J. Bacteriol*. 172:519-24.

Ito, K., K. Kawakami, and Y. Nakamura. (1993) Multiple control of *E. coli* lysis-tRNA synthetase expression involved a transcriptional repressor and a translational enhancer element. *Proc. Natl. Acad. Sci. U. S. A*. 90: 302-306.

Inokuchi, Y., Yuyama, N., Hirashima, A., Nishikawa, S., Ohkawa, J., Taira, K. (1994) A hammerhead ribozyme inhibits the proliferation of an RNA coliphage SP in *Escherichia coli*. *J Biol Chem*. 15;269(15):11361-6.

- Iseman, M.D.** (1993) Treatment of multidrug-resistant tuberculosis. *N Engl J Med.* 9;329(11):784-91. Review.
- Iserentant, D., and W. Fiers.** (1980) Secondary structure of mRNA and efficiency of translation initiation. *Gene.* 9(1-2):1-12.
- Ish-Horowicz, D. and Burke, J.F.** (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Research.* 9: 2989-2998
- Ivey-Hoyle, M., Steege, D.A.** (1992) Mutational analysis of an inherently defective translation initiation site. *J. Mol. Biol.* 224(4):1039-54.
- Jackett, P.S., Aber, V.R., Lowrie, D.B.** (1978) Virulence of *Mycobacterium tuberculosis* and susceptibility to peroxidative killing systems. *J. Gen. Microbiol.* 107(2):273-8.
- Jackson, M., F. X. Berthet, I. Otal, J. Rauzier, C. Martin, B. Gicquel, C. Guilhot** (1996) The *Mycobacterium tuberculosis* purine biosynthetic pathway: isolation and characterization of the *purC* and *purL* genes. *Microbiology* 142: 2439-2447.
- Jackson, M., Crick, D.C., Brennan, P.J.** (2000) Phosphatidylinositol is an essential phospholipid of mycobacteria. *J Biol Chem.* 275(39):30092-9.
- Jacob, W. F., M. Santer, and A. E. Dahlberg.** (1987) A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proc. Natl. Acad. Sci. U. S. A.* 84(14): 4757-61.
- Jacobs, W.R. Jr., Tuckman, M., Bloom, B.R.** (1987) Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature.* 327(6122):532-5.
- Jacobs, W. R. Jr.** (1992) Advances in mycobacterial genetics: new promises for old diseases. *Immunobiology* 184: 147-156.
- Jacobs, W. R., Jr, Kapalana, G. V., Cirillo, J. D., Pascopella, L., Snapper, S. B., Udani, R. A., Jones, W., Barletta, R. G. and Bloom, B. R.** (1991) Genetic systems for *Mycobacteria*. *Methods Enzymol.* 204, 537-555.
- James, W., Cowe, E.** (1997) Computational approaches to the identification of ribozyme target sites. *Methods Mol Biol.* 74:17-26
- Janssen, G. R., Ward, J. M. and Bibb, M. J.** (1989) Unusual transcriptional and translational features of the aminoglycoside phosphotransferase gene (*aph*) from *Streptomyces fradiae*. *Genes Dev.* 3(3):415-29.
- Ji, Y., B. Zhang, S. F. Van Horn, P. Warren, G. Woodnutt, M. K. Burnham, and M. Rosenberg** (2001) Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 293: 2266-2269
- Jones, P.G., and Inouye, M.** (1994) The cold shock response – a hot topic. *Mol. Microbiol.* 11:811-818.
- Jones, P.G., and Inouye, M.** (1996) RbfA, a 30S ribosomal binding factor, is a cold-shock protein whose absence triggers the cold-shock response. *Mol. Microbiol.* 21:1207-1218.
- Kapalana, G.V., B. R. Bloom, W. R. Jr. Jacobs** (1991) Insertional mutagenesis and illegitimate recombination in mycobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 88: 5433-5437.
- Kastelein, R. A., B. Berkhout, G. P. Overbeek, and J. van Duin.** (1983) Effect of the sequences upstream from the ribosome-binding site on the yield of protein from the cloned gene for phage MS2 coat protein. *Gene.* 23(3):245-54.

- Kato-Maeda, M., Rhee, J. T., Gingeras, T.R., Salamon, H., Drenkow, J., Smittipat, N., Small, P.M. (2001) Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res.* 11(4):547-54.
- Kawa, D., Wand, J., Yuan, Y., Liu, F. (1998) Inhibition of viral gene expression by human ribonuclease P. *RNA* 4 (11): 1397-1406.
- Kirsebom, L.A., Svard, S.G. (1994) Base pairing between *Escherichia coli* RNase P RNA and its substrate. *EMBO J.* 17;13(20):4870-6.
- Kleckner, N. (1988) Transposon Tn10. In *Mobile DNA* ed. D. E. Berg, M. M. Howe, Washington: Am. Soc. Microbiol.
- Klock, G. and Hillen, W. (1986) Expression, purification and operator binding of the transposon Tn1721-encoded Tet repressor. *J Mol Biol.* 189(4):633-41.
- Kochi, A. (1994) Tuberculosis: distribution, risk factors, mortality. *Immunobiology* 191(4-5):325-36.
- Knipfer, N., Seth, A., Shrader, T.E. (1997) Unmarked gene integration into the chromosome of *Mycobacterium smegmatis* via precise replacement of the *pyrF* gene. *Plasmid.* 37(2):129-40.
- Kobayashi, M., Nagata, K., Ishihama, A. (1990) Promoter selectivity of *Escherichia coli* RNA polymerase: effect of base substitutions in the promoter -35 region on promoter strength. *Nucleic Acids Res.* 18(24):7367-72
- Kochi, A. (1991). The global tuberculosis situation and the new control strategy of the World Health Organisation. *Tubercle* 72: 1-6.
- Kolattukudy, P. E., N. D. Fernandes, A. K. Azad, A. M. Fitzmaurice, and T. D. Sirakova (1997) Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol. Microbiol.* 24(2):263-70.
- Kolb, F.A., C. Malmgren, E. Westhof, C. Ehresmann, B. Ehresmann, E. G. Wagner, P. Romby (2000) An unusual structure formed by antisense-target RNA binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment. *RNA.* 6:311-24.
- Kole, R., Baer, M.F., Stark, B.C., Altman, S. (1980) *E. coli* RNase P has a required RNA component. *Cell.* 19(4):881-7.
- 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(5):913-22.
- Kues, U., U. Stahl (1989) Replication of plasmids in gram-negative bacteria. *Microbiol. Rev.* 53: 491-516.
- Kufel, J., and Kirsebom, L.A. (1996) Different cleavage sites are aligned differently in the active site of M1 RNA, the catalytic subunit of *Escherichia coli* RNase P. *Proc. Natl. Acad. Sci. U.S.A.* 93(12):6085-90.
- Kushner. S. In: Neidhardt. F. C., Ingraham. J. L., Low. K.B., Magasanik. B., Schaechter. M., Umberger. H. E. eds. *E. coli* and Salmonella. Cellular and Molecular Biology: 2nd edn. Volume 1. Chapter 56: ASM Press, Washington DC 1996
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227: 680-685.
- Lancini, G., Pallanza, R., Silvestri, L.G. (1969) Relationships between bactericidal effect and inhibition of ribonucleic acid nucleotidyltransferase by rifampicin in *Escherichia coli* K-12. *J. Bacteriol.* 97(2):761-8.

- Lang, V., Gualerzi, C., McCarthy, J.E.** (1989) Ribosomal affinity and translational initiation in *Escherichia coli*. *In vitro* investigations using translational initiation regions of differing efficiencies from the *atp* operon. *J. Mol. Biol.* 210(3):659-63.
- Lange, R., and R. Hengge-Aronis.** (1994) The cellular concentration of the σ^s subunit of RNA polymerase in *E. coli* is controlled at the levels of transcription, translation, and protein stability. *Genes. Dev.* 8: 1600-1612.
- La Teana, A., Brandi, A., O'Connor, M., Freddi, S., Pon, C.L.** (2000) Translation during cold adaptation does not involve mRNA-rRNA base pairing through the downstream box. *RNA.* 6(10):1393-402.
- Lease, R. A., M. E. Cusick, M. Belfort** (1998) Riboregulation in *Escherichia coli*: *dsrA* RNA acts by RNA:RNA interactions at multiple loci. *Proc. Natl. Acad. Sci. U. S. A.* 95:12456-61.
- Lee, Y., and F. C. Schmidt** (1985) Characterization of the *in vivo* RNA product of the pOUT promoter of IS10R. *J. Bacteriol.* 164:556-62.
- Lee, N.S., Bertrand, E., Rossi, J.** (1999) mRNA localization signals can enhance the intracellular effectiveness of hammerhead ribozymes. *RNA.* 5(9):1200-9.
- Legocki, R. P., Legocki, M., Baldwin, T. O., and Szalay, A. A.** (1986) Bioluminescence in Soybean root nodules: Demonstration of a general approach to assay gene expression *in vivo* by using a bacterial luciferase. *Proc. Natl. Acad. Sci. USA* 83: 9080-9084
- Levin, M.E., Hatfull, G.F.** (1993) *Mycobacterium smegmatis* RNA polymerase: DNA supercoiling, action of rifampicin and mechanism of rifampicin resistance. *Mol Microbiol.* 8(2):277-85.
- Levine, C., Hiasa, H., Marians, K.J.** (1998) DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication, and drug sensitivities. *Biochim Biophys Acta.* 1;1400(1-3):29-43. Review.
- Li, Y., Guerrier-Takada, C., Altman, S.** (1992) Targeted cleavage of mRNA *in vitro* by RNase P from *Escherichia coli*. *Proc Natl Acad Sci U S A.* 15;89(8):3185-9.
- Liljenstrom, H., von Heijne, G.** (1987) Translation rate modification by preferential codon usage: intragenic position effects. *J. Theor. Biol.* 124(1):43-55.
- Lisser, S., Margalit, H.** (1993) Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res.* 21(7):1507-16.
- Liu, F., Altman, S.** (1996) Requirements for cleavage by a modified RNase P of a small model substrate. *Nucleic Acids Research* 24; 2690-2696
- Lodish, H.F.** (1969) Species specificity of polypeptide chain initiation. *Nature.* 224(222):867-70.
- Lodish, H.F.** (1970) Specificity in bacterial protein synthesis: role of initiation factors and ribosomal subunits. *Nature.* 226(247):705-7.
- Ma, C. and R. W. Simons.** (1990) The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site. *EMBO J.* 9:1267-74.
- Ma, M. Y., B. Jacob-Samuel, J. C. Dignam, U. Pace, A. R. Goldberg, S. T. George** (1998) Nuclease-resistant external guide sequence-induced cleavage of target RNA by human ribonuclease P. *Antisense Nucleic Acid Drug Dev.* 8: 415-426.
- Mahmoudi, A., Iseman, M.D.** (1993) Pitfalls in the care of patients with tuberculosis. Common errors and their association with the acquisition of drug resistance. *JAMA.* 7;270(1):65-8.

- Majdalani, N., C. Cuning, D. Sledjeski, T. Elliott, S. Gottesman (1998) DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc. Natl. Acad. Sci. U. S. A.* 95:12462-7.
- Mandell, G.L., and Sande, M.A. (1990) In: *The Pharmacological Basics of Therapeutics*, eds. Gilman, A.G., Rall, T.W., Nies, A.S., and Taylor, P. (Pergamon, New York), pp. 1146-1164.
- Marklund, B. I., D. P. Speert, R. W. Stokes (1995) Gene replacement through homologous recombination in *Mycobacterium intracellulare*. *J. Bacteriol.* 177: 6100-6105.
- Martin, L.D., and Rothschild, B.M. (1989) Frequency of pathology in a large natural sample from Natural Cave Trap (Late Pleistocene). *J. Vert. Paleontol.* 9:31A
- Martin, L.D., and Rothschild, B.M. (1998) Earth history and the evolution of sickness. In: Greenblatt, C.L., ed. *Digging for Pathogens: Ancient emerging diseases:their evolutionary, anthropological and archeological context*. Jerusalem: Balaban Publishers, pp. 15-46.
- Matteucci, M. D., R. W. Wagner (1996) In pursuit of antisense. *Nature* 7; 384(6604 Suppl):20-22.
- Mc Adam, R. A., Weisbrod, T.R., Martin, J., Scuderi, J. D., Brown, A. M., Cirillo, J. D., Bloom, B. R., Jacobs, W. R. (1995) *In vivo* growth characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. *Infect. Immun.* 63: 1004-1012
- Mc Carthy, J. E. G. and G. Gualerzi (1990) Translational control of prokaryotic gene expression. *Trends Genet.* 3:78-85.
- Mc Carthy, J. E. and Brimacombe, R. (1994) Prokaryotic translation: the interactive pathway leading to initiation. *Trends Genet.* 10(11): 402-7.
- McClain, W.H., Guerrier-Takada, C., Altman, S. (1987) Model substrates for an RNA enzyme. *Science.* 23:238(4826):527-30.
- McCray, E., and Onorato, I.M. (2000) In: *Multi-drug resistant tuberculosis*, I. Bastian, F. Fortaels, eds., (Kluwer, Dordrecht, Netherlands), pp. 45-57.
- McDermott, P. F., D. G. White, I. Podglajen, M. N. Alekshun, S. B. Levy (1998) Multidrug resistance following expression of the *Escherichia coli* *marA* gene in *Mycobacterium smegmatis*. *J. Bacteriol.* 180: 2995-2998.
- McFadden, J. (1996) Recombination in mycobacteria. *Mol. Microbiol.* 21: 205-211.
- McKinney, J., Guerrier-Takada, C., Wesolowski, D., Altman, S. (2001) Inhibition of *Escherichia coli* viability by external guide sequences complementary to two essential genes. *Proc Natl Acad Sci U S A.* 98(12):6605-10.
- Meighen, E. A. (1988) Enzymes and genes from the *lux* operons of bioluminescent bacteria. *Annu. Rev. Microbiol.* 42: 151- 176
- Middlebrook, G., and Cohn, M. L. (1953) Some observations on pathogenicity of isoniazid-resistant variants of tubercle bacilli. *Science* 118: 297-299
- Miller, J. H. (1992) *A Short Course in Bacterial Genetics*. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory.
- Mitta, M., Fang, L., and Inouye, M. (1997). Deletion analysis of *cspA* of *E. coli*: requirement of the AT-rich UP element for *cspA* transcription and the downstream box in the coding region for its cold shock-induction. *Mol. Microbiol.* 26: 321-335.

- Mizuno, S., Yamazaki, H., Nitta, K., Umezawa, H. (1968) Inhibition of DNA-dependent RNA polymerase reaction of *Escherichia coli* by an antimicrobial antibiotic, streptovaricin. *Biochim Biophys Acta*. 157(2):322-32.
- Moll, I., Huber, M., Grill, S., Sairafi, P., Mueller, F., Brimacombe, R., Londei, P., Blasi, U. (2001) Evidence against an Interaction between the mRNA downstream box and 16S rRNA in translation initiation. *J. Bacteriol.* 183(11):3499-505.
- Moll, I., S. Grill, C. Gualerzi, and U. Blasi (2002). Leaderless mRNA in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Microbiol.* 43 (1), 239-246.
- Morita, M. T., Tanaka, Y., Kodama, T. S., Kyogoku, H., Yanagi, H., and Yura, T. (1999) Translational induction of heat shock transcription factor σ^{32} : evidence for a built-in RNA thermosensor. *Genes Dev.* 13:655-665.
- Movahedzadeh, F., Colston, M.J., and Davis, E. O. (1997) Characterization of *M. tuberculosis* LexA: Recognition of a Cheo (Bacillus-type SOS) box. *Microbiology* 143:929-936.
- Mulder, M.A., Zappe, H., Steyn, L.M. (1997) Mycobacterial promoters. *Tuber Lung Dis.* 78(5-6):211-23. Review.
- Mulder, M. A., Zappe, H., and Steyn, L.M. (1999a) The *Mycobacterium tuberculosis katG* promoter region contains a novel upstream activator. *Microbiol.* 145, 2507-18.
- Mulder, N. J., Powles, R. E., Zappe, H., Steyn, L.M. (1999) The *Mycobacterium tuberculosis mysB* gene product is a functional equivalent of the *Escherichia coli* sigma factor, KatF. *Gene* In Press.
- Mulder, M.A., Nair, S., Abratt, V.R., Zappe, H., Steyn, L.M. (1999) Involvement of the N- and C-terminal domains of *Mycobacterium tuberculosis KatG* in the protection of mutant *Escherichia coli* against DNA-damaging agents. *Microbiology.* 145 (Pt 8):2011-21.
- Munson, L.M., Stormo, G.D., Niece, R.L., Reznikoff, W.S. (1984) *lacZ* translation initiation mutations. *J Mol Biol.* 177(4):663-83.
- Nagai, H., H.Yuzawa, T. Yura. (1991) Interplay of two cis-acting mRNA regions in translational control of sigma 32 synthesis during the heat shock response of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 88(23):10515-9.
- Nakahigashi, K., Yanagi, H., Yura, T. (1995) Isolation and sequence analysis of *rpoH* genes encoding sigma 32 homologs from gram negative bacteria: conserved mRNA and protein segments for heat shock regulation. *Nucleic Acids Res.* 23(21):4383-90.
- Nakashima, K., Kanamaru, K., Mizuno, T., and Horikushi, K. (1996) A novel member of the *cspA* family of genes that is induced by cold shock in *E. coli*. *J. Bacteriol.* 178: 2994-2997.
- Newton, G.L., Unson, M.D., Anderberg, S.J., Aguilera, J.A., Oh, N.N., delCardayre, S.B., Av-Gay, Y., Fahey, R.C. (1999) Characterization of *Mycobacterium smegmatis* mutants defective in 1-d-myo-inosityl-2-amino-2-deoxy-alpha-d-glucopyranoside and mycothiol biosynthesis. *Biochem Biophys Res Commun.* 255(2):239-44.
- Nigou, J., Besra, G.S. (2002) Cytidine diphosphate-diacylglycerol synthesis in *Mycobacterium smegmatis*. *Biochem J.* 10
- Noll, M., Noll, H. (1972) Mechanism and control of initiation in the translation of R17 RNA. *Nat New Biol.* 238(86):225-8
- Noller, H. F. (1993) On the origin of the ribosome: coevolution of subdomains of tRNA and rRNA, p. 137-156. In R. F. Gesteland and J. F. Atkins (ed.), *The RNA World*. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.

- Noller, H. F., and Nomura, M. (1996) Ribosomes, p. 167-187. In Neidhardt, F. C. (ed.), *E. coli and salmonella: Cellular and molecular biology*. ASM Press, Massachusetts, Washington, D. C.
- Nomura, M., Gourse, R., Baughman, G. (1984) Regulation of the synthesis of ribosomes and ribosomal components. *Annu. Rev. Biochem.* 53:75-117.
- Nordstrom, K., S. Molin, J. Light (1984) Control of replication of bacterial plasmids: genetics, molecular biology, and physiology of the plasmid R1 system. *Plasmid* 12:71-90.
- Nordstrom, K., and E. G. Wagner (1994) Kinetic aspects of control of plasmid replication by antisense RNA. *Trends Biochem. Sci.* 19:294-300.
- Norman, E., O. A. Dellagostin, J. McFadden, J. W. Dale (1995) Gene replacement by homologous recombination in *Mycobacterium bovis* BCG. *Mol. Microbiol.* 16: 755-760.
- Novick, R. P., S. J. Projan, C. C. Kumar, S. Carleton, A. Gruss *et al.* (1985) Replication control for pT181, an indirectly regulated plasmid. Pp 290-320. In *Plasmids in Bacteria*, ed, D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson, A. Hollaender. New York: Plenum.
- O'Connor, M., Asai, T., Squires, C. L. and Dahlberg, A. E. (1999) Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA. *Proc Natl Acad Sci U S A.* 96(16):8973-8.
- O' Donnell, S. M., and G. R. Janssen (2001). The initiation codon affects ribosome binding and translational efficiency in *E. coli* of cI mRNA with or without 5' untranslated leader. *J. Bacteriol.* 183: 1277-1283
- Oh, B.K., Pace, N.R. (1994) Interaction of the 3'-end of tRNA with ribonuclease P RNA. *Nucleic Acids Res.* 11;22(20):4087-94.
- Okamoto, K., and M. Freundlich (1986) Mechanism for the autogenous control of the *crp* operon: transcriptional inhibition by a divergent RNA transcript. *Proc. Natl. Acad. Sci. U S A.* 83:5000-4.
- Ordway, D.J., Sonnenberg, M.G., Donahue, S.A., Belisle, J.T., Orme, I.M. (1995) Drug-resistant strains of *Mycobacterium tuberculosis* exhibit a range of virulence for mice. *Infect Immun.* 63(2):741-3.
- Pace, N. R., Brown, J. W., (1995) evolutionary perspective on the structure and function of Ribonuclease P, a Ribozyme. *J. Bacteriol.* 177: 1919-1931.
- Pal, B.K., Scherer, L., Zelby, L., Bertrand, E., Rossi, J.J. (1998) Monitoring retroviral RNA dimerization in vivo via hammerhead ribozyme cleavage. *J Virol.* 72(10):8349-53.
- Pallen, M. J. (1999) Micobacterial genomes. *Mol. Microbiol.* 32: 907-912.
- Parish, T., Stoker, N.G. (1997) Development and use of a conditional antisense mutagenesis system in mycobacteria. *FEMS Microbiol Lett.* 154(1):151-7
- Parish, T., Liu, J., Nikaido, H., Stoker, N.G. (1997) A *Mycobacterium smegmatis* mutant with a defective inositol monophosphate phosphatase gene homolog has altered cell envelope permeability. *J Bacteriol.* 179(24):7827-33.
- Parish, T., Gordhan, B.G., McAdam, R.A., Duncan, K., Mizrahi, V., Stoker, N.G. (1999) Production of mutants in amino acid biosynthesis genes of *Mycobacterium tuberculosis* by homologous recombination. *Microbiology.* 145 (Pt 12):3497-503
- Parish, T., Stoker, N.G. (2000) Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. *Microbiology.* 146 (Pt 8):1969-75

- Parish, T., Lewis, J., Stoker, N.G. (2001) Use of the mycobacteriophage L5 excisionase in *Mycobacterium tuberculosis* to demonstrate gene essentiality. *Tuberculosis (Edinb)*. 81(5-6):359-64.
- Pavelka, M.S. Jr, Jacobs, W.R. Jr. (1996) Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of *Mycobacterium smegmatis*. *J Bacteriol*. 178(22):6496-507.
- Pavelka, M.S. Jr., W. R. Jr. Jacobs (1999) Comparison of the construction of unmarked deletion mutations in *Mycobacterium smegmatis*, *Mycobacterium bovis* bacillus Calmette-Guerin, and *Mycobacterium tuberculosis* H37Rv by allelic exchange. *J. Bacteriol*. 181:4780-4789.
- Pelcic, V., Reyrat, J.M., Gicquel, B. (1996a) Generation of unmarked directed mutations in mycobacteria, using sucrose counter-selectable suicide vectors. *Mol Microbiol*. 20(5):919-25.
- Pelcic, V., Reyrat, J.M., Gicquel, B. (1996b) Positive selection of allelic exchange mutants in *Mycobacterium bovis* BCG. *FEMS Microbiol Lett*. 144(2-3):161-6.
- Pelcic, 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(20):10955-60
- Pelcic, V., Reyrat, J.M., Gicquel, B. (1998) Genetic advances for studying *Mycobacterium tuberculosis* pathogenicity. *Mol Microbiol*. 28(3):413-20. Review.
- Perreault, J.-P., Altman, S. (1992) Important 2'-hydroxyl groups in model substrates for M1 RNA, the catalytic RNA subunit of RNase P from *Escherichia coli*. *J. Mol. Biol*. 226, 399-409.
- Persson, C, E. G. Wagner, K. Nordstrom (1990) Control of replication of plasmid R1: structures and sequences of the antisense RNA, CopA, required for its binding to the target RNA, CopT. *EMBO J*. 9:3767-75.
- Petersen, G. B., P. A. Stockwell, and D. F. Hill. (1988) Messenger RNA recognition in *Escherichia coli*: a possible second site of interaction with 16S ribosomal RNA. *EMBO J*. 7(12):3957-62.
- Petersen, C. (1989) Long-range translational coupling in the *rplJL-rpoBC* operon of *Escherichia coli*. *J. Mol. Biol*. 206:323-32.
- Petrelli, D., A. La Teana, C. Garolo, R. Spurio, C. L. Pon, and C. O. Gualerzi (2001) Translation initiation factor IF3: Two domains, five functions, one mechanism. *EMBO J*. 20:4560-4569.
- Plehn-Dujowich, D., Altman, S. (1998) Effective inhibition of influenza virus production in cultured cells by external guide sequences and ribonuclease P. *Proc. Natl. Acad. Sci. USA*. 95 (13): 7327-7332.
- Pohlner, J., J. Kramer, T. F. Meyer. (1993) A plasmid system for high-level expression and *in vitro* processing of recombinant proteins. *Gene*. 130(1):121-6.
- Polisky, B., X. Y. Zhang, T. Fitzwater (1990) Mutations affecting primer RNA interaction with the replication repressor RNA I in plasmid CoIE1: potential RNA folding pathway mutants. *EMBO J*. 9:295-304.
- Putlitz, J., Yu, Q., Burke, J.M., Wands, J.R. (1999) Combinatorial screening and intracellular antiviral activity of hairpin ribozymes directed against hepatitis B virus. *J Virol*. 73(7):5381-7.
- Queen, C., M. Rosenberg. (1981) Differential translation efficiency explains discoordinate expression of the galactose operon. *Cell*. 25(1):241-9.
- Rao, N. N. and Kornberg, A. (1996) Inorganic polyphosphate supports resistance and survival of stationary-phase *E. coli*. *J. Bact* 178: 1394-1400.

- Rapaport, E., A. Levina, V. Metelev, P. C. Zamecnik (1996) Antimycobacterial activities of antisense oligodeoxynucleotide phosphorothioates in drug-resistant strains. *Proc. Natl. Acad. Sci. U. S. A.* 93: 709-713.
- Raue, H. A., Klootwijk, J. and Musters, W. (1988) Evolutionary conservation of structure and function of high molecular weight ribosomal RNA. *Prog Biophys Mol Biol.* 51(2):77-129.
- Reddy P, A. Peterkofsky, K. McKenney (1985) Translational efficiency of the *Escherichia coli* adenylate cyclase gene: mutating the UUG initiation codon to GUG or AUG results in increased gene expression. *Proc Natl Acad Sci U S A.* 82(17):5656-60.
- Resch, A., Tedin, K., Grundling, A., Mundlein, A. and Blasi, U. (1996) Downstream box-anti-downstream box interactions are dispensable for translation initiation of leaderless mRNAs. *EMBO J.* 15(17):4740-8.
- Reyrat, J. M. , F. X. Berthet, B. Gicquel (1995) The urease locus of *Mycobacterium tuberculosis* and its utilization for the demonstration of allelic exchange in *Mycobacterium bovis* bacillus Calmette-Guerin. *Proc. Natl. Acad. Sci. U. S. A.* 92: 8768-8772.
- Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G.D., Gold, L. (1992) Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol.* 6(9):1219-29.
- Ringquist, S., Jones, T., Snyder, E. E., Gibson, T., Boni, I. and Gold, L. (1995) High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. *Biochemistry.* 34(11):3640-8.
- Rocha, E.P., Danchin, A., Viari, A. (1999) Translation in *Bacillus subtilis*: roles and trends of initiation and termination, insights from a genome analysis. *Nucleic Acids Res.* 27(17):3567-76.
- Rosenberg, A.H., Goldman, E., Dunn, J.J., Studier, F.W., Zubay, G. (1993) Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. *J. Bacteriol.* 175(3):716-22.
- Rothschild, B.M., Martin, L.D., Lev, G., Bercovier, H., Bar-Gal, G.K., Greenblatt, C., Donoghue, H., Spigelman, M., Brittain, D. (2001) *Mycobacterium tuberculosis* complex DNA from an extinct bison dated 17,000 years before the present. *Clin Infect Dis.* 33(3):305-11.
- Ruvkun, G. B., F. M. Ausubel (1981) A general method for site-directed mutagenesis in prokaryotes. *Nature* 89: 85-88.
- Sager, P., M. Schalimtzek, and V. Moller-Christensen (1972) A case of tuberculosis spondylosa in the Danish Neolithic age. *Danish Med. Bul.* 19: 176-180.
- Saito, H., and C. C. Richardson (1981) Genetic analysis of gene 1.2 of bacteriophage T7: isolation of a mutant of *Escherichia coli* unable to support the growth of T7 gene 1.2 mutants. *J. Virol.* 37:343-51.
- Salo, W.L., Aufderheide, A.C., Buikstra, J., Holcomb, T.A. (1994) Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci U S A.* 91(6):2091-4.
- Sander, P., Meier, A., Bottger, E.C. (1995) rpsL+: a dominant selectable marker for gene replacement in mycobacteria. *Mol Microbiol.* 16(5):991-1000.
- Sandler, P., Weisblum, B. (1988) Erythromycin-induced stabilization of *ermA* messenger RNA in *Staphylococcus aureus* and *Bacillus subtilis*. *J Mol Biol.* 203(4):905-15
- Sandler, P., Weisblum, B. (1989) Erythromycin-induced ribosome stall in the *ermA* leader: a barricade to 5'-to-3' nucleolytic cleavage of the *ermA* transcript. *J Bacteriol.* 171(12):6680-8.

- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad Sci USA* 74, 5463-5467.
- Sambrook, J., Fritsh, E. F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory.
- Sarkis, G.J., Jacobs, W.R. Jr., Hatfull, G.F. (1995) L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. *Mol Microbiol.* 15(6):1055-67.
- Schmetterer, G., Wolk, C.P., Elhai, J. (1986) Expression of luciferases from *Vibrio harveyi* and *Vibrio fischeri* in filamentous cyanobacteria. *J Bacteriol.* 167(1):411-4.
- Schneider, T. D., Stormo, G. D., Gold, L. and Ehrenfeucht, A. (1986) Information content of binding sites on nucleotide sequences. *J Mol Biol.* 188(3): 415-31.
- Schumperli, D., McKenney, K., Sobieski, D.A., Rosenberg, M. (1982) Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell.* 30(3):865-71.
- Seth, A. (1984) A new method for linker ligation. *Gene Analysis Techniques.* 1: 99-103
- Shean, C. S. and Gottesman, M. E. (1992) Translation of the prophage lambda *cI* transcript. *Cell* 70(3):513-22.
- Shiba, T., Tsutsumi, K., Yano, H., Ihara, Y., Kameda, A., Tanaka, K., Takahashi, H., Munekata, M., Rao, N. N. and Kornberg, A. (1997) Inorganic polyphosphate and the induction of *rpoS* expression. *Proc. Natl. Acad. Sci. USA.* 94; 11210-11215.
- Shine, J. and Dalgarno, L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A.* 71(4):1342-6
- Shinedling, S., Green, L., Barrick, G., Stormo, G. D. and Gold, L. Manuscript in prep
- Shinnick, T.M., King, C.H., Quinn, F.D. (1995) Molecular biology, virulence, and pathogenicity of mycobacteria. *Am. J. Med. Sci.* 309(2): 92-8. Review.
- Simons, R. W., and N. Kleckner (1983) Translational control of IS10 transposition. *Cell.* 34:683-91.
- Simons, R. W., and N. Kleckner (1988) Biological regulation by antisense RNA in prokaryotes. *Annu. Rev. Genet.* 22:567-600.
- Sippel, A., Hartmann, G. (1968) Mode of action of rifamycin on the RNA polymerase reaction. *Biochim Biophys Acta.* 157(1):218-9.
- Small, P.M., Hopewell, P.C., Singh, S.P., Paz, A., Parsonnet, J., Ruston, D.C., Schechter, G.F., Daley, C.L., Schoolnik, G.K. (1994) The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med.* 16;330(24):1703-9.
- 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(18):6987-91.
- Sprengart, M. L., Fatscher, H. P. and Fuchs, E. (1990) The initiation of translation in *E. coli*: apparent base pairing between the 16srRNA and downstream sequences of the mRNA. *Nucleic Acids Res.* 18(7):1719-23.
- Sprengart, M. L., Fuchs, E. and Porter, A. G. (1996) The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. *EMBO J.* 15(3):665-74.

- Sprengart, M. L., and A. G. Porter** (1997) Functional importance of RNA interactions in selection of translation initiation codons. *Mol. Microbiol.* 24: 19-28
- Springer, B., S. Master, P. Sander, T. Zahrt, M. McFalone, J. Song, K. G. Papavinasasundaram, M. J. Colston, E. Boettger, V. Deretic** (2001) Silencing of oxidative stress response in *Mycobacterium tuberculosis*: expression patterns of *ahpC* in virulent and avirulent strains and effect of *ahpC* inactivation. *Infect. Immun.* 69: 5967-5973.
- Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M.** (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. U. S. A.* 94(18): 9869-74.
- Stead, W.W., Eisenach, K.D., Cave, M.D., Beggs, M.L., Templeton, G.L., Thoen, C.O., Bates, J.H.** (1995) When did *Mycobacterium tuberculosis* infection first occur in the New World? An important question with public health implications. *Am J Respir Crit Care Med.* 151(4):1267-8.
- Steitz, J. A.** (1969) Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature.* 224(223):957-64
- Stephenson, M.L., P. C. Zamecnik** (1978) Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proc. Natl. Acad. Sci. U. S. A.* 75: 285-288.
- Stormo, G. D., Schneider, T. D. and Gold, L. M.** (1982a) Characterization of translational initiation sites in *E. coli*. *Nucleic Acids Res.* 10(9):2971-96.
- Stormo, G. D., Schneider, T. D., Gold, L. and Ehrenfeucht, A.** (1982b) Use of the 'Perceptron' algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Res.* 10(9): 2997-3011.
- Stormo, G. D.** In "Maximizing gene expression" (W. Reznikoff and L. Gold, eds.), p. 195. Butterworth, Stoneham, Massachusetts, 1986.
- Stover, C.K., de la Cruz, V.F., Fuerst, T.R., Burlein, J.E., Benson, L.A., Bennett, L.T., Bansal, G.P., Young, J.F., Lee, M.H., Hatfull, G.F., et al.** (1991) New use of BCG for recombinant vaccines. *Nature.* 351(6326):456-60.
- Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M..** (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A.* 94(18):9869-74.
- Sullenger, B.A., Cech, T.R.** (1993) Tethering ribozymes to a retroviral packaging signal for destruction of viral RNA. *Science.* 3;262(5139):1566-9.
- Suzuki, Y., Nagata, A., Yamada, T.** (1991) Analysis of the promoter region in the rRNA operon from *Mycobacterium bovis* BCG. *Antonie Van Leeuwenhoek.* 60(1):7-11.
- Tabatadze, M. A. Horwitz** (2000) Treatment of *Mycobacterium tuberculosis* with antisense oligonucleotides to glutamine synthetase mRNA inhibits glutamine synthetase activity, formation of the poly-L-glutamate/glutamine cell wall structure, and bacterial replication. *Proc. Natl. Acad. Sci. U. S. A.* 97: 418-423.
- Tanabe, H., Goldstein, J., Yang, M., and Inouye, M.** (1992) Identification of the promoter region of the *E. coli* major cold shock gene, *cspA*. *J. Bacteriol.* 174: 3867-3873.
- Tapprich, W. E., Goss, D. J., Dahlberg, A. E.** (1989) Mutation at position 791 in *Escherichia coli* 16S ribosomal RNA affects processes involved in the initiation of protein synthesis. *Proc Natl Acad Sci U S A* 86(13):4927-31

Tedin, K., Moll, I., Grill, S., Resch, A., Graschopf, A., Gualerzi, C. O. and Blasi, U. (1999) Translation initiation factor 3 antagonizes authentic start codon selection on leaderless mRNAs. *Mol Microbiol.* 31(1):67-77.

Thanaraj, T. A., and M. W. Pandit. (1989) An additional ribosome-binding site on mRNA of highly expressed genes and a bifunctional site on the colicin fragment of 16S rRNA from *Escherichia coli*: important determinants of the efficiency of translation-initiation. *Nucleic Acids Res.* 17(8):2973-85.

Thomas, D. Y., G. Dubuc, and S. Narang, (1982) *Escherichia coli* plasmid vectors containing synthetic translational initiation sequences and ribosome binding sites fused with the *lacZ* gene. *Gene.* 19(2):211-9

Tomizawa, J. (1986) Control of ColE1 plasmid replication: binding of RNA I to RNA II and inhibition of primer formation. *Cell* 47:89-97.

Tomizawa, J. (1990) Control of ColE1 plasmid replication. Intermediates in the binding of RNA I and RNA II. *J. Mol. Biol.* 212:683-94.

Trias, J., and Benz, R. (1994) Permeability of the cell wall of *Mycobacterium smegmatis*. *Mol. Microbiol.* 14(2):283-90.

Umezawa, H., Mizuno, S., Yamazaki, H., Nitta, K. (1968) Inhibition of DNA-dependent RNA synthesis by rifamycins. *J Antibiot (Tokyo).* 21(3):234-6.

Van Etten, W. J., and G. R. Janssen (1998) An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in *E. coli*. *Mol. Microbiol.* 27 (5):987-1001.

Vellanoweth, R.L., and J. C. Rabinowitz. (1992) The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. *Mol Microbiol.* 6(9):1105-14.

Verano, J.W., and Ubelaker, D.H. (1992) Disease and demography in the Americas. Washington, D.C.: Smithsonian Institution Press.

Vernon, A., Burman, W., Benator, D., Khan, A., Bozeman, L. (1999) Acquired rifamycin monoresistance in patients with HIV-related tuberculosis treated with once-weekly rifapentine and isoniazid. Tuberculosis Trials Consortium. *Lancet.* 29;353(9167):1843-7.

von Gabain, A., Belasco, J.G., Schottel, J.L., Chang, A.C., Cohen, S.N. (1983) Decay of mRNA in *Escherichia coli*: investigation of the fate of specific segments of transcripts. *Proc Natl Acad Sci U S A.* 80(3):653-7.

Vosloo, W., Tippoo, P., Hughes, J.E., Harriman, N., Emms, M., Beatty, D.W., Zappe, H., Steyn, L.M. (1997) Characterisation of a lipoprotein in *Mycobacterium bovis* (BCG) with sequence similarity to the secreted protein MPB70. *Gene.* 188(1):123-8.

Vosloo, W. (1998) Molecular studies on a novel lipoprotein from *Mycobacterium bovis* BCG. PhD. Thesis.

Wagner, E. G. H., J. von Heijne, and K. Nordstrom (1987) Control of replication of plasmid R1: translation of the 7k reading frame in the RepA mRNA leader region counteracts the interaction between *copA* RNA and *copT* RNA. *EMBO J.* 6:515-22.

Wagner, G. E., S. Brantl (1998) Kissing and RNA stability in antisense control of plasmid replication. *Trends Biochem. Sci.* 23:451-4.

Wagner, E. G. and Simons, R. W. (1994) Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.* 48:713-42.

- Wang, N., Yamanaka, K., and Inouye, M. (1999) CspI, the ninth member of the CspA family of *E. coli*, is induced upon cold shock. *J. Bacteriol.* 181:1603-1609.
- Wards, B.J., G. W. de Lisle, D. M. Collins (2000) An *esat6* knockout mutant of *Mycobacterium bovis* produced by homologous recombination will contribute to the development of a live tuberculosis vaccine. *Tuber. Lung Dis.* 80: 185-189.
- Wassarmen, K. M., A. Zhang, and G. Storz (1999) Small RNAs in *Escherichia coli*. *Trends Microbiol.* 7:37-45.
- WHO Report 2001. Global tuberculosis control, communicable diseases, World Health Organisation Geneva. <http://www.who.int/gtb/publications/globrepoi/index.html>.
- Williams, T.M., Burlein, J.E., Ogden, S., Kricka, L.J., Kant, J.A. (1989) Advantages of firefly luciferase as a reporter gene: application to the interleukin-2 gene promoter. *Anal Biochem.* 176(1):28-32.
- Wilson, T., G. W. de Lisle, J. A. Marcinkeviciene, J. S. Blanchard, D. M. Collins (1998) Antisense RNA to *ahpC*, an oxidative stress defence gene involved in isoniazid resistance, indicates that *AhpC* of *Mycobacterium bovis* has virulence properties. *Microbiology* 144: 2687-2695.
- Wimberley, B. T., D. E. Brodersen, W. M. Clemons, R. J. Warren, A. P. Carter, C. Vornhelm et al. (2000) Structure of the 30S ribosomal subunit. *Nature* 407:327-339.
- Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., Usman, N. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.* 23(14):2677-84.
- Winzeler, E. and Shapiro, L. (1997) Translation of the leaderless *Caulobacter dnaX* mRNA. *J Bacteriol.* 179(12):3981-8.
- Wittman, H. G. (1982) Components of bacterial ribosomes. *Annu. Rev. Biochem.* 51: 155-183.
- Woese, C. R., Gutell, R., Gupta, R., and Noller, H. F. (1983) Detailed analysis of the higher-order structure of 16S-like rRNA. *Microbiol. Rev.* 47: 621-669.
- Womble, D. D., X. Dong, R. H. Rownd (1987) Changes in RNA secondary structure may mediate the regulation of IncFII plasmid gene expression and DNA replication. In *Molecular biology of RNA: New perspectives*, ed. M. Inouye, B. S. Dudock, pp. 225-247. New York :Academic.
- Wong, H.C., Chang, S. (1986) Identification of a positive retroregulator that stabilizes mRNAs in bacteria. *Proc Natl Acad Sci U S A.* 83(10):3233-7.
- Wong, E. M. , and B. Polisky (1985) Alternative conformations of the ColE1 replication primer modulate its interaction with RNA I. *Cell* 42:959-66.
- Wood, K.V., de Wet, J.R., Dewji, N., DeLuca, M. (1984) Synthesis of active firefly luciferase by in vitro translation of RNA obtained from adult lanterns. *Biochem Biophys Res Commun.* 1984 124(2):592-6.
- Wood, K.V. (1990) Luc genes: introduction of colour into bioluminescence assays. *J Biolumin Chemilumin.* 5(2):107-14.
- Wood, K. V. (1991a) Recent advances and prospects for use of Beetle luciferases as genetic reporters. In: P. Stanley and L. Kricka (eds.). *Bioluminescence and chemiluminescence: current status*. John Wiley and sons, Chichester.
- Wood, K. V. (1991b) The origins of Beetle luciferases. In: P. Stanley and L. Kricka (eds.). *Bioluminescence and chemiluminescence: current status*. John Wiley and sons, Chichester.

- Yamanaka, K., Fang, L., and Inouye, M.** (1998). The CspA family in *E. coli*: Multiple gene duplication for stress adaptation. *Mol. Microbiol.* 27:247-255.
- Yamanaka, K., Mitta, M., and Inouye, M.** (1999). Mutation analysis of the 5'-UTR of the cold shock *cspA* mRNA of *E. coli*. *J. Bacteriol.* 181, 20: 6284-6291.
- Yanisch-Perron, C., Vieira, J., Messing, J.** (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene.* 33(1):103-19.
- Young, D. B., and S. T. Cole** (1993) Minireview: Leprosy, tuberculosis, and the new genetics. *J. Bacteriol.* 175: 1-6.
- Yu, Q., Pecchia, D.B., Kingsley, S.L., Heckman, J.E., Burke, J.M.** (1998) Cleavage of highly structured viral RNA molecules by combinatorial libraries of hairpin ribozymes. The most effective ribozymes are not predicted by substrate selection rules. *J Biol Chem.* 4;273(36):23524-33.
- Yuan, Y., E. S. Hwang, S. Altman** (1992) Targeted cleavage of mRNA by human RNase P. *Proc. Natl. Acad. Sci. U. S. A.* 89: 8006-8010.
- Yuzawa, H., Nagai, H., Mori, H., Yura, T.** (1993) Heat induction of sigma 32 synthesis mediated by mRNA secondary structure: a primary step of the heat shock response in *Escherichia coli*. *Nucleic Acids Res.* 21(23):5449-55.
- Zabeau, M. and Stanley, K. K.** (1982) Enhanced expression of cro- β -galactosidase fusion proteins under the control of the P_R promoter of bacteriophage λ . *EMBO J.* 1, 1217-1224.
- Zamecnik, P.C., M. L. Stephenson** (1978) Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proc. Natl. Acad. Sci. U. S. A.* 75: 280-284.
- Zhang, A., S. Altuvia, A. Tiwari, L. Argaman, R. Hengge-Aronis, G. Storz** (1998) The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J.* 17:6061-8.
- Zimmerman, M. R.** (1979) Pulmonary and osseous tuberculosis in an Egyptian mummy. *Bul. New York Acad. Med.* 55: 604-608.
- Zhou, T., Kim, J., Kilani, A.F., Kim, K., Dunn, W., Jo, S., Nepomuceno, E., Liu, F.** (2002) *In vitro* selection of external guide sequences for directing RNase P-mediated inhibition of viral gene expression. *J Biol Chem.* 277(33):30112-20.
- Zuker, M.** (1994) Computer Analysis of Sequence Data, Part II, A.M. Griffin & H.G Griffin, Eds., Volume 25, chapter 23, pages 267-294. CRC Press, Inc., Totowa, NJ

APPENDIX A

BACTERIAL STRAINS AND PLASMIDS

A.1. Bacterial Strains

The bacterial strains used in this study are summarized in table A.1.

Table A.1 Bacterial strains used in this study.

Strain	Genotype	Reference
<i>E. coli</i> LK111	<i>recA, lacI, lacZ, lacY</i> ⁺	Zabeau & Stanley, 1982
<i>E. coli</i> JM110	<i>F', traD36, lac^qΔ (lacZ) M15, proA⁺B⁺/rpsL (Str^r), thr, leu, thi, lacy, galK, galT, ara, fhuA, dam, dcm, supE44Δ (lac-proAB)</i>	Yanisch-Perron <i>et al.</i> , 1985
<i>M. smegmatis</i> LR222	High frequency transformation mutantf	Beggs <i>et al.</i> , 1995
<i>M. bovis</i> BCG (Tokyo) strain 172	Attenuated vaccine strain 172	State Vaccine Institute, Pinelands, S.A.

A.2. Plasmids

The vectors used in this study are summarized in table A.2, and the recombinant plasmids are listed in tables A.3, A.4, A.5, A.6, and A.8.

Table A.2. Vectors and cloning systems used in this study

Plasmid	Size (bp)	Markers	Description	Reference
pUC18	2686	Amp ^R , lacZ ⁺	M13-based cloning vector	Yanisch-Perron <i>et al.</i> , 1985
pUC19	2686	Amp ^R , lacZ ⁺	M13-based cloning vector	Yanisch-Perron <i>et al.</i> , 1985
pUCBM21	2725	Amp ^R , lacZ ⁺	M13-based cloning vector	Roche
pGem [®] -luc	4933	Amp ^R , lacZ ⁺	M13-based cloning vector	Promega
pJC86	5720	Kan ^R , lacZ ⁺	Luciferase cassette vector	Jack Crawford, CDC Atlanta, U.S.A.

			E. coli-mycobacterial shuttle vector	
--	--	--	--------------------------------------	--

Table A.3 Recombinant plasmids described in Chapter 3

Plasmid	Key Features	Technical Notes
pJCLuc	Promoterless luciferase gene in pJC86.	Michelle Mulder, PhD thesis, University of Cape Town.
pJCKatG	Luciferase gene under the control of <i>M. tuberculosis katG</i> promoter in pJCLuc.	Michelle Mulder, PhD thesis, University of Cape Town.
pUCBM21	M13-based cloning vector	Roche
pAT100B	PUC19 containing 1.8 kb region of <i>M. bovis</i> BCG <i>mpb83</i> gene	P Tippoo, University of Cape Town
pUC21Oligo17	trp terminator cloned into the <i>HindIII/SphI</i> restriction sites of pUCBM21	A plasmid used to subclone MPB83 Sau3AI fragments upstream of the trp terminator, creating pUC21Oligo17-5, 36, 24, 25
pJCKatG-MPB83-5	PJCKatG harbouring region 1 of <i>mpb83</i> in the sense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-36	PJCKatG harbouring region 1 of <i>mpb83</i> in the antisense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-24	PJCKatG harbouring region 2 of <i>mpb83</i> in the sense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-25	PJCKatG harbouring region 2 of <i>mpb83</i> in the antisense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>

Table A.4 Recombinant plasmids described in Chapter 4

Plasmid	Key Features	Technical Notes
pJCKatG-MPB83-5	PJCKatG harbouring region 1 of <i>mpb83</i> in the sense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-36	PJCKatG harbouring region 1 of <i>mpb83</i> in the antisense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>
pJCKatG-MPB83-24	PJCKatG harbouring region 2 of <i>mpb83</i> in the sense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>

pJCKatG-MPB83-25	PJCKatG harbouring region 2 of <i>mpb83</i> in the antisense orientation	Transcription vector introduced into <i>M. bovis</i> to generate sense/antisense <i>mpb83</i>

Table A.5 Recombinant plasmids described in Chapter 5

Plasmid	Key Features	Technical Notes
pJCLuc	Promoterless luciferase gene in pJC86.	Michelle Mulder, University of Cape Town.
pJCKatG	Luciferase gene under the control of <i>M. tuberculosis katG</i> promoter in pJCLuc	Michelle Mulder, University of Cape Town.
p283HH	Transcription vector containing hammerhead sequences and transcription terminator	Dr Celia Guerrier-Takada, Yale University
pJCImpCEGS1/2	Mycobacterium transcription vector expressing <i>M. smegmatis</i> ImpC EGS 1 or 2	Transcription controlled by <i>katG</i> promoter. EGSs cleaved from transcripts <i>in cis</i> by flanking hammerhead sequences.
pMtbImpCEGS1	External guide sequence complementary to nucleotides 22-36 of <i>M.tb impC</i> mRNA.	EGS synthesised <i>in vitro</i> using T7 polymerase. Complementary to a region of <i>impC</i> mRNA that is unlikely to form secondary structures.
pMtbImpCEGS2	External guide sequence complementary to nucleotides 136-150 of <i>M.tb impC</i> mRNA.	EGS synthesised <i>in vitro</i> using T7 polymerase. Complementary to a region of <i>impC</i> mRNA that is unlikely to form secondary structures.
pGemTMPB83	<i>mpb83</i> lipoprotein gene cloned in pGemT.	<i>mpb83</i> synthesised <i>in vitro</i> from pGemTMPB83 using T7 polymerase (819 base transcript).
pGem3Zf(+)-MPB83EGS1	External guide sequence complementary to nucleotides 10-25 of <i>mpb83</i> cloned in pGemT.	EGS synthesised <i>in vitro</i> using T7 polymerase. Complementary to a region of <i>mpb83</i> mRNA that is unlikely to form secondary structures (200 base full-length transcript; 74 b 5', 42 b EGS, 84 b 3').

pGem3Zf(+)-MPB83EGS2	External guide sequence complementary to nucleotides 138-153 of <i>mpb83</i> cloned in pGemT.	EGS synthesised <i>in vitro</i> using T7 polymerase. Complementary to a region of <i>mpb83</i> mRNA that is unlikely to form secondary structures (200 base full-length transcript; 74 b 5', 42 b EGS, 84 b 3').
pUC β -lactamase	<i>bla</i> gene cloned in pUC18.	<i>bla</i> gene synthesised <i>in vitro</i> with authentic 5' and 3' ends (1121 base transcript).

Table A.6 Recombinant plasmids described in Chapter 6

Plasmid	Key Features	Technical Notes
pJCLuc	Promoterless luciferase gene in pJC86.	Michelle Mulder, University of Cape Town.
pJCImpCEGS1/2	Mycobacterium transcription vector expressing <i>M. smegmatis</i> ImpC EGS 1 or 2	Transcription controlled by <i>katG</i> promoter. EGSs cleaved from transcripts <i>in cis</i> by flanking hammerhead sequences.

Table A.7 Recombinant plasmids described in Chapter 8

Plasmid constructs	Size (bp)	Description
pJCLuc	7420	<i>E. coli</i> -mycobacterial shuttle vector containing luciferase gene (pGem®-luc)
pJCLucKatG	7956	pJCLuc with <i>M. tb katG</i> promoter upstream of luciferase gene
pJCLucKatGΔSD	7945	pJCLucKatG with SD sequence upstream of <i>luc</i> gene removed
pEcLuc	7984	SD upstream of <i>luc</i> removed and DB optimized for binding to <i>E. coli</i> 16S rRNA
pMsLuc	7984	SD upstream of <i>luc</i> removed and DB optimized for binding to <i>M. smegmatis</i> 16S rRNA

A.3. CULTURE CONDITIONS

A.3.1 *E. coli*

E. coli cultures were grown aerobically in 2 X YT (yeast tryptone) medium. Agar was added to the medium at a final concentration of 1.5% (w/v). The medium was supplemented with kanamycin (50µg/ml), ampicillin (50µg/ml), X-GAL (100µg/ml), and IPTG (0.3 mM) where necessary (see Appendix C).

2X YEAST TRYPTONE (YT) LIQUID MEDIA:

16g Bacto-tryptone

10g Bacto-yeast extract

5g Sodium chloride

Add water to 1l and autoclave.

For agar plates: add 15g agar to the above mixture and autoclave.

A.3.2 *Mycobacteria*

Mycobacterial cultures were grown aerobically in Middlebrook 7H9 medium supplemented with catalase-free ADC and Tween 80. Agar was added to the medium at a final concentration of 1.5% (w/v). The medium was supplemented with kanamycin (40µg/ml), and cyclohexamide (0.01% (w/v) where necessary.

Middlebrook 7H9 medium:

2g/l Asparagine

0.5g/l Casitone

2.5g/l Sodium phosphate

1g/l Potassium phosphate

50mg/l Ferric ammonium citrate

10mg/ml Magnesium sulphate

0.5mg/ml Calcium chloride

0.1mg/l Zinc sulphate

0.1mg/l Copper sulphate

Make up to 900ml distilled water and autoclave.

ADC (albumin dextrose catalase):

5g Bovine serum Albumin (BSA) - fraction V

2g Glucose

Add water to 100ml and filter sterilize (Millipore filter - 20 μ m)

Use 100ml ADC per 900ml Middlebrook 7H9 (10%).

Middlebrook-ADC-Tween

900 ml Middlebrook 7H9

100ml ADC supplement

0.05% Tween 80

University of Cape Town

APPENDIX B

GENERAL TECHNIQUES

The enzymes and kits used in this study were supplied by Roche (Germany), Promega (Madison, U.S.A.), Amersham (UK), Pharmacia (Uppsala, Sweden). Chemicals used in this study were supplied by Melford Laboratories, Biorad laboratories, BDH Chemicals Ltd, Sigma Chemicals, Fluka Biochemicals and Merck.

B.1. CLONING

B.1.1. Restriction endonuclease digestion

Restriction endonuclease digestions were carried out according to the method of Sambrook *et al.* (1989), in a 20µl volume containing 0.5 – 10 µg of DNA in a solution of the supplied restriction endonucleases incubation buffer, and 10 units of restriction endonuclease. The digests were incubated at 37°C for between 1 and 18 hrs.

B.1.2. Blunt-ending with T4 DNA polymerase and Mung Bean nuclease

Cohesive ends resulting from restriction endonuclease digestion were blunt-ended using the Amersham DNA Blunting kit (Amersham Life Science, UK), according to the manufacturers instructions. The T4 DNA polymerase from the kit fills in 5' protruding ends with its 5'-3' polymerase activity and removes 3' protruding ends with its 3'-5' exonuclease activity. Alternatively, cohesive ends were removed by digestion with Mung Bean nuclease (Roche). The DNA was digested in a 50µl reaction containing 5µl of 10 X Mung Bean digestion buffer (0.5M sodium acetate pH5, 0.3 M sodium chloride, 0.01 M zinc sulphate), and 10 units of Mung Bean nuclease. The reaction mix was incubated at 30°C for 30 minutes, and the DNA was recovered by phenol/chlorophorm/isoamyl alcohol (25:24:1) extraction and precipitation with ethanol (see B.1.4).

B.1.3. Agarose gel electrophoresis

For agarose gel electrophoresis, a horizontal, submerged gel system was used, as described by Sambrook et al. (1989), with a tris-acetate buffer (see Appendix C). Agarose concentrations varied from 0.8% (w/v) to 3% (w/v). Gel tracking dye (see Appendix C) was added to the DNA sample prior to electrophoresis. The DNA was stained with ethidium bromide (0.2 µg/ml), visualized using a 310 nm transilluminator and photographed using a Kodac DC120 digital camera and software.

B.1.4. Phenol squeeze extraction of DNA from agarose gels

DNA was purified from agarose gels using a phenol extraction procedure (Seth, 1984). The gel band was finely chopped using a Feather surgical blade, an equal volume of phenol was added, and the sample was immediately frozen on dry ice for 10 minutes. The sample was centrifuged at 14 000 rpm for 10 minutes and the aqueous phase recovered. The DNA was extracted twice more by adding an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1), vortexing and centrifuging for 5 minutes. The aqueous phase was recovered and 1/10th of a volume of 3M sodium acetate (pH 5.2) and 2.5 volumes of ice cold absolute ethanol were added to precipitate the DNA. After incubation at -70°C for 15 minutes, the DNA was collected by centrifugation at 14 000 rpm for 15 minutes. The DNA was washed with 70% ethanol (v/v), dried, and resuspended in a minimum volume of sterile distilled water.

B.1.5. Ligation reactions

Insert and vector DNA were recovered from restriction endonuclease digestion or blunt-end reactions by extraction with phenol/chloroform/isoamyl alcohol (25:24:1) and precipitation with ethanol. The DNA was ligated according to the method of Sambrook et al. (1989) in a 15 µl reaction containing 10 X T4 DNA ligation buffer and 10 units of T4 ligase (Roche). Ligations were carried out at 16°C for 3-4 hours or overnight. In each ligation experiment, reactions were included which had no DNA or the vector self-ligated, as controls.

B.1.6. Preparation and transformation of competent *E. coli* cells

E. coli cells were made competent for the uptake of DNA according to the method of Chung and Millar (1988). Cells were grown in 5 ml 2 X YT broth to stationary phase. The culture was diluted 1/22 into 100 ml of 2 X YT broth and grown to early exponential phase ($OD_{600} = 0.5$) at 37°C. The cells were collected by centrifugation at 4000 rpm for 5 minutes at 4°C, resuspended in 1/10th of the volume in ice cold TSB (2 X YT broth containing 10% (w/v) PEG 4000, 5% (w/v) DMSO, 10 mM magnesium chloride, and 10 mM magnesium sulphate), and held on ice for 20 minutes before transformation. The competent *E. coli* cells were transformed by the addition of 2-5µl of ligated DNA to 50µl of cells. The cells were held on ice for 20 minutes and heat shocked at 42°C for 2 minutes. TSB (450µl) containing 20 mM glucose was added to the transformation mix, and the cells were incubated at 37°C for 1 hour to facilitate expression of the antibiotic resistance gene. A 200µl aliquot of the cells was plated onto 2 X YT agar plates with the appropriate antibiotic selection. Competent cells were frozen in dry-ice and ethanol and stored at -70°C.

B.1.7. Plasmid preparation from *E. coli*

Plasmids were isolated from *E. coli* using the alkaline lysis method of Ish-Horowicz and Burke (1981). A 5 ml overnight culture was harvested by centrifugation and the pellet was resuspended in 200µl of solution 1 (10 X stock: 0.25 M tris-Cl pH 8, 0.5 M glucose, 0.1 M EDTA pH8). The sample was kept at room temperature for 5 minutes, and 400µl of freshly prepared solution 2 (0.2 M sodium hydroxide, 1% (w/v) SDS) was added. The sample was placed on ice for 5 minutes. This was followed by the addition of 300µl of solution 3 (3 M potassium acetate, 2 M acetic acid, pH 4.8), and incubation on ice for a further 5 minutes. The precipitated proteins were removed by centrifugation at 14 000 rpm for 40 minutes and the DNA in the supernatant was precipitated by the addition of 0.8 volumes of isopropanol and recovered by centrifugation. The pellet was resuspended in water and re-precipitated with 1/10th of a volume of 3 M sodium acetate (pH 4.5), and 2.5 volumes of ice-cold 100% (v/v) ethanol. The pellet was washed with 0% (v/v) ethanol and re-suspended in water. Large scale DNA preparations were performed using the Nucleobond AX PC-kit 100 (Macherey-Nagel), according to manufacturers instructions.

B.2. Isolation of chromosomal DNA from mycobacterial cells

A 1000 ml culture of Mycobacterial cells was harvested by centrifugation at 3000 rpm for 10 minutes at 4°C, and the pellet resuspended in 500µl of TES (10 mM tris, 1 mM EDTA, 0.15 M sodium chloride) (Jacobs *et al.*, 1991). The cells were disrupted by 2 cycles in a biospec Mini-Bead beater at 4200 rpm for 30 seconds each. The cell debris was removed by centrifugation. The supernatant was extracted twice with phenol/chlorophorm/isoamyl alcohol (25:24:1) and the top phase was recovered. The DNA was precipitated with 1/10th of a volume of 3M sodium acetate (pH 5.2) and 2.5 volumes of ice cold absolute ethanol. The pellet was washed with 70% (v/v) ethanol and resuspended in sterile distilled water. The RNA was removed by digestion with RNase A, and the DNA was extracted twice more with phenol/chlorophorm/isoamyl alcohol (25:24:1). The aqueous phase was collected and the DNA was precipitated as above. The DNA pellet was resuspended in 150-200µl of sterile distilled water. The quality of the chromosomal DNA was checked by electrophoresis on a 0.8% (w/v) agarose gel, and the concentration of the DNA was determined by measuring the absorbance at 260 nm on a Beckman DU-40 Spectrophotometer. The concentration was calculated using the relationship $OD_{600} = 1$ for 50µg of dsDNA.

B.3. Plasmid isolation from mycobacteria

Plasmids were isolated from mycobacteria using the method of Jacobs et al. (1991). A 5 ml culture was grown to late exponential phase and the cells harvested by centrifugation. The pellet was resuspended in 150µl of GTE (25 mM tris-CL pH 8, 10 mM EDTA, 50 mM glucose), containing 10 mg/ml lysozyme, and incubated at 37°C overnight. Two volumes of a freshly prepared 0.2 M sodium hydroxide, 1% (w/v) SDS solution was added

, and the suspension was incubated on ice for 10 minutes. After incubation, µl of 3 M potassium 225acetate was added and the sample was incubated on ice for a further 10 minutes. The proteins and cell debris were removed by centrifugation at 14 000 rpm for 30 minutes, and the DNA in the supernatant was extracted with phenol/chlorophorm/isoamyl alcohol (25:24:1). The DNA was precipitated with 2 volumes of ethanol, and the pellet was washed with 70 % (v/v) ethanol and resuspended in 25-50µl of TE.

B.4. Bio-Rad assay

Protein concentration was determined using a Bio-Rad DC Protein Assay kit. The protein standards, 0.2 mg/ml, 0.5 mg/ml, 0.75 mg/ml, and 1.0 mg/ml, were prepared by dilution of the Bio-Rad bovine serum albumin protein standard (1.5 mg/ml) in the appropriate buffer. For each assay, 100µl of solution A' (20µl of solution S/ml of solution A) was mixed with 20µl of protein sample, and 800µl of solution B. The samples were incubated at room temperature for 15-20 minutes, and the OD at 750 nm determined. The OD₇₅₀ values of the Bio-Rad protein standards were plotted against concentration to give a linear graph. The concentration of protein in the samples was determined from the graph.

B.5 SDS-Polyacrylamide gel electrophoresis.

SDS-Polyacrylamide gels were poured and electrophoresed in a Hoefer gel pouring stand and a vertical electrophoresis apparatus. The gel preparation and electrophoresis were performed according to the method of Laemmli (1970). A resolving gel of acrylamide in 1.5 M tris (pH 8.8), 0.4% (w/v) SDS buffer was poured first and allowed to polymerize for at least 1 hour. The stacking gel in a 0.5 M tris (pH6.8), 0.4% (w/v) SDS buffer was then poured and allowed to polymerize overnight. Gels were electrophoresed at a constant current of 40 mA in a tris-glycine based buffer (10 X solution: 15.15 g tris, 72.05 g glycine, 1% (w/v) SDS in 1 litre). Protein samples were mixed with a sample buffer (see Appendix C) and boiled for 5-10 minutes before loading. Non-denaturing polyacrylamide gel electrophoresis was performed as described for SDS-PAGE, except that SDS was omitted from all buffers. SDS-polyacrylamide gels were stained with Coomassie[®] Brilliant Blue R-250 (Bio-Rad Laboratories) (1.25 g Coomassie[®] Brilliant Blue R-250 in 50% (v/v) methanol, 10% (v/v) acetic acid) for 4-8 hours, and destained with 30% (v/v) methanol, 10% (v/v) acetic acid, and 3% glycerol.

B.6. Immunoblotting

Immunoblotting was performed using a method adapted from Sambrook et al. (1989). The proteins of interest were electrophoresed in an SDS-polyacrylamide gel, as described above. The proteins were transferred to an Amersham Hybond-C Super

nitrocellulose membrane in a transfer buffer (22.4 g glycine, 4.8 g tris, and 400 ml methanol in 2 litres of water), using an Omeg Scientific semi-dry blotter, at a constant current of 200 mA. After transfer, the membrane was washed twice in TBS (0.05 M tris-Cl pH7.5, 0.15 M sodium chloride) and incubated in 1% (w/v) blocking solution (Chemiluminescence Blotting Substrate POD kit, Roche) for 1 hour. This was followed by incubation in 0.5% (w/v) blocking solution containing the primary antibody for 1 hour. The unbound antibody was removed by washing the membrane twice in TBST (TBS + 0.1% (v/v) Tween 20) for 10 minutes each. The membrane was soaked in 0.5% blocking solution for 2 X 10 minutes, then incubated with POD-labeled anti-Rabbit-IgG (Roche) diluted 1 in 10000 in 0.5 % (w/v) blocking solution for 30 minutes. The secondary antibody was removed by washing the membrane 4 times in TBST for 15 minutes each. The kit detection solution was prepared by mixing reagents A and B in a 100:1 ratio and allowing them to stand at room temperature for 30 minutes before use. The bound antibodies on the membrane were detected by incubating in detection reagent for 1 minute and exposure to X-ray film for 30 to 60 seconds.

APPENDIX C

BUFFERS AND SOLUTIONS

Chemicals used in this study were supplied by Melford Laboratories (England), Bio-Rad Laboratories (CA, U.S.A.), BDH Chemicals Ltd (Poole, England), Sigma Chemicals (St. Louis, U.S.A.), Fluka Biochemicals, and Merck (Darmstadt, Germany). The compositions of the commonly used solutions are give below.

Ampicillin

100 mg/ml stock solution

2 g in 20 ml of water

Filter sterilize and store at -20°C

Cell Culture Lysis Reagent (Promega)

25 mM tris-phosphate (pH 7.8)

2 mM DTT

2 mM 1,2-diaminocyclohexane-N,N,N',N'-tetraacetic acid

10% (v/v) glycerol

1% (v/v) Triton X-100

Luciferase/Luciferin L/L reagent (Promega)

Purified luciferase

d-luciferin-BSA

DTT

Tris-acetate (pH7.75)

EDTA

Magnesium acetate

L/L Buffer

0.1 M tris-acetate (pH 7.75)

0.2 2 mM EDTA

Luciferase Assay Reagent (Promega)

20 mM tricine

33.3 mM DTT

270 μ M coenzyme A
470 μ M luciferin
530 μ M ATP
0.1 μ M EDTA
1.07 μ M magnesium carbonate
2.67 μ M magnesium sulphate

IPTG

0.056 g IPTG in 2 ml water
Filter sterilize and store at -20°C

DNA Analysis

Ethidium bromide

10 mg/ml stock solution:
0.2 g ethidium bromide in 20 ml of water
store in the dark

Gel tracking Dye

0.25 g bromophenol blue
40.0 g sucrose
4.0 ml 0.5 EDTA (pH 8.0)
Make up to 100 ml with water

10X NNB (1l):

162g Tris
27.5g Boric acid
9.3g EDTA

10X Klenow buffer:

0.1M Tris.HCl pH 7.6
0.1M Magnesium chloride
0.5M Sodium chloride
10mM DTT
0.5mg/ml BSA
Distilled water

50X TRIS ACETATE BUFFER (TAE) (1l):

242g Tris
57.1ml Glacial acetic acid
100ml 0.5M EDTA pH 8.0
Use 1X TAE when running agarose gels.

20X SSPE (stock) (1l):

174g Sodium chloride

27.6g Sodium dihydrogen phosphate

7.4g EDTA, pH 7.4.

50X DENHARDTS SOLUTION (stock):

1% Ficoll

0.1% Polyvinylpyrrolidone

1% BSA fraction v.

HYBRIDIZATION BUFFER:

50% Formamide

5XSSPE

0.1% SDS

5X Denhardts solution

0.1% SDS

100µg/ml Herring sperm DNA.

RNA analysis:**10X MOPS buffer (1l):**

41.8g MOPS (Morpholinepropanesulfonic acid)

800ml DEPC-treated water

Adjust to pH 7.0

16.6ml 3M DEPC-treated sodium acetate

20ml 0.5M DEPC-treated EDTA (pH 8.0)

20X SSC (stock) (1l):

175.3g Sodium chloride

88.2g Sodium citrate

Adjust to pH 7.0

LOADING DYE :

0.3% Bromophenol blue

0.3% Xylene cyanol

10mM EDTA pH 7.5

97.5% Formamide

ELECTROPHORESIS SAMPLE BUFFER (RNA):

0.75ml Formamide
0.15ml 10X MOPS/EDTA buffer
50mM Sodium acetate, pH 7.0.
0.1ml Glycerol
0.08ml Bromophenol blue (10%)
0.1ml DEPC-treated water.

3x AQUEOUS HYBRIDIZATION SOLUTION:

3M Sodium chloride
0.5M HEPES pH 7.5
1mM EDTA pH 8

HYBRIDIZATION BUFFER:

50% Formamide
5XSSPE
0.1% SDS
100ug/ml Herring sperm DNA
5X Denhardts solution
0.1% SDS
100µg/ml herring sperm DNA.

1.5% DENATURING AGAROSE/FORMALDEHYDE GEL (100ml):

1.5g Agarose
10ml 10X MOPS buffer
87ml DEPC-treated water
5.1ml 37% Formaldehyde

6% ACRYLAMIDE/ 8M UREA DENATURING GEL (100ml):

48g Urea
39ml Distilled water
15ml 40% (w/v) Acrylamide/ 2% (w/v) bis-acrylamide
10ml 10X NNB

To polymerise the acrylamide add 350µl 13% (w/v) ammonium persulphate and 80µl of N,N,N',N'-tetramethyl-ethylenediamine (TEMED)

Protein analysis:**TRANSFER SOLUTION (2l):**

4.8g Tris
22.4g Glycine

400ml Methanol

TRIS-BUFFERED SALINE (TBS):

50mM Tris

150mM Sodium chloride

Adjust pH to 7.5

PHOSPHATE-BUFFERED SALINE (PBS) (pH 7.2):

137mM Sodium chloride

2.7mM Potassium chloride

4.3mM Di-sodium hydrogen phosphate

1.4mM Potassium di-hydrogen phosphate

SDS LOADING BUFFER:

3% (w/v) SDS

5% (v/v) β -mercaptoethanol

10% (v/v) Glycerol

0.625M Tris.Cl (pH 6.8)

0.01% (w/v) Bromophenol blue

Non denaturing polyacrylamide gel mix

8 ml 5 X tris-glycine stock

6.67 ml 30 % (w/v) acrylamide

1 ml 2% (w/v) bis-acrylamide

2 ml 50% glycerol

22.37 ml water