Transcription of rDNA in mycobacteria

Thesis

How to cite:

For guidance on citations see FAQs.

© 2001 Kristine Bourke Arnvig

Version: Version of Record

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.21954/ou.ro.0000fcfd

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.

oro.open.ac.uk
Transcription of rDNA in Mycobacteria

Kristine Bourke Arnvig, B. Sc., Cand. Scient.

Thesis submitted to Open University
for the degree of Doctor of Philosophy in Life Sciences

August 2001

Division of Mycobacterial Research
National Institute for Medical Research
Mill Hill, London
Abstract

The genus *Mycobacterium* consists of numerous species, which display large variations in growth requirements and growth rates. They do however have certain common characteristics including growth rates that are lower than those of many other bacteria and a low number of *rrn* operons. As a rule the slower growing species, including the human pathogen *Mycobacterium tuberculosis*, have a single *rrn* operon, *rrnA*, and the faster growing species have two operons, *rrnA* and *rrnB*. The promoters of these operons vary between species and they display differential regulation upon a change in growth conditions. Nothing is known about the regions responsible for the regulation of *rrn* expression. In the present study DNA fragments from the promoter and leader regions of mycobacterial *rrn* operons were fused to a promoter-less *lacZ* gene and the resulting reporter constructs were assayed for β-galactosidase activity. This investigation led to the identification of a number of regulatory regions, which act at different stages of gene expression and with different effects. The results demonstrate that the two operons have certain features in common. These include a highly activating upstream region and a downstream/leader region that suppresses *rrn* expression. *In vitro* interactions between the *M.tbc rrn* leader and *M.tbc* Nus factors furthermore suggest that the *M.tbc rrn* anti-termination mechanism may differ significantly from the *E. coli* paradigm.
1 Introduction

1.1 Tuberculosis and the Mycobacteria .............................................. 10
1.2 Mycobacterial Growth ................................................................. 11
1.3 Macromolecular Synthesis ............................................................ 13
1.4 Transcription ................................................................................ 15
  1.4.1 RNA polymerase .................................................................. 15
    1.4.1.1 The core enzyme ....................................................... 15
    1.4.1.2 The holoenzyme ........................................................ 19
    1.4.1.3 The σ factors of Mycobacteria ..................................... 23
  1.4.1.4 Activated transcription – the targets ............................... 25
  1.4.2 Bacterial promoters ................................................................. 27
  1.4.3 The promoters of Mycobacteria .............................................. 29
  1.4.4 Mechanisms of transcription ................................................... 33
    1.4.4.1 Initiation ................................................................. 33
    1.4.4.2 Elongation ............................................................ 35
    1.4.4.3 Termination of transcription ....................................... 42
1.5 Translation – the ribosome ............................................................. 46
1.6 The \(rrn\) operons ............................................................................. 46
  1.6.1 The \(rrn\) promoters ............................................................... 49
  1.6.2 The \(rrn\) leader region ............................................................ 50
  1.6.3 The \(rrn\) promoters and leader regions of Mycobacteria .......... 52
  1.6.4 Regulation of ribosome synthesis (lessons from \(E.\ coli\)) ...... 56
    1.6.4.1. Balancing rRNA and ribosomal protein ...................... 56
    1.6.4.2 Feedback regulation .................................................. 57
    1.6.4.3 The Stringent Response ............................................ 59
    1.6.4.4 Stringent Response in Mycobacteria .......................... 63
    1.6.4.5 Growth Rate Dependent Regulation ......................... 64
    1.6.4.6 Growth Rate Dependent Regulation in Mycobacteria? 65
  1.6.4.7 Activated \(rrn\) transcription ............................................. 67
  1.6.4.8 \(rrn\) Anti-termination ...................................................... 71
    1.6.4.8 Anti-termination – the mechanism .............................. 81
2 Aim of the study ................................................................................. 84
3 Materials and methods ......................................................................... 85
  3.1 Cloning of mycobacterial promoter fragments ........................... 85
  3.2 Site directed mutagenesis ............................................................. 87
  3.3 Ligations ..................................................................................... 88
  3.4 Competent cells of \(M.\ sm\) .............................................................. 89
3.5 Transformations______________________________________________89
3.6 DNA isolation________________________________________________90
3.7 RNA isolation________________________________________________91
3.8 Growth of bacteria___________________________________________92
3.9 Preparation of cell free extracts______________________________92
3.10 Protein assays______________________________________________93
3.11 DNA sequencing____________________________________________94
3.12 Primer extension____________________________________________95
3.13 Expression and purification of M.tb RNAP subunits_______________96
3.14 In vitro transcription________________________________________99
3.15 Electrophoretic mobility shift assays___________________________100

4 The expression of rrn operons (Results)__________________________102
4.1 Introduction________________________________________________102
4.2 Activity of rrn operons_______________________________________104
4.3 The transcription start point of reporter constructs______________109
4.4 Feedback regulation of rrn expression___________________________111
4.5 The rrn operons: Discussion__________________________________115

5 Core promoter activities (Results)_______________________________118
5.1 Introduction________________________________________________118
5.2 The effect of the native transcription start point________________119
5.3 Comparison of minimal promoters______________________________120
5.3.1 Mutations in the M.tb PCL1 promoter________________________122
5.3.2 An extended -10 promoter___________________________________123
5.4 Mycobacterial core promoters: Discussion_______________________124

6 Regulatory elements of rrn operons (Results)____________________128
6.1 The M.tb rrnA operon: Introduction____________________________128
6.2 The P1 promoter____________________________________________131
6.3 The PCL1 promoter___________________________________________135
6.4 Regulatory regions of the rrnA operon: Discussion________________137
6.5 The M.sm rrnB operon: Introduction____________________________140
6.6 The downstream and leader regions of the rrnB operon___________143
6.7 The rrnB upstream activating region___________________________146
   6.7.1 Phase of the helix dependent activation_____________________150
6.7.2 Interaction between the *rrnB* UAR and the RNA polymerase α subunit 153
6.8 Species specificity of the *rrnB* regulatory elements 158
6.9 Regulatory regions of the *M.sm* *rrnB* operon: Discussion 162

7 The *M.tb* *rrn* leader region (Results) 164
7.1 Introduction to *rrn* leader functions 164
7.2 The mycobacterial *rrn* leader’s influence on *lacZ* expression 165
7.3 Hybrid reporter constructs 171
7.4 Interactions between Nus factors and the *rrn* leader 173
  7.4.1 NusA and the *rrn* leader 177
  7.4.2 NusB and the *rrn* leader 179
7.5 Functions of the *M.tb* *rrn* leader region: Discussion 182

8 Towards an *M.tb* in vitro transcription system (Results) 186
8.1 Species specificity of RNA polymerases 186
8.2 Expression of the *M.tb* RNA polymerase subunits 188
8.3 The *M.tb* in vitro transcription system: Discussion 190

9 Discussion 195
9.1 Mycobacteria and their *rrn* operons 195
9.2 Expression of *rrn* operons 198
9.3 Feedback regulation of *rrn* operons 200
9.4 The mycobacterial minimal and core promoters 202
9.5 The promoters of the *M.tb* *rrnA* operon 204
9.6 The *M.tb* *rrnA* leader region 208
9.7 The *M.sm* *rrnB* operon 212
9.8 The *rrn* operons of Mycobacteria – a comparison 214

10 Future prospects 217

Acknowledgements 219

References 220

Appendices 251

Appendix A: Promoter and leader sequences 252
  *M.tb* *rrnA* 253
  *M.sm* *rrnA* 254
  *M.sm* *rrnB* 255

Appendix B: List of reporter plasmids 256
List of figures

1.1 Functional regions of σ70-like σ factors 21
1.2 The concentration of ppGpp determines the mRNA CGR 41
1.3 General structure of mycobacterial rrn operons 53
1.4 Sequence alignment of RNAP αCTDs 70
4.1 The promoter probe vector, pEJ414 103
4.2 The rrn operons of M.tb and M.sm 105
4.3 The transcription start point of reporter constructs 110
4.4 Growth rates of M.sm wild type and rrn \* strains 112
4.5 Activity of rrn operons in rrn knockout background 114
6.1 The M.tb rrnA reporter constructs 129
6.2 Activities of the M.tb P1 reporter constructs 133
6.3 Activities of the M.tb PCL1 reporter constructs 136
6.4 Core promoter regions of M.tb P1, PCL1, and M.sm P_B 139
6.5 The initial dissection of the M.sm rrnB operon 141
6.6 The regulatory effects of the M.sm rrnB downstream region 144
6.7 The upstream activating region of the M.sm rrnB operon 148
6.8 Phase of the helix dependent activation of transcription 152
6.9 Gel shift assays with the rrnB UAR and M.tb RNAP 157
6.10 Species specificity of the rrnB regulatory elements 160
7.1 Truncation of the \textit{M.\textit{tb rrn}} leader region inhibits \textit{lacZ} expression 167

7.2 The \textit{M.\textit{tb rrn}} leader region 169

7.3 \textit{\beta}-gal activities of hybrid reporter constructs 172

7.4 \textit{In vitro} transcription of the \textit{M.\textit{tb rrn}} leader region 175

7.5 \textit{In vitro} interactions between \textit{M.\textit{tb}} NusA and the \textit{rrnA} leader 178

7.6 \textit{In vitro} interactions between \textit{M.\textit{tb}} NusB and the \textit{rrnA} leader 180

8.1 Overexpression of \textit{M.\textit{tb}} RNAP subunits in \textit{E. coli} 189

9.1 A model of the interaction between \textit{M.\textit{tb}} NusB and the \textit{rrn} leader 211

9.2 The mycobacterial model \textit{rrn} operon 216
# List of Tables

1.1 RNAP subunit homologies

1.2 Growth rates and $rrn$ gene dosage

1.3 The $rrn$ operons of different mycobacterial species

4.1 Activities of full-length $rrn$ reporter constructs

5.1 Activities of minimal and core promoter constructs

5.2 Activities of minimal promoter constructs

5.3 Activities of mutant minimal promoter constructs

6.1 The *M.tb* $rrnA$ reporter constructs

6.2 The *M.sm* $rrnB$ reporter constructs

6.3 The effects of the $rrnB$ upstream and downstream regions

6.4 Phase of the helix dependent activation of transcription

6.5 Species specificity of the $rrnB$ regulatory elements

7.1 Expression of the $rrnA$ leader region in *M.sm* and *E.coli.*

9.1 The activating potential of the $rrn$ promoters' upstream regions
**List of abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APS</td>
<td>Ammonium Persulphate</td>
</tr>
<tr>
<td>BCG</td>
<td>Bacille Calmette-Guérin</td>
</tr>
<tr>
<td>bp</td>
<td>basepair</td>
</tr>
<tr>
<td>B.s</td>
<td><em>B. subtilis/Bacillus subtilis</em></td>
</tr>
<tr>
<td>β-gal</td>
<td>β-galactosidase</td>
</tr>
<tr>
<td>CAP</td>
<td>Catabolite Activator Protein</td>
</tr>
<tr>
<td>CGR</td>
<td>Chain Growth Rate</td>
</tr>
<tr>
<td>CTD</td>
<td>carboxy-terminal domain</td>
</tr>
<tr>
<td>αCTD</td>
<td>carboxy-terminal domain of α subunit</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>DOTS</td>
<td>Direct Observation Treatment, Short-course</td>
</tr>
<tr>
<td>E.c</td>
<td><em>E. coli/Escherichia coli</em></td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Dalton</td>
</tr>
<tr>
<td>KO</td>
<td>Knockout</td>
</tr>
<tr>
<td>M.ab</td>
<td><em>Mycobacterium abscessus</em></td>
</tr>
<tr>
<td>M.ch</td>
<td><em>Mycobacterium chelonae</em></td>
</tr>
<tr>
<td>M.sm</td>
<td><em>Mycobacterium smegmatis</em></td>
</tr>
<tr>
<td>M.tb</td>
<td><em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MP</td>
<td>Minimal promoter</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>NTD</td>
<td>amino-terminal domain</td>
</tr>
<tr>
<td>αNTD</td>
<td>amino-terminal domain of α subunit</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleoside triphosphate</td>
</tr>
<tr>
<td>ppGpp</td>
<td>di-phospho-guanosine-diphosphate</td>
</tr>
<tr>
<td>Rif</td>
<td>Rifampicin (and Rifamycin etc)</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA Polymerase</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>r-protein</td>
<td>ribosomal protein</td>
</tr>
<tr>
<td>S.c</td>
<td>S. coelicolor/Streptomyces coelicolor</td>
</tr>
<tr>
<td>SDM</td>
<td>Site-directed mutagenesis</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>SDS polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Taq</td>
<td>Thermus aquaticus</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>T.t</td>
<td>T. thermophilus</td>
</tr>
<tr>
<td>UAR</td>
<td>Upstream Activating Region</td>
</tr>
<tr>
<td>UP element</td>
<td>Upstream element</td>
</tr>
<tr>
<td>wt</td>
<td>wild type</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Tuberculosis and the Mycobacteria

*Mycobacterium tuberculosis* (*M.tb*) is the causative agent of tuberculosis. On an annual basis this bacterium is responsible for more human deaths than any other infectious agent and the human, social, and economical implications are therefore vast (Snider *et al.*, 1994). For many years the illness was considered almost eradicated in the industrialised part of the world. However, over the last five to ten years the situation has changed and the disease has re-entered this part of the world. Tuberculosis is still mainly associated with poor social circumstances and the majority of infected individuals are found in third world countries. The rise in tuberculosis incidence correlates with the emergence of multi-drug resistant strains of the bacterium as well as the increased incidence of HIV and AIDS which greatly increases the susceptibility to tuberculosis (Snider *et al.*, 1994). Another consequence of HIV infection and AIDS is increased susceptibility to infection by otherwise non-pathogenic mycobacteria. Furthermore the Calmette-Guérin (BCG) vaccine is not as efficient as was initially believed and the efficacy displays strong regional variations (e.g. Bloom and Fine, 1994). The reasons for these variations remain controversial but factors such as the presence of
environmental mycobacteria, differences in the strain used for vaccination, and different ways of administering the vaccine are all believed to play a role (Bloom and Fine, 1994; Grange, 1996). A relatively new approach in the battle against tuberculosis is the introduction of the so-called Direct Observation Treatment, Short-course or DOTS, which has been advised by the WHO. The strategy involves either daily/weekly visits to a clinic where the drugs are administered by a health-care worker or even admission of patients to the clinic in order to make sure that the drugs are taken. Alternatively relatives of the patient can be assigned to observe the correct administration of the drugs. When carried out correctly the strategy is very efficient but it is still subject to major debate because of its social implications and impracticalities in some situations (Dye et al., 1998; Walley et al., 2001). The present situation therefore requires an increased effort in the general understanding of the bacterium in order to be able to identify new putative drug targets or vaccination strategies.

1.2 Mycobacterial Growth

The genus *Mycobacterium* can be divided into slow growing species and fast (or faster) growing species. Another means of dividing the mycobacterial species is according to their pathogenicity. Thus, there are the obligate pathogens, opportunistic pathogens and non-pathogenic mycobacteria. Interestingly all of the
obligate pathogens belong to the slow growing category. These include *M. tb*, *M. africanum*, *M. microti*, and *M. bovis*, which make up the so-called tuberculosis-complex (Goodfellow and Wayne, 1982), as well as the more distantly related *M. leprae*. *M. bovis* is the parental strain of *M. bovis* BCG, an attenuated strain used for vaccination. Members of the *M.tb* complex have doubling times of 10 to 30 hours *in vitro* whereas the *in vitro* culturing of *M. leprae* has been unsuccessful so far. Among the fast growing species is *M. smegmatis* (*M.sm*) with a minimal doubling time of two to three hours *in vitro*. Because of its fast growth, easy culturing, non-pathogenicity, and an easily transformed strain, mc²155 (Snapper *et al.*, 1990), *M.sm.* is often used as a surrogate host for expression of genes from the pathogenic mycobacteria, in particular *M.tb*. Apart from the slow growth, *M.tb* (and mycobacteria in general) are characterised by the ability to enter a state of dormancy (Wayne, 1976; Dick *et al.*, 1998; Lim *et al.*, 1999). The cells are able to persist in this non-replicating state in an infected individual for decades without provoking any symptoms in the host. When circumstances are favourable for the bacterium e.g. in the case where the infected individual is immune-compromised, the bacterium subsequently emerges from its dormancy and the result is an active infection (Snider *et al.*, 1994). Dormancy gives rise to at least two major problems. The bacterium loses susceptibility to rifampicin and isoniazid, antibiotics, which only have an effect on actively growing cells. It is therefore
vital that the patient be treated with a combination of antibiotics in order to target both growing as well as dormant bacteria (Wayne and Sramek, 1994; Young, 1994). The medical treatment is lengthy and compliance is often poor, with many patients terminating their therapy when acute symptoms subside, believing they are cured. This favours the emergence of multi-drug resistant strains. As mentioned above mycobacterial pathogenicity correlates with slow growth, and the transition into dormancy is also believed to be related to a change in growth conditions, possibly by a gradual reduction in available oxygen (Wayne, 1976; Dick et al., 1998; Lim et al., 1999). It is therefore of relevance to study growth and growth related mechanisms in mycobacteria in order to get a better understanding of the bacterium as such and to identify potential new drug targets.

1.3 Macromolecular Synthesis

It is still not known how the growth of pathogenic mycobacteria is maintained at its characteristically low rate or what the implications are for the pathogenicity of the bacteria. Is the slow growth a requirement for infection, virulence, or persistence? All aspects of macromolecular synthesis, i.e. transcription, translation, and DNA replication, as well as cell wall synthesis and cell division are tightly linked to the growth and growth stage of the cell. It has been suggested that slow growth is a consequence of the low number of rrn operons (Wheeler and
Ratledge, 1994), but for a number of reasons this may be confusing cause and effect. First, in *E. coli* it is the growth rate that determines the expression of *rrn* operons not *vice versa* (i.e. growth rate dependent regulation, see section 1.6.4.5). Second, it is possible to reduce *rrn* gene dosage without affecting the total amount of rRNA synthesised (Ellwood and Nomura, 1980; Condon *et al.*, 1993); in other words the *rrn* operons are operating at sub-maximal levels most of the time. A similar situation has been observed in *M. sm*, where the deletion of either of the two *rrn* operons (i.e. a 50% reduction in gene dosage) does not have a dramatic effect on growth (Sander *et al.*, 1996) (and this work). Finally it is likely that the slow growing mycobacteria would have duplicated their rDNA if this were the only reason for slow growth. Alternatively the synthesis of the complex cell wall has been suggested as being rate limiting for growth (Winder, 1982). Yet another possibility lies in the replication rate of the genome. The *M.tb* replication rate has been estimated to be approximately 3000 basepairs per minute which is roughly ten times slower than that of *M.sm* and almost 20 times slower than that of *E. coli* (Hiriyanna and Ramakrishnan, 1986). Transcription elongation has been measured to be approximately ten times slower in *M.tb* than in *E. coli* (Harshey and Ramakrishnan, 1977). Judging by the homology between the *M.tb* and *M.sm* enzymes and the *in vitro* elongation rate of the latter (Chamberlin *et al.*, 1979), the *M.tb* RNA polymerase is probably capable of faster elongation rates. Finally, since transcription and translation are generally tightly coupled in prokaryotes, the
peptide elongation rate in *M.tbc* is presumably correspondingly slow as well. In other words it appears that all aspects of macromolecular synthesis in *M.tbc* are very slow but the question remains: which is cause and which is effect? In a mixed population, a fast growing strain would quickly out-compete a slow growing one were there not distinct advantages in the slow growth. So the slow growth is likely to be a trait which has been carefully selected for.

1.4 Transcription

Transcription is the first step of gene expression and the level at which most gene expression is regulated. Transcription is carried out by the RNA polymerase (RNAP), a multisubunit enzyme that shares extensive sequence homologies between bacteria (Table 1.1) and which also displays structural and functional similarities with both archaeal and eukaryotic RNAPs (Ebright, 2000).

1.4.1 RNA Polymerase

1.4.1.1 The core enzyme

The bacterial RNAP core enzyme consists of four different subunits with the composition α₂ββ'ω (e.g. Zhang *et al.*, 1999; Minakhin *et al.*, 2001a). Other factors become associated with the enzyme at different stages of transcription and
the boundaries between core components, holoenzyme, and transcription factors (elongation, termination, or anti-termination) are not unequivocal.

The \textit{in vivo} assembly of the core enzyme begins with a dimerisation of the two $\alpha$ subunits via a specific dimerisation domain in the N-terminal part of the protein (Ishihama, 1981; Igarashi \textit{et al.}, 1991a). The assembly pathway is:

$$ \begin{array}{c} \beta \\ \beta'\omega \\ 2\alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \alpha_2\beta\beta'\omega \end{array} $$

(Ishihama, 1981; Minakhin \textit{et al.}, 2001b). Each $\alpha$ monomer binds one of the large subunits, $\beta$ and $\beta'$ implying that the two $\alpha$ monomers have different structures and roles in the final complex (Heyduk \textit{et al.}, 1996; Zhang \textit{et al.}, 1999). Apart from initiating the assembly of the polymerase, the $\alpha$ subunits are involved in the regulation of transcription on a number of levels, i.e. activated transcription, transcriptional pausing, termination, and anti-termination (see below).

The $\beta$ and $\beta'$ subunits are encoded in a single operon, \textit{rpoBC} (Burgess \textit{et al.}, 1987). The two subunits comprise the catalytic centre of the RNAP and interact with each other throughout, with most of both molecules sharing numerous functions (Zhang \textit{et al.}, 1999). Both subunits are involved in extensive DNA contacts downstream of the promoter elements (i.e. $-35$ and $-10$ regions) during transcription initiation (Naryshkin \textit{et al.}, 2000). Suppressors of polyauxotrophy associated with ppGpp deficient strains (see section 1.6.4.3) map to both \textit{rpoB} and
rpoC and recently it has been shown that ppGpp interacts with both subunits (Cashel et al., 1996; Toulokhonov et al., 2001a). Furthermore a number of mutations altering the "stringency" of the RNAP or the responsiveness to ppGpp have been described and these are also found in both subunits (Oostra et al., 1981; Nene and Glass, 1983; Tedin and Bremer, 1992; Zhou and Jin, 1998).

Both \( \beta \) and \( \beta' \) are also involved in elongation, pausing and termination and mutations resulting in altered pausing and termination patterns map to both rpoB and rpoC (Landick et al., 1990; Weilbaecher et al., 1994). The \( \beta \) and \( \beta' \) subunits do, however, display quite distinct interactions with the other constituents of the elongation complex, i.e. downstream DNA, DNA:RNA hybrid, upstream DNA, and RNA transcript (Nudler, 1999; Zhang et al., 1999; Korzheva et al., 2000).

The \( \beta \) subunit contains the substrate binding site and is involved in the formation of the promoter open complex as well as promoter clearance (Godovikova et al., 1987; Jin and Turnbough, 1994; Severinov and Darst, 1997; Naryshkina et al., 2001). Furthermore a specific, highly conserved domain of \( \beta \), the "flap" domain, is believed to interact with the RNA hairpin in transcriptional pausing and possibly termination (Zhang et al., 1999; Toulokhonov et al., 2001b). The \( \beta \) subunit harbours the rifampicin (Rif) binding domain and all mutants conferring Rif resistance map to the \( \beta \) subunit (e.g. Jin and Gross, 1988). Curiously it has been suggested that the \( \beta' \) subunit is required for the formation of the Rif binding
site (Naryshkina et al., 2001). Since it is evident from the recently published Rif-RNAP structure that there are no contacts between the Rif molecule and β', perhaps this subunit induces an allosteric change of the Rif binding site on the β subunit (Campbell et al., 2001). Finally, the β subunit has been proposed to interact with NusA (section 1.6.4.7).

A distinct feature of the β' subunit are two zinc binding motifs which are both found in highly conserved segments of β'. The most N-terminal of the two motifs is involved in anti-termination of phage HK022 but an actual binding of zinc remains to be demonstrated (Clerget et al., 1995). The second site has been demonstrated to bind zinc and this chelation is important for the proper assembly of the polymerase. Mutating the residues that make up the binding site (or in vitro assembly of the polymerase in the absence of zinc) leads to a misfolding of the β' subunit and subsequently to an enzyme without activity (Markov et al., 1999). The latter site has been confirmed as zinc binding in the Taq crystal structure (Zhang et al., 1999).

The ω subunit is the only non-essential constituent of the E. coli RNAP core enzyme and it is believed to act as a β'-specific chaperone. The RNAP purified from an ω deletion mutant is associated with and dependent on GroEL for
function (Mukherjee et al., 1999). Furthermore this small subunit has been found to solely interact with the β' subunit (Zhang et al., 1999). Finally, overexpression of ω can suppress an assembly defect of a mutant β' subunit (Minakhin et al., 2001a). So far the ω subunit does not appear to be involved in RNAP function as such, since reconstituted RNAP with and without ω are indistinguishable in in vitro transcription assays (Zalenskaya et al., 1990; Minakhin et al., 2001a).

Bacillus spp have been reported to have two ω subunits associated with RNAP purified from exponentially growing cells (Achberger et al., 1982b). Recently Ebright and co-workers have suggested a mycobacterial ω homologue (Minakhin et al., 2001a) but whether this protein (Rv1390) actually functions as such in M.tb remains to be demonstrated. The two proteins do not display extensive sequence homology (Table 1.1), but the fact that the suggested M.tb rpoZ lies downstream of the same gene (gmk) as the E. coli rpoZ, substantiates that Rv1390 could indeed be a true ω homologue. An in vitro reconstitution of the M.tb RNAP with and without ω would address this question. It should be noted that in contrast to ω the eukaryotic counterpart, RPB6, is essential (Minakhin et al., 2001a).

1.4.1.2 The holoenzyme

The RNAP core enzyme is capable of initiating transcription from non-specific DNA sequences whereas sequence specific transcription initiation is conferred by
incorporation of one of a number of σ factors resulting in the RNAP holoenzyme (Burgess et al., 1969; Gross et al., 1992). Hence σ factors are the main determinants in which genes are to be transcribed, and the exchange of one σ factor for another leads to the differential expression of entire regulons under various growth conditions. Different organisms have different numbers of σ factors. Which σ factor gets incorporated into the RNAP is determined by the relative abundance of the factor, the affinity between RNAP core and the factor, and the presence of anti-σ factors that interfere with the association of σ factor and RNAP core enzyme (Ishihama, 2000). The primary σ factors are highly conserved between species, and the protein can be divided into functional regions 1 to 4 with further subdivisions (Gross et al., 1992; Record et al., 1996; Bown et al., 1997), (Figure 1.1). Region 2.1, which is conserved between primary σ factors (as well as some alternative σ factors) has been shown to be both necessary and sufficient for binding to RNAP core (Lesley and Burgess, 1989; Gross et al., 1992). However, additional regions are also involved in the σ-RNAP core interaction (Sharp et al., 1999). The σ factors interact with the core enzyme via the N-terminal half of both β and β’ (Arthur and Burgess, 1998; Owens et al., 1998). The σ factor is furthermore involved in the isomerisation from closed to open complex (Sen et al., 1998; Bown et al., 1999; Burr et al., 2000). The interactions with the promoter DNA are in E. coli σ70 conferred by regions 2.4
Table 1.1: RNAP subunit homologies.

<table>
<thead>
<tr>
<th>Species</th>
<th>α</th>
<th>β</th>
<th>β'</th>
<th>ω</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. smegmatis</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>92</td>
</tr>
<tr>
<td>E. coli</td>
<td>60</td>
<td>58</td>
<td>58</td>
<td>37</td>
</tr>
<tr>
<td>B. subtilis</td>
<td>85</td>
<td>71</td>
<td>65</td>
<td>53</td>
</tr>
<tr>
<td>S. coelicolor</td>
<td>86</td>
<td>83</td>
<td>N.I.</td>
<td>87</td>
</tr>
</tbody>
</table>

N.I. not identified

The table illustrates the degree of conservation of RNAP subunits expressed as percent similarity to the M.tb subunits. The genome sequence of S. coelicolor has not been finished and the β' homologue could not be identified. The species listed in the table are mentioned at different stages in the text.

Figure 1.1: Functional regions of primary σ factors.

The figure illustrates how σ70-like σ factors of bacteria can be divided into functional regions interacting with RNA polymerase (binding to the RNAP core enzyme, region 2.1), transcription factors (activation, region 4.2), promoter elements (-10 box, region 2.4; extended -10 box or TG motif, region 2.5; -35 box, region 4.2). Other functional regions have been omitted for simplicity. Figure adapted from (Record et al., 1996; Bown et al., 1997).
(-10 region), 2.5 (the extended -10 box, section 1.4.2) and 4.2 (-35 box) (Record et al., 1996; Bown et al., 1997) (Figure 1.1).

The RNAP, both core and holoenzyme, purified from a number of Bacillus spp includes an additional factor, δ, encoded by the rpoE gene (Achberger et al., 1982b; Lampe et al., 1988). This factor alters the interactions between the RNAP and the promoter in vitro but sequence specific initiation of transcription still depends on the presence of σ (Achberger et al., 1982a; Hyde et al., 1986; Dobinson and Spiegelman, 1987; Lopez de Saro et al., 1999). Some lines of evidence point towards the δ subunit being of particular importance in growth phase transitions but deletion of rpoE does not confer any apparent phenotype and the overall function of the δ subunit is not fully understood (Lampe et al., 1988; Lopez de Saro et al., 1999).

A BLAST search with the B. subtilis RNAPδ sequences indicates sequence homologues in e.g. Lactococcus lactis and Staphylococcus aureus (p values 5 x 10^{-12} and 6 x 10^{-12} respectively). Functional homologues with poorer sequence homology are found in Mycoplasma genitalium and Mycoplasma pneumoniae (p values 0.46 and 0.92, respectively). No homologues are found in mycobacteria but a functional homologue with different sequences is a possibility.
Knowledge about mycobacterial RNAP is relatively scarce compared to that of the *E. coli* enzyme. Overall the core enzyme has the same subunit composition as other bacterial polymerases, $\alpha_2\beta\beta'\omega$, and the individual subunits share a high degree of homology with those of other species (Predich *et al.*, 1995; Cole *et al.*, 1998) (Table 1.1).

1.4.1.3 The $\sigma$ factors of Mycobacteria

$\sigma$ factors confer the promoter specificity to the RNAP (Burgess *et al.*, 1969). *M. tb* has a total of 13 different $\sigma$ factors, all belonging to the $\sigma^{70}$ family and most of them belonging to the extracytoplasmic class of $\sigma$ factors (Cole *et al.*, 1998; Manganelli *et al.*, 1999). The $\sigma$ factors have been identified by sequence homology or by function and they display quite different expression patterns depending on the growth conditions (Doukhan *et al.*, 1995; Cole *et al.*, 1998; Manganelli *et al.*, 1999). SigA and SigB are the two principal sigma factors, but of the two only SigA is essential in *M. sm* and this is likely to be the case in *M. tb* as well (Cole *et al.*, 1998; Gomez *et al.*, 1998). A single amino acid substitution in this protein (R522H) is responsible for rendering a virulent strain of *M. bovis* avirulent (Collins *et al.*, 1995). Whether the attenuation of the strain is due to an altered RNAP-promoter interaction or a reduced (missing) interaction with a transcriptional activator has not been established.
SigF, and recently SigE, are the only σ factors that have been experimentally deleted in *M.tb* to date (Chen *et al.*, 2000; Manganelli *et al.*, 2001). The sigF deletion strain shows very little difference in survival when exposed to stresses that induce the expression of this σ factor. The mutant strain does, however, appear to have alterations in the cell envelope, and is less virulent than its wild type counterpart. Interestingly the SigF knockout strain grows to a much higher optical density before entering stationary phase (Chen *et al.*, 2000). SigF displays homology to sporulation specific σ factors from *B. subtilis* and *S. coelicolor* and is likely to be involved in the transition to and persistence of dormancy (DeMaio *et al.*, 1996). The “delay” in entering stationary phase, i.e. closing down the cell machinery, seen with the SigF knockout strain, is consistent with this hypothesis. Interestingly this particular σ factor appears to initiate transcription from one of the two promoters of the *M.tb* *rrnA* operon, P1 (Chen and Bishai, 1998). SigF has been suggested to be specific for the slow growing mycobacteria but a BLAST search revealed a SigF homologue in *M.sm* (81% identity to *M.tb* SigF, but not yet assigned as such).

SigE is involved in the expression of sigB, and the sigE deletion mutant appears to be less virulent than its wild type counterpart, H37Rv (Manganelli *et al.*, 2001).
1.4.1.4 Activated transcription – the targets

Activated transcription refers to an acceleration of one or more steps in the transcription initiation pathway (see section 1.4.4.1) (Rhodius and Busby, 1998). The activation can involve direct interaction with any of the RNAP subunits but in the majority of cases the C-terminal domain of the $\alpha$ subunit ($\alpha$CTD) is the target or the mediator (reviewed by Busby and Ebright, 1994; Rhodius and Busby, 1998; Severinov, 2000).

The primary $\sigma$ factor of $E. coli$, $\sigma^{70}$, acts as a target for a number of activators, an interaction that usually involves additional contacts with RNAP (Li et al., 1994; Artsimovitch et al., 1996; Lonetto et al., 1998; Rhodius and Busby, 2000). Although it has only been demonstrated for some of these activators, it seems likely that they all have dual interactions with the RNAP due the spatial arrangements between activator and RNAP.

There are few examples of the $\beta$ and $\beta'$ subunits interacting with transcriptional activators. The $C_4$-dicarboxylic acid transporter protein (DCTD) of $R. meliloti$ interacts with the $\beta$ subunit in $\sigma^{54}$-RNAP (Lee and Hoover, 1995), whereas the N4SSB protein of bacteriophage N4 interacts with the $\beta'$ subunit (Miller et al., 1997).
The N-terminal domain of α (αNTD) is also targeted by activators. CAP (Catabolite Activator Protein) is an interesting example in that it can activate transcription in two mechanistically distinct ways contacting either the αCTD or both the αNTD, αCTD, and σ^{70} (reviewed by Busby and Ebright, 1999). The difference is determined by the structure of the promoter, which also gives rise to the classification as either class I or class II (Busby and Ebright, 1999). At a class I CAP-dependent promoter, the CAP binding site is located upstream of the –35 region, activation is via interactions between CAP and αCTD only, and the result is an increase in the binding constant (K_B, section 1.4.4.1) between template and RNAP, i.e. a recruitment of the enzyme to the promoter. Activators acting in the same fashion include *E. coli* OmpR, *A. tumefaciens* VirG, and *B. subtilis* phage Φ29 p4. This type of activation involves highly specialised factors and can be species specific regarding both factor and the αCTD (Igarashi *et al.*, 1991b; Mencia *et al.*, 1998; Lohrke *et al.*, 1999). Due to the flexible inter-domain linker of the RNAPα subunit, a characteristic of activation mediated by the αCTD is “phase of the helix dependency”. Because the αCTD is able to stretch approximately 40Å, i.e. more than an entire helical turn, from the αNTD, an activating element can be shifted upstream from its original location, provided that it remains on the same face of the DNA helix (Blatter *et al.*, 1994; Zhou *et al.*, 1994; Jeon *et al.*, 1997). This can be visualised by inserting nucleotides
corresponding to integral or non-integral helical turns between the element and the core promoter. Thus the insertion of half a helix abolishes activity whereas the insertion of the second half (i.e. a full helix) restores activity. Phase of the helix dependency is observed when the αCTD interacts with a transcription factor as well as when it interacts directly with the DNA (Ross et al., 1993).

At class II CAP-dependent promoters the situation is more complex. Here the CAP binding site overlaps the −35 region of the promoter and the protein is wedged in between the αCTD and the rest of the RNAP, making contacts with αCTD, αNTD, and region 4.2 of σ⁷₀. The interaction with αCTD has no activating effect (rather a neutralising effect) whereas the activation happens via the αNTD. In this situation the rate of open complex formation (kᵣ, section 1.4.4.1) is increased. The interaction with σ is likely to act as a compensation for the fact that the −35 region of this type of promoters is not optimal due to the overlapping CAP binding site (Bown et al., 2000; Rhodius and Busby, 2000).

1.4.2 Bacterial promoters

The DNA determinants of transcription initiation, the promoters, usually consist of two promoter elements, which in the case of the primary σ factors are centred around −10 and −35 relative to the transcription start point. The two elements are recognised by one of the many σ factors within the cell. In the case of σ⁷₀ the
regions responsible for the –10 and –35 contacts have been identified as regions 2.4 and 4.2, respectively (Gross et al., 1992). Both hexamers contribute to the binding constant, $K_B$, between DNA and RNAP, i.e. the closer to the consensus sequences, the higher the binding constant (Ellinger et al., 1994a; Record et al., 1996). Both elements are, however, also involved in additional steps in the initiation pathway (Record et al., 1996). The region between the two elements, the spacer, does not conform to a consensus as much as to a specific length which in *E. coli* is 17 basepairs (reviewed by Record et al., 1996). Changes in the length of the spacer influences promoter strength as well as regulation (Ellinger et al., 1994a; Josaitis et al., 1995; McKane and Gussin, 2000; Voulgaris et al., 2000). There is, however, in some promoters an additional motif, the extended –10 region. This motif is a TG di-nucleotide located upstream of the –10 region which is recognised by a distinct region (2.5) of $\sigma^{70}$ (Barne et al., 1997). Until recently it was believed that this motif was located exclusively at position –14/–15 relative to the transcription start point (Bown et al., 1997). However, a detailed study of *E. coli* promoters revealed that the TG di-nucleotide is slightly over-represented at positions –16 and –17 as well (Burr et al., 2000). The TG motif enhances transcription several fold, the –14/–15 element having the largest effect (Burr et al., 2000). A number of promoters containing the TG motif do not conform to a specific –35 region although they are still recognised by $\sigma^{70}$ (Bown et al., 1997).
The TG motif at position -14/-15 has been demonstrated to promote the isomerisation from closed to open complex (see 1.4.4.1) but whether the TG at positions -16/-17 acts in the same step of initiation remains to be determined (Burr et al., 2000).

In some strong promoters, such as the E. coli rrn promoters, an additional promoter element, the UP element, is located upstream of the -35 region (Ross et al., 1993), see below). Both the extended -10 region and the UP element were believed to be present in much higher frequencies in Gram-positive than Gram-negative bacteria (Graves and Rabinowitz, 1986; Helmann, 1995). However, at least in the case of the extended -10 region this has now been revised (Burr et al., 2000).

The region downstream of the transcription start point is involved in promoter clearance and hence of importance for initiation by the following RNAP but very little is known about its structural requirements (Kammerer et al., 1986; Record et al., 1996).

1.4.3 The promoters of Mycobacteria

Mycobacteria have a genomic GC content of between 66 and 72% (Grange, 1982) which is likely to have implications for molecular mechanisms of the bacteria. A high GC content could result in different sequences in promoters and other Cis-
acting elements compared to other, more intensively studied organisms such as *E. coli* or *B. subtilis*. The high GC content is reflected in the codon bias. The initiation codon is often found to be GTG instead of ATG (sometimes TTG) and internal codons very often have G or C at the wobble position (Andersson and Sharp, 1996; Mulder *et al*., 1997). This makes it difficult to compare promoter activities by means of reporter genes expressed in mycobacteria and in e.g. *E. coli* since codon usage and hence translational efficiency is bound to be different in the two species.

Mycobacterial promoter sequence and structure is still relatively unknown territory. A promoter consensus for *M.tb* and for *M.sm* −10 regions has been suggested by Bashyam and co-workers but the number of promoters studied (14 *M.sm* and 10 *M.tb* promoters) was low (Bashyam *et al*., 1996). It was not possible to deduce a consensus for the −35 region for either species because of the differences between the promoters. Furthermore it was shown that one promoter could in fact function without its −35 region, although this promoter did not posses an extended −10 region (Bashyam *et al*., 1996). However, mycobacterial promoters do contain this motif and it enhances promoter activity between 4 and 54 fold depending on the context (Bashyam and Tyagi, 1998). Thus mycobacterial promoters appear to have less stringent requirements for their −35 regions allowing larger sequence diversity in this region, a phenomenon also seen in
Streptomyces spp (Strohl, 1992). Mulder et al. have compiled a number of mycobacterial promoters from several studies and suggested a consensus common to mycobacteria in general (Mulder et al., 1997). The two consensus sequences, (TTGACG/A and TATA/GA/CT for -35 and -10 regions, respectively) are in perfect agreement with the E. coli consensi. This is somewhat surprising given the different nucleotide composition in the two species and the fact that promoters from either species are recognised with reduced efficiency in the other (see below). Furthermore the purified RNAP from E. coli and M.sm display different promoter preferences on the same template in vitro (Wiggs et al., 1979). Finally the compiled promoters are not recognised by the same σ factor, which makes it questionable to deduce a consensus. Along the same lines a promoter consensus common to all mycobacteria is perhaps not a valid measure. The most recent report based on a higher number of promoters is by Gomez and Smith and they suggest consensus sequences for -10 and -35 regions, which are very close to the ones suggested by Mulder et al. but with a less stringent sequence requirement (Gomez and Smith, 2000). A number of mycobacterial promoters resemble the well-known structure of E. coli promoters, and are in fact recognised by the E. coli polymerase, although with reduced efficiency (Kieser et al., 1986; Ward et al., 1986; Suzuki et al., 1991; Das Gupta et al., 1993). However, there seems to be a tendency towards M.sm promoters being more active than M.tb promoters when expressed in E. coli (Das Gupta et al., 1993). This difference could probably be
accounted for by the fact that \textit{M.sm} promoters are, on average, stronger than \textit{M.tb} promoters when expressed in either \textit{M.sm} or \textit{M.tb} (Bashyam \textit{et al.}, 1996). It is worth noting that the strength of a given promoter in the \textit{M.sm} background does not always correlate with the strength of the same promoter when recognised in \textit{E. coli}, thus emphasising the importance of a homologous expression system (Das Gupta \textit{et al.}, 1993). The reverse situation, \textit{E. coli} promoter recognition by \textit{M.sm}, also seems to happen with reduced efficiency (Bashyam \textit{et al.}, 1996), although there is the possibility that the differences in CAT reporter expression could, at least to some extent, reflect differences in post-transcriptional events. An older study in which genes from \textit{M. bovis} BCG were expressed in \textit{S. lividans} and \textit{E. coli} substantiates this finding and demonstrates that \textit{S. lividans} is a better surrogate host than \textit{E. coli} (Kieser \textit{et al.}, 1986). However, with the emergence of mycobacterial cloning systems and an easily transformed strain of \textit{M.sm} (Snapper \textit{et al.}, 1990), the most obvious surrogate host for mycobacterial genes is \textit{M.sm}. Here it is worth noting that Bashyam and co-workers have demonstrated that the majority of investigated \textit{M.sm} promoters are recognised with equal efficiency in \textit{M.sm} and \textit{M.tb} and vice versa (Bashyam \textit{et al.}, 1996). Another mycobacterial species that has been suggested as surrogate host is \textit{M. aurum} but so far this species has not been used as extensively as \textit{M.sm} (Srivastava \textit{et al.}, 1997).
Other regulatory elements such as repressors, activators, and terminators are poorly defined in mycobacteria. In fact transcriptional terminators specific for mycobacteria have not yet been identified, although some terminators from other organisms function as such in mycobacteria (Timm et al., 1994).

1.4.4 Mechanisms of transcription

Transcription can be divided into initiation, elongation, and termination; events that are separate but tightly linked. The transition from one to the other depends on a variety of factors and I have outlined the major determinants below.

1.4.4.1 Initiation

The initiation event itself can be divided into a multistep process described as

\[ K_B \quad k_f \]

\[ R + P \rightarrow RP_C \rightarrow RP_O \rightarrow RP_{NTP} \rightarrow R_E \]

where R designates RNAP holoenzyme, P is the promoter, RP is the complex between the promoter and the enzyme in closed or open conformation, or with bound initiating nucleotide (subscripts C, O, and NTP, respectively); \( R_E \) is the elongation complex, \( K_B \) is the binding constant of RNAP to promoter, and \( k_f \) is the rate constant for open complex formation (figure adapted and simplified from Record et al., 1996). The strength of a promoter is a complex balance of all the
parameters. The binding constant ($K_B$) is important for the initial binding but in the case of too high affinity between the promoter and the RNAP this has a suppressing effect on promoter activity due to difficulties in promoter escape (Ellinger et al., 1994b; Ellinger et al., 1994a; Strainic et al., 1998). Conversely it has been shown that promoters with low binding constants can be highly efficient in vivo (Kammerer et al., 1986).

The transition from closed to open complex involves a melting of template DNA around the transcription start site and the nucleotide composition in this region is therefore important for this transition (see 1.6.4.3). Interactions upstream of the core promoter, i.e. upstream of the –35 region, are all made by the α-subunit, contacts in the region between –40 and –7 are primarily made by the σ subunit, whereas the two catalytic subunits, β and β’, mainly interact with sequences downstream of –10 (Naryshkin et al., 2000). In the presence of the appropriate, initiating nucleotide, the open complex forms a transcriptionally competent complex (RP$_{NTP}$) which is capable of initiating and terminating several rounds of transcription without translocating along the DNA template (Record et al., 1996). This so-called abortive initiation produces transcripts of 7 to 12 nucleotides in size which are released from the initiation complex and only a fraction of the initiation complexes undergo the transition to elongation complexes (Record et al., 1996).
1.4.4.2 Elongation

As the RNAP starts to move along the template the growing RNA chain displaces σ from the enzyme (Daube and von Hippel, 1999). As soon as σ has dissociated, the NusA protein associates with the core enzyme. The association of σ and NusA are mutually exclusive and σ can in fact displace NusA from the RNAP (in the absence of RNA) (Greenblatt and Li, 1981).

The elongation of RNA does not happen in a continuous movement. Within a gene there are several pause and termination signals which can be read or bypassed depending on a number of factors. The strategy of using immobilised RNAP has made it possible to transcribe DNA templates in steps of single nucleotides (Kashlev et al., 1993). Obviously this system is highly artificial especially since it slows down the rate of transcriptional elongation dramatically (probably several thousand-fold) and this alone may well introduce artefacts. Nevertheless this approach has shed some light on the mechanisms of elongation as well as on pausing and termination, especially in combination with the increasing amount of information on RNAP structure (Nudler, 1999; Korzheva et al., 2000).

The first set of experiments revealed an irregular relationship between the length of a transcript and the position of the front end of the RNAP. This was initially explained by the so-called inchworming model in which the front of the polymerase was believed to stall while still synthesising RNA, i.e. the catalytic
centre of the enzyme moved forward. The accumulation of transcript within the enzyme would generate a strain and the RNAP would eventually “leap” forward, shifting several nucleotides forward in one move, thus restoring the “proper” distance between the front and the centre of the enzyme. This particular movement only happened at specific DNA sequences (inchworming signals) including those found at intrinsic terminators (Nudler et al., 1994; Nudler et al., 1995). This model was, however, replaced by the current model which involves a continuous forward motion interrupted by occasional backtracking of the enzyme, i.e. no jumping (Komissarova and Kashlev, 1997b; Komissarova and Kashlev, 1997a; Nudler et al., 1997). The 3’ end of the transcript extrudes from the RNAP and depending on the nucleotide sequence the enzyme can get trapped in an arrested, inactive state or it can simply oscillate back and forth on the template maintaining an active state (Komissarova and Kashlev, 1997b; Komissarova and Kashlev, 1997a; Nudler et al., 1997). The inactive state can be reactivated by cleavage of the transcript induced by GreA or GreB (or TFIIS in eukaryotes) (Borukhov et al., 1993; Komissarova and Kashlev, 1997b; Nudler et al., 1997). The backtracking of RNAP is promoted by weak interactions in the RNA:DNA hybrid within the enzyme as shown by nucleotide substitutions that either weaken or strengthen the hybrid. This implies that nucleotide mismatches promote backtracking, and this could therefore provide a mechanism for transcriptional fidelity (Komissarova and Kashlev, 1997b; Nudler et al., 1997). This model of
transcriptional elongation also explains how the RNAP seemed capable of identifying sequences believed to be ahead of the RNAP according to the inchworming model (Nudler et al., 1994; Nudler et al., 1995). The backtracking model and the inchworming model are not mutually exclusive, since only the latter can explain how an elongation complex can survive an encounter with a replication complex according to (Landick, 1997). The RNAP backtracking has important implications for transcriptional pausing. The simplest pause signal is a stretch of U-residues which destabilise the DNA:RNA hybrid and thus leads to RNAP backtracking, a so-called class II pause (Nudler et al., 1997; Artsimovitch and Landick, 2000). This type of pause signal is highly conserved across species and it has even been demonstrated that a human pause signal can in fact halt both *E. coli* and SP6 RNAP’s probably by inducing backtracking, as shown by the reactivation of the elongation complex by transcript cleavage (Mote and Reines, 1998). Gusarov and Nudler suggest that the $K_m$ for UTP nucleotides could play a more direct role in U-stretch induced pausing as well (Gusarov and Nudler, 1999).

There is another type of pausing, a class I pause, which is both operon and species specific (Chan et al., 1997; Artsimovitch et al., 2000; Artsimovitch and Landick, 2000). This type of signal consists of a hairpin structure followed by a U-rich tail and thus resembles an intrinsic terminator, although there are slight differences between the two (see 1.4.4.3). This type of pausing involves a direct interaction
between the hairpin structure of the transcript and a specific region of the RNAP (Artsimovitch and Landick, 1998; Toulokhonov et al., 2001b). The interaction induces a conformational change in the RNAP, which is believed to prevent further addition of nucleotides to the transcript (Toulokhonov et al., 2001b). The elongation/termination factor, NusA, increases hairpin dependent pausing as well as intrinsic termination (Sigmund and Morgan, 1988b; Artsimovitch and Landick, 2000; Toulokhonov et al., 2001b). At least in the case of pausing this has been suggested to occur via a stabilisation of the interaction between the RNAP flap-domain and the hairpin (Toulokhonov et al., 2001b). Interestingly both elongation factors NusA and NusG have different effects on the two types of pauses: NusA enhances class I pauses but has no effect on class II pauses, whereas NusG has no effect on class I pauses but diminishes class II pausing (Artsimovitch and Landick, 2000).

The rate of transcriptional elongation or RNA chain growth rate (CGR) varies tremendously from one species to another, as well as within a species depending on the template and the growth conditions. Thus, in E. coli there is an inverse correlation between the concentration of the regulatory nucleotide ppGpp and the elongation rate of mRNA (Figure 1.2). In contrast, stable RNA (i.e. rRNA and tRNA) is elongated at a relatively constant rate but much faster than mRNA (Bremer and Dennis, 1996). M.tb cells have a low RNA:DNA ratio characteristic of slow growing bacteria (Winder, 1982). The average RNA CGR (both stable
RNA and mRNA) in *M.tb* has been estimated to be between four and ten nucleotides per second. This is approximately a tenth of the rate in *E. coli* but in order to account fully for the low RNA:DNA ratio it has been suggested that the frequency of transcription initiation is also extremely low (Harshey and Ramakrishnan, 1977). The CGR in *E. coli* is 40 to 55 nucleotides per second for mRNA (depending on growth rate) and 80 to 90 nucleotides per second for stable RNA (independent of growth rate) (Bremer and Dennis, 1996). The reasons for these differences between *E. coli* and *M.tb* have not been determined but a number of related parameters could perhaps shed more light on the question. The CGR of stable RNA in *E. coli* remains unchanged at approximately 85 nucleotides per second over a range of growth rates given that the temperature and the number of *rrn* operons remain unchanged (Bremer and Dennis, 1996). However, the *E. coli* RNAP is in fact capable of a CGR of at least 135 nucleotides per second (Condon *et al.*, 1993). The *in vitro* elongation rate of *M.sm* RNAP is 74% of the *E. coli* elongation rate under the same conditions and using the same template, indicating similar functional potentials of the two enzymes (Chamberlin *et al.*, 1979). Since the *M.sm* and *M.tb* RNAP enzymes are nearly identical in amino acid composition (Table 1.1) it is unlikely that the difference in CGR between *M.tb* and *E. coli* is due to differences in the RNAP enzyme itself. Rather it is a question of differences in the template, (e.g. GC-content, modified bases, and pause sites), or in the cytoplasmic concentration of one or more factors (ribo-nucleotides,
ppGpp or an additional, regulatory factor). There are no reports on the intracellular NTP concentrations in *M. tb* or *M. sm*. Polyamines (i.e. spermine, spermidine, and putrescine) have been suggested to play a part in the slow transcription of *M. tb* DNA (Jain and Tyagi, 1987; Sarkar *et al.*, 1995). These compounds are present in very low concentrations in some of the mycobacteria (compared to *E. coli*) and they have furthermore been shown to have a stimulatory effect on transcription *in vitro* (Sarkar *et al.*, 1995). Whether a higher polyamine concentration would in fact increase the transcriptional activity *in vivo* remains to be determined.

In contrast to the rRNA CGR, the mRNA CGR does change with the growth rate in *E. coli* (Bremer and Dennis, 1996) and the low CGR of *M. tb* could in part be explained by simple extrapolation of this relation. The CGR for *M. tb* is an average of all RNA species and not just mRNA and the relation should therefore be regarded carefully (Figure 1.2 bottom panel). However, even though there appears to be a correlation, the effector(s)/regulator(s) may not be the same. The slow RNA CGR in *M. tb* is likely to have implications for other molecular mechanisms such as peptide elongation rate and termination of transcription. There are no reports on mycobacterial peptide elongation rates to date but presumably the peptide elongation rate follows the mRNA CGR in order to prevent Rho-dependent termination (see section 1.4.4.3). The frequency of termination would probably also be affected but it is impossible to predict in what way since intrinsic
Figure 1.2: The concentration of ppGpp determines the mRNA CGR.
The top panel illustrates the inverse correlation between the ppGpp concentration and the average elongation rate of mRNA in the *E. coli* cell at various growth rates. The bottom panel shows how the chain growth rate is inversely correlated to the logarithm of the generation time. The red point corresponds to the values for *M.tuberculosis* obtained from (Harshey and Ramakrishnan, 1977). The remaining values (*E. coli*) are obtained from (Bremer and Dennis, 1996).
termination can be either increased or decreased with decreasing CGR, whereas Rho-dependent termination seems to increase with decreasing CGR (see below). The difference in CGR between mRNA and rRNA (and other stable RNA species) within a species is due to differences in the template (i.e. pause and anti-termination signals, see below).

1.4.4.3 Termination of transcription

Termination, in combination with anti-termination is another widespread means of regulating gene expression, and the mechanisms involved are numerous (reviewed by Greenblatt et al., 1993; Richardson and Greenblatt, 1996). The termination of transcription can be divided into two separate mechanistic events: pausing of the RNAP and release of the transcript from the elongation complex. The pausing of the RNAP appears to be a prerequisite for the dissociation of the elongation complex since all terminators induce RNAP pausing (Platt, 1986; Artsimovitch and Landick, 1998).

Two mechanisms of transcriptional termination have been characterised in *E. coli*: Rho-dependent and intrinsic termination (Platt, 1986; Richardson and Greenblatt, 1996). Rho-dependent termination involves the interaction between a termination factor, Rho (also known as NusD), and a so-called *rut* (Rho utilisation) site on the RNA. Rho binds to the *rut* site and translocates along the nascent chain by ATP
utilisation until it encounters and dissociates the paused elongation complex (Richardson and Greenblatt, 1996). When mRNA is in the process of being translated, the ribosomes shield the RNA from Rho action. As soon as the ribosomes encounter a stop codon and dissociate from the RNA, Rho has free access to the naked transcript and transcription is terminated. In the case of a nonsense mutation the result is premature termination of transcription and genes within the same operon but downstream of the nonsense mutation will not be expressed (Condon et al., 1995b). The latter is termed polarity, a well-known phenomenon in *E. coli*. Rho interacts with both NusA and NusG but with different affinities (Schmidt and Chamberlin, 1984b; Li et al., 1993). As mentioned earlier rRNA is elongated approximately twice as fast as mRNA. This is believed to be the main reason that the *rrn* elongation complex is inert to Rho action in spite of not being translated (Condon et al., 1995b; Vogel and Jensen, 1995). At least one Rho-dependent terminator is located within the promoter proximal fifth of the mature *E. coli* 16S RNA (Aksoy et al., 1984) but the significance of this has not been established. It is possible that not all RNAPs that traverse the anti-termination motifs are modified into anti-terminated elongation complexes and hence would be susceptible to this termination site. The dissociation of these “non-anti-terminated” elongation complexes would allow the anti-terminated complexes to proceed at the higher rate characteristic of rRNA elongation. Alternatively an RNAP transcribing the upstream gene and continuing into the *rrn*
gene may not be (as efficiently) modified, i.e. the terminator would prevent RNAP read-through. In *B. subtilis* Rho protein is present in considerably lower concentrations than in *E. coli* and consequently Rho-dependent termination as well as polarity are not as widespread (Ingham *et al.*, 1999) Furthermore the *rho* gene can be deleted in *B. subtilis* perhaps for the same reason (Quirk *et al.*, 1993). Unfortunately nothing is known about Rho-dependent termination in mycobacteria or even if Rho is essential.

Intrinsic terminators function independently of accessory factors. These terminators are made up of regions of GC-rich dyad symmetry capable of forming very stable stem-loop structures in the nascent transcript, followed by a stretch of T’s (or U’s in the transcript). The spacing between the base of the hairpin and the U-stretch as well as the number of U residues determines whether the signal is a pause or a terminator (Chan *et al.*, 1997). The mechanism of termination essentially involves a competition between the G:C interactions in the hairpin structure outside the RNAP and the much weaker A:U interactions in the RNA:DNA hybrid inside the elongation complex. The formation of the hairpin disrupts the RNA:DNA interactions in the upstream part of the hybrid provided that the U-stretch is located immediately at the base of the hairpin; hence nucleotide substitutions which stabilise the RNA:DNA hybrid in this region prevent the formation of the hairpin (Gusarov and Nudler, 1999). At the same time interactions between the RNAP and the single stranded RNA are disrupted.
due to the hairpin formation, and together this makes the elongation complex
dissociate (Gusarov and Nudler, 1999). Interestingly intrinsic terminators lacking
the U-rich tail are present in *Streptomyces spp* (Ingham *et al.*, 1995). However,
these terminators appear to have a more complex structure but are in fact
recognised by the *E. coli* RNAP as well. In the light of the close homology
between the RNAP’s from mycobacteria and *Streptomyces spp* (Table 1.1) a
similar mechanism may be found in mycobacteria.

The coupling between transcription and translation is vital in prokaryotes. Once
the nascent mRNA becomes exposed, Rho factor binds and terminates
transcription. However, since rRNA is not translated this mechanism must be
circumvented by other means, i.e. anti-termination. An anti-termination system
requires several anti-termination factors, which upon interaction with a specific
element downstream of the promoter modify the RNAP in a way that it is able to
read through Rho-dependent terminators. Proteins believed to act in the *E. coli rrn*
anti-termination system include NusA, NusB, NusE (same as ribosomal protein
S10), NusG, ribosomal protein S4 and the RNAP core enzyme (Squires *et al.*, 1993; Torres *et al.*, 2001). Anti-termination enables the cell to regulate the rRNA
synthesis in response to signals other than transcriptional activation. Since the
anti-termination complex includes ribosomal proteins the synthesis of rRNA
appears to be regulated by the availability of free ribosomal protein as well as *vice
versa*.
1.5 Translation – the ribosome

The synthesis of proteins is carried out by the ribosomes using an mRNA template. Apart from translation the ribosomes are also involved in a number of sensory/regulatory mechanisms. For instance in E. coli it has been demonstrated that permanent occupancy of the ribosomal A-site, either by antibiotics or by aminoacyl-tRNA, induces the cold shock regulon. Conversely an antibiotic induced vacancy of the A-site leads to a heat shock response (VanBogelen and Neidhardt, 1990). Thus different antibiotics, which all block protein synthesis can result in completely opposite responses depending on the state of the ribosome. In B. subtilis a number of antibiotic resistance mutations mapping to different ribosomal genes are linked to defects in various stages of sporulation, i.e. the ribosome is closely involved in determining whether the cells should continue growth or sporulate (Chambliss, 1979). This could perhaps be of relevance regarding M.tb dormancy.

1.6 The rrn operons

As the numbers in Table 1.2 demonstrate the M.tb rrn gene dosage is low with only one gene encoding ribosomal RNA compared to the closely related species S. coelicolor (and other Streptomyces spp), or E. coli, and B. subtilis, which have six, seven, and ten rrn operons, respectively (Baylis and Bibb, 1988; Gourse et
However, the phenomenon of low rDNA copy number is not only seen in mycobacteria, but is also found in other species e.g. mycoplasmas and archaea (Amikam et al., 1984; Hui and Dennis, 1985; Fegatella et al., 1998). In addition to being present in only one or two copies per genome, mycobacterial *rrn* genes are located far away from the replication origin and hence do not have the advantage of an increase in gene dosage early in the replication cycle (Cole et al., 1998). However, it remains unknown whether the mycobacterial (*M.tbc*) cell has numerous replication forks as seen in *E. coli* at maximum growth rates (Bremer and Dennis, 1996).

Table 1.2: Growth rates and *rrn* gene dosage.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Doubling time*</th>
<th>No. of <em>rrn</em> operons</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td><em>S. coelicolor</em></td>
<td>140</td>
<td>6</td>
</tr>
<tr>
<td><em>M. smegmatis</em></td>
<td>160</td>
<td>2</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>900</td>
<td>1</td>
</tr>
</tbody>
</table>

*The doubling time is approximate and given in minutes*

The table compares the approximate growth rates of various bacteria with the number of *rrn* operons in the genome. The numbers illustrate a tendency towards higher gene dosage with higher growth rates but the correlation is not strict.
The organisation of the rDNA genes is very well conserved between bacterial species. The genes are primarily found in a single operon in the order 16S-23S-5S with transcribed spacers between the individual genes (reviewed by Condon et al., 1995b). The *rrn* genes are transcribed as a single unit, which is subsequently processed into the individual components (King et al., 1986). This ensures balanced amounts of the three rRNA species, which are incorporated into the small (16S) and the large (23S and 5S) ribosomal subunits. The spacer regions vary somewhat more, and in some species these harbour tRNA genes that differ from one operon to the next within a species (e.g. *E. coli* *rrn* operons (Condon et al., 1995b)). Mycobacteria do not have tRNA genes in their spacer regions and the same is the case for *S. coelicolor* (and other *Streptomyces spp*) which are more closely related to the mycobacteria than *E. coli* (van Wezel et al., 1991).

The expression of the individual *rrn* operons is regulated at the level of transcription initiation as well as termination. Furthermore, part of the processing of the precursor rRNA depends on the incorporation of the pre-rRNA into ribosomes and this in turn depends on the availability of ribosomal proteins, i.e. protein synthesis and hence growth (King et al., 1986).
1.6.1 The *rrn* promoters

The seven *rrn* operons in *E. coli* all have the same promoter organisation with two tandemly arranged $\sigma^{70}$ promoters separated by approximately 120 basepairs (Gourse *et al.*, 1996). The upstream promoter, P1, is highly regulated and by far the most active at medium to high growth rates. At medium to high growth rates the downstream promoter, P2, is weaker than P1 and more or less constitutive unless taken out of its native context (Sarmientos and Cashel, 1983; Sarmientos *et al.*, 1983; Liang *et al.*, 1999). Isolated P2 promoters display features such as upstream activation and stringent control and have a higher promoter activity but the *in vivo* relevance (if any) is not known (Josaitis *et al.*, 1990; Gafny *et al.*, 1994). Sequences homologous to the $\sigma^{32}$ (heat-shock $\sigma$) recognition elements are found in all seven *E. coli rrn* operon P1 regions, and experiments performed with the *rrnB* P1 have demonstrated that $\sigma^{32}$ can in fact initiate transcription from this putative promoter *in vitro* (Newlands *et al.*, 1993). Whether there is promoter activity *in vivo* remains to be determined. It is, however, entirely possible that these sequences comprise *bona fide* heat shock promoters and thus are important for maintaining the expression of *rrn* genes following a temperature up-shift. This is substantiated by the finding that rRNA operons in *S. lividans* are expressed at high temperatures *in vivo* (de Leon and Mellado, 1997). The *E. coli rrn* core promoters are characterised by having a 16 basepair spacer instead of the
consensus 17 basepairs. The insertion of the "missing" basepair increases activity dramatically (between seven and ten fold) in a construct that spans -48 to +1. However, the increase in activity was less than two fold in constructs which included the full upstream activating region (UAR) (Gaal et al., 1989). This is consistent with the finding of Ellinger et al., that promoters with a high homology score (compared to consensus) have a very high $K_B$ which inhibits promoter clearance (Ellinger et al., 1994a). Changing the spacer to 17 basepairs changes the regulation of the promoter on a number of levels (Josaitis et al., 1995; Voulgaris et al., 2000), and it has been suggested that the *E. coli* *rrn* promoters have evolved towards a higher degree of regulation rather than just maximum activity (Gourse et al., 1996).

1.6.2 The *rrn* leader region

The *rrn* leader region, that is, the region between the transcription start point and the 5' end of the mature 16S RNA, contains a number of elements involved in transcriptional and post-transcriptional events. In *E. coli* these include pausing and termination sites (Kingston and Chamberlin, 1981), anti-termination signals (Berg et al., 1989; Albrechtsen et al., 1990), and RNase cleavage sites (Young and Steitz, 1978; King et al., 1986). In a number of organisms the *rrn* leader basepairs with the 3' end of the 16S gene thereby forming a stem with an RNase III
cleavage site. The 16S with some additional flanking sequences comprise a "loop" in this structure (King et al., 1986). There are, however, indications that different structures form transiently which act as a scaffold for the proper folding of the 16S rRNA (Pardon and Wagner, 1995; Besancon and Wagner, 1999). The E. coli rrn leader also contains a single termination site, tL, located between the anti-termination sequences and the mature 16S RNA, and its function as a terminator depends on NusA but not on Rho. The termination effect of NusA is furthermore enhanced by ppGpp (Kingston and Chamberlin, 1981). Deletion of this terminator/pause site from the leader region results in transcriptional polarity of the rrn operon as demonstrated by an increased ratio between 16S and 23S RNA (Zacharias and Wagner, 1987). However, point mutations in this region leads to instability of promoter proximal transcripts, as demonstrated by a decreased ratio of 16S to 23S (Theissen et al., 1990). The actual in vivo function of this terminator remains unknown but it could act as a discriminator between elongation complexes, which have and have not been anti-terminated as suggested by (Zacharias and Wagner, 1987). Alternatively it could act as a regulator of rrn expression as suggested by Kingston and Chamberlin (i.e. turnstile attenuation, (Kingston and Chamberlin, 1981). Finally it could act as a simple pause signal in order for the RNAP to stall, thereby allowing the leader RNA to fold into the proper conformation. Mutations throughout the rrn leader, including the tL region
and the *nut* site, as well as mutations in various *nus* genes furthermore affect the maturation of the ribosomes, indicating important post-transcriptional regulatory functions in this region (Theissen *et al.*, 1993).

### 1.6.3 The *rrn* promoters and leader regions of Mycobacteria

Transcription of the mycobacterial *rrn* genes stems from a varying number of promoters depending on the operon and the species. All of these promoters have been identified by primer extension and subsequent sequence comparison between a number of mycobacterial species (Gonzalez-y-Merchand *et al.*, 1996a; Gonzalez-y-Merchand *et al.*, 1997). Unlike the situation in *E. coli* where all seven *rrn* operons have the same general structure, the two mycobacterial operons have very different structures (Figure 1.3). The *rrnA* operon, defined as the operon downstream of the *murA* gene, has been found in all species investigated to date. Furthermore, the two promoters present in the *rrnA* operon of *M.tb*, P1 and PCL1, are found in all *rrnA* operons (Gonzalez-y-Merchand *et al.*, 1997). Interestingly the relation between these promoters is somewhat inverse compared to the situation in *E. coli*, with the downstream promoter accounting for most of the activity (~75%) at the majority of growth stages (Gonzalez-y-Merchand *et al.*, 1998). There are several indications that the P1 promoter is recognised by a σ factor different from the one, which recognises PCL1.
Figure 1.3: General structure of mycobacterial *rrn* operons.

The figure illustrates the promoter and leader regions of the two *rrn* operons, *rrnA* and *rrnB*, of the genus *Mycobacterium*. All species investigated to date have the *rrnA* operon (top), which always contains the two promoters, P1 and PCL1. The *rrnA* operon of faster growing species have promoters in addition to P1 and PCL1. In fast growing species with one *rrn* operon, the extra promoters are inserted downstream of the PCL1 promoter (transcription start points indicated by green arrows). Some faster growing species have a second operon, *rrnB* (bottom), which contains only one promoter. In the faster growing species with two operons, the extra promoter of the *rrnA* operon is inserted between P1 and PCL1 (indicated by a blue arrow). The yellow boxes indicate anti-termination motifs, BoxB, BoxA and BoxC. Note that the *rrnB* operon does not contain a BoxB motif. Figure adapted from (Gonzalez-y-Merchand *et al.*, 1997).
Expression of the *M.bovis* *rrn* operon in *E. coli* reveals that only PCL1 is recognised as a promoter in this organism (Suzuki *et al*., 1991) (this work). The sequence of the P1–35 region does not resemble that of a conventional ς₇₀ promoter in contrast to the remaining mycobacterial *rrn* promoters (see Table 5.2). In concordance with this fact Chen and co-workers have demonstrated that an alternative ς factor, SigF recognises the P1 promoter *in vitro* (Chen and Bishai, 1998). Finally it is highly likely that at least one of the promoters in an operon such as the *rrn* operons is recognised by the primary ς factor of that particular organism. In other mycobacterial species the *rrnA* operon can have insertions of up to three additional promoters between P1 and PCL1 or downstream of PCL1 (Table 1.3 and Figure 1.3). Furthermore the spacing between these promoters varies from 67 (between P2 and PCL1 in *M. neoaurum*) to as much as 255 (between PCL1 and P2 in *M. abscessus*) (Gonzalez-y-Merchand *et al*., 1997).

The *rrnB* operon, defined as the operon downstream of the *tyrS* gene, is only present in some species, and so far in all cases only one promoter has been identified in this operon (Gonzalez-y-Merchand *et al*., 1997). A small fraction of the rRNA molecules have been initiated further upstream in both operons. Whether this is due to read-through from the upstream genes or additional promoters located within these genes has not been determined (Gonzalez-y-Merchand *et al*., 1999). The diversity in promoter content and spacing implies that
the leader regions of the different \textit{rrn} transcripts vary dramatically in size. They do however, all contain sequences capable of extensive secondary structures which could act as pausing or termination signals \textit{in vivo} (Ji et al., 1994a; Gonzalez-y-Merchand et al., 1997).

### Table 1.3: The \textit{rrn} operons of different mycobacterial species.

<table>
<thead>
<tr>
<th>Species</th>
<th>No. of promoters in \textit{rrnA}</th>
<th>\textit{rrnB} operon (Y/N)</th>
<th>Growth rate*</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{M. tuberculosis}</td>
<td>2</td>
<td>N</td>
<td>&gt;14</td>
</tr>
<tr>
<td>\textit{M. smegmatis}</td>
<td>3</td>
<td>Y</td>
<td>2-4</td>
</tr>
<tr>
<td>\textit{M. chelonae}</td>
<td>5</td>
<td>N</td>
<td>3-4</td>
</tr>
<tr>
<td>\textit{M. abscessus}</td>
<td>5</td>
<td>N</td>
<td>3-4</td>
</tr>
<tr>
<td>\textit{M. phlei}</td>
<td>3</td>
<td>Y</td>
<td>2-5</td>
</tr>
<tr>
<td>\textit{M. neoaurum}</td>
<td>3</td>
<td>Y</td>
<td>&lt;4</td>
</tr>
<tr>
<td>\textit{M. fortuitum}</td>
<td>4</td>
<td>Y</td>
<td>2-4</td>
</tr>
</tbody>
</table>

* Growth rate indicates the time required to obtain visible growth on solid medium

The table indicates the number of promoters in the \textit{rrnA} operon and whether the species contains the \textit{rrnB} operon. The last column compares the growth rates of the different species. Adapted from (Gonzalez-y-Merchand et al., 1997).

The individual mycobacterial \textit{rrn} promoters have different numbers of nucleotides separating the two hexamers: 16 basepairs in \textit{rrnA} P1, 17 basepairs in \textit{rrnB} P1, and 18 basepairs in \textit{rrnA} PCL1. The number of basepairs in the P1 spacer is uncertain, though, since the –35 region has a different appearance than other $\sigma^{70}$ promoter elements and therefore could have different boundaries than the ones
that have been proposed. In general it seems that either the spacing between the promoter elements is not as critical as in *E. coli* or the boundaries of the elements have not been precisely assigned (Mulder *et al.*, 1997; Gomez and Smith, 2000).

**1.6.4 Regulation of ribosome synthesis (lessons from *E. coli*)**

The synthesis of ribosomes in bacteria is regulated on a variety of levels. Overall it is the amount of rRNA that in turn regulates the synthesis of the remaining components of the ribosome, the ribosomal proteins (Course *et al.*, 1996). The maximum synthesis of rRNA depends on gene dosage as well as the promoter content and strength in the individual operons. The number of *rrn* operons varies from one to several depending on the bacterial species, and there is a tendency towards a correlation between an organism's maximum growth rate and the number of *rrn* operons in the genome (Table 1.2). However, the correlation is not strict and furthermore, due to the highly tuned regulatory mechanisms of the cells, the number can in fact be changed either way without dramatically affecting growth (section 1.6.4.2).

**1.6.4.1 Balancing rRNA and ribosomal protein**

In *E. coli*, several mechanisms linking the synthesis of ribosomes to growth rate or growth conditions have been identified. These include growth rate dependent
regulation, feedback inhibition and the stringent response (reviewed by Condon et al., 1995b; Cashel et al., 1996; Gourse et al., 1996; Keener and Nomura, 1996). Since the synthesis of rRNA and ribosomal protein (r-protein) accounts for a big fraction of the total macromolecular synthesis it is vital that this synthesis is strictly regulated in order for the cell not to waste its resources. The synthesis of rRNA and r-protein is furthermore coordinated in a way that the individual components of the ribosomes are balanced. This is managed by a network of regulatory mechanisms. The synthesis of a number of r-proteins is inhibited by feedback inhibition of translation by a protein encoded within the operon. Additional levels of regulation exist for some of the r-protein operons as well (Keener and Nomura, 1996). The S10 protein participates in the anti-termination of rrn transcription either as free protein or as ribosome constituent (Das et al., 1985). Recent findings suggest that additional r-proteins participate in anti-termination as well, although the mechanism is probably not specific for rrn anti-termination (Torres et al., 2001).

1.6.4.2 Feedback regulation

A change in the number of rrn operons has little or no effect on the amount of rRNA synthesised within certain limits of both copy number and growth rate (Jinks-Robertson et al., 1983; Condon et al., 1993; Stevenson and Schmidt, 1998).
Premature termination of *rrn* transcription (as seen in the *nusB5* mutant) does not reduce the amount of ribosomes synthesised (Sharrock *et al.*, 1985a). Deletion of *fis*, encoding Fis protein, which activates *rrn* promoter activity during some growth stages has little or no effect on growth (Ross *et al.*, 1993). At the same time the ribosome content of the cells remains closely adjusted to the growth rate (see section 1.6.4.5). This tight regulation of rRNA synthesis can in part be attributed to the mechanism of feedback regulation of rRNA synthesis. A number of experiments have demonstrated that *E. coli* during normal balanced growth does not utilise its full potential for rRNA synthesis (Condon *et al.*, 1993; Condon *et al.*, 1995a). With as few as three functional *rrn* operons the cells have only 20% less ribosome content and are still able to maintain a generation time of 24 minutes (Condon *et al.*, 1995a). Squires and co-workers have shown that *rrn* inactivation (or deletion) results in an increased expression from the remaining, intact operons (Condon *et al.*, 1993). Interestingly this is brought about by an increase in both the frequency of initiation as well as transcription elongation. The increase in initiation frequency indicates that the *rrn* promoters are sub-saturated with RNAP under normal conditions, i.e. the concentration of free RNAP molecules is limiting. The increase in elongation rate, being as much as 135 nucleotides per second or 50% more than usual, indicates that RNAP works at sub-maximal efficiency under normal conditions. This adjustment of *rrn* operon activity is independent ppGpp levels (Condon *et al.*, 1993). How the feedback
regulation works on the molecular level has still not been resolved but it appears to be an excess translational capacity of the cell that acts on the ribosomal promoters (Cole et al., 1987). In an attempt to identify the DNA determinants for feedback regulation, Squires and co-workers found that the feedback activation had a different determinant than the feedback inhibition. In other words the feedback regulation of rrn operons may in fact consist of two distinct mechanisms (Voulgaris et al., 2000).

1.6.4.3 The Stringent Response

When wild type E. coli cells are subject to amino acid starvation, a series of events can be observed. The scarcity of amino acid(s) leads to the occupancy of uncharged, cognate tRNAs in the ribosomal A-site. This in turn triggers the synthesis of the hyperphosphorylated nucleotides ppGpp and pppGpp commonly referred to as ppGpp. The nucleotide is synthesised by the relA gene product from ATP and GTP, and a concomitant drop in the concentration of these nucleotides can be observed although not all can be accounted for by the conversion (Cashel et al., 1996). The increase in ppGpp concentration is rapid and massive and the nucleotide exerts its effect on a number of cellular mechanisms. ppGpp binds to the RNAP via the β and β’ subunits (Owens et al., 1987; Reddy et al., 1995; Chatterji et al., 1998; Touloukhonov et al., 2001a) and induces a conformational
change in the enzyme (Woody et al., 1987; Toulokhonov et al., 2001a). The now modified RNAP has at least three major effects. One is a decrease in the average mRNA CGR due to sequence-specification pausing of the RNAP, i.e. the elongation of synthetic polydeoxyribonucleoside templates is not affected by the presence of ppGpp (Kingston et al., 1981; Woody et al., 1987). The second effect is an immediate cessation of stable RNA synthesis, and the third is the activation of specific amino acid biosynthetic operons (Cashel et al., 1996). The mRNA CGR appears to be regulated by ppGpp during all stages of growth since there is an inverse correlation between ppGpp concentration and CGR in vivo (Vogel et al., 1992; Vogel and Jensen, 1995; Sørensen et al., 1994; Bremer and Dennis, 1996) (Figure 1.2). Furthermore the mRNA CGR is slightly elevated in a relA spoT double mutant (i.e. a ppGpp^0 mutant) compared to wild type (Vogel and Jensen, 1995). Not much is known about the mechanism but it appears that this effect of ppGpp is mediated by, and thus dependent on the NusA protein (Kingston and Chamberlin, 1981; Vogel and Jensen, 1997). An obvious question is why only mRNA elongation is affected, and the answer probably lies in two characteristic features of rRNA (and other stable RNA species). One is the presence of the BoxA sequence in the leader and the spacer of rRNA genes (Berg et al., 1989). By traversing this sequence the elongating RNAP becomes inert to the effect or presence of ppGpp. The other is that rDNA may be devoid of ppGpp dependent
pause sites, since rRNA sequences are transcribed faster than mRNA sequences even in the absence of anti-termination signals (Vogel and Jensen, 1995). The block in stable RNA synthesis happens at the level of transcription initiation. It has been known for a long time that the ppGpp-bound RNAP displays weaker interactions with specific (stringent) promoters (Kingston et al., 1981; Woody et al., 1987) and specifically destabilises the promoter open complex at these promoters (Hamming et al., 1980). Recently it was demonstrated that ppGpp in fact decreases the half-life of all promoter open complexes and that the difference lies in the fact that the stringently controlled promoters form unusually unstable open complexes at all growth stages and conditions. The additional destabilisation brought about by ppGpp reduces the half-life of the complex beyond that which is required to proceed into the elongation step (Barker et al., 2001b). The DNA determinant, the discriminator, is a GC-rich sequence between the −10 box and the transcription start point, and it is believed that the strong GC interactions are at least in part responsible for the mechanism (Josaitis et al., 1995; Barker et al., 2001b). An additional effect of the stringent response may be a reduction in the purine nucleotide pool which could further destabilise the promoter open complexes at the rrn promoters (Gaal et al., 1997; Barker et al., 2001b). During the stringent response the synthesis of ATP and GTP is inhibited and the pools are drained due to the synthesis of ppGpp (Lazzarini et al., 1971). However, the in vivo relevance and effect of the NTP concentrations are controversial (Liang et
al., 1999). According to Gourse and co-workers (Barker et al., 2001a) the reduced initiation from stable RNA promoters could also explain the increased activity of the biosynthetic operons during the stringent response. The promoters of these operons depend on a relatively high concentration of free RNAP for initiation. By blocking stable RNA synthesis, which accounts for the majority of transcription in the cell, a substantial number of RNAP molecules would become available for other promoters/operons such as those involved in amino acid biosynthesis (Barker et al., 2001a). This model explains why the relA spoT double mutant is polyauxotrophic (Xiao et al., 1991), since the level of rRNA synthesis is considerably higher at all times thereby limiting the available RNAP (Gaal and Gourse, 1990; Barker et al., 2001b). It could also account, at least in part, for the observation made by Stevenson et al., that overexpression of rrn operons from a multi-copy plasmid reduces the growth rate and that the effect is smaller at higher growth rates, i.e. when the growth medium contains amino acid supplements (Stevenson and Schmidt, 1998).

The advantages of the stringent response are obvious. The specific inhibition of mRNA CGR would maintain a coupling between transcription and translation in a situation where the peptide elongation rate is reduced due to amino acid scarcity. This would prevent Rho-dependent termination due to naked mRNA, as well as mistranslation in a situation where noncognate tRNAs and mRNA are present in abundance (Sørensen et al., 1994). The cessation of rRNA synthesis would also
redirect the cells resources towards other targets where they were more needed, i.e. the biosynthesis of amino acids. Whether the instability of the open complexes and the pausing during elongation are in fact carried out by the same mechanism remains unanswered.

1.6.4.4 Stringent Response in Mycobacteria

*M. tb* has a single RelA/SpoT homologue, Rel* _{Mtb}*, responsible for both the synthesis and hydrolysis of ppGpp. The existence of a single RelA/SpoT protein with dual functions, i.e. analogous to the *E. coli* SpoT, is a common feature of Gram-positive bacteria (Sun *et al.*, 2001). The Rel* _{Mtb}* bears more resemblance to *E. coli* SpoT than RelA and furthermore complements the *E. coli* relA spoT double mutant (Avarbock *et al.*, 1999). The trigger for the synthesis of ppGpp in *M. tb* is an increase in the ratio of uncharged to charged tRNA as well as complete starvation (Avarbock *et al.*, 2000; Primm *et al.*, 2000). The Rel* _{Mtb}* has been shown to be of importance for the growth and long-term survival of the bacterium *in vitro* whereas the survival during intracellular growth appears to be unaffected by the deletion of the gene (Primm *et al.*, 2000). A stringent response as such has not been demonstrated and the only mycobacterial *rrn* promoter bearing resemblance to the discriminator of *E. coli* is the PCL1 promoter of the *rrnA* operon (Gonzalez-y-Merchand *et al.*, 1997).
1.6.4.5 Growth Rate Dependent Regulation

Growth rate dependent regulation refers to a strict correlation between the growth rate of the cells ($\mu$), the number of ribosomes per total cell protein (proportional to $\mu$), and the rate of ribosome synthesis (proportional to $\mu^2$) (Maaløe and Kjeldgaard, 1966; Gausing, 1979). The molecular mechanisms behind this phenomenon remain elusive although several suggestions have been proposed. It has been argued that growth rate dependent regulation and the stringent response are one and the same (Ryals et al., 1982; Baracchini and Bremer, 1988; Zacharias et al., 1991). This cannot be the case since a) the DNA determinants for the two mechanisms can be separated although they are physically closely linked or even overlapping (Bartlett and Gourse, 1994; Josaitis et al., 1995), and b) relA spoT double mutants, which do not synthesise (p)ppGpp still display growth rate dependent regulation of ribosome synthesis (Gaal and Gourse, 1990; Bartlett and Gourse, 1994). A study by Gourse and co-workers describes how the activity of the $rrnB$ and $rrnD$ P1 promoters (and probably the remaining $rrn$ P1 promoters) is highly sensitive to the concentration of the initiating nucleotide (but not to any of the remaining three NTPs) (Gaal et al., 1997). The open complexes formed at the $rrnB$ and $rrnD$ P1 promoters are unusually unstable but stability increases with increasing concentrations of ATP or GTP, respectively. The increase in stability correlates with the observed increase in promoter activity supporting the notion
that there is a direct effect between open complex stability and promoter activity. The authors argue that the intracellular pools of purines (ATP and GTP) increase with the growth rate and that this provides the long sought for direct link between the growth rate and rRNA synthesis (Gaal et al., 1997). However attractive this model may seem it is being questioned for two main reasons. A recent study by Petersen and Møller demonstrates that the nucleotide pools do not change with the growth rate in E. coli strains that display growth rate dependent regulation of rRNA synthesis (Petersen and Møller, 2000). They argue that the discrepancy is due to partial pyrimidine starvation which results in an increase in the purine nucleotide pools (Vogel et al., 1991; Petersen and Møller, 2000). The other is that the NTP-sensing model implies that the concentration of nucleotides and not RNAP molecules is limiting which, according to Bremer and co-workers is not the case (Liang et al., 1999).

1.6.4.6 Growth Rate Dependent Regulation in Mycobacteria?

The relation between growth rate and ribosome content and synthesis in bacteria applies to growth rates as low as 0.2 generations per hour (Maaløe and Kjeldgaard, 1966). Thorough investigations at lower growth rates have not been performed. M. tb has a maximum growth rate of approximately 0.07 generations per hour at optimal in vitro conditions. The question is therefore obvious: is the
synthesis of rRNA growth rate regulated in the slow growing mycobacteria? It appears that the relation between growth rate and ribosome content applies to the fast growing mycobacteria but not the slower growing species (Winder, 1982). It has been demonstrated that the expression of the entire \textit{M.tb rrnA} operon increases with increasing growth rates (Verma \textit{et al.}, 1999). However, a strict correlation such as the \textit{E. coli} growth rate dependent regulation has not been demonstrated. Also the relative expression from the individual promoters changes with growth rate and during the different stages of growth (Gonzalez-y-Merchand \textit{et al.}, 1998). According to Cox and co-workers, the synthesis of rRNA increases well into stationary phase. However, this conclusion is based on an observed increase in the amount of unprocessed rRNA and the assumption that rRNA synthesis and processing are balanced at all stages of growth (Gonzalez-y-Merchand \textit{et al.}, 1998). Since rRNA processing depends on the incorporation into ribosomes and on protein synthesis, as seen by a sensitivity to chloramphenicol treatment (Schlessinger, 1979), this increase in the level of pre-rRNA in mid to late stationary phase is more likely to be due to reduced processing than to increased transcription. It is not known whether the observed changes in \textit{rrn} activity rely on the same mechanisms as those in \textit{E. coli}.
1.6.4.7 Activated *rrn* transcription

The *E. coli* *rrn* promoters are very strong promoters. The initiation frequency at these promoters is around once per second and the transcription of stable RNA accounts for approximately 80% of total RNA synthesis during rapid exponential growth (Bremer and Dennis, 1996). The strength is at least in part achieved by the presence of activating elements upstream of the *rrn* promoters. The upstream activating region (UAR) consists of two distinct elements. Both activate via the αCTD but one is factor dependent and the other is factor independent.

Factor dependent activation involves binding of the Fis protein to the region upstream of the P1 promoter. There are three binding sites for this protein of which the promoter proximal site seems to have the largest effect on *rrn* promoter activity *in vivo* (Ross *et al.*, 1990). The synthesis of Fis is induced dramatically upon a nutritional up-shift (either from poor to rich media or from stationary to exponential phase) and the protein can thus be regarded as a cell cycle regulator (Thompson *et al.*, 1987; Ball *et al.*, 1992). However, the observed increase in Fis levels does not account for the increase in *rrn* promoter activity following the transition from stationary to exponential phase (Appleman *et al.*, 1998). Fis activation at *rrn* promoters is phase-of-the-helix-dependent and involves interaction with the αCTD (Newlands *et al.*, 1992; Bokal *et al.*, 1997).

A BLAST search revealed no apparent Fis homologue in *M.tb* but a possible candidate was found in the *M.sm* genome (62% homology compared to the C-
terminal half of *E. coli* Fis, p=0.0027). The possibility of a different protein with related function remains.

Factor independent activation refers to a direct interaction between the RNAP and the DNA UP element. Factor independent activation, like Fis, is mediated by the αCTD. The Cis-element consists of two phased AT-rich stretches located approximately −60 to −40 relative to transcription start point of the P1 promoter (Ross *et al.*, 1993). The two α monomers are arranged in tandem and the promoter proximal site has the largest effect on activation (Murakami *et al.*, 1997; Estrem *et al.*, 1999). The precise location of the UP element is crucial since the activation is phase-of-the-helix-dependent and an “out-of-phase” element inhibits promoter clearance, possibly by inducing a conformational change in RNAP (Newlands *et al.*, 1992; Tagami and Aiba, 1999).

There are indications that the UP element in fact acts as a compensatory mechanism for the down-effect of the odd 16-basepair spacer, characteristic of *E. coli* rrn promoters, and vice versa. Thus the activity of the core promoter without the UP element and with a 17 basepair spacer has an activity compared to that of the promoter with UP element but wild type spacer (Gaal *et al.*, 1989). The mutant with wild type consensus in both core promoter elements was not tested in the presence of the UP element in this study. Subsequent studies have demonstrated that the UP element in combination with a consensus promoter has an extraordinarily high affinity for RNAP. However, this increase in $K_B$ has a
negative effect on subsequent steps of initiation such as promoter clearance (Ellinger et al., 1994b; Strainic et al., 1998).

A mycobacterial homologue to the *E. coli* UP element has not been identified, but this could be due to the high GC content of the mycobacterial genome; in other words, a mycobacterial UP element could have the same function (and location) but a different nucleotide composition. This possibility is sustained by the recent finding that the extremely GC rich organism (70% GC content), *T. thermophilus*, has an UP element analogue located upstream of its 16S gene. This element bears no resemblance to the *E. coli* or *B. subtilis* UP elements but probably acts as such judging from affinity measurements using NMR spectroscopy (Wada et al., 1999). All seven amino acid residues identified in the *E. coli* RNAP α subunit as most crucial for DNA binding and UP element function (Gourse et al., 2000) are conserved between *E. coli*, *M.tb*, *S. coelicolor* and *B. subtilis*, whereas in *T. thermophilus* only four of these are present (Figure 1.4). The size of the spacers in the mycobacterial *rrn* promoters vary from 16 to 18 basepairs (Gonzalez-y-Merchand et al., 1997) which at least to some extent argues against the idea of an UP element with the same function.
Figure 1.4: Sequence alignment of αCTDs.

The figure illustrates the extensive sequence homology between the α-CTDs from a number of distantly related organisms. The CTD shown includes residues 248 to 329 in the *E. coli* sequence (according to Yasuno et al., 2001). The upper sequence is that of the *M. tb* protein and the bottom sequence is that of the *E. coli* (*E.c*) protein, which is the most well-characterised of the five shown. The remaining α proteins are from *S. coelicolor* (*S.c*), *B. subtilis* (*B.s*) and *T. thermophilus* (*T.t*). The asterisks above the sequence indicate amino acids important for UP element function in *E. coli*.
1.6.4.8 *rrn* Anti-termination

There are numerous examples of anti-pausing and anti-termination mechanisms in both prokaryotes and eukaryotes (reviewed by Greenblatt *et al*., 1993; Henkin, 1996; Richardson and Greenblatt, 1996). Probably the best characterised system is the λN mediated anti-termination of bacteriophage λ (Richardson and Greenblatt, 1996). The *rrn* anti-termination system is closely related to the λN system although there are at least two major differences. One is the λ encoded protein, N, which is obviously not part of the *rrn* system, and the second is that λN anti-termination acts on both Rho-dependent and intrinsic terminators whereas *rrn* anti-termination only has an effect on Rho-dependent terminators (Gottesman *et al*., 1980; Albrechtsen *et al*., 1990; Greenblatt *et al*., 1993). More accurately an elongation complex, which transcribes through an *rrn* leader region will subsequently read through Rho-dependent terminators. These modified elongation complexes are only terminated by (strong) intrinsic terminators (Aksoy *et al*., 1984; Holben and Morgan, 1984; Albrechtsen *et al*., 1990). Anti-termination (both *rrn* and λN) is signalled by the presence of a specific sequence, the *nut* site, located downstream of the promoters (reviewed by Condon *et al*., 1995b; Richardson and Greenblatt, 1996). As the RNAP transcribes the *rrn* BoxA sequence, a complex consisting of RNAP, NusA, NusB, NusE, NusG, and possibly additional ribosomal proteins assembles, and as a result the RNA
elongation rate is almost doubled (Squires et al., 1993; Vogel and Jensen, 1997; Zellars and Squires, 1999; Torres et al., 2001). Whether this is the reason for, or just a related feature of the anti-termination is not known for sure.

**NusA**

NusA is an essential protein in *E. coli*. (Nakamura and Uchida, 1983). However, in strains with reduced Rho-dependent termination, the *nusA* gene can be deleted (Zheng and Friedman, 1994). In *B. subtilis* the *nusA* gene is essential in *rho*+ as well as *rho*− strains (Ingham et al., 1999). The *E. coli* protein contains three RNA-binding domains, namely one S1 and two KH domains (Mah et al., 1999) and the *M.tbc* protein shares these features (B. Gopal personal communication). Worbs et al. have suggested that these three domains interact to form specific binding sites for RNA from non-specific sites (Worbs et al., 2001). The *E. coli* NusA binds to RNA but there is some controversy as to the sequence to which it binds, whether it binds specifically, and whether it can bind on its own (Tsugawa et al., 1985; Pardon et al., 1994; Liu and Hanna, 1995; Mogridge et al., 1995; Mah et al., 2000). Furthermore the binding affinities of NusA in these experiments appear to be extremely low, further indicating that the binding may in fact be non-specific. (Tsugawa et al., 1985; Mah et al., 2000). NusA has several functions. It is involved in transcriptional pausing as well as both Rho-dependent and intrinsic
termination (Kassavetis and Chamberlin, 1981; Kingston and Chamberlin, 1981; Lau et al., 1983; Sigmund and Morgan, 1988a; Sigmund and Morgan, 1988b; Burns et al., 1998; Artsimovitch and Landick, 2000). Finally NusA is essential for *rrn* anti-termination *in vivo* (Vogel and Jensen, 1997). The protein can thus be regarded as an elongation factor, a termination factor, and an anti-termination factor.

The *E. coli* NusA protein becomes associated with RNAP core after the release of σ; whether it remains associated with RNAP or undergoes cycles of association and dissociation during the process of elongation remains controversial (Greenblatt and Li, 1981; Schmidt and Chamberlin, 1984a). Perhaps this is a question of additional factors associating with the elongation complex. NusA interacts with both β, β', and α via both its NTD and the CTD, hence having two independent regions for RNAP interaction (Ito and Nakamura, 1996; Liu et al., 1996; Mah et al., 1999).

*In vitro* NusA reduces the RNA elongation rate in both a sequence dependent and a sequence independent manner. In other words, NusA enhances RNAP pausing at specific pause sites (Kassavetis and Chamberlin, 1981; Kingston and Chamberlin, 1981; Lau et al., 1983; Sigmund and Morgan, 1988a; Sigmund and Morgan, 1988b; Artsimovitch and Landick, 2000) as well as reduces the enzyme’s elongation rate *per se*. The latter is a result of either competitive inhibition of NTP binding or a NusA induced conformational change in the RNAP (Schmidt and
Chamberlin, 1984a; Zhang and Hanna, 1994). However, *in vivo* experiments with
the *nusA10cs* mutant demonstrate similar elongation rates of *lacZ* mRNA in wild
type cells and cells depleted of NusA (Vogel and Jensen, 1997).

As mentioned earlier, NusA is believed to stabilise the interaction between the
pause-hairpin of nascent transcripts and the RNAP. Hairpin dependent pause sites
and intrinsic terminators have many features in common. They are structurally
very alike, NusA has an enhancing effect on both mechanisms (Greenblatt *et al*.,
1981; Schmidt and Chamberlin, 1987; Sigmund and Morgan, 1988a; Sigmund and
Morgan, 1988b; Liu *et al*., 1996), and both involve interactions with the CTD of
RNAPα (Liu *et al*., 1996). It is therefore tempting to speculate that the
mechanism for intrinsic termination involves the same interaction between RNAP
and NusA, although this has not yet been demonstrated.

NusA has been reported to both enhance and suppress Rho-dependent termination
*in vitro* (Greenblatt *et al*., 1981; Sigmund and Morgan, 1988a; Sigmund and
Morgan, 1988b). But since Rho-dependent termination is intimately linked to the
coupling of transcription and translation (or to anti-termination) it is difficult to
draw any conclusions from isolated transcriptional studies. In fact it does seem
that results obtained with *in vitro* coupled transcription-translation systems are in
better agreement with the *in vivo* results (Kung *et al*., 1975; Nakamura *et al*.,
1986a). Another possibility is that Rho-dependent terminators of λ and of *E. coli*
react differently towards the action of NusA. This is supported by the fact that the
NusAII mutant enhances the former whereas it suppresses the latter (Nakamura et al., 1986a).

NusA is furthermore implicated in ppGpp mediated reduction in mRNA CGR, but there is some dispute as to whether it is required for, or just enhances ppGpp action (Kingston and Chamberlin, 1981; Vogel and Jensen, 1997; Burns et al., 1998).

Finally, NusA is essential for λ anti-termination (Richardson and Greenblatt, 1996) and for rrn anti-termination in vitro (Squires et al., 1993). It has furthermore been demonstrated that NusA is essential for the rrn BoxA-mediated increase in elongation rate of rRNA, i.e. rrn anti-termination in vivo (Vogel and Jensen, 1997). This is puzzling since NusA is otherwise believed to decrease the elongation rate. Perhaps in the case of the anti-termination complex the NusA protein function is modulated and it may just act as a linker between the RNAP and the other anti-termination specific factors. Alternatively the protein has two distinct and separable functions as suggested by a number of groups (Schmidt and Chamberlin, 1984a; Nakamura et al., 1986b; Sigmund and Morgan, 1988a; Sigmund and Morgan, 1988b).

The E. coli NusA CTD (i.e. the last 79 amino acids) plays an interesting role in NusA actions. This region is not required for NusA-enhanced intrinsic termination or λN-mediated anti-termination (Liu et al., 1996; Mah et al., 1999). It is, however, required for the interaction with the RNAPα subunit, which in turn is
involved in or required for enhancement of intrinsic termination (Liu et al., 1996; Mah et al., 2000). Together these statements appear somewhat conflicting. Finally, according to Greenblatt and co-workers the NusA protein cannot bind on its own to RNA unless the NusA CTD has been deleted (Mah et al., 2000). This CTD region is absent in *M.tb*, *B. subtilis*, and other Gram-positive NusA proteins but the implications are not known. The *M.tb* NusA has an additional feature, a poly-proline stretch consisting of five uninterrupted and two additional proline residues towards the end of the protein. The function of these residues is still unknown.

In summary, the actions and targets of NusA are multiple and not fully understood. Furthermore it appears that there is ongoing dispute as to what functions NusA has. Future experiments will hopefully resolve these differences.

**NusB**

The *E. coli* NusB is not essential but the inactivation of *nusB* leads to a cold sensitive phenotype (Taura et al., 1992). The *nusB5* mutant, which causes loss of function furthermore displays premature termination of *rrn* transcription as shown by an increase in the ratio between promoter proximal and distal *rrn* transcripts (Sharrock et al., 1985a). The protein binds to the *rrn* BoxA but not to the *λ* BoxA, and only in the presence of NusE (Nodwell and Greenblatt, 1993;
Mogridge et al., 1998). The protein is, however, part of the λN anti-termination complex and does bind to the elongation complex in the presence of the remaining anti-termination factors (Mason and Greenblatt, 1991; Mogridge et al., 1995; Mogridge et al., 1998). In vitro, NusB is necessary for rrn BoxA-mediated anti-termination (Squires et al., 1993), and in vivo NusB is necessary for rrn BoxA-mediated increase in elongation rate (Zellars and Squires, 1999). It has been suggested that NusB may play a role in translation as well as in transcription (Friedman and Court, 1995). The NusB proteins from E. coli and M.tb display a relatively high degree of conservation. There is, however, one major difference between the two proteins: the E. coli NusB is a monomer whereas the M.tb NusB appears to be a dimer (Altieri et al., 2000; Gopal et al., 2000). The implications for this difference are still unknown. The M.tb NusB displays weak interactions with M.tb NusE but a binding to RNA has not yet been demonstrated (Gopal et al., 2001).

**NusE**

NusE is identical to ribosomal protein S10 and the *nusE* gene is essential. NusE binds directly to the RNAP core and to NusB and hence acts a physical link between the BoxA and the RNAP (Mason and Greenblatt, 1991; Nodwell and Greenblatt, 1993). In vitro NusE can function in anti-termination both as an
isolated protein and as a ribosome constituent, but an in vivo requirement for the protein has only been demonstrated for λN mediated anti-termination, not for rrn anti-termination (Friedman et al., 1981; Das et al., 1985; Zellars and Squires, 1999). Like NusB, NusE does not bind to BoxA on its own (Nodwell and Greenblatt, 1993).

**NusG**

The *E. coli nusG* gene is essential for *E. coli* growth and like NusA the NusG protein has been shown to have multiple functions in the process of transcription (Downing et al., 1990). The *E. coli* NusG protein has the opposite effect of NusA in many situations but the two proteins do not compete for the same binding sites on RNAP and their targets vary (Burns et al., 1998). Like NusA, NusG interacts with both RNAP and Rho, although the binding to RNAP is weaker than that of NusA (Mason and Greenblatt, 1991; Li et al., 1993; Burns and Richardson, 1995). In contrast NusG interacts with Rho with high affinity (Li et al., 1993; Pasman and von Hippel, 2000). *In vitro* NusG increases the mRNA CGR by decreasing the half-life of class II pausing, possibly by preventing RNAP backtracking (Burova et al., 1995; Burns et al., 1998; Artsimovitch and Landick, 2000). Results obtained with cells depleted of NusG indicate that the mRNA elongation rate is not affected by NusG in vivo but that the protein may have an effect on one or
more post-transcriptional processes (Zellars and Squires, 1999). Thus the mRNA elongation rate, determined by the induction-lag of β-galactosidase expression, does not correlate with the mRNA elongation rate, determined by direct measurements on lacZ mRNA (Burova et al., 1995; Zellars and Squires, 1999). NusG enhances Rho-dependent termination at some terminators although overexpression of NusG actually inhibits Rho-dependent termination (Sullivan et al., 1992; Burova and Gottesman, 1995; Burns et al., 1999). The NusG protein is, however, essential for the rrn BoxA-mediated increase in RNA chain elongation rate, i.e. rrn anti-termination (Zellars and Squires, 1999). NusG is furthermore found in λN-modified transcription complexes in vitro (Mason and Greenblatt, 1991; Mogridge et al., 1995), but a direct in vivo requirement for NusG in λN-anti-termination has not been demonstrated (Sullivan et al., 1992). A further indication that this gene is not essential for λN-mediated anti-termination is the fact that a mutant of nusG was, unlike the other Nus factors, not identified as a mutation inhibiting λ growth but rather as a suppressor of the nusAl mutation (Sullivan et al., 1992). Curiously, in B. subtilis nusG is not essential but its deletion does have effects on growth (Ingham et al., 1999). Whether the fact that nusG is not essential is linked to the lower degree of Rho-dependent termination in B. subtilis remains to be determined. Whether the M.tb nusG can be deleted is an open question since the B. subtilis nusG can be, whereas the S. coelicolor nusG cannot (Ingham et al., 1999; Puttikhunt et al., 1995)
The nut site

The λ nut site consists of two elements, BoxA and BoxB (Greenblatt et al., 1993; Richardson and Greenblatt, 1996). An additional conserved sequence, BoxC resides downstream of BoxB but the function of this remains unknown (Berg et al., 1989). In the E. coli rrn operons the order of the BoxA and BoxB has been reversed but the functional importance of this inversion has not been determined (Li et al., 1984; Berg et al., 1989). The rrnB operon of mycobacteria does not contain a BoxB motif (Ji et al., 1994b). The motifs act on RNA level, in other words, the region has to be transcribed but must not be translated in order to function (Horwitz et al., 1987). BoxA is a highly conserved motif found in a number of bacterial (and archaeal) rrn operons and it appears to be the only essential part of the rrn anti-termination motif (Berg et al., 1989; Squires et al., 1993; Vogel and Jensen, 1997). Interestingly only the rrn BoxA and not the λ BoxA is capable of binding the NusB-NusE heterodimer in vitro (Nodwell and Greenblatt, 1993; Mogridge et al., 1998). BoxB is a region of dyad symmetry capable of hairpin formation and this motif interacts with the λN protein (Chattopadhyay et al., 1995). This region is essential in λ anti-termination possibly because of the λ BoxA’s inability to bind the NusB-NusE dimer (Greenblatt et al., 1993). The importance and function of BoxC in anti-termination has not been established to date. Expression of the λ mutL from a multi-copy plasmid specifically inhibits rRNA transcription. This inhibition is,
however, only partly caused by defective anti-termination as demonstrated by an increased ratio of leader RNA to 23S RNA (Sharrock et al., 1985b).

1.6.4.8 Anti-termination – the mechanism

The *rrn* anti-termination mechanism involves a number of the above mentioned factors as well as an unidentified cellular component (Squires et al., 1993). As the σ factor dissociates from the RNAP the NusA protein associates with the core enzyme (Greenblatt and Li, 1981) and the presence of NusA in the elongation complex is probably necessary for the further assembly of the anti-termination complex. Whether NusG is also associated with the RNAP before the enzyme encounters the BoxA signal is not known. As the RNAP transcribes the *rrn* BoxA sequence, the NusB-NusE heterodimer joins the complex by binding to the BoxA RNA sequence (Nodwell and Greenblatt, 1993). NusA, NusB, and NusG have been shown to be essential for the increased elongation rate characteristic of rRNA transcription (Vogel and Jensen, 1997; Zellars and Squires, 1999). Whether this also applies to NusE remains unresolved since the *nusE71* mutant does not have an effect on the BoxA mediated increase in elongation rate. On the other hand *nusE71* is a missense mutation, which does not cause loss of function or NusE deficiency, whereas the experiments with NusA, NusB, and NusG involved
a deficiency of functional protein (Vogel and Jensen, 1997; Zellars and Squires, 1999).

The question remains: Is the anti-terminated elongation complex anti-terminated simply by virtue of its elongation rate; that is by escaping Rho, as suggested by e.g. Condon et al. (1995b)? This model is supported by an inverse correlation between RNAP elongation rate and Rho-dependent termination (Jin et al., 1992), a relation that does not apply to intrinsic termination, (Chamberlin et al., 1979; McDowell et al., 1994; Weilbaecher et al., 1994). On the other hand it seems that the interactions are more complex. NusG increases the elongation rate but the protein also enhances Rho-dependent termination unless overexpressed (Burova and Gottesman, 1995; Burns et al., 1998; Burns et al., 1999). Conversely a protein such as NusA, which reduces the elongation rate is capable of decreasing Rho-dependent termination (Burns et al., 1998). Finally, Rho is capable of terminating transcription carried out by phage SP6 and T7 RNAP’s which both have elongation rates that are much faster than E. coli RNAP (Pasman and von Hippel, 2000). These findings obviously do not support the model of kinetic coupling. Alternatively, do the proteins in the anti-termination complex inhibit Rho action by direct protein-protein contact? This possibility is supported by the finding that NusA inhibits the ATPase function of Rho in vitro (Schmidt and Chamberlin, 1984b) and that NusG overexpression inhibits Rho-dependent termination possibly by altering the stoichiometry of the two proteins in Rho-NusG complexes.
(Burova and Gottesman, 1995). Finally, how much of this applies to the situation in *M. tb*? As described in section 1.4.4.2 the *M. tb* RNAP elongation rate is much lower than that of *E. coli*, so perhaps the kinetic considerations do not have any applications in *M. tb*. A number of the anti-termination factors appear different in organisms such as *E. coli*, *B. subtilis*, and *M. tb*. Mycobacterial *rrn* anti-termination motifs have been suggested by Ji *et al.* (Ji *et al.*, 1994b; Ji *et al.*, 1994a) but no function has been demonstrated.
2 Aim of the study

The overall aim of this study was to investigate the two \textit{rrn} operons, \textit{rrnA} and \textit{rrnB}, of \textit{M.tb} and \textit{M.sm}, respectively. This was done in order to get a further understanding of the elements involved in the regulation of mycobacterial rRNA expression.

The specific objectives were to

1. Use reporter constructs to investigate the expression of \textit{rrn} operons
2. Use reporter constructs to identify core promoter elements in \textit{rrn} operons
3. Identify sequences of importance within core promoters
4. Use reporter constructs to identify regulatory elements flanking the core promoters
5. Characterise these regulatory elements
6. Investigate the regulatory functions of the \textit{rrn} leader region including protein/protein and protein/RNA interactions, possibly involved in the anti-termination of the \textit{M.tuberculosis rrn} operon
3 Materials and methods

3.1 Cloning of mycobacterial promoter fragments

The strategy used in the promoter study was to fuse a series of mycobacterial promoter fragments to the lacZ gene of E. coli. The promoter probe vectors were pEJ414 and pEJ425, kindly provided by E.O.Davis, NIMR. These plasmids are derivatives of the integrating vector pMV361 (Stover et al., 1991) with a promoterless lacZ gene and kanamycin\(^R\) or hygromycin\(^R\), respectively. The plasmids integrate by means of the plasmid-encoded integrase in a site-specific manner via the attB site. The plasmids have an oriE derived replication origin but no mycobacterial replication origin (Figure 4.1). Hence integration of the plasmid into the mycobacterial genome is required for transformation of the cells. All reporter constructs were cloned as Xbal-HindIII, Xbal-Xbal or HindIII-HindIII fragments (details of each promoter insert is given in Appendix B). M.sm strain numbers equal the number of the integrated plasmid (i.e. KAM42 harbours the integrated pKA42).

pKA58 was created by amplifying the region from ~400 basepairs upstream of the M.tb rrnA P1 transcription start point to 219 basepairs downstream of the PCL1 transcription start point. The fragment was digested with PstI and HindIII and
ligated into pUC19. The resulting plasmid (pMtbA) was digested with *Xba*I and *Hind*III and the *rrnA* fragment was ligated into pEJ414 between *Xba*I and *Hind*III.

pKA59 was created by PCR amplification of *M.sm* DNA using the same primers and cloning strategy as in pKA58.

pKA75 was created by digesting pKA72 with *Xba*I, filling in the overhangs with Klenow fragment (Promega) according to standard procedures. The *rrnB* upstream *Xba*I fragment from pKA69 was likewise filled in with Klenow and ligated into the now blunt-ended pKA72 resulting in pKA75.

pKA76 was created by deleting 2 basepairs in pKA75 by means of site directed mutagenesis (see appendices B and D for details on primers and plasmids).

Promoter inserts of less than 100 basepairs in length (and that of pKA87) were created by annealing complementary oligonucleotides. The oligonucleotides were 5'-phosphorylated either before or after annealing with T4 polynucleotide kinase (Promega) according to manufacturer's instructions. Complementary oligonucleotides with the appropriate overhangs for direct ligation were annealed at a concentration of 2 µM in buffer (1x restriction buffer H, Roche). The oligonucleotides were mixed and heated to 90°C, followed by slow cooling to room temperature.

All other promoter inserts (except that of pKA87) were created by PCR using:

10 ng genomic DNA (*M.tuberculosis* or *M.sm*), 0.3 µM each primer (Appendix B-D), 200 µM dNTP's in 1x *Pfu* buffer, and 2.5 units of *Pfu* polymerase (Stratagene) in a 50
μl reaction. Cycling conditions: 3 min at 95°C, 30 x (1 min at 95°C, 1 min at 60°C, 1 min/kb at 72°C), 7 min. at 72°C

PCR products were digested with the appropriate enzymes, and purified on PCR spin columns (QIAGEN) before ligation into pUC19, pEJ414 or pEJ425.

3.2 Site directed mutagenesis

Site directed mutagenesis was performed by using the Quickchange site directed mutagenesis protocol (Stratagene). Two complementary primers containing the desired mutation(s) and 10 to 15 flanking nucleotides were designed. The primers had a $T_M$ of more than 78°C according to the formula:

$$T_m = 81.5 + 0.41(GC\%) - 675/N - \%\text{mismatch}$$

where N is the primer length in basepairs. Primers were purified by HPLC (Oswell). The mutagenesis reaction was performed with *Pfu turbo* according to the Quickchange protocol using 0.5μl, 1.0μl, and 2.0 μl of mini-prep DNA as template. After the mutagenesis reaction the DNA was treated with *DpnI* for one hour in order to digest the template and the mutagenised plasmid was transformed into competent *E. coli* DH5α (Sambrook *et al.*, 1989) sub-cloning efficiency (or supercompetent *Epicurian coli*, from Stratagene, if sub-cloning efficiency did not generate transformants). Plasmids were isolated and the desired fragment was cut
out and ligated into a new vector. In the case of reporter constructs the mutagenised fragment was ligated into the same vector backbone; this was done in order to avoid the risk of mutations in the $\text{lacZ}$ reporter during the PCR cycling. In the case of the RNAP subunits, the desired fragment was ligated into an expression vector (see below).

3.3 Ligations

Vectors were dephosphorylated before ligation. The digested vector was incubated 2 x 30 minutes with 1 unit calf intestine alkaline phosphatase (Roche) in restriction enzyme buffer. After dephosphorylation the DNA was purified with PCR purification columns (QIAGEN) or run on a 0.8% agarose gel and purified with Qiaquick Gel purification system (QIAGEN). The ligations were performed by using Rapid DNA Ligation Kit (Roche): Vector and insert were mixed and incubated between 15 and 120 minutes at room temperature according to manufacturer’s instructions.

The strategy for cloning of the RNAP subunits was different and is indicated below in a separate section.
3.4 Competent cells of *M.sm*

A culture of *M.sm* was grown overnight to an OD of $0.8 < A_{600} < 1.0$. Cells were cooled on ice for 30 min and centrifuged in a GSA rotor at 5000 rpm for 10 min. The pellet was washed three times in 10% ice-cold glycerol reducing the volume by 50% each time until finally it was resuspended in 1/10 the original volume. The cells were kept at 4°C for at least one hour before electroporation.

3.5 Transformations

The ligated DNA was transformed into competent *E. coli*, DH5α (subcloning efficiency from Life Technologies) according to standard procedures. The plasmids were isolated from *E. coli* and the sequence of the insert was determined. Correct clones were transformed by electroporation into competent *M.sm mc²155* (Snapper *et al.*, 1990), *M.sm SMR5*, *M.sm KO14* or *M.sm KO16* (Sander *et al.*, 1996). The transformation of mycobacteria is described elsewhere (Papavinanasundaram *et al.*, 1998). The sequence of the integrated promoter fragment was verified by isolating genomic DNA from each strain and PCR amplification using primers Pmint and lac518R (Appendix C). The PCR product was purified on a YM-50 microcon spin column (Millipore) and sequenced by means of the d-Rhodamine sequencing kit (Applied Biosystems) in an automatic sequencing reaction. Mostly both strands were sequenced using lac518R and
Pmnt. If only one strand was sequenced lac518R was used. The samples were run on an ABI377 automatic sequencer.

3.6 DNA isolation

Genomic DNA was isolated by two different methods. For sequence determination of integrated plasmids the InstaGene matrix from Bio-Rad was used. A loopful of mycobacteria was washed in 1 ml of water, centrifuged, and resuspended in 0.2 ml InstaGene matrix. The mixture was incubated at 56°C for 20-30 minutes, vortexed for 10 seconds, and boiled for 8 minutes. The mixture was vortexed again and centrifuged, and 10 µl of the supernatant was used in a standard PCR reaction.

Higher quality DNA was isolated by a modified version of the method described by (Gonzalez-y-Merchand et al., 1996b). A 10-25 ml overnight culture was harvested by centrifugation at 5000 rpm in an HS-4 rotor. The pellet was resuspended in 1 ml 3mg/ml lysozyme and lipase (SIGMA) and incubated at 37°C for one hour. The suspension (which was now very viscous) was freeze-thawed three times by shifting from dry ice/ethanol to 70°C waterbath for ten minutes. The samples were extracted twice with phenol/chloroform/isoamylalcohol (125:25:1), once with chloroform and precipitated with two volumes of ethanol.
The pellet was washed with 70% ethanol, dried and resuspended in the desired volume of water.

Small scale plasmid DNA was isolated from 5 ml overnight *E. coli* culture by means of QIAprep Spin Miniprep kit (QIAGEN). Larger amounts of plasmid DNA was prepared with QIAGEN midi or maxi prep kit.

3.7 RNA isolation

50 ml of culture was grown to an OD of 0.5< $A_{600}$ <0.8. The culture was mixed with crushed ice in order to cool the cells instantly. The cultures were centrifuged for 15 minutes at 10000 rpm in a GSA rotor at 4°C. The pellet was resuspended in 0.6 ml 6M guanidinium buffer (Hybaid RNA isolation kit), and mixed with one volume of phenol/chloroform (5:1). The cells were broken by means of glassbeads in a Ribolyser (Hybaid) for 20 seconds at speed setting 6.0. The pellet was extracted once more with phenol/chloroform and the rest of the procedure is exactly as recommended by Hybaid and according to standard procedures for isolation of nucleic acids. After washing twice in 70% ethanol, the pellet was resuspended in 100 μl water and purified by means of RNeasy columns (QIAGEN) according to manufacturer's instructions. The RNA was treated twice with 10 units of RNase-free DNasel (Roche) for 30 to 60 minutes at 37°C in DNasel buffer (Promega) in the presence of 1u/μl RNasin. After each DNase
treatment the RNA was purified on RNeasy columns. The RNA was kept in an aqueous solution with 1u/μl RNasin (Promega) at -20°C.

3.8 Growth of Bacteria

*E. coli* was grown in liquid L broth (Appendix E) with ampicilin (100 μg/ml), kanamycin (50 μg/ml), hygromycin (250 μg/ml), or chloramphenicol (34μg/ml). The same concentrations of antibiotics were used for L plates. *M. sm* was grown in modified Dubos broth (Appendix E) with 15 μg/μl kanamycin or 50 μg/ml hygromycin. Sub-culturing of some of the strains resulted in very long lag phase, variation in cell morphology and β-gal expression. Therefore all strains were inoculated directly from the plates into 30 ml Dubos broth.

All cultures had been growing exponentially for at least five generations (overnight) to ensure balanced growth, and were in mid-log phase (OD 0.4 < A<sub>600</sub> < 0.8) by the time of harvest. Plates for mycobacteria were 7H11 agar (Difco) with either 25 μg/ml kanamycin or 50 μg/ml hygromycin.

3.9 Preparation of cell free extracts

30 ml of culture was cooled on ice and harvested by centrifugation for 10 minutes at 5000 rpm in an HS-4 rotor. The cell pellet was washed three times in cold Z
buffer without β-mercaptoethanol (β-mercaptoethanol goes off quickly and interferes with the BCA assay, see below). Finally the cells were resuspended in 1 ml of Z buffer and transferred to a 2 ml screw cap tube with ~ 0.5 ml glassbeads (150-212 microns, acid washed, SIGMA). The cells were disrupted in a Ribolyser (Hybaid) for 30 seconds at speed setting 6.5 and the cell debris was removed by centrifugation for 10 minutes (13000 rpm at 4°C). A small aliquot of the cleared lysate was removed for protein estimation and the rest was mixed with β-mercaptoethanol to a final concentration of 38 mM.

3.10 Protein assays

The β-galactosidase (β-gal) assay was carried out essentially as described by Miller (Miller, 1972). The assay is a colourimetric determination of the conversion of ONPG to ONP, measured on an ELISA plate reader at 405 nm (yellow colour). The β-gal activity is calculated as 380 x A₄₀₅/minute x mg total protein and expressed as units/mg protein. 500 μl cell extract (diluted or undiluted) is mixed with 100 μl ONPG solution (4mg/ml) and the reaction mixture is incubated at 28°C until a yellow colour develops. The reaction is stopped by increasing pH by the addition of 250 μl NaCO₃ (1M). β-gal assays
were performed in duplicate or triplicate within each experiment and at least three independent experiments were performed on each strain.

The protein concentration was determined by means of the BCA Protein Assay Kit (PIERCE) according to manufacturer's instructions.

3.11 DNA sequencing

Two different methods of DNA sequencing were used in this study. For primer extension analysis manual sequencing was carried out by means of "Sequenase kit" from USB. 5μg of plasmid DNA was denatured by adding 0.1 volume of 2M NaOH/2mM EDTA and incubating at 37°C for 30 minutes. The DNA was precipitated by adding 0.1 volume of sodium acetate (pH 5.0) and 2.5 volumes of 100% ethanol. The samples were incubated at -80°C for 30 to 60 minutes and centrifuged at 13000 rpm for 15 minutes. The pellet was washed in 70% ethanol and resuspended in 7 μl water including 1 pmol of the desired primer. (For the sequencing reaction the primer is unlabelled, although it is possible to use the same primer, labelled with $^{32}$P, as in the primer extension. This results in less sharp bands and different precautions will have to be made using $^{32}$P rather than $^{35}$S). 2 μl of T7 sequenase buffer was added and the mixture was heated to 65°C for 2 minutes and cooled slowly to room temperature before leaving it on ice. The
rest of the procedure is exactly as recommended by the manufacturer and was carried out with $^{35}$S-dATP. Samples were run on a 6% acrylamide sequencing gel at 55 W for 3 hours.

All other sequence reactions were performed as automatic sequencing using the d-Rhodamine sequencing kit from Applied Biosystems with 8-10 μl miniprep DNA as template. The reactions were carried out according to manufacturer’s instructions except for an addition of 5% DMSO to facilitate DNA denaturation during the PCR cycling.

### 3.12 Primer extension

The primer used for the primer extension, lac518R, anneals approximately 90 basepairs downstream of the $\textit{lacZ}$ initiation codon (sequence can be found in Appendix D).

10 pmol of primer was labelled with 30 μCi $\gamma$-$^{32}$P ATP using T4 PNK (Promega) in a total volume of 10 μl 1 x PNK buffer. The mixture was incubated at 37° for 10 minutes and the enzyme was heat inactivated for 5 minutes at 90°C and subsequently stored at -20°C. The primer extensions were carried out by mixing 10 to 50 μg of total RNA and 1 pmol primer in a total volume of 6μl. The mixture was denatured at 70°C for 10 minutes and cooled slowly to 58°C, which is the melting temperature of the primer. The samples were centrifuged briefly, 5μl of 2
x AMV RT buffer (Promega) was added and the mix was incubated at 58°C for an additional 20 to 30 minutes. The tubes were cooled at room temperature for 10 minutes and a mixture of 5 µl 2 x AMV buffer, 1.4 µl Sodium pyrophosphate (40mM), 1.6 µl water, and 1 unit of AMV RT (Promega) was added. The mixture was incubated at 42°C for 30 minutes. The volume was brought to 200µl with water and extracted once with phenol/chloroform/isoamylalcohol. The samples were ethanol precipitated and resuspended in 10 µl formamide loading buffer (98% formamide, 10 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenolblue). The samples (2 µl) was loaded onto a 6% acrylamide urea sequencing gel (1 x TBE) and run for 3 hours at 55 W. The gel was made with Long Ranger acrylamide solution (50% stock Solution; JT Baker, see Appendix E).

3.13 Expression and purification of *M.tb* RNAP subunits

The gene of interest was PCR amplified using *Pfu Turbo* (Stratagene). The primers for each amplification are listed in appendices C and D.

In the case of pKAα the PCR product was digested with *NdeI* and *BamHI* and cloned into pET15b (Studier *et al.*, 1990).

pKAω was created by ligating the PCR product directly into pCR-Blunt TOPO II according to manufacturer’s (Invitrogen) instructions resulting in plasmid TOPO-
ω. The TOPO-ω plasmid was subsequently digested with NdeI and EcoRI and the fragment was ligated into the same sites in pET15b.

pKAβ was created by ligating the PCR product into pCR-Blunt TOPO II. The resulting plasmid (TOPO-β) was mutagenised by site directed mutagenesis with primers F52118 and F52119 (see Appendix D for sequences). The primers introduced an Ncol site in the junction between the TOPO -vector and the upstream end of the rpoB insert, and at the same time changed the M.tb rpoB TTG initiation codon to an ATG codon. The resulting plasmid (TOPO-βmut) was digested with Ncol and EcoRI and the rpoB insert was ligated into pET15b digested with the same enzymes.

Approximately same strategy was employed for generating pKAβ'. The β' coding region was PCR amplified with flanking DNA and ligated into pCR-blunt TOPO II. The plasmid was subject to site directed mutagenesis with primers F53396 and F53397 (Appendix C and D), introducing an ATG initiation codon internal of an Ndel site. After retrieving the mutagenised plasmid the rpoC insert was excised with Ndel and BamHI and ligated into pET15b into the same same sites.

M.tb sigA was PCR amplified and the product was ligated into pCR-blunt TOPO II. The resulting TOPO-σ was digested with AflIII and EcoRI and ligated into the NcoI and EcoRI sites in pET15b (AflIII was used for the digest since the DNA overhangs generated with AflIII and NcoI are compatible and the M.tb sigA gene contains a number of NcoI but no AflIII sites).
All pET15b ligations were transformed into *E. coli* DH5α (sub-cloning efficiency), the resulting plasmids were isolated, and a number of clones with the correct insert were transformed into *E. coli* BL21 DE3 pLysS (Studier *et al.*, 1990). Small-scale expression cultures were made by diluting an overnight culture 25-fold (grown in L broth with 100 µg/ml ampicillin + 34 µg/ml chloramphenicol) in 2 ml of the same medium (pre-warmed to 37°C). Growth was continued for another 90 minutes at 37°C before inducing with 1mM IPTG. Growth was continued for three hours before the cells were harvested by centrifugation. The pellet was resuspended in 50 ml denaturing lysis buffer (100 mM NaH$_2$PO$_4$ + 250 mM NaCl, pH 7.0, 8 M urea) and heated for 5 to 10 minutes before mixing with half a volume of SDS sample buffer. The lysates (5 µl) were subjected to SDS-PAGE on 12% gels (Appendix E) at 20 mA for two hours in 1 x Tris/Glycine/SDS buffer (Bio-Rad).

Only RNAPα was purified. Cells were grown at 37°C to OD 0.5, induced with IPTG (1mM final concentration) and incubated at 30°C for five to six hours. The cells were harvested by centrifugation and kept at -80°C overnight. The pellet was resuspended in 10 ml native lysis buffer (LBf, 50 mM K$_2$HPO$_4$ + 250 mM NaCl, pH 7.0), mixed with EDTA-free protease inhibitor (Complete(R) protease inhibitor cocktail tablets, Roche Diagnostics) and sonicated in bursts of one minute to shear the DNA. The supernatant was cleared by centrifugation in an SS-
34 rotor at 8000 rpm for 30 minutes and purified on a cobalt affinity column (Talon, Clontech). The affinity matrix was prepared by decanting the supernatant (i.e. storage buffer) and washing twice in LBf. The supernatant from the centrifugation was mixed with the Talon matrix and incubated at 4°C for one hour with gentle mixing and loaded onto a Gradi-frac (Pharmacia). The column was washed twice in wash buffer (i.e. LBf with 5 mM imidazol) and eluted with LBf, 250 mM imidazol at a flow rate of 1.0 ml/minute. The eluate was collected in 3 ml fractions. The protein was purified a second time by size exclusion using S-75 sepharose columns (Pharmacia). The protein was concentrated on Centricon YM-10 (Millipore) to a final concentration of ~30 mg/ml.

Selected UV absorbing fractions were loaded on a 10% polyacrylamide SDS gel and run for 2 hours at 20mA. The gel was stained with Coomassie Brilliant blue R-250 (SIGMA).

3.14 In vitro transcription

The probes for the RNA gel shifts corresponded roughly to PCL1 generated transcripts ending at +104 or +219. The region was PCR amplified with primers F30406 and F3677 or F13396 and the resulting fragments were digested with HindIII and cloned into pGEM3Zf. These templates were linearized with XbaI and used in an in vitro transcription reaction using: 0.5 µg template, 2 µl 10 x SP6
transcription buffer, 4 µl 3 NTP mix (2 mM each), 1 µl UTP (300 µM), 50 µCi α-32P-UTP (2.5 µl), 1 µl RNasin (Promega), 1 µl SP6 RNA polymerase (Roche).

The reaction mix was incubated at 37°C for 30 to 60 minutes. Template DNA was removed by incubating with 10 u DNasel (Roche RNase-free) for 15 minutes, and the probes were purified using RNeasy spin columns (QIAGEN). The radioactivity of each sample was determined by scintillation counting. Using the obtained activity and the known specific activity of the UTP in the transcription reaction an estimated concentration of each probe was calculated.

3.15 Electrophoretic mobility shift assays

The probe for the full-length upstream region of rRN B (referred to as usB) was obtained cloning a PCR fragment into the XbaI site of pUC19 resulting in plasmid pusB. The fragment was excised with SalI and EcoRI and subsequently labelled by a fill-in reaction using 1 pmol DNA, 50 µCi α-32P-dATP (Amersham 3000 Ci/mmol), 33 µM each CTP, GTP, TTP, 1 unit Klenow in 10µl 1 x Klenow buffer (Promega). The binding reaction between the usB-DNA and the RNAPα subunit was performed in a low-salt buffer (DNA-EMSA buffer, Appendix E). 1/5 volume of glycerol loading buffer without dye (15 µl TBE, 500 µl 80% glycerol, 485 µl water) was added prior to loading. Samples were run on a native 8%
acrylamide gel in 1/2 x TBE at 25mA at 4°C, dried and exposed to autoradiography or phosphor imaging.

The RNA binding reactions were incubated at room temperature in 20 µl 1 x RNA-EMSA buffer (Appendix E) with the indicated amounts of protein. 1/5 volume of glycerol loading buffer without dye was added prior to loading onto a native 8% acrylamide/ 1 x TBE gel. The gels were run at 35mA for 2.5 hrs in 1x TBE, dried and exposed to autoradiography or phosphor imaging.
4 The expression of rrn operons (Results)

4.1 Introduction

In prokaryotes the expression of rrn operons is tightly linked to the growth rate of the cells by a number of mechanisms (Condon et al., 1995b). To date very little is known about transcriptional regulation and rrn expression in mycobacteria. The primary objective of the current study was to investigate the transcription of rRNA in mycobacteria; more specifically to identify regions in the rrn operons that either enhance or repress the expression. The majority of manipulations in rrn genes are either masked by compensatory activity by other rrn operons, if the organism contains more than one rrn operon (Condon et al., 1993), or lethal if the organism contains only one rrn operon. Therefore a reporter gene system was applied. Use of a reporter gene, i.e. a gene that is translated, could generate some artefacts in the expression pattern of otherwise untranslated RNA. These artefacts would, however, for the most be associated with post-transcriptional events although a reporter gene system will always be prone to artefacts for a number of reasons (see Chapter 9). A series of mycobacterial promoter fragments were fused to a promoterless lacZ reporter gene. M.sm does not contain a lacZ gene and the expression of β-gal is driven only by the cloned promoter fragment. In order to achieve maximum stability of the reporter constructs, the fusions were made with
Figure 4.1: The promoter probe vector, pEJ414.

This vector and its derivative, pEJ425, which has a hygromycin resistance cassette instead of a kanamycin resistance cassette, were used for the reporter gene fusions in this study. The vectors are derivatives of pMV361 (Stover et al., 1991) and harbour the attachment site (att) and integrase gene (int) of the mycobacteriophage L5. Integration is site-specific and in single copy. The plasmids have a pUC19 derived replication origin, oriE, but no replication origin for mycobacteria. Integration into the mycobacterial genome is therefore essential for transformation of the cells. Promoter fragments were cloned upstream of a promoter-less lacZ gene between the XbaI and HindIII sites, indicated in bold italics (top panel), and transformed into M.sm strains. The sequence in the top panel shows the region around the polylinker region of pEJ414. Blue letters indicate lacZ coding sequence and the blue arrow, primer lac518R used for primer extension assays. Lower case, red t-residues indicate transcriptional terminators mentioned in the text. Plasmids that were transformed into the M.sm rrr knock out strains were derived from pEJ425.
integrating vectors (pEJ414 or pEJ425, kindly provided by E.O.Davis NIMR). A schematic overview of pEJ414 has been shown in Figure 4.1. Because the vectors contain no mycobacterial origin of replication, transformation of mycobacteria depends on the integration of the plasmid into the bacterial genome.

The integration occurs in a site-specific manner via the attB site in *M.sm* and the chromosomal context is therefore the same in all constructs (Lee et al., 1991). Both vectors contain transcriptional terminators in all three reading frames immediately upstream of the polylinker to prevent read-through from upstream vector sequences.

Transformed cells of *M.sm* were grown overnight to an OD of 0.4 < A<sub>600</sub> < 0.8 before harvest. Promoter activity was determined by assaying the β-gal activity and normalising the values to the total protein content of each sample. In *M.sm* during balanced growth, the activity of the promoterless *lacZ* gene was consistently determined to be between 0.5 and 1.0 units/mg protein (data not shown), and therefore not taken into account.

### 4.2 Activity of rrn operons

The structures of the promoter and leader regions of the three rrn operons, *M.tb rrnA*, *M.sm rrnA*, and *M.sm rrnB* are shown schematically in Figure 4.2. These operons contain 2, 3, and 1 promoters, respectively as well as putative
Figure 4.2: The *rrn* operons of *M. tb* and *M. sm*.  
The figure is a schematic illustration of the *rrn* operons of *M. tb* and *M. sm* investigated in this study. The *rrnA* operon of *M. tb* contains two promoters, P1 and PCL1, whereas the *rrnA* operon of *M. sm* contains three promoters, P1, PCL1, and P2. *M. tb* does not contain the *rrnB* operon. Transcription start points of the different promoters have been indicated by arrows. The anti-termination motifs, BoxB, BoxA, and BoxC are shown as yellow boxes. Distances between various elements of the promoter and leader regions have been indicated. The figure has been drawn according to Gonzalez-y-Merchand *et al.* (1997).
anti-termination signals BoxB, BoxA, and BoxC. No sequences have been assigned as BoxB in the *M.sm rrnB* operon (Ji *et al.*, 1994b). The promoter plus leader regions of the three operons were cloned into the promoter probe vector and transformed into *M.sm mc²155*. The sizes of the flanking regions in each construct are indicated in Table 4.1. The two constructs representing the *M.tb rrnA* operon (KAM58 and KAM60) have the same 3' end 219 basepairs downstream of the transcription start point of the PCL1 promoter, which is 29 basepairs into the mature portion of the 16S RNA.

The *M.tb rrnA* construct, KAM58, was created according to a recent report by Verma *et al.* for comparative reasons (Verma *et al.*, 1999) and extended 307 basepairs upstream of the P1 core promoter (i.e. 345 basepairs upstream of the P1 transcription start point). KAM60 is slightly shorter with 262 basepairs upstream of P1, corresponding to 300 basepairs upstream of the transcription start point.

The fragment containing the *M.sm rrnA* (KAM59) operon was created using the same PCR primers as those used for amplification of *M.tb rrnA* (KAM58). The extent and sequence of the upstream regions of these constructs are therefore the same, whereas the downstream region varies by 10 basepairs (Table 4.1). The primary difference between the *M.tb rrnA* (KAM58) and *M.sm rrnA* promoter regions is therefore an insertion of an extra promoter (P2) in the *M.sm rrnA* construct as is the case *in vivo* (Gonzalez-y-Merchand *et al.*, 1997). The construct
representing the *M.sm rrnB* operon (KAM55) includes 103 basepairs upstream of the core promoter (i.e. 140 basepairs upstream of the transcription start point) and 207 basepairs downstream of the transcription start point. The *M.sm rrnB* construct differs from the others in that it does not extend into the mature 16S but has its 3' end 93 basepairs upstream of the mature 16S. Because of the extent of the flanking regions, these constructs are referred to as the full-length constructs and in this study represent the activities of the entire *rrn* operons. Cultures of the full-length reporter strains displayed some difficulties in growth, especially when sub-cultured from an overnight stationary phase culture. Hence all cultures were subsequently inoculated directly from the plate into 30 ml of modified Dubos broth and grown overnight to mid-log phase.

The four full-length constructs were assayed for β-gal activity and found to have similar activities, i.e. between 1095 and 1548 units/mg protein during balanced growth (Table 4.1). There was a small but reproducible difference between the four activities, which to some extent reflects the different number of promoters in the three operons. Thus the highest activity is conferred by the *M.sm rrnA* operon with three promoters and the second highest is conferred by the larger of the *M.tb rrnA* constructs with two promoters. The activity of the shorter of the *M.tb rrnA* and the *M.sm rrnB* constructs (KAM60 and KAM55) are almost identical (~1100 units/mg protein, Table 4.1). The difference in activity between KAM58 and KAM60 (362 units/mg protein) was surprising since the deleted region is located
more than 250 basepairs upstream of the P1 core promoter. The region was, however, not investigated further and subsequent constructs with the P1 upstream region all begin at position –300 relative to the P1 transcription start point (262 basepairs upstream of the P1 core promoter) and are thus comparable to KAM60.

Table 4.1: Activities of the full-length rrn reporter constructs.

<table>
<thead>
<tr>
<th>Strain  (KAM number)</th>
<th>55</th>
<th>58</th>
<th>59</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operon</td>
<td>M.sm rrnB</td>
<td>M.tb rrnA</td>
<td>M.sm rrnA</td>
<td>M.tb rrnA</td>
</tr>
<tr>
<td>Upstream flanking region (bp)</td>
<td>140</td>
<td>345</td>
<td>345</td>
<td>300</td>
</tr>
<tr>
<td>Downstream flanking region (bp)</td>
<td>207</td>
<td>219</td>
<td>229</td>
<td>219</td>
</tr>
<tr>
<td>Number of promoters</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>β-gal activity</td>
<td>1106±18</td>
<td>1457±135</td>
<td>1548±70</td>
<td>1095±138</td>
</tr>
</tbody>
</table>

The table shows the sizes of the flanking regions, the number of promoters identified in each operon, and the β-gal activity of the reporter constructs investigated in the present study. Upstream region is relative to the transcription start point of the most upstream promoter. Downstream flanking region is counting from the transcription start point of the most downstream promoter. The M.tb rrnA operon is represented in two different constructs, KAM58 and 60 with different upstream regions.
4.3 The transcription start point of reporter constructs

In order to verify that the presence of the reporter gene did not alter the transcription start site of the promoters, primer extension experiments were carried out using RNA from strain KAM56, the reporter strain with the highest expression level (~24000 units/mg protein). Using the strain with the highest expression level ensured a strong signal, which should also reveal the presence of secondary/cryptic promoters. The primer was specific for the lacZ gene and should therefore only generate extension products from the rrn-reporter fusion and not from the native rrn transcripts. RNA from KAM56 (10, 20 and 50 μg) was used in a standard primer extension reaction, which was run on a 6% polyacrylamide sequencing gel. Figure 4.3 shows the result of a representative experiment, which demonstrates that the transcription start point in the reporter construct is the same as that of the native rrnB operon (Gonzalez-y-Merchand et al., 1996a). Therefore, the presence of the lacZ reporter does not influence promoter recognition or transcription start site.
Figure 4.3: The transcription start point of reporter constructs.

Primer extension analysis were performed on RNA from KAM56 (reporter construct of \textit{rrnB} (-140 to +10). The primer extension was carried out with 10, 20, and 50 \textmu g of total RNA (lanes 1-3) and run alongside a sequence reaction shown on the left. The sequence from the -10 box to the \textit{HindIII} site is shown in bold below the gel with the transcription start point marked with an asterisk. The figure shows the single extension product obtained with RNA from this strain, which is in accordance with the transcription start point reported for the native \textit{rrnB} operon (Gonzalez-y-Merchand \textit{et al.}, 1996). The experiment demonstrate that the fusion to the reporter gene does not alter the original or introduce additional transcription start points.
4.4 Feedback regulation of rrrn expression

In E. coli the expression of rrrn operons is regulated on a number of levels. One mechanism, which is still not fully understood, is the mechanism of feedback inhibition or activation (recent investigations suggest that this is in fact two distinct mechanisms, (Voulgaris et al., 2000)). A number of rrrn operons can be inactivated in E. coli without affecting growth or rRNA levels dramatically (Condon et al., 1993). This is because the remaining intact rrrn operons increase their expression levels thus compensating for the reduced gene dosage.

In order to investigate the effect a reduction in rrrn gene dosage would have on the expression of rrrn genes in M.sm, two rrrn knockout strains were transformed with reporter constructs. The strains are derivatives of M.sm SMR5 (streptomycin resistance), in which either the rrrnA or the rrrnB operon has been disrupted by homologous recombination (Sander et al., 1996). The growth rates of the two knockout strains and the parental strain were monitored to determine whether the reduced gene dosage has an effect on growth. Figure 4.4 shows the growth curves and it is evident that the growth rates of the three strains are virtually identical.

When the cultures were left to grow for longer periods of time, a slight difference in growth could, however, be observed and the growth rates were in the order

\[ \text{SMR5} > \text{rrnB} > \text{rrnA} \]
Figure 4.4: Growth rates of *M. sm* wild type and *rrn* strains.

Growth of three strains of *M. sm* was monitored over a period of six hours. The strains are: *M. sm* SMR5, which has a mutation in the *rpsL* gene conferring streptomycin resistance (black trace); KO14, a derivative of SMR5 in which the *rrnA* operon has been inactivated by inserting a kanamycin resistance cassette (blue trace); KO16, a derivative of SMR5 in which the *rrnB* operon has been inactivated analogous to KO14 (red trace). All strains are described in Sander *et al.*, (1996). The figure demonstrates that the growth rates of the three strains are almost identical under the used growth conditions.
indicating that a deletion of the *M.sm rrnA* operon has a larger effect than a deletion of the *M.sm rrnB* operon. The three strains were transformed with two full-length constructs, *M.sm rrnB* (pKA55\textsubscript{B}) and *M.tb rrnA* (pKA58\textsubscript{A}) and assayed for β-gal activity. These plasmids have the same inserts as pKA55 and pKA58, respectively but the promoter inserts were made in pEJ425 (hygromycin resistance) since the *rrn* knockout strains were already kanamycin resistant (Sander et al., 1996). The three strains were also transformed with reporter constructs containing minimal promoter fragments, i.e. the two promoter hexamers and the spacer of a number of *rrn* promoters. As expected the reporter activities of the minimal promoters were unaffected by the reduced gene dosage, i.e. the β-gal activity of a given promoter construct was independent of the genetic background (data not shown). The two full-length constructs did, however, display interesting phenotypes. The β-gal activities of both *rrn* operons showed an almost two-fold increase in β-gal activity when assayed in SMR5 background compared to the wild type background (Figure 4.5). This was surprising since SMR5 is wild type regarding the *rrn* genes. Furthermore the two full-length *rrn* reporter constructs, i.e. the *M.tb rrnA* and the *M.sm rrnB* operons, display very different behaviour in the three mutant backgrounds (SMR5, *rrnA*, and *rrnB*'). The β-gal activity of the *rrnA* reporter construct was in the same range in the three backgrounds indicating an absence of feedback activation of the promoter in the
Figure 4.5: Activity of \textit{rrn} operons in \textit{rrn} knockout background

The two reporter constructs \textit{M.tb} \textit{rrnA} and \textit{M.sm} \textit{rrnB} were transformed into \textit{M.sm} mc²155, \textit{M.sm} SMR5 (str^5), \textit{M.sm} \textit{rrnA}^- and \textit{M.sm} \textit{rrnB}^- . The strains were assayed for \( \beta \)-gal activity and the different expression levels are depicted above. Expression of both \textit{rrn} reporter constructs is lower in wild type background than in SMR5 (compare blue bars with bordeaux bars). \textit{M.tb} \textit{rrnA} is unaffected by the reduced \textit{rrn} gene dosage (compare expression levels of \textit{M.tb} \textit{rrnA} in SMR5, \textit{rrnA}^- and \textit{rrnB}^- backgrounds). In \textit{rrnA}^- background(yellow bar) the \textit{M.sm} \textit{rrnB} reporter has an expression level which is twice as high as in SMR5 background and almost four times as high as in wild type background. In \textit{rrnB}^- background the same reporter has an expression level comparable to wild type background, i.e. lower than in the SMR5 background.
reporter construct. In contrast the \textit{rrnB} reporter displayed a two-fold increase in activity in the \textit{rrnA} background. However, the same reporter construct showed an approximately two-fold decrease in the \textit{rrnB} background. The latter finding was very surprising and presumably an artefact of the reporter system (see Discussion).

\subsection*{4.5 The \textit{rrn} operons: Discussion}

The expression of three mycobacterial \textit{rrn} operons, one of the \textit{rrnB} type and two of the \textit{rrnA} type, were investigated using reporter gene fusions. The promoter-reporter fusions were integrated into the \textit{M.sm} genome in single copy in a site-specific manner thereby ensuring a high degree of stability (Kumar \textit{et al}., 1998). Primer extension analysis demonstrated that the reporter gene did not introduce or alter the existing transcription start point of the cloned promoter fragments. Finally one of the \textit{M.tb} constructs (KAM58) was made so that it corresponded to a \textit{lacZ} reporter construct used by Verma \textit{et al}. A comparison between the reporter activities obtained in the present study and in the report by Tyagi’s group demonstrated that the obtained expression levels differed approximately three-fold. This is in excellent agreement with the fact that the construct used by Tyagi and co-workers was harboured on a replicating vector (copy number~3) (Verma \textit{et al}.)
Therefore the reporter system used in this study was considered reliable and reproducible.

The expression levels of the investigated *rrn* operons are all in the same range, i.e. between 1095 and 1548 units/mg protein in spite of having one, two, and three promoters respectively (Table 4.1). Similar *in vivo* expression levels between the two *M.sm* *rrn* operons have been reported previously as shown by primer extension analysis (Gonzalez-y-Merchand *et al*., 1998). The small difference between expression levels of the four reporter constructs could to some extent be explained by the different numbers of promoters in the three operons. However, this correlation is not strict since the lowest activity is seen in KAM60, i.e. a construct with two promoters.

In order to investigate the phenomenon of feedback regulation in mycobacteria two of the operons, *M.sm rrnB* and *M.tb rrnA*, were expressed in *rrn* knockout background. Curiously only the *rrnB* operon responded to the reduced gene dosage with a two-fold increase in activity in *rrnA* background. The two-fold decrease in the *rrnB"* background is considered to be an artefact since a down-regulation of *rrn* expression as a result of reduced *rrn* gene dosage does not make sense. It is possible that a mutation in *lacZ* affected the expression or the activity of β-galactosidase illustrating one of the drawbacks of reporter gene assays (see Chapter 9).
The conclusion is that the *M.sm* *rrnB* operon displays feedback activation of expression as is the case with the *E. coli* *rrn* operons (Condon *et al.*, 1993). The *M.tb* *rrnA* reporter construct indicated no feedback activation in either background.

Inactivation of the *M.sm* *rrnA* operon has a larger effect than inactivation of the *M.sm* *rrnB* operon judging by the growth of the knockout strains. This could probably be explained by the number of promoters in the two operons. Thus inactivation of the *M.sm* *rrnB* operon corresponds to the deletion of one promoter whereas inactivation of the *M.sm* *rrnA* operon removes three promoters of which one, the P2 promoter, is as active as the PB promoter during exponential growth (Gonzalez-y-Merchand *et al.*, 1998).
5 Core promoter activities (Results)

5.1 Introduction

A promoter can be divided into a core constituent and regulatory regions. The bacterial core promoter is usually defined as the sequences required for σ binding, i.e. the -10 and -35 regions, the spacer between these elements, and the region to the transcription start point (e.g. Record et al., 1996). Many studies on core promoters have focused on the two promoter elements because point mutations in these have a dramatic effect on promoter activity. However, the initial contacts between RNAP and the promoter involves a much larger region (Record et al., 1996). In the present study three rrn core promoters were cloned into the promoter probe vector with and without their native transcription start point in order to determine the effect of the region between the -10 box and the transcription start point and to compare core promoter activity. In addition to these core promoters a series of promoter elements spanning the -35 and -10 region, i.e. core promoters without the native transcription start points, were cloned in order to compare the strength of the -35/spacer/-10 regions. These promoters are referred to as minimal promoters. The minimal promoters contained CC di-nucleotide inserts between the promoter elements and the restriction site in order to separate these.
5.2 The effect of the native transcription start point

The mycobacterial core promoters *M.sm* P_B, *M.tb* P1, and *M.tb* PCL1 were cloned both with and without their native transcription start point. The transcription start points and the promoter elements have been identified by primer extensions and sequence comparison between a number of mycobacterial *rrn* operons (Gonzalez-y-Merchand *et al.*, 1996a; Gonzalez-y-Merchand *et al.*, 1997). The β-gal activities of reporter constructs with core promoters and minimal promoters were measured and the results are shown in Table 5.1:

<table>
<thead>
<tr>
<th>Promoter</th>
<th>minimal promoter</th>
<th>core promoter</th>
<th>% increase in activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M.tb</em> P1</td>
<td>13.2 ± 0.9</td>
<td>61 ± 6</td>
<td>369</td>
</tr>
<tr>
<td><em>M.tb</em> PCL1</td>
<td>195 ± 22</td>
<td>217 ± 17</td>
<td>11</td>
</tr>
<tr>
<td><em>M.sm</em> P_B</td>
<td>16.6 ± 3.0</td>
<td>34 ± 2</td>
<td>100</td>
</tr>
</tbody>
</table>

*β-gal activity is in units/mg protein*

The table indicates the activity of each of the two core promoters from the *M.tb* *rrnA* operon and the P_B promoter from the *M.sm* *rrnB* operon. The promoter-reporter constructs were assayed for β-gal activity with and without their native transcription start point (core promoter and minimal promoter respectively).

The main observation is that the β-gal activity varies substantially (approximately six-fold) between the three core promoters. The core promoters can therefore be ordered according to strength as follows:
PCL_{1b} > P1_{1b} > P_B

In each case the core promoter gave higher β-gal activities than the corresponding minimal promoter, indicating that the region between the −10 box and the native transcription start point had a positive effect on the activity of all three promoters. In the case of the PCL_{1b} promoter, the effect is negligible with only ~10% increase in activity, whereas the activity of the P_B promoter increased 100%. The largest effect was seen on the P1_{1b} promoter with an increase in activity of around 400% (Table 5.1). Thus, there is an inverse correlation between the activity of the minimal promoter and the effect of including the native transcription start point.

5.3 Comparison of minimal promoters

In order to compare the strength of different promoter hexamers and spacer regions, and to investigate the sequence requirements for the individual hexamers, a number of additional minimal promoters were cloned into the promoter probe vector and assayed for activity. The additional wild type minimal promoters were the PCL1 promoter from \textit{M. chelonae} (identical to the PCL1 promoter from \textit{M. abscessus}, PCL_{1ab(ch)}) and a consensus P2 promoter, P2_{con}, i.e. consensus hexamers and \textit{M.sm} spacer according to (Gonzalez-y-Merchand et al., 1997). The sequences of the −10 and −35 regions of the cloned minimal promoters and the β-gal activities of the constructs are shown in Table 5.2. From the derived β-gal
activities listed in this table the minimal promoters can be ordered according to strength as follows:

$$PCL_{lab/ch} > PCL_{tb} > P_{2sm} > P_B > P_{1tb}$$

with a thirty-fold difference in activity between $PCL_{lab/ch}$ and $P_{1tb}$. Thus, the minimal promoters shown in Table 5.2 can be divided into two groups: strong minimal promoters with activities in the range of hundreds, which include the two wild type $PCL1$ promoters and the $P2_{con}$ promoter, the latter being the weakest of these promoters. The remaining promoters with activities in the range of tens are defined as weak and only two promoters, $P_{1tb}$ and $P_B$, are found in this category.

Table 5.2: Activities of mycobacterial minimal promoters.

<table>
<thead>
<tr>
<th>Minimal promoter</th>
<th>-35</th>
<th>-10</th>
<th>β-gal activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus*</td>
<td>TTGACG/A</td>
<td>TATA/GA/CT</td>
<td>N.D.</td>
</tr>
<tr>
<td>$PCL1$ ($M.ch/ab$)</td>
<td>TTGACT</td>
<td>TAGACT</td>
<td>389 ± 22</td>
</tr>
<tr>
<td>$PCL1$ ($M.tb$)</td>
<td>TTGACT</td>
<td>TAGACT</td>
<td>195 ± 22</td>
</tr>
<tr>
<td>$P2$ (cons)</td>
<td>TTGACT</td>
<td>TATTCT</td>
<td>134 ± 10</td>
</tr>
<tr>
<td>$P_B$ ($M.sm$)</td>
<td>TTGACT</td>
<td>TAACTT</td>
<td>17 ± 3</td>
</tr>
<tr>
<td>$P1$ ($M.tb$)</td>
<td>GTCTCGGT</td>
<td>TATGCT</td>
<td>13 ± 1</td>
</tr>
</tbody>
</table>

* According to (Mulder et al., 1997); N.D. not determined

The cloned fragments contain the region from the 5' end of the -35 box to the 3' end of the -10 box. Sequences of each element are shown and according to (Gonzalez-y-Merchand et al., 1997). The β-gal activities (units/mg protein) of the reporter constructs are shown in the last column in each table.
5.3.1 Mutations in the *M.tb* PCL1 promoter

Of the core promoters in the two operons investigated in detail, *M.sm rrnB* and *M.tb rrnA*, the strongest was the *M.tb* PCL1 promoter. Therefore this promoter was chosen for a study of the effects of various mutations within the core promoter sequence. It should be noted that these mutated promoters did not include the native transcription start point and were therefore comparable to the PCL1 minimal promoter. First the boundaries of the −10 region were verified by introducing a T to C change in the first and last nucleotide of the presumed −10 hexamer (KAM38: T−7C and KAM39: T−12C). Two additional mutants were made each of which had one of the two promoter hexamers completely deleted (KAM40: Δ−10 and KAM41: Δ−35). The β-gal activities of the four mutant promoters are listed in Table 5.3. From the shown β-gal values it can be concluded that T(−7) and T(−12) are essential for promoter activity, and therefore the boundaries of the −10 hexamer are likely to be correctly assigned. Furthermore it appears that the sequence requirements for the −10 hexamer are significantly more strict than the requirements for the −35 hexamer since a deletion of the former abolishes activity, whereas deletion of the latter results in approximately 10% of the wild type activity (Table 5.3). This finding is substantiated by the observation that mycobacteria in general tolerate a wider variety of sequences in the −35 region (Bashyam *et al.*, 1996). However, a closer
look at the sequences of the mycobacterial rrn promoters studied here reveals a higher degree of homology in the -35 regions than in the -10 regions (apart from the P1 promoter, Table 5.2).

Table 5.3: Activities of mutant minimal promoter constructs.

<table>
<thead>
<tr>
<th>Promoter (MP)</th>
<th>-35</th>
<th>-10</th>
<th>β-gal activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCL1 (M.tb) wt</td>
<td>TTGACT</td>
<td>TAGACT</td>
<td>195 ± 22</td>
</tr>
<tr>
<td>PCL1 T-7C (M.tb)</td>
<td>TTGACT</td>
<td>TAGACC</td>
<td>1.5 ± 0.1</td>
</tr>
<tr>
<td>PCL1 T-12C (M.tb)</td>
<td>TTGACT</td>
<td>CAGACT</td>
<td>5.7 ± 0.8</td>
</tr>
<tr>
<td>PCL1 Δ -10 (M.tb)</td>
<td>TTGACT</td>
<td>(CCAAGC)</td>
<td>0</td>
</tr>
<tr>
<td>PCL1 Δ -35 (M.tb)</td>
<td>(TAGACC)</td>
<td>TAGACT</td>
<td>18 ± 1</td>
</tr>
<tr>
<td>PCL1 (M.ch/ab) wt</td>
<td>TTGACT</td>
<td>TGGTACAGT</td>
<td>389 ± 22</td>
</tr>
<tr>
<td>PCL1 (TG) (M.ch/ab)</td>
<td>TTGACT</td>
<td>cCGTACAGT</td>
<td>26 ± 2</td>
</tr>
</tbody>
</table>

A number of mutant minimal promoters (MP) have been compared to their wild type counterparts. The indicated sequences are according to (Gonzalez-y-Merchand et al., 1997) and mutations in the promoter elements are underlined. The β-gal activity of each construct is shown in the last column and is given in units/mg protein. The sequences in brackets in the deletion mutants are sequences conferred by the vector.

5.3.2 An extended -10 promoter

The strongest of the minimal promoters, that of the PCL1_{ab/ch} promoter contains a TGN motif located immediately upstream of the -10 box. This motif has been shown to enhance the activity of a number of promoters and it is abundant in Gram-positive organisms such as B. subtilis, C. pasteurianum, and mycobacteria.
(Helmann, 1995; Graves and Rabinowitz, 1986; Bashyam and Tyagi, 1998). In order to examine the effect of the TGN motif in this particular context, a mutant promoter with a TG to CC change was created and fused to the lacZ reporter gene. A comparison of the β-gal activities from wild type and mutant demonstrated a 15-fold reduction in promoter activity in the mutant (Table 5.3). Thus, this PCL1 promoter is the strongest of the assayed minimal promoters but only as long as the extended -10 region is intact. Deleting the TG motif reduces the activity of the promoter so that it becomes almost as weak as the M.tb P1 and M.sm P_B promoters (compare Tables 5.2 and 5.3).

5.4 Mycobacterial core promoters: Discussion

The mycobacterial rrn core promoter sequences differ in several positions but none of them confer to the consensus suggested by Mulder et al. (Gonzalez-y-Merchand et al., 1997; Mulder et al., 1997).

The experiments in the previous section demonstrated a number of features regarding mycobacterial rrn core promoters. Initially it was shown that the native transcription start point of a core promoter increases the expression compared to the minimal promoter and that there seems to be an inverse correlation between the strength of a minimal promoter and the advantage of including the native transcription start point. This correlation would, however have to be confirmed by
the investigation of more promoters in order to make a final conclusion. The –35 region of the \textit{M.tb} P1 promoter differs significantly from those of the other \textit{rrn} promoters (Table 5.2), and \textit{in vivo} this promoter is probably recognised by a different \(\sigma\) factor than the PCL1 and \(P_B\) promoters (Chen and Bishai, 1998; Gomez and Smith, 2000). Whether this is the reason that the activity of this particular promoter is so much more dependent on its native transcription start point has not been established. An alternative explanation could be that the DNA sequence \textit{per se}, i.e. the region between the –10 box and the transcription start point accounts for the difference. Reporter constructs with hybrid promoters, i.e. PCL1 minimal promoter and the P1 initiation region, should address this question. The \textit{rrn} promoters in this study (apart from the \textit{M.tb} P1 promoter) all share the same –35 region suggesting very stringent sequence requirement for this element. This seemed curious since a complete deletion of this region has a smaller effect on transcriptional activity than single point mutations in the –10 region (Table 5.3). However, a closer look at the sequences conferred by the vector and the CC inserts in the \(\Delta\) –35 mutant reveals that a new, albeit very poor, –35 region can be deduced (Table 5.3). In contrast the deletion of the –10 box resulted in a new –10 region with only a single consensus nucleotide, thus explaining the complete lack of activity in this construct (Table 5.3).
An extended –10 region was identified in the PCL1 promoters of *M.ab* and *M.ch*. This TG motif at positions –15/-14 increased promoter activity approximately 15-fold. Furthermore it was demonstrated that the mutant promoter without the TGN motif was relatively weak. Previously the extended –10 region has been shown to increase activity in *M.sm* between 4 and 54 fold depending on promoter context (Bashyam and Tyagi, 1998). Interestingly *M.ch* and *M.ab* are the only mycobacteria that contain *rrn* promoters with the TGN motif (Gonzalez-y-Merchand *et al.*, 1997). In comparison, the majority of the *B. subtilis rRN* promoters contain the TG motif (Helmann, 1995). *M.ch* and *M.ab* have been categorised as fast growing species with a single *rrn* operon in their genome (Gonzalez-y-Merchand *et al.*, 1997). This operon in turn contains a relatively high number (5) of promoters (Gonzalez-y-Merchand *et al.*, 1997). Comparing the *M.ch/ab* PCL1 promoter with the *M.tb* PCL1 promoter reveals that the –10 region of the *M.tb* promoter is closer to the proposed consensus sequence for this particular element (Mulder *et al.*, 1997). This in turn suggests that the *K_B* of the *M.tb* promoter is likely to be higher than the *K_B* of the *M.ch/ab* PCL1 promoter (Ellinger *et al.*, 1994a). In *E. coli* promoters the extended –10 region increases the rate of open complex formation, *k_f* (Burr *et al.*, 2000). It seems plausible that an increase in *k_f* rather than *K_B* would be an advantage in the context of numerous promoters in order to avoid promoter occlusion.
The comparison between different minimal promoters and core promoters demonstrates large variations in activities, sequence requirements, and strategies for increasing promoter activity. Thus, some promoters obtain their strength from the hexamers, while others obtain their strength from sequences outside the hexamers, i.e. the regions flanking the –10 box, indicating that promoter strength is achieved by increasing different parameters in the initiation pathway.
6 Regulatory elements of \textit{rrn} operons (Results)

As expected the core promoter activities could not account for the observed activity of the full \textit{rrn} constructs (compare Tables 4.1 and 5.1). Furthermore, the \( P_B \) core promoter was by far the weakest of the assayed core promoters although this promoter is the only promoter identified in and therefore solely responsible for, the expression of the \( M.sm \) \textit{rrnB} operon. The activity of the full \textit{rrnB} operon construct was, however, comparable to the activities of both \( M.sm \) \textit{rrnA} and \( M.tb \) \textit{rrnA} operon constructs, which harbour at least one strong core promoter each. These findings indicate the presence of additional activating elements in the \( M.tb \) \textit{rrnA} and \( M.sm \) \textit{rrnB} operons, in particular at least one highly activating element in the \( M.sm \) \textit{rrnB} operon. In order to identify the regulatory elements of the \( M.tb \) \textit{rrnA} and the \( M.sm \) \textit{rrnB} promoter regions, the promoter/leader regions were dissected into constructs harbouring a single promoter plus upstream and/or downstream regions. These fragments were fused to the \textit{lacZ} reporter and subsequently assayed for \( \beta \)-gal activity.

6.1 The \textit{M.tb rrnA} operon: Introduction

The structure of the \( M.tb \) \textit{rrnA} operon and the reporter constructs derived from this operon are outlined in Figure 6.1. Previous investigations on the expression of
Figure 6.1: The \textit{M.tbc} \textit{rrnA} reporter constructs.

The figure illustrates, schematically, the reporter constructs derived from the \textit{M.tbc} \textit{rrnA} operon. The \textit{rrn} sequences are indicated in green with the two promoters, P1 and PCL1, as green boxes with the transcription start point as asterisks. Anti-termination motifs are shown as yellow boxes. The sizes of the upstream and downstream flanking regions and the spacing between the two minimal promoters are indicated. \(\beta\)-gal activities of the longer reporter constructs are shown in the column on the right.
the *M.tb rRNA* operon have demonstrated that the PCL1 promoter accounts for the majority of rRNA transcription during exponential growth (Gonzalez-y-Merchand *et al.*, 1997, Gonzalez-y-Merchand, 1998). The β-gal activities of the two core promoter constructs, P1 and PCL1, have already been described in section 5.2. The full-length promoter/leader construct (harbouring both of these promoters) extending 300 basepairs upstream of the P1 transcription start point (strain KAM60) has an activity of 1095 units/mg protein, which is four times higher than the combined activity of the P1 and PCL1 promoters. In order to evaluate the contribution of each of the promoters including their flanking regions, the *M.tb rRNA* promoter/leader region was dissected into P1 fragments and PCL1 fragments. The individual reporter constructs have been outlined in Figure 6.1, whereas Table 6.1 indicates the exact size of each fragment along with its β-gal activity.

Due to the short distance between the two promoters, P1 constructs extend 41 basepairs downstream of the transcription start point in order not to include core promoter sequences from PCL1. Conversely, PCL1 constructs have 48 basepairs upstream of the −35 region (up to but not including the −10 box of the P1 promoter). All the constructs were transformed into *M.sm* and subsequently assayed for β-gal activity.
Table 6.1: The *M.tb* *rrnA* reporter constructs.

<table>
<thead>
<tr>
<th>Strain (KAM number)</th>
<th>Promoter fragment</th>
<th>β-gal activity (units/mg protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td><em>M.tb rrnA</em></td>
<td>1095 ±1 38</td>
</tr>
<tr>
<td>49</td>
<td>P1 (-300 to +41)</td>
<td>4388 ± 300</td>
</tr>
<tr>
<td>50</td>
<td>P1 (-300 to -8)</td>
<td>2438 ± 475</td>
</tr>
<tr>
<td>51</td>
<td>P1 (-37 to +41)</td>
<td>74 ± 10</td>
</tr>
<tr>
<td>35</td>
<td>P1 (-37 to -8 with CC)</td>
<td>13.2 ± 0.9</td>
</tr>
<tr>
<td>88</td>
<td>P1 (-37 to +1)</td>
<td>61 ± 6</td>
</tr>
<tr>
<td>53</td>
<td>PCL1 (-84 to -7)</td>
<td>448 ± 21</td>
</tr>
<tr>
<td>61</td>
<td>PCL1 (-84 to +219)</td>
<td>174 ± 11</td>
</tr>
<tr>
<td>62</td>
<td>PCL1 (-36 to +219)</td>
<td>40 ± 9</td>
</tr>
<tr>
<td>34</td>
<td>PCL1 (-36 to -7 with CC)</td>
<td>195 ± 22</td>
</tr>
<tr>
<td>89</td>
<td>PCL1 (-36 to +1)</td>
<td>217 ± 17</td>
</tr>
</tbody>
</table>

* See table 4.1 for details

The table lists all the reporter constructs of the *M.tb* *rrnA* operon used in this study. Strain numbers are indicated on the left and the boundaries of each fragment, relative to the transcription start point of the indicated promoter, are shown in the middle column. The β-gal activities of individual P1 and PCL1 promoter fragments are shown in Figures 6.2 and 6.3.

### 6.2 The P1 promoter

None of the reporter constructs driven exclusively by the P1 promoter of the *M.tb* *rrnA* operon were expressed in *E. coli* (data not shown). This suggests that the P1 promoter in *M.sm* is recognised by a σ factor that does not have a homologue in *E. coli*. This observation is in agreement with a previous report stating that the
*M. bovis* (BCG) *rrn* P1 is not recognised in *E. coli* (Suzuki *et al.*, 1991). This in turn supports the notion that the two promoters of the *M.tb rrnA* operon are recognised by different σ factors in *M.sm*.

The activities of the P1 reporter constructs expressed in *M.sm* vary more than 300-fold from 13 units/mg protein to 4388 units/mg protein (Table 6.1 and Figure 6.2). Starting with the minimal promoter (−37 to −8) and gradually extending the P1 downstream region increases activity moderately (Figure 6.2). The minimal promoter (−37 to −8) has an activity of 13 units/mg protein, the core promoter (−37 to +1) has an activity of 61 units/mg protein, and KAM 51 (−37 to +41) has an activity of 74 units/mg protein. Thus, the region between −8 and +1 increases activity five-fold whereas the region from +1 to +41 confers less than a two-fold increase in reporter activity. In other words it appears that in the context of the P1 promoter without upstream sequences, the region between the −10 box and the transcription start point is more important for promoter activity than the region between +1 and +41. The total activating potential of the region between −8 and +41 is approximately six-fold in the absence of the upstream region (from 13 units/mg protein to 74 units/mg protein).

The activating potential of the P1 downstream region is slightly different when the 262 basepairs upstream of the P1 core promoter are included in the constructs. When the P1 upstream region is present the downstream region confers only two-fold increase in activity (compare KAM49 and KAM50 in Figure 6.2).
Figure 6.2: Activities of the *M. tb* P1 reporter constructs.
The figure illustrates the regions covered by the different reporter constructs with the corresponding strain number indicated above each construct. The constructs and their activities have been compared to KAM60 comprising the entire promoter/leader region of the *M. tb* *rrnA* operon. On the right side of the figure is a diagram showing the individual β-gal activities of the reporter constructs. Constructs without the P1 upstream region have very low activity compared to constructs including the upstream region. The bottom panel illustrates the activation conferred by the upstream and downstream regions and their influence on each other, respectively.
The β-gal activities of these two constructs are 4388 and 2438 units/mg protein, which is higher than any of the remaining P1 constructs and even higher than the full length M.tb rrnA reporter construct (Figure 6.2). Thus, the region downstream of the −10 box of the M.tb P1 promoter has a small enhancing effect on the activity of the promoter.

The region upstream of the P1 promoter enhances lacZ expression. The highest activity (4388 units/mg protein) is found in the construct that covers the largest region, namely −300 to +41 relative to the P1 transcription start point. When the upstream region is deleted while maintaining the same 3’end (+41, KAM51) the activity is reduced 60-fold. Constructs with the 3’end at −8 but with or without the upstream region display an almost 200-fold difference in activity, i.e. 2438 units/mg protein compared to 13 units/mg protein (Table 6.1 and Figure 6.2). Thus it is evident that a highly activating region resides upstream of the P1 core promoter, whereas a weakly activating region is found downstream of P1. The bottom panel in Figure 6.2 illustrates how the activating potential of the upstream and downstream regions influence each other such that the presence of the upstream region reduces the activation potential of the downstream region and vice versa.
6.3 The PCL1 promoter

In contrast to the observations made with the P1 promoter, all of the PCL1 derivatives display lower β-gal activity than the full-length M.tb rrnA construct. This is somewhat surprising given the relative strength of the core promoters and minimal promoters as well as the relative contribution of the two promoters in vivo (Gonzalez-y-Merchand et al., 1997; Gonzalez-y-Merchand et al., 1998). In all these cases the PCL1 promoter appears to be the stronger of the two.

The PCL1 constructs shown in Figure 6.3 display activities between 40 and 512 units/mg protein. In this case the highest activity is not found in the construct covering the largest region but in the construct that does not contain any sequences downstream of the -10 box, i.e. KAM53 which spans -84 to -7 relative to transcription start point (Figure 6.3). The constructs used for investigating the effect of the (small) upstream region are KAM34/ KAM53 and KAM62/KAM61 (Table 6.1 and Figure 6.3). The KAM34 and KAM53 constructs both have their 3’ end at -7 and the 5’ end at -36 and -84, respectively. The effect of the 47 upstream basepairs is small with less than three-fold difference in activity between the two constructs (Table 6.1). The KAM62 and KAM61 reporter constructs have the 3’ end at +219 and the same 5’ ends as KAM34 and KAM53, respectively. In these two strains the PCL1 upstream region enhances promoter activity little more than four-fold (Table 6.1), which is not significantly different
Figure 6.3: Activities of the *M.tb* PCL1 reporter constructs.

The figure outlines the various PCL1 reporter constructs, their boundaries and the β-gal activity obtained with each. Numbers above the constructs indicate strain numbers and exact β-gal values can be found in Table 6.1. The construct covering the entire *rrnA* promoter and leader region (KAM60) has been included for comparison. The figure illustrates how all of PCL1 reporter constructs display a lower activities than the full-length construct.
from the increase observed with KAM34 and KAM53. The weak activation potential of the *M.tb* PCL1 upstream region may be due to the limited size of this fragment compared to the much larger upstream region in the P1 reporter constructs.

The difference in activity between KAM53 and KAM61 suggests that the region between $-7$ and $+219$ suppresses rather than increases promoter activity. This notion is sustained by comparing KAM34 ($-36$ to $-7$) with KAM62 ($-36$ to $+219$). Between these two constructs the presence of the downstream region reduces the activity approximately five-fold (Table 6.1). Since it has already been established that the region between $-7$ and $+1$ increases promoter activity, it is evident that the region downstream of the PCL1 transcription start point, i.e. the *M.tb* *rrn* leader region suppresses *lacZ* expression under the growth conditions used in this study. One explanation for this could be the presence of one or more pause sites within this region. A more detailed investigation of the *M.tb* *rrn* leader region is presented in Chapter 7.

### 6.4 Regulatory regions of the *rrnA* operon: Discussion

In conclusion the *M.tb* *rrnA* operon contains a highly activating region upstream of its two promoters, P1 and PCL1. This region is capable of activating the proximal P1 promoter almost 200-fold and the activating potential is thus
comparable to the *E. coli* rrn upstream activating region (UAR) (Condon et al., 1995b; Gourse et al., 1996). The region between the two promoters, i.e. upstream of the PCL1 promoter harbours a limited activating potential possibly because of its restricted size in the constructs used in this study. Reporter constructs in which a larger upstream region is included and where the P1 promoter has been knocked out by single point mutations may result in a different picture of the activation of the PCL1 promoter. The sequence of the *M.tb* UAR, i.e. upstream of the P1 promoter contains two perfect palindromic sequences (−69 to −60 and −46 to −40) relative to P1 transcription start point and apparently located on the same face of the DNA helix as the P1 −35 region (Figure 6.4). One of these is also found upstream of the PCL1 promoter, although in this case it appears a little out of phase with the −35 region (−44 to −38 relative to the PCL1 transcription start point, Figure 6.4). These sequences could in theory act as binding sites for transcriptional activators but require further investigations.

The region downstream of the P1 promoter has a weakly activating effect on the P1 promoter (between two- and six-fold), the activation being weakest when the UAR is also present. Conversely the UAR loses some of its activating potential when the P1 downstream region is included in the construct. Together these findings suggest that the two regions/elements activate the same step in the transcription initiation pathway and that a saturating level of activation has been obtained in KAM49 (4388 units/mg protein).
Figure 6.4: Core promoter regions of *M. tb* P1, *M. tb* PCL1, and *M. sm* P_B.

The core promoter elements are underlined and bold, and the transcription start point is indicated by asterisk. These regions contain two different palindromic sequences, I and II (boxed). Palindrome II is present upstream of all three promoters with a single nucleotide change in the *M. sm* rrnB promoter region, indicated in red. In the *M. tb* rrnA operon all three palindromes are perfect. Palindrome I was only found upstream of P1. Orange arrows indicate approximate integral helical turns. The significance of the sequences is unknown and they are not found in all mycobacterial rrn promoters (Gonzalez-y-Merchand et al., 1997). An AT-rich sequence in the rrnB upstream region is indicated in green.
6.5 The *M.sm* *rrnB* operon: Introduction

Unlike the *rrnA* operon the *rrnB* operon is not found in all mycobacterial species and it invariantly contains a single promoter (Gonzalez-y-Merchant *et al.*, 1997). The PB promoter in its native context displays large variations in activity under different growth conditions and in different stages of growth (Gonzalez-y-Merchant *et al.*, 1998). However, the elements involved in these variations are not known. A detailed investigation of the flanking regions of a promoter is facilitated by the absence of other promoters whose elements could otherwise interfere with the elements of the promoter being studied. As the PB promoter is the sole promoter of the *rrnB* operon it was chosen for such an investigation. The minimal PB promoter was constructed with and without the flanking CC di-nucleotides; β-gal activity increased from 17 to 24 units/mg protein in the absence of CC; i.e. a 40% increase in activity (compare KAM37 with KAM71, Table 6.2). The minimal promoter used for comparison in this part of the study is KAM71 (without the CC inserts). An initial dissection of this operon resulted in constructs spanning −140 to +207 (full-length, KAM55), −140 to +10 (KAM56), and −40 to +207 (KAM57), shown in Figure 6.5. The β-gal activities of these constructs including the minimal promoter and the core promoter range from 15 units/mg
Figure 6.5: The initial dissection of the *M. sm* *rrnB* operon.
The figure shows an outline of the initial *M. sm* *rrnB* reporter constructs. The *rrnB* promoter/leader region is shown in green, fused to the *lacZ* reporter (blue). Relevant elements are: promoter (green), transcription start point (asterisk), anti-termination motifs, BoxA and BoxC (yellow). The arrows above the *rrnB* region indicate the sizes of the flanking regions (Table 4.1). The extents of the constructs are indicated and numbers above each construct refer to strain numbers. β-gal activity of each construct is shown on the right.
<table>
<thead>
<tr>
<th>Strain (KAM number)</th>
<th>Promoter fragment</th>
<th>β-gal activity (units/mg protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td><em>M. sm rrnB</em> (-140 to +207)</td>
<td>1106 ± 17</td>
</tr>
<tr>
<td>56</td>
<td>-140 to +10</td>
<td>24025 ± 3142</td>
</tr>
<tr>
<td>57</td>
<td>-40 to +207</td>
<td>15 ± 3</td>
</tr>
<tr>
<td>72</td>
<td>-39 to -8</td>
<td>35 ± 3</td>
</tr>
<tr>
<td>73</td>
<td>-53 to -8</td>
<td>673 ± 112</td>
</tr>
<tr>
<td>77</td>
<td>-41 to -8</td>
<td>52 ± 13</td>
</tr>
<tr>
<td>78</td>
<td>-52 to -8</td>
<td>668 ± 132</td>
</tr>
<tr>
<td>79</td>
<td>-60 to -8</td>
<td>993 ± 96</td>
</tr>
<tr>
<td>80</td>
<td>-70 to -8</td>
<td>2211 ± 142</td>
</tr>
<tr>
<td>81</td>
<td>-80 to -8</td>
<td>1690 ± 242</td>
</tr>
<tr>
<td>84</td>
<td>-36 to +10</td>
<td>62 ± 17</td>
</tr>
<tr>
<td>85</td>
<td>-80 to +10</td>
<td>10999 ± 931</td>
</tr>
<tr>
<td>87</td>
<td>-140 to -8</td>
<td>7636 ± 228</td>
</tr>
<tr>
<td>37</td>
<td>-36 to -8 with flanking CC</td>
<td>17 ± 3</td>
</tr>
<tr>
<td>71</td>
<td>-36 to -8 w/o flanking CC</td>
<td>24 ± 1</td>
</tr>
<tr>
<td>90</td>
<td>-36 to +1</td>
<td>34 ± 2</td>
</tr>
</tbody>
</table>

The table indicates the strain numbers of all the *M. sm rrnB* reporter constructs, and the cloned promoter fragments with their boundaries relative to the transcription start point. The last column shows the β-gal activities of the constructs. Activities of individual constructs have been compared in figures 6.5 to 6.7.
protein to 24000 units/mg protein, i.e. a 1600 fold difference in activity (Figure 6.5). These results indicate the presence of a highly activating region upstream of the core promoter and a suppressing effect of the region downstream of the promoter. The upstream and downstream flanking regions of the $P_B$ promoter were investigated in more detail in order to determine the contributions of each flanking region. All the $M.sm$ $rrnB$ reporter constructs, their boundaries, and their $\beta$-gal activities are listed in Table 6.2.

6.6 The downstream and leader regions of the $rrnB$ operon

The region downstream of $-8$ relative to the transcription start point of the $P_B$ promoter have two opposite effects. A comparison between minimal promoter ($-36$ to $-8$), core promoter ($-36$ to $+1$) and KAM84 ($-36$ to $+10$) demonstrates that the region between $-8$ and $+10$ confers a gradual increase in promoter activity from 24 to 62 units/mg protein (i.e. a three-fold increase, Figure 6.6). However, the region further downstream, from $+10$ to $+207$, suppresses the activity four-fold, (compare KAM84 and KAM57 Table 6.2), assuming the nucleotides from $-36$ to $-40$ do not suppress activity (which is confirmed by reporter strains KAM71, 72, and 77, Table 6.2). Thus, the region downstream of $+10$, i.e. most of the $rrnB$ leader region, confers a reduction in reporter activity (Figure 6.6).
Figure 6.6: The regulatory effects of the *M. sm rrnB* downstream region.
The diagram shows how the β-gal activity increases when the 3'end of the *rrn* fragment is extended from -8 to +10 in the reporter constructs. A further extension from +10 to +207 reduces the activity. Individual strain numbers are indicated below each bar and the 3' end of each strain pair is indicated below each pair. The increase in activity conferred by the region between -8 and +10 is the same regardless of the presence or absence of the *rrnB* upstream region (without upstream region: blue bars and numbers; with upstream region: bordeaux bars and numbers). The suppressing effect conferred by the region downstream of +10 increases with increasing promoter activity (compare the difference between KAM84 and KAM57 with the difference between KAM56 and KAM55).
A comparison of reporter constructs with the same 3' ends as above (-8, +10, and +207) but including the entire \( rrnB \) upstream region to -140 demonstrate a similar effect, although quantitatively there is a difference illustrated in Figure 6.6. When the \( rrnB \) upstream region is included, the region between -8 and +10 increases promoter activity approximately three-fold (compare KAM87 with KAM56), which is the same as the increase observed in constructs without the \( rrnB \) upstream region. However, the difference in activity conferred by the region between +10 and +207 is a 22-fold reduction when the constructs included the \( rrnB \) upstream region (compare KAM56, and KAM55) as opposed to the four-fold reduction in constructs without the upstream region. Figure 6.6 demonstrates how each segment (-8 to +10 and +10 to +207) confers similar effects in the presence and absence of the \( rrnB \) upstream region. From this diagram it becomes evident that within each fragment the relative effects are almost the same in spite of the roughly 300-fold difference in absolute reporter activity. Thus, the \( M.sm \) \( rrnB \) leader region (i.e. downstream of +10) harbours an element that suppresses the expression of the \( lacZ \) reporter. Furthermore this suppression increases with increasing promoter activity. Since the region conferring the reduced activity is located downstream of +10 the most likely explanation is that a post-initiation event is responsible for the reduced expression. In other words, the \( rrnB \) leader region possibly contains one or more pause sites analogous to the \( E. coli \) \( rrn \)
leader. Alternatively the region between +10 and +207 confers a destabilisation of the transcript or of the elongation complex.

6.7 The \textit{rrnB} upstream activating region

The initial dissection of the \textit{rrnB} operon (Figure 6.5) suggests the presence of a highly activating region upstream of the P_B promoter (–140 to –36 relative to the transcription start point of the P_B promoter). Furthermore all reporter constructs devoid of upstream region (apart from the five nucleotides from –37 to –41) display very low β-gal activities (between 15 and 62 units/mg protein) regardless of the size of the downstream region. These include the minimal promoter (KAM71), the core promoter (KAM90), KAM84 (–36 to +10), KAM72 (–39 to –8), KAM77 (–41 to –8), and KAM57 (–40 to +207). Of these constructs KAM84 (–36 to +10) has the highest activity with 62 units/mg protein (Table 6.2). In order to rule out the possibility of cryptic promoters being present in the \textit{rrnB} upstream region, the fragment (–140 to –37) was cloned into the promoter probe vector in both orientations. Subsequent β-gal assays revealed that the \textit{M.sm} \textit{rrnB} upstream region has no promoter activity on its own (data not shown). This implies that the \textit{rrnB} upstream region only has a function in the presence of the already identified P_B core promoter. Thus, the \textit{rrnB} upstream region harbours at least one highly
activating element. A number of constructs in which the 3' end was maintained at -8 while the 5' end was gradually extended were made. Figure 6.7 compares the region that each reporter construct covers with the obtained β-gal activity, and from these results it becomes evident that the activity increases in parallel to the increase in upstream DNA. The first significant increase in activity is observed when the 5' end is extended from -41 to -52 (compare KAM77 KAM78, Figure 6.7). This 11-basepair fragment activates transcription approximately 15-fold. The regions from -52 to -60 and from -60 to -70 each result in less than two-fold activation but in combination the region from -36 to -70 enhances activity almost 100-fold (compare KAM71 and KAM80, Figure 6.7). There is no increase in activity (perhaps even a small decrease) when the construct is extended from -70 to -80 suggesting that the 5' boundary of an element is located before, or at -70. However, the region from -80 to -140 enhances the activity another three-fold. The total increase in activity obtained with the rrnB upstream region (-140 to -37), when the 3' end was at -8 is thus 318-fold (compare minimal promoter activity with KAM87, Table 6.2).

Except for an overall increase in activity, the picture remains very much the same when the 3' end of the reporter constructs are extended from -8 to +10. KAM84 (-36 to +10) has an activity of 62 units/mg protein but when the upstream region up to -80 is included the activity increases to 10999 units/mg protein, i.e. ~180-fold. The increase in activity conferred by this upstream segment in constructs
Figure 6.7: The upstream activating region of the \textit{M.sm} rrnB operon.

The left side of the figure outlines the boundaries of a number of reporter constructs which all had the 3' end at -8 beginning with the minimal promoter (KAM71). The numbers above each construct refer to strain number and the exact boundaries and β-gal values of each construct can be found in Table 6.2. The figure illustrates how a gradual extension of the 5' end of the \textit{rrn} fragment results in a gradual increase in reporter activity shown as bars on the right side of the figure. The numbers above the constructs refer to the strain number. The total activation from minimal promoter to KAM87 is 318-fold.
ending at -8 is 70-fold in comparison (KAM71 and KAM81). The region from -80 to -140 increases activity two-fold increase compared to three-fold above. The total activation conferred by the region from -140 to -36 when the 3' end is at +10 is 388-fold, which is in the same range as the 318-fold increase noted above. Finally, when the downstream region is extended to +207, the activation potential of the rrnB upstream region seems to diminish. The difference in β-gal activity between KAM55 and KAM57 is only 73-fold compared to the more than 300-fold in the other two contexts. The effect of the rrnB upstream region in the context of different 3' ends has been illustrated in Table 6.3:

Table 6.3: The effects of the rrnB upstream and downstream regions.

<table>
<thead>
<tr>
<th>5'end</th>
<th>3'end</th>
<th>-8</th>
<th>+10</th>
<th>+207</th>
<th>Suppression (+10 to +207)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-36</td>
<td>24 ± 1</td>
<td>62 ± 17</td>
<td>15 ± 3*</td>
<td>4-fold</td>
</tr>
<tr>
<td></td>
<td>-80</td>
<td>1690 ± 242</td>
<td>10999 ± 931</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>-140</td>
<td>7636 ± 228</td>
<td>24025 ± 3142</td>
<td>1106 ± 17</td>
<td>22-fold</td>
</tr>
<tr>
<td>Activation (-140 to -36)</td>
<td>318-fold</td>
<td>388-fold</td>
<td>73-fold</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*) The 3' end of KAM57 is located at -40; N.D. Not determined; activities are units/mg protein

The table illustrates how the rrnB upstream and downstream regions affect lacZ expression and how the effect of each region depends on the presence of the other region. The β-gal activities of various reporter constructs, with 5' and 3' ends as indicated, were compared in order to gain information about the interactions between the rrnB upstream and downstream regions. From these results it can be observed that the rrnB upstream region looses some of its activating potential when the 3' end is extended from +10 to +207. Alternatively, the suppressing effect of the rrnB downstream region (+10 to +207) increases with increasing promoter activity.
Thus there appears to be at least two discrete elements in the \textit{rrnB} upstream region; one located between $-140$ and $-80$ and the other between $-70$ and $-41$. Together they are capable of activating more than 300-fold depending on the context. It is possible that both regions ($-140$ to $-80$ and $-70$ to $-40$) consist of more than one element.

\textbf{6.7.1 Phase of the helix dependent activation}

In the majority of situations, activated transcription involves the C-terminal domain of the RNAP\(\alpha\) subunit (\(\alpha\)CTD). Since this domain is connected to the \(\alpha\)NTD via a highly flexible interdomain linker, the \(\alpha\)CTD can move as much as 44Å away from the rest of the RNAP molecule (Blatter \textit{et al.}, 1994; Zhou \textit{et al.}, 1994). This in turn means that an activating element, whether it binds a transcription factor or interacts with \(\alpha\)CTD directly, can be shifted further upstream relative to the core promoter, and still retain most of its activity as long as it remains on the same face of the DNA helix. If, on the other hand, the element is shifted away from the initial face of the helix, activity will be abolished. The shifts can be accomplished by inserting nucleotides between the core promoter and the upstream region.

To further the understanding of the mechanism(s) behind the upstream activation of the \textit{rrnB} promoter, a phasing experiment was designed. The strategy was to
separate the \textit{rrnB} core promoter from the upstream region with an increasing number of nucleotides. The insertions were made between $-37$ and $-36$ relative to the transcription start point, and the resulting mutant promoters were fused to the \textit{lacZ} reporter (see Materials and Methods for details). The inserts correspond to a half, three quarters, a full, and little more than a full turn of the DNA helix (6, 8, 11 and 13 basepairs, respectively). The resulting reporter strains were measured for $\beta$-gal activity and the results are shown in Table 6.4.

Inserting half or three quarters of a helical turn, i.e. six or eight basepairs, respectively, virtually abolishes the activation from the \textit{rrnB} upstream region, the eight-basepair insert having the largest effect (less than 0.5\% of wild type). The activity is, however, partially restored by inserting another half helical turn resulting in a full or little more than a full helical turn relative to wild type. Thus, an 11-basepair insert results in 38\% of wild type activity whereas a 13-basepair insert results in as much as 74\% of wild type activity (Table 6.4). These results suggest that the activation of the \textit{rrnB} core promoter by the upstream element depends on the spatial arrangement of the two regions rather than proximity. Curiously, the promoter activity is closest to wild type activity with a 13 basepair insert (one turn + 2 basepairs), and conversely lowest with the 8 basepair insert (half turn + 2 basepairs). These values suggest that there may be a slight twist between the positions of the promoter elements and the activating element. This led me to speculate whether the spacing between the core promoter and the
Table 6.4: Phase of the helix dependent activation of transcription.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Insert (bp)</th>
<th>Helical turns</th>
<th>Activity (% of wt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>87</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>74</td>
<td>6</td>
<td>1/2</td>
<td>1.2</td>
</tr>
<tr>
<td>70</td>
<td>8</td>
<td>3/4</td>
<td>0.4</td>
</tr>
<tr>
<td>76</td>
<td>11</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>75</td>
<td>13</td>
<td>&gt;1</td>
<td>74</td>
</tr>
<tr>
<td>83</td>
<td>1</td>
<td>---</td>
<td>23*</td>
</tr>
<tr>
<td>82</td>
<td>2</td>
<td>---</td>
<td>14*</td>
</tr>
</tbody>
</table>

*The values for these two strains are relative to KAM56 (see text)

The PB core promoter and the rrnB upstream region were separated by the indicated inserts in basepairs (bp). β-gal activities of the resulting reporter strains was determined and compared to the corresponding construct(s) without insert.

Figure 6.8: Phase of the helix dependent activation of transcription.

The diagram is a more visual presentation of the numbers listed in Table 6.4 above. Both illustrate how reporter activity decreases as the rrnB UAR is twisted to the opposite face of the DNA helix whereas activity is partially restored by shifting the UAR back to the initial face of the DNA helix. The values for the 1 and 2 bp inserts are relative to a different strain than the remaining constructs (see text).
activating region is sub-optimal in the wild type construct, and hence if the insertion of one or two basepairs would increase activity relative to wild type. Two additional constructs with one and two basepair inserts were made and the β-gal values obtained with the resulting reporter strains are listed in Table 6.4. These constructs differ in their 3’end from the previous ones in that they end at +10 rather than –8. However, the values in Table 6.4 are relative to the reporter constructs without inserts and KAM82 and KAM83 have thus been compared to KAM56. From these results it is evident that separating the core promoter and the UAR with one or two basepairs does not increase the activity. In fact the promoter activity decreases significantly by these insertions. The compiled results of these experiments are summarised in Figure 6.8, which clearly illustrates the changes in promoter activity as the size of the insert increases. This phasic behaviour of promoter activation strongly suggests that the αCTD is involved, either directly or indirectly, in the activation of the PB promoter: directly by interacting with a DNA element, indirectly via protein-protein interactions with a transcription factor which in turn binds to the DNA.

6.7.2 Interaction between the rrnB UAR and the RNA polymerase α subunit

The previous sections demonstrated that the rrnB upstream region harbours at least one highly activating element and furthermore that this element displays
phase of the helix dependency. Therefore the DNA-protein interactions in this region were investigated. A probe corresponding to the \textit{rrnB} upstream region was made and used in electrophoretic mobility shift assays with cell extracts from \textit{M. sm}. The \textit{rrnB} operon displays a dramatic increase in activity when a stationary phase culture is diluted into fresh medium, i.e. a shift from stationary phase to exponential growth (Gonzalez-y-Merchand \textit{et al.}, 1998). In case a putative transcription factor displays variations in expression or activity during different stages of growth, such as the \textit{E. coli} Fis protein and in parallel with the variations in \textit{rrnB} activity, cell extracts were harvested at different stages of growth (four hours after dilution into fresh medium and in balanced growth). Concentrations up to 2.0 $\mu$g/$\mu$l total protein were used and the reactions were subjected to electrophoresis on an 8\% native acrylamide gel, but no retardation of the probe could be observed (data not shown). The apparent lack of interaction between the probe and a putative transcription factor can be attributed to a number of reasons: the binding affinity of the factor is extremely low, the factor is present in very low concentrations in the cell or the activation is factor independent. In factor independent activation the RNAP$\alpha$ subunit interacts directly with the \textit{rrnB} upstream region in a sequence specific manner (reviewed by Gourse \textit{et al.}, 2000).

Apart from a short AT-stretch, which is not in phase with the promoter elements (–60 to –55, Figure 6.4), the \textit{rrnB} upstream region does not harbour sequences that resemble the well-known structure of the \textit{E. coli} or \textit{B. subtilis} UP elements.
However, a recent report describes an UP element with a very different structure/sequence identified in the GC-rich organism *T. termophilus* (Wada *et al.*, 2000). A sequence alignment of five α subunits from *M.tb*, *S. coelicolor*, *B. subtilis*, *E. coli* and *T. thermophilus* is shown in Figure 1.4. The alignment reveals a remarkable conservation of the seven amino acids essential for UP element function in *E. coli*. In *T. thermophilus* three of these have been substituted but nevertheless a mycobacterial UP element with a different composition seems a possibility (Murakami *et al.*, 1996). Since α-DNA interactions are relatively weak, (e.g. Ross *et al.*, 1998) it was assumed that this interaction would not be sufficient for the entire polymerase to bind to the probe, which does not include the core promoter elements. Therefore it was decided to investigate whether RNAPα as a free subunit, would be able to interact directly with the *rrnB* upstream region.

Assuming that the RNAP subunits from *M.sm* and *M.tb* display very high sequence homology and since sequence information on the *M.sm* protein was not available at the time, the gene encoding the *M.tb* RNAP α subunit (*M.tb* *rpoA*) was PCR-amplified from genomic DNA and cloned into an *E. coli* pET15b expression vector, described in chapter 8 (the sequence for the *M.sm* RNAPα has now become available and the two subunits display 95% homology, Table 1.1).
A set of Electrophoretic Mobility Shift Assays (EMSA) were carried out with the \textit{rrnB} upstream region ($-140$ to $-37$) as probe. Figure 6.9 shows two experiments, which both demonstrate an interaction between the RNAP\textalpha protein and the \textit{rrnB} upstream region. The interaction is weak and could only be detected at high protein concentration, $\geq 30 \mu M$ (panel A). The high concentration necessary for generating a shift of the probe raised the question whether the interaction between RNAP\textalpha and the DNA is sequence specific and a competition experiment was performed. Figure 6.9 panel B shows an EMSA using \textit{rrnB} UAR or pUC19 DNA as competitor. From this experiment it is evident that the pUC19 DNA is as efficient in competing for binding of RNAP\textalpha as the \textit{rrnB} upstream region itself (Figure 6.9 panel B, compare lanes 3-5 with lanes 6-8). The conclusion is therefore that the interaction between the \textit{rrnB} upstream region and the RNAP\textalpha subunit is not sequence specific but due to a general DNA binding property of \textalpha by virtue of being an RNAP subunit. The lack of a sequence specific interaction between RNAP\textalpha and \textit{rrnB} UAR in turn indicates that the \textit{rrnB} activation is dependent on a transcription factor. In case the interaction between the \textit{rrnB} UAR and this putative transcriptional activator display cooperativity with RNAP\textalpha, a second series of EMSA's with \textit{M.sm} cell free extracts were performed in which the cell extracts were used in combination with the purified RNAP\textalpha subunit. In these experiments the concentration of RNAP\textalpha was kept below 10 $\mu M$, which did
Figure 6.9: Gel shift assay with the rrnB UAR and the M.tb RNAP\(\alpha\).
The figure shows electrophoretic mobility shift assays using the rrnB UAR as probe and the M.tb RNAP\(\alpha\) protein. Panel A shows how an increasing protein concentration shifts increasing amounts of the probe. Lane 1: free probe, lanes 2-7: 20, 30, 40, 50, 75, 100 \(\mu\)M RNAP\(\alpha\). Panel B shows a competition experiment in which the UAR/\(\alpha\) complex is disrupted by adding increasing amounts of specific or non-specific unlabelled DNA. Lane 1: free probe, lane 2 probe + 50 \(\mu\)M RNAP\(\alpha\), lanes 3-5 same as lane 2 with 1, 10, and 50 fold excess unlabelled rrnB UAR DNA; lanes 6-8 same as lane 2 with 1, 10, and 50 fold excess unlabelled pUC19 DNA.
not generate a shift under the conditions used in the previous experiments. Total cell extracts from *M.sm* with or without RNAPα generated no visible change in the mobility of the *rrnB* probe (data not shown). The conclusion from these experiments is therefore that the activation is most likely due to interactions between the *rrnB* upstream region and one or more transcriptional activator(s) of unknown identity. An *in vitro* transcription system with purified (recombinant) components should be able to determine beyond doubt whether the activation of the *rrnB* operon is factor dependent or independent (or both if the *rrnB* UAR harbours more than one element).

**6.8 Species specificity of the *rrnB* regulatory elements**

The mechanism by which the regulatory elements of the *rrnB* operon operate is not known. In order to investigate the activity of the reporter constructs in a different organism, selected reporter constructs were assayed in an *E. coli* background (*E. coli* DH5α, which is lacZΔM15 (Sambrook *et al.*, 1989)). In *E. coli* the reporter constructs are harboured on a multi-copy plasmid with unknown but moderate copy number. Furthermore the *lacZ* reporter is an *E. coli* gene and codon bias is likely to have an effect on the translational efficiency of this gene between *M.sm* and *E. coli*. For these reasons the β-gal activities cannot be
compared directly between the two organisms. It is, however, possible to regard
the values on a relative basis.

Three constructs of the *M. sm* *rrnB* operon were expressed and assayed in *E. coli*. 
These are pKA55 (−140 to +207), pKA56 (−140 to +10), and pKA57 (−40 to 
+207). Figure 6.10 illustrates the difference in activities of the constructs in the
two species. Curiously the values for the entire promoter/leader region are almost
identical when expressed in *E. coli* and in *M. sm* (1149 units/mg protein in *E. coli*
compared to 1106 units/mg protein in *M. sm*). KAM55 and KAM56 (pKA55 and 
pKA56) both harbour the core promoter and the upstream region but the region
downstream of +10 is absent from the KAM/pKA56 constructs. Between KAM56
and KAM55 the activity decreases from 24025 units/mg protein to 1106 units/mg 
protein. In other words the downstream region confers a 22-fold reduction in β-
gal activity in *M. sm*. However, when the same constructs are expressed in *E. coli*,
the activity increases from 304 units/mg protein to 1149 units/mg protein; in other 
words the downstream region confers a four-fold increase in β-gal activity in *E. coli*. Thus the *rrnB* downstream region (+10 to +207) has opposite effects in the
two species.

The *rrnB* upstream region was investigated in the same manner. KAM55 and 
KAM57 differ in their upstream region. Both constructs end at +207 but KAM57
begins at −40 whereas KAM55 begins at −140. When expressed in *M. sm* the 
upstream region confers a 74-fold increase in activity from 14 units/mg protein to
Figure 6.10: Species specificity of the *rrnB* regulatory elements.

Three *rrnB* reporter constructs with the indicated boundaries were expressed in *M. sm* (blue bars) and *E. coli* (bordeaux bars). The full-length construct, pKA55, gives rise to similar levels of β-gal activity in the two organisms (centre columns). Deleting the region downstream of +10 (i.e. the *rrnB* leader region) increases activity in *M. sm* but reduces activity in *E. coli* (-104 to +10 fragment, left hand side of diagram). Conversely a deletion of the upstream region reduces activity in *M. sm* but increases activity in *E. coli* (-40 to +207 fragment right hand side).
1106 units/mg protein. However, when the same two constructs are expressed in *E. coli* (pKA55 and pKA57), the same upstream region confers a two-fold reduction in activity from 2616 units/mg protein to 1149 units/mg protein. The compiled values of these experiments are listed in Table 6.5:

**Table 6.5: Species specificity of the *rrnB* regulatory elements.**

<table>
<thead>
<tr>
<th></th>
<th><em>M. sm</em></th>
<th><em>E. coli</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Upstream region</td>
<td>74</td>
<td>0.4</td>
</tr>
<tr>
<td>Downstream region</td>
<td>0.05</td>
<td>4</td>
</tr>
</tbody>
</table>

The table illustrates the difference in activation/suppression conferred by the *rrnB* upstream and downstream regions in *M. sm* and *E. coli*, respectively. Reporter constructs were expressed in the two backgrounds and the activation or suppression conferred by the flanking regions compared. The numbers refer to activating potential of the indicated region; green numbers indicate an activating effect, and red numbers indicate a suppressing effect. The results demonstrate how the upstream and downstream regions both have opposite effects when expressed in the two organisms.

Thus in the case of the *rrnB* operon it appears that the regulatory effects are reversed between *M. sm* and *E. coli*. In *M. sm*, the downstream region has a repressing activity whereas the same region increased reporter activity when expressed in *E. coli*. Conversely the upstream region increases promoter activity in *M. sm* but the same region suppresses the activity in *E. coli*. These findings indicate that the mycobacterial regulatory signals investigated here are not
recognised in *E. coli*. This in turn suggests that these elements are specific to mycobacteria and perhaps other closely related organisms.

### 6.9 Regulatory regions of the *M.sm rrnB* operon: Discussion

The dissection of the *M.sm rrnB* operon resulted in the following observations. As is the case with the promoters of the *M.tb rrnA* operon, the P_B core promoter displays higher activity than the corresponding minimal promoter. A further extension of the 3' ends of the *rrnB* reporter constructs reveals that the region between +1 and +10 also increases promoter activity. The region between +1 and +10 is involved in the initial contacts between RNAP and promoter region as well as in promoter clearance which could explain the small increase in activity conferred by this region (Record *et al.*, 1996). The region between +10 and +207, which represents most of the *rrnB* leader region suppresses activity between four and 22-fold depending on the upstream context. Table 6.3 compares the activation and suppression levels conferred by each region and in the different contexts.

The *rrnB* upstream region contains the most highly activating region (UAR) identified in the present study. The *rrnB* UAR can activate the downstream promoter almost 400-fold, again depending on the context, i.e. the remaining sequences included in the reporter constructs (Table 6.3). The activation conferred by the *rrnB* UAR furthermore depends on helical phasing, meaning that the
spatial arrangement between core promoter and upstream region is more important than the distance *per se*. Electrophoretic mobility shift assays with the RNAPα subunit indicate that the activation is factor dependent but an activator could not be identified. A likely candidate for a binding site for this activator is the palindromic sequence, which is also located upstream of the two promoters in the *M.tbc rrnA* operon (palindrome II in Figure 6.4). The sequence is, however, imperfect in the *M.sm rrnB* operon. Finally it seems that the two most prominent regulatory regions in the *rrnB* operon, i.e. the UAR and the leader are species specific as shown by an inversion of the effects of each element when expressed in *E. coli*. Deletion of the *rrnB* UAR increases the expression level in *E. coli*. This suggests that the *rrnB* UAR suppresses *lacZ* expression or that the sequences of the promoter probe vector (which are immediately upstream of the promoter in KAM57) interfere with and increase *lacZ* expression. The presence of a DNA sequence in this region that would be recognised by an *E. coli* repressor is very unlikely. Therefore it seems more likely to be an artefact of the vector sequences (see Chapter 9).
7 The *M. tb* rRN leader region (Results)

7.1 Introduction to rRN leader functions

Ribosomal RNA is not translated. In an organism such as *E. coli* where Rho dependent termination is widespread and Rho factor is abundant, the cells must have a means to prevent premature termination of rRNA transcription by Rho. This is achieved by anti-termination, which refers to a modification of the elongation complex such that it becomes inert to Rho action and therefore reads through Rho-dependent terminators (Richardson and Greenblatt, 1996). It is still unknown exactly how the anti-termination mechanism works but it is possible that the increased elongation rate of RNAP transcribing rDNA is the sole explanation. The increase in elongation rate is signalled upon transcription of the *nut* site, which is contained in the rRN leader region (Albrechtsen *et al.*, 1990). The *nut* site transcript acts as a nucleation point for the assembly of the anti-termination complex, which consists of the elongation complex and Nus factors A, B, E, and G as well as another, unidentified cellular component, possibly ribosomal proteins (Squires *et al.*, 1993; Richardson and Greenblatt, 1996; Torres *et al.*, 2001). Apart from anti-termination, the *E. coli* rRN leader has been implicated in pausing, termination, transcript stability, and ribosome folding/assembly. These mechanisms all seem to involve the so-called *t* <sub>L</sub> region located 24 to 48 basepairs.
upstream of the mature *E. coli* 16S (Kingston and Chamberlin, 1981; Theissen *et al.*, 1990, Theissen *et al.*, 1993). Point mutations in the *E. coli* tL region reduces the stability of 16S RNA. However, deleting most or all of the tL structure in *E. coli* leads to transcriptional polarity, suggesting reduced or defective anti-termination (Theissen *et al.*, 1990).

The mycobacterial *rrn* leader regions are very heterologous in sequence and all of them appear to have extensive secondary structures as well as sequence homologues of the *E. coli* or *B. subtilis* nut sites but nothing is known about their intrinsic functions (Ji *et al.*, 1994b; Ji *et al.*, 1994a; Gonzalez-y-Merchand *et al.*, 1997). A mycobacterial analogue of the *E. coli* tL region is a possibility that has not been investigated.

### 7.2 The mycobacterial *rrn* leader's influence on lacZ expression

In section 6.3 it was demonstrated that the *M.tb* *rrnA* leader region confers a reduction in the expression of the *lacZ* reporter. Thus the region between the transcription start point and +219 reduced the activity of the PCL1 core promoter approximately five-fold (from 217 units/mg protein to 40 units/mg protein, Table 6.1). A number of reporter constructs, which contain a truncated leader or a full version of the *M.tb* *rrn* leader region were made in different promoter contexts. The truncated leader ends at +104 relative to PCL1 transcription start point, which
is approximately halfway through the *rrn* leader region, 41 basepairs downstream of BoxC (Figure 7.1). The full leader ends at +219 relative to PCL1 transcription start point, which is 29 basepairs into the mature portion of the 16S gene. The various constructs with different promoter contents have been shown schematically in Figure 7.1 with the β-gal activities of each reporter strain. The region upstream of the leader contains either the PCL1 core promoter, the PCL1 promoter with upstream region (up to but not including the P1 promoter), or the entire *rrnA* promoter region 300 basepairs upstream of the P1 transcription start point (Figure 7.1). Constructs with the full leader (KAM62, 61, and 60) display a gradual increase in β-gal activity as the region upstream of the PCL1 promoter is extended, i.e. 40, 174, and 1095 units/mg protein, respectively (Figure 7.1). This was expected as it had already been established that the PCL1 upstream region has a weak activating effect (chapter 6) and since P1 and the entire *rrnA* UAR is included in KAM60. In contrast the constructs with the truncated leader display very low activities as well as very little variations in these activities regardless of the extent of upstream region or promoter content, i.e. 11, 20, and 19 units/mg protein. In other words the expression of the *lacZ* reporter is severely impaired by the truncation of the *rrn* leader. The proposed secondary structure of the *M.tb* leader region is shown in Figure 7.2A. This figure reveals that +104 is located almost halfway through the stem-loop structure L1, and that constructs ending at +104 are devoid of this structure. Since this large stem-loop structure is conserved
Figure 7.1: Truncation of the M.tb rrn leader region inhibits lacZ expression. The diagram illustrates the differences in expression levels when the rrn fragment of the reporter constructs ends at +104 and at +219, relative to the transcription start point of the PCL1 promoter. Reporter constructs ending at +219 display variations in activity whereas reporter constructs ending at +104 display virtually same low level of lacZ expression (11 to 19 units/mg protein). The distances between BoxC and +104 and +219 are 41 and 156 basepairs in the two leader fragments, respectively. Numbers above each construct refers to the strain number.
in a number of mycobacteria (Gonzalez-y-Merchand et al., 1997) it was decided to investigate if the L1 structure has an effect on the expression of the lacZ reporter\(^1\). The nucleotides in the loop itself have been proposed to interact with a sequence in the 16S-23S spacer region, which is not present in the reporter constructs (Ji et al., 1994a) but an alternative or additional regulatory function is possible. A construct was made which has the same upstream region as KAM48 and KAM60 but the 3' end is located exactly at the base of the stem of the L1 structure, i.e. at position +127 relative to PCL1 transcription start point (Figure 7.2A). The β-gal activities of reporter constructs with 3’ ends at +104, +127, and +219 relative to the PCL1 transcription start point were compared and the results are shown in Figure 7.2B. The strain harbouring the \textit{rrn}A fragment to +127, has an activity of 155 units/mg protein, which is eight-fold higher than that of the truncated leader (19 units/mg protein) but still seven-fold lower than that of the full-length leader (1095 units/mg protein). Thus, the L1 structure has some effect on reporter activity but it is not entirely responsible for the difference in activity between truncated and full-length leader constructs. These results suggest that sequences throughout the entire \textit{rrn} leader region are of importance for proper expression of the \textit{M. tb} \textit{rrn}A operon.

The constructs containing the entire \textit{rrn}A promoter region as well as the truncated (+104) or the full (+219) leader region (i.e. pKA48 and pKA60) were expressed in

\(^{1}\) The L1 stem has 5 not 6 G residues as suggested by (Gonzalez-y-Merchand et al., 1997)
Figure 7.2: The *M. tb* *rrnA* leader region.

Panel A shows the proposed secondary structure of the *M. tb* *rrn* leader region (adapted from Gonzalez-y-Merchand et al., 1997). The numbers are relative to the transcription start point of the PCL1 promoter. Panel B illustrates the β-gal activity obtained with three reporter constructs containing the entire *M. tb* *rrnA* promoter region (i.e. P1, PCL1 and UAR). The three reporter constructs differ in the length of the leader region, which ends at positions +104, +127 and +219, respectively. Positions +104 and +127 have been indicated with arrows and in red in panel A. Position +190 (also indicated) is the last nucleotide of the leader transcript and +219 is thus 29 nucleotides into the mature 16S and is therefore not shown. The differences in β-gal activity between the three constructs are eight-fold, between +104 and +127, and seven-fold, between +127 and +219.
E. coli. As already demonstrated the β-gal activity of these constructs display an almost 60-fold difference in activity when expressed in M.sm. However, when the same two constructs are expressed in E. coli the difference in β-gal activity is less than two-fold, Table 7.1:

Table 7.1: Expression of the *M.tb* *rrnA* leader region in *M.sm* and in *E.coli*.

<table>
<thead>
<tr>
<th>Leader Region</th>
<th>Activity*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full-length leader (pKA60)</strong></td>
<td>M.sm: 1095 E.coli: 8856</td>
</tr>
<tr>
<td><strong>Truncated leader (pKA48)</strong></td>
<td>M.sm: 19   E.coli: 5499</td>
</tr>
<tr>
<td><strong>Ratio (full/truncated)</strong></td>
<td>58        1.6</td>
</tr>
</tbody>
</table>

*Activity is in units/mg protein

Two constructs, pKA48 and pKA60, were expressed in *M.sm* and in *E.coli*. Both constructs contain the entire *M.tb* *rrnA* promoter region but pKA48 has a truncated leader region, whereas pKA60 has the entire leader region and ~30 basepairs of the mature 16S gene. The table compares the effect conferred by the *rrnA* leader region in the two species. In *M.sm* truncation of the leader region results in a severe reduction in the expression of the *lacZ* reporter. In *E.coli* the same truncation results in less than two-fold reduction in *lacZ* expression.

Since none of the leader-less constructs were expressed in *E. coli* it could not be established whether a full leader region confers a reduction in β-activity, as is the case in *M.sm*. However, the truncated leader does not cause an impairment of *lacZ* expression in *E. coli*. In conclusion, one or more of the regulatory elements of the *M.tb* *rrn* leader are not recognised in *E. coli*. 
7.3 Hybrid reporter constructs

In order to investigate whether the phenotypes associated with the truncated and the full leader fragments (+104 and +219) could be transferred to another promoter context two hybrid reporter constructs were made. Of all the reporter strains in the present study, KAM56 spanning -140 to +10 of the *M.sm* *rrnB* operon displays the highest β-gal activity (24025 units/mg protein), and was therefore chosen for this investigation. The *M.tb* *rrnA* PCL1 fragments spanning -8 to +104 and -8 to +219 were fused to the *M.sm* *rrnB* promoter. The inserts were made at position +10 in KAM56 at the *HindIII* site and the constructs are outlined in Figure 7.3. The β-gal values were compared to that of the leader-less construct (KAM56) and two observations were made. Fusing the full leader (-8 to +219) to the *rrnB* construct (-140 to +10) reduces β-gal activity seven-fold compared to the leader-less construct (from 24025 to 3458 units/mg protein). This supports the earlier finding that the *rrnA* (and *rrnB*) leader regions both suppress activity (Sections 6.3 and 6.6). Fusing the truncated leader fragment (-8 to +104) to the *rrnB* promoter results in approximately 800-fold reduction in β-gal activity (from 24025 units/mg protein to 31 units/mg protein), and the difference in β-gal activity between the two hybrid constructs is 112-fold. Therefore the phenotypes associated with the truncated and the full leader region are transferable from one *rrn* operon to another and hence the elements of the *M.tb* *rrnA* leader are still
Figure 7.3: β-gal activities of hybrid reporter constructs.
The *M.sm* rrnB fragment from -140 to +10 was fused to the *M.tb* rrn leader region, i.e. -8 to +104 or -8 to +219 relative to the transcription start point of the PCL1 promoter. The figure illustrates the differences in β-gal activities obtained with the three constructs. The highest activity is obtained with the leader-less construct (KAM56), a truncated leader gives the lowest activity (KAM67), and a full-length leader region gives intermediate activity (KAM68). The inserts corresponding to the probes used for EMSA are indicated in red above the constructs.
functional in a different promoter context. Furthermore these elements are functional in spite of being shifted 24, i.e. 8 (from -8 to +1) + 6 (HindIII) +10 (from +1 to +10), basepairs further downstream relative to the new transcription start point (Figure 7.3). The results obtained with the rrnB-rrnA hybrid constructs therefore indicate that the phenomena of moderate suppression by the full-length leader and strong suppression by the truncated leader are not artefacts but due to *bona fide* transferable elements.

7.4 Interactions between Nus factors and the *rrn* leader

The *E. coli rrn* leader regions contain sequences that are important for the expression of the *rrn* operons. These include the *nut* site, which interacts with a number of *E. coli* proteins in order to modify the elongation complex to an anti-termination complex. In the previous section it was concluded that the *M.tb rrn* leader suppresses the expression of the *lacZ* reporter. Furthermore it was shown that a truncation of the leader region reduces the expression to extremely low levels. The molecular mechanisms behind these phenotypes were not established but they could involve an anti-termination system analogous to *E. coli rrn* anti-termination. Some of the proteins involved in *rrn* anti-termination display variations between *M.tb* and *E. coli*. The *M.tb* NusA contains the same RNA-binding domains as the *E. coli* homologue but lacks the entire CTD, which in *E.*
coli is responsible for interacting with the RNAPα. This interaction is required for
NusA to bind RNA, but a truncation of the CTD renders E. coli NusA capable of
RNA binding on its own (Mah et al., 2000). E. coli NusA has been shown to bind
to the λ P_L transcript upstream of BoxA but with very low affinity (Tsugawa et
al., 1985). Other experiments demonstrate that NusA cannot bind to the nut site
RNA on its own and that BoxA RNA does not alter the interaction between RNA
and NusA (Mogridge et al., 1995; Liu and Hanna, 1995).
The M.tb NusB protein forms a homodimer and the interaction with NusE is very
weak. Furthermore the M.tb NusB-NusE complex shows no binding to a BoxA
RNA oligonucleotide (Gopal et al., 2000; B. Gopal personal communication). The
E. coli NusB forms a relatively stable heterodimer with NusE, an interaction that
is stabilised by the interaction with an rrn nut site RNA oligonucleotide. Neither
E. coli NusB nor NusE can bind to the nut site alone (Nodwell and Greenblatt,
1993).
In order to get a further understanding of the function of the M.tb leader region
and the phenotypes associated with the truncated and the full-length leader,
respectively, the in vitro interactions between the rrnA leader and various Nus
factors were investigated. A set of ribo-probes corresponding to the M.tb rrnA
truncated and full leaders, respectively was designed. The DNA fragments
spanning -8 to either +104 or +219 (relative to PCL1 transcription start point)
Figure 7.4: In vitro transcription of the *M.tb* rrn leader region.
Panel A outlines the template used for the *in vitro* transcription of the *M.tb* rrnA leader region. The plasmids contain an *rrn* leader fragment (truncated, i.e. -8 to +104 relative to PCL transcription start point or full, i.e. -8 to +219). The DNA fragments (112 and 227 basepairs) were cloned into the *HindIII* of pGEM3Z in both orientations. The plasmids were linearised with *SalI* and used in an *in vitro* transcription reaction with SP6 RNAP. This generated four different probes: truncated sense and anti-sense and full sense and anti-sense. The two truncated leader probes were run on a 6% acrylamide sequencing gel and the result is shown in panel B (lane 1: sense transcript and lane 2: anti-sense transcript).

**Figure 7.4:** In vitro transcription of the *M.tb* rrn leader region.
were cloned into pGEM3z in both orientations in order to obtain sense and anti-sense transcripts. The plasmids were linearised with SalI and each probe was transcribed in vitro with SP6 RNAP to generate run-off transcripts of 150 nt and 265 nt, respectively (Figure 7.4A). The activity of the probes was determined by scintillation counting and an approximate concentration was calculated.

Preliminary gel runs revealed that the two truncated leader transcripts (sense and anti-sense) had different mobility on 8% native acrylamide gels (not shown). Therefore both probes were run on a denaturing 6% polyacrylamide gel in order to compare the mobility of the transcripts in the absence of secondary structures. This experiment revealed that sense transcript is shorter than the anti-sense transcript (Figure 7.4B). A difference in size between these two transcripts indicates that the sense probe is not a run-off transcript but that transcription has been (prematurely) terminated. This in turn indicates the presence of a strong pause signal or an intrinsic terminator within the first half of the rrnA leader region, but relatively close to the intended 3' end (+104) judging by the difference in size between the two transcripts (Figure 7.4B).

A series of EMSAs were performed with the four probes and different combinations of M.tb Nus factors.
7.4.1 NusA and the \textit{rrn} leader

In \textit{E. coli} NusA is essential for \textit{rrn} anti-termination (Vogel and Jensen, 1997). Furthermore the RNAP\(\alpha\) subunit is required for NusA to bind RNA (Mah et al., 2000). To determine whether \textit{M.tb} NusA bound to the \textit{M.tb} \textit{rrn} leader and whether RNAP\(\alpha\) was required, each probe (1 nM) was mixed with 10 nM NusA, 10 nM RNAP\(\alpha\) or 10 nM NusA+10 nM RNAP\(\alpha\) and run on native 8\% acrylamide gels. The result of one such experiment is shown in Figure 7.5. The four panels each represent one of the four different probes and a number of complexes can be observed in each panel. The complexes that are generated with both sense and anti-sense probes are considered to be unspecific and these are indicated by black double-headed arrows. The complexes generated only with the sense probes are considered specific, indicated by red, single-headed arrows (Figure 7.5). From this experiment it is concluded that \textit{M.tb} NusA interacts in a sequence-specific manner with the truncated as well as the full \textit{M.tb} \textit{rrn} leader. Furthermore it can be observed that RNAP\(\alpha\) does not generate any shifts, indicating that the protein does not bind to any of the transcripts (lanes 3). Adding RNAP\(\alpha\) to the NusA reaction did not alter the mobility of the NusA-RNA complexes indicating that there is no interaction between NusA and RNAP\(\alpha\), which would have generated a supershift (panels A and C lanes 4). The conclusion from these experiments is that NusA binds in a sequence specific manner to the \textit{rrn} leader with high affinity and
Figure 7.5: *In vitro* interactions between *M.tbc* NusA and the *rrnA* leader.
The figure illustrates electrophoretic mobility shift assays performed with *M.tbc rrn* leader RNA and NusA/RNAPα. Panel A is with the truncated leader in sense orientation. Panel B is truncated leader, anti-sense orientation. Panel C is full-length leader, sense orientation. Panel D: full-length leader, anti-sense orientation. Lanes, 1: 1 nM probe; 2: 1 nM probe+10 nM NusA; 3: 1 nM probe+10 nM RNAPα; 4: 1 nM probe+10 nM NusA and RNAPα. The arrows indicate the different complexes generated; black double-headed arrows indicate that the complex is formed with both sense and anti-sense transcripts, whereas red single-headed arrows indicate that the complexes are specific for the sense transcripts.
independently of RNAPα. Ribosomal protein S1 interacts in a sequence specific manner with the BoxA RNA, which could suggest that the S1 domain is responsible for the specific NusA-rrn leader interaction (Mogridge and Greenblatt, 1998). However, a recent report suggests that the three individual RNA binding domains of NusA interact to form a sequence specific binding site (Worbs et al., 2001). *In vitro* assays with mutant NusA proteins should address this question. The non-specific interaction between NusA and RNA was probably due to the fact that the protein contains three distinct RNA binding domains (Mah et al., 1999; Worbs et al., 2001).

**7.4.2 NusB and the rrn leader**

NusB is required for *rrn* anti-termination in *E. coli* (Sharrock et al., 1985a; Squires et al., 1993; Zellars and Squires, 1999). The protein forms a heterodimer with NusE, which binds to the *rrn nut* site (Nodwell and Greenblatt, 1993). In order to investigate the *in vitro* interaction between the *M.tb* NusB and NusE and the *rrn* leader region in a manner analogous to the NusA-RNA interaction (previous section), the two proteins were mixed with the truncated and the full-length *rrn* leader probes, and the samples were run on 8% native acrylamide gels. In this case 1 nM probe was used in combination with 15 nM NusB, 15 nM NusE, or both (Figure 7.6). The most prominent feature of this experiment is that only
Figure 7.6: *In vitro* interactions between *M.tb* NusB and the *rrnA* leader.
The figure illustrates a series of electrophoretic mobility shift assays with the truncated leader (panels A and B) and the full-length *rrn* leader (panels C and D). Each fragment was transcribed in sense and anti-sense orientation as indicated in the figure. Lanes are the same in each panel. Lane 1, 1nM probe; 2, 1nM probe+15 nM NusB; 3, 1nM probe+15nM NusE; 4, 1 nM probe+15nM each NusB+NusE. The figure demonstrates a specific interaction between the full-length leader region and *M.tb* NusB, indicated by a red arrow. The black arrow in panel A indicates a very faint complex between the truncated leader and NusB.
the full-length sense probe is capable of generating a significant shift upon the addition of Nus factors (panel C). Furthermore *M. tb* NusB shifts the probe in the absence of NusE (panel C lane 2). *M. tb* NusE does not generate a shift on its own and does not alter the mobility of the NusB-RNA complex (panel C compare lanes 2 and 4). A faint complex is barely visible with the truncated sense probe (panel A lanes 2 and 4) but none of the anti-sense probes show any sign of interaction with the proteins indicating that the interaction is very sequence-specific. The complexes in panel C appear somewhat smeared suggesting that some dissociation occurs during electrophoresis. This in turn indicates that the interaction between the *M. tb* *rrnA* leader region and NusB is weaker than the corresponding interaction with NusA (Figure 7.5). The conclusions from this experiment are that the full-length *rrn* leader transcript is required in order for *M. tb* NusB to bind properly to the *rrn* leader region. Furthermore that the *M. tb* NusB binds to the leader in the absence of NusE, the latter which does not appear to influence the interaction between NusB and RNA. The faint complex generated with the truncated leader transcript indicates a weak binding between this transcript and NusB. This in turn suggests that the NusB binding site may be within the first half of the leader and that binding is stabilised by sequences in the second half of the leader. Alternatively a low-affinity binding site could be present in the first half of the leader and a high-affinity binding site in the second half. Electrophoretic mobility shift assays with the second half of the *rrn* leader
and NusB should address this question. Either way sequences in the second half of the \textit{rrn} leader are required for proper interaction between NusB and RNA.

\textbf{7.5 Functions of the \textit{M.tb} rrn leader region: Discussion}

Reporter gene assays with various promoters and different lengths of the \textit{M.tb} \textit{rrn} leader region were assayed for \(\beta\)-gal activity. The results demonstrated that \textit{lacZ} expression is reduced in the presence of the leader region compared to the leader-less constructs. Furthermore it became clear that a truncation of the leader reduces the activity even further such that \textit{lacZ} expression was almost eliminated. The small reduction in activity conferred by the full leader region could be explained by the presence of one or more pause sites within this region. This is to some extent supported by the proposed secondary structure of the \textit{M.tb} leader (Figure 7.2) in which a number of hairpin structures followed by U-rich tails can be observed (Gonzalez-y-Merchand \textit{et al.}, 1997). These structures could act as pause sites or terminators and hence reduce \textit{lacZ} (and \textit{rrn}) expression. Furthermore it is known that the \textit{E. coli} \textit{rrn} leaders contain several pause sites as well as one major terminator, \(t_L\), immediately upstream of the 16S gene (Kingston and Chamberlin, 1981). The functions of these pause sites and the \(t_L\) region are not fully understood.
The reporter constructs containing the truncated leader display extremely low β-gal activities and in one case, the \textit{rrnB-rrnA} hybrid construct, expression is reduced more than 100-fold compared to the full-length leader. Several lines of evidence indicate that a post-initiation event is responsible for the low expression associated with the truncated leader. All the \textit{rrnA} constructs contain the PCL1 promoter, which, as a core promoter, has an activity of 210 units/mg protein. This is between 10- and 20-fold higher than any of the activities obtained with the truncated leader constructs. The activity of the \textit{rrnB} reporter construct could be reduced almost 800-fold by fusing the \(-8/+104\) fragment to position +10 of the \textit{rrnB} fragment. Most importantly, in the hybrid constructs the \textit{rrnA} sequences are all shifted 24 basepairs further downstream relative to the new transcription start point. This implies that in the hybrid constructs, the element(s) must be located downstream of the region where the RNAP makes initial promoter contacts. In other words, the RNAP must be in elongation phase as it encounters the relevant sequence if a transcriptional signal is involved. Alternatively a post-transcriptional event is responsible for the observed phenotype. Two possibilities have been considered: the truncation of the \textit{rrn} leader destabilises either the transcript or the elongation complex. Point mutations in the \textit{E. coli} \textit{tL} region reduce the stability of 16S RNA. If the stability of the \textit{M.tb rrn}-reporter transcript were affected in the same way by truncating the \textit{M.tb} leader region, this would obviously reduce the
expression of the lacZ reporter and this could explain the extremely low β-gal activity obtained with the reporter constructs ending at +104. The stability of the transcript was not investigated but a truncated leader would possibly mimic an incorrectly folded leader, which could act as a target for degradation.

Instability of the elongation complex would require a modification compared to the normal elongation complex, i.e. the one which transcribes the leader-less constructs and which is stable. The E. coli rrn leaders contain sequences, which signal the assembly of an anti-termination complex (Condon et al., 1995b). A possibility is that the truncated rrn leader in M.tb initiates but does not complete the assembly of an anti-termination complex. A partially assembled complex could be intrinsically unstable or be dissociated by accessory factors (see discussion).

In conclusion there are a number of indications that rrn anti-termination could function differently in M.tb than in E. coli. The two Nus factors reported to interact directly with the nascent transcript in E. coli are NusA and NusB (Nodwell and Greenblatt, 1993; Mah et al., 2000). In E. coli both proteins require the interaction with other factors, i.e. the RNAPα and NusE, respectively (Nodwell and Greenblatt, 1993; Mah et al., 2000). In M.tb, however, both proteins are not only capable of binding to the rrn leader on their own but they are also unaffected by the presence of RNAPα and NusE, respectively and they interact with separate parts of the leader region. Together with the facts that M.tb NusB
forms a homodimer and that the \textit{rrnB} leader does not contain a BoxB element, the present findings suggest that an \textit{M.tb} anti-termination mechanism does exist but it may have diverged significantly from the \textit{E. coli} paradigm.
8 Towards an *M.tb* in vitro transcription system (Results)

8.1 Species specificity of RNA polymerases

Bacterial RNA polymerases display extensive sequence homologies between species (Table 1.1), but they still harbour species specific differences in sequence, structure and function. For example, the RNAP holoenzyme of a number of Gram-positive organisms contain an additional subunit, the δ subunit, the significance of which is still not fully understood (e.g. Lopez de Saro *et al.*, 1999). Thus in order to identify species specific mechanisms of transcription, i.e. initiation, elongation and termination, it is essential to work with a homologous system, since a heterologous system could never provide an unambiguous answer.

Apart from differences in enzyme-promoter interactions, conferred by the σ factors, a number of species specific mechanisms reside in the core enzyme. The different RNAP core enzymes display species-specific variations in functions and interactions with transcription factors. The α subunit has been shown to interact in a species-specific manner with at least one transcriptional activator (*Mencia et al.*, 1998) as well as with an upstream activating DNA element (*Wada et al.*, 2000). *In vitro* transcription with enzymes from a number of species demonstrate differences in elongation rates and termination efficiencies, properties that are related, and linked to the large RNAP subunits (*Chamberlin et al.*, 1979;
Weilbaecher et al., 1994; Nudler, 1999). Furthermore the *E. coli* and *B. subtilis* enzymes behave differently towards identical pause and termination sites *in vitro*, signals that also involve the β and β' subunits (Artsimovitch et al., 2000). Finally, the *M.sm* RNAP displays an elongation rate that is approximately 75% of the *E. coli* RNAP elongation rate. However, the *in vivo* elongation rate of the *M.tb* enzyme is only 10% and not 75% of the *E. coli* elongation rate and yet the catalytic subunits of the *M.sm* and *M.tb* enzymes are 95% similar (Table 1.1) (Harshey and Ramakrishnan, 1977). The reason for the large difference in elongation rate between *M.tb* RNAP *in vivo* and *M.sm* RNAP *in vitro* is not known but an *in vitro* transcription system with *M.tb* RNAP may resolve this question.

In order to obtain an *in vitro* transcription system specific for *M.tb* and in order to gain further insight into the mechanisms of *M.tb* (and mycobacterial) transcription, an *in vitro* reconstitution of the *M.tb* RNAP has been initiated. The strategy is to assemble the enzyme *in vitro* from recombinant proteins rather than a purification of the enzyme from whole cell extracts since the latter would require vast amounts of *M.tb* culture. The *in vitro* approach furthermore has the possibility of incorporating otherwise lethal mutations into the enzyme.
8.2 Expression of the \textit{M.tb} RNA polymerase subunits

The individual RNAP subunits (\(\alpha\), \(\beta\), \(\beta'\), \(\omega\) and SigA) were amplified by PCR from genomic DNA and cloned into \textit{E. coli} expression vectors resulting in plasmids pK\(\alpha\), pK\(\beta\), pK\(\beta'\) and pSigA (see Materials and Methods for details). All constructs were transformed into \textit{E. coli} BL21 DE3 pLysS (Studier et al., 1990) and small scale expression cultures were prepared. Whole cell extracts were subjected to SDS-PAGE on 10-12\% Tris/Glycine gels. Figure 8.1 panel A shows a 12\% gel with the uninduced and induced strains carrying expression plasmids pK\(\beta\), pK\(\beta'\) and pSigA.

The \(\beta\) subunit and SigA result in single highly expressed protein bands (lanes 2 and 6) whereas the \(\beta'\) subunit appears as weak, multiple bands with approximately same mobility as \(\beta\) (lane 4). The apparent size of the \(\beta'\) subunit is therefore slightly lower than expected, since the predicted M_{w} of \(\beta\) is 129 kDa and that of \(\beta'\) is 147 kDa (plus the His\(_6\)-tag). This indicates that the \textit{M.tb} \textit{rpoC} gene may require a different cloning/expression strategy in order to obtain the correct, full-length protein. Affinity purification should address which method is required.

The \(\beta\) subunit and SigA (expected size 59 kDa) appear to have the correct size and are thus ready for large-scale expression and purification.

The \textit{M.tb} \textit{rpoA} was expressed on a small scale, like the remaining RNAP subunits, in order to verify the size and expression of the protein (Figure 8.1 panel B).
Figure 8.1: Overexpression of *M. tb* RNAP subunits in *E. coli*.

The individual subunits of the *M. tb* RNAP (α, β, β’, ω and SigA) were cloned into pET15b and expressed in *E. coli* BL21 DE3 pLysS. Cells were harvested and the lysates were run on Tris/Glycine/SDS polyacrylamide gels. Panel A shows a 12% gel with whole cell extracts from clones expressing β, β’ and SigA, uninduced (u) and induced (i). Panels B and C illustrate the expression and purification of the α subunit. Panel B shows a 12% gel with whole cell extracts from three individual pKαα clones, induced (i) and uninduced (u). One clone was selected for large scale expression and purification, illustrated in Panel C, which shows a 10% gel with the following samples: Lane 1, input fraction; 2, flowthrough; 3-14 eluate fractions 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36.
One clone was selected for expression on a larger scale and the His6-tagged protein was purified on a cobalt affinity matrix (see Materials and Methods for details). Figure 8.1 panel B and C illustrates the expression and purification of RNAPα. Panel B shows a 12% Tris/Glycine/SDS acrylamide gel with cell lysates from small scale expression cultures of three individual clones (induced and uninduced). Panel C shows a 10% Tris/Glycine/SDS acrylamide gel with samples from different stages of the purification (large scale) of *M.tb* RNAPα. The figure illustrates the efficiency of the purification method and shows that there is very little contamination from other proteins (panel C, lane 5). The RNAPα protein was used in electrophoretic mobility shift assays and displayed DNA binding activity, indicating that it had retained its proper structure and function (Chapter 6).

The ω subunit of *M.tb* is less than 8 kDa in size and could not be visualised in a normal small-scale expression culture lysate. A purification step is therefore necessary in order to verify that the protein is being expressed.

### 8.3 The *M.tb* *in vitro* transcription system: Discussion

During the course of this study it became evident that an *in vitro* transcription system would address a number of important questions. Since regulatory elements are often species specific a homologous *in vitro* transcription system is required.
The *in vitro* reconstitution of the *M.tbc* RNAP is still at a very early stage. The genes for the individual subunits have been cloned and the proteins overexpressed in *E.coli*. The β′ subunit has approximately the same mobility as the β subunit when subjected to SDS-PAGE. This suggests that the *rpoC* gene product is smaller than expected and that part of the C-terminal or the N-terminal is missing. This could be due to degradation, alternative translation initiation site or premature termination of transcription or translation. A different (earlier) cloning strategy generated a plasmid that had the original *M.tbc* ribosome binding site and translation initiation codon. This clone showed no expression, which argues against the presence of an internal secondary initiation codon within the *rpoC* coding region. Since the N-terminus includes a His6-tag, affinity chromatography should determine if the N-terminus is intact. If the protein is present in full-length as well as an N-terminally truncated version, the affinity purification will separate the full-length His6-tagged protein from the truncated. If, however, part of the C-terminal is missing due to degradation or premature termination of either transcription or translation, a different approach is required. If a nonsense mutation is present at the 3′ end of the cloned *rpoC* insert this would lead to a truncation of the protein and a different clone would possibly generate a protein with the correct size. If premature termination of transcription is the reason for the truncation of the protein, and if there is a small fraction of full-length protein, an
N-terminal His6-tag may solve the problem, but this would require a new cloning strategy.

To date only the α subunit has been purified but this protein is highly soluble and makes a good foundation for an in vitro assembly of the entire complex. A number of reports describe the purification of individual subunits and reconstitution of the entire RNAP by different methods (e.g. Zalenskaya et al., 1990; Igarashi and Ishihama, 1991; Kashlev et al., 1993). However, Ebright and co-workers have developed a method that relies entirely on the intrinsic ability of the RNAPα to act as an assembly platform for the rest of the RNAP (Tang et al., 1995). In their system the presence of the α protein ensures that the remaining core constituents fold into the proper conformation after isolation and solubilisation of the proteins from inclusion bodies. The α, β and β' subunits are mixed in a 1:1:2 ratio in the presence of 6M urea. Renaturation of the enzyme is achieved by stepwise dialysis into buffer with decreasing urea concentration. By having a His6-tag on the α subunit (or the β' subunit) the entire complex can be purified by affinity chromatography (Tang et al., 1995).

Initially a reconstitution with and without the ω fragment should address whether the proposed ω protein has the same function in M.tb as in E.coli. Assembly of the E. coli RNAP in the absence of ω leads to misfolding of the enzyme (possibly the β' subunit) and hence loss of activity (Mukherjee et al., 1999).
A fully functional *M.tb* RNAP would be a powerful tool in the characterisation of various elements of transcription, i.e. $\sigma$ factor and promoter relations, elongation rates, pause and termination sites could be determined. Especially the elongation rate is an intriguing issue, since the *M.tb* enzyme displays an elongation rate, which is only 10% of the *E. coli* RNAP elongation rate *in vivo* (Harshey and Ramakrishnan, 1977). The reason for the extremely slow elongation rate is not known but a comparison between different enzymes, templates and accessory factors may resolve this question. This could in turn perhaps shed more light on what mechanisms correlate the RNA CGR to the growth rate of the bacterium.

The role of the *M.tb* RNAP$\alpha$ subunit in transcriptional activation could be further investigated. In *E. coli* the RNAP$\alpha$ subunit plays a significant role in transcriptional activation, factor dependent as well as independent (Ebright and Busby, 1995). More specifically related to the present study an *in vitro* transcription system would determine whether the activation of the *M.tb* *rrnA* and *M.sm* *rrnB* operons is factor dependent or independent. Is there an *M.tb* UP element analogue and if so, what is the sequence? An *in vitro* transcription system would also help determine whether a putative transcriptional activator is more abundant in *M.sm* than in *M.tb*, i.e. does an *M.sm* cell extract activate transcription more than an *M.tb* cell extract?

Another interesting aspect of developing an *M.tb* *in vitro* transcription system would be the prospect of a further characterisation of the *rrn* leader region and the
*M. tb* anti-termination complex. An *in vitro* transcription assay would enable the characterisation of pause sites in the *rrn* leader region, their location and their requirement for accessory factors such as NusA and ppGpp. It would aid in the characterisation of termination in general, intrinsic as well as Rho-dependent. Finally it would make it possible to determine the Nus factor requirements for *rrn* anti-termination, i.e. RNAP read-through at Rho-dependent terminators.
9 Discussion

9.1 Mycobacteria and their \textit{rrn} operons

The genus \textit{Mycobacterium} consists of numerous species, which display large variations in growth requirements and growth rates (Goodfellow and Wayne, 1982). They do however have certain common characteristics including growth rates that are lower than those of many other bacteria, and a low number of \textit{rrn} operons. As a rule (with exceptions) the slower growing species have a single \textit{rrn} operon, \textit{rrnA}, whereas the faster growing species have two operons, \textit{rrnA} and \textit{rrnB}. An interesting feature of the mycobacterial \textit{rrn} operons is that the structure varies from one operon to the next, as well as between species (Gonzalez-y-Merchand et al., 1997). Thus the \textit{rrnA} operon has between two and five promoters depending on the species, whereas the \textit{rrnB} operon has a single promoter in all cases studied so far (Gonzalez-y-Merchand et al., 1997). The individual promoters are recognised by at least two different \sigma factors. The P1 promoter of the \textit{rrnA} operon is recognised by the alternative \sigma factor, SigF, which is present in very low amounts during exponential growth, judging by the level of \textit{sigF} mRNA (DeMaio et al., 1996; Chen and Bishai, 1998; Manganelli et al., 1999). Although it remains to be demonstrated, SigA presumably recognises a number of, if not all of the remaining \textit{rrn} promoters. The sequences of the –10 and –35 regions of
these promoters display high homology to *E.coli* $\sigma^{70}$ promoters, which suggest that they would be recognised by SigA or SigB (Gomez *et al.*, 1997; Gomez and Smith, 2000). In *M.sm*, and probably in *M.tb* as well, SigB is not essential (Gomez *et al.*, 1998; Manganelli *et al.*, 2001), which makes SigA the more likely candidate since at least one *rrn* promoter in any given organism would be assumed to be recognised by the primary $\sigma$ factor of that particular organism. Thus, in *M.tb*, which contains a single *rrn* operon with the two promoters P1 and PCL1, only a single promoter is recognised by the primary $\sigma$ factor. This is obviously very different from the situation in e.g. *E.coli* were 14 promoters in seven operons are all recognised by $\sigma^{70}$ (Condon *et al.*, 1995b; Gourse *et al.*, 1996). The expression of the mycobacterial *rrn* operons changes with the growth rate (Gonzalez-y-Merchand *et al.*, 1999; Verma *et al.*, 1999). However, the effects conferred by individual regions in the *rrn* operons are not known. One aim of the present study was therefore to investigate specific features of various mycobacterial *rrn* promoters and operons in order to identify regulatory elements/regions involved in the expression of the mycobacterial *rrn* operons. Furthermore an investigation of minimal promoters and core promoters from *rrn* operons was carried out to gain further knowledge about mycobacterial transcription in general. In order to be able to manipulate the *rrn* regulatory regions, a reporter system was employed. The majority of these manipulations
would have been deleterious or interfered substantially with the growth of the cells were they performed on the native rrn operons. Various promoter-containing fragments were fused to a promoter-less lacZ gene and the resulting constructs were integrated into the genome of M.sm. Cultures of each strain were in balanced growth at the time of harvest and the β-gal activity was determined. A number of experiments demonstrated that the reporter system is reliable and gives reproducible results. However, reporter systems have drawbacks. One is that mutations in the reporter gene will not give rise to the same selection pressure as mutations in the native gene. This situation is illustrated in the experiments with the rrn knockout strains. In one strain the reporter gene appeared to have lost some of its activity. Since the promoter region had been sequenced and found correct, the mutation responsible for the reduced β-gal activity probably resides in the lacZ gene. The M.sm transposon, IS1096, has a preference for the lacZ coding sequence, and loss of β-gal activity as a result of its transposition is a possibility (Tyagi et al., 2000). Alternatively a missense mutation in lacZ could result in loss of β-gal activity without reducing the expression level. A comparison between β-gal activity and β-gal levels, the latter determined by Western blotting, could perhaps have resolved this.

Additional artefacts can arise with reporter systems as a consequence of high promoter activity. It has been demonstrated that induction of lacZ from high copy number plasmids interferes with the translational machinery of the cells in E.coli.
(Vind et al., 1993). Furthermore there is the possibility of toxicity of the β-galactosidase protein since this is a heterologous protein for M.sm. High levels of β-galactosidase does not appear to be a problem as such, since expression levels of more than 20000 units/mg protein could be obtained reproducibly. However, some strains display difficulties in resuming growth after reaching stationary phase. Since this phenotype for the most part is associated with the larger constructs and does not necessarily correlate with high lacZ expression levels it is likely to be a result of specific rrn regulatory elements harboured in some reporter constructs rather than the lacZ gene per se. Expression of a λnut site from a multi-copy plasmid inhibits rrn expression and growth in E.coli (Sharrock et al., 1985b). A shift from stationary phase to exponential phase increases the activity of the M.sm P2 and P_B promoters dramatically (Gonzalez-y-Merchand et al., 1998). It seems possible that a similar increase in rrn promoter activity in the full-length reporter constructs upon sub-culturing a stationary-phase culture, could interfere with the expression of the native rrn genes e.g. by sequestration of Nus factors required for expression of the native rrn genes.

9.2 Expression of rrn operons

The expression levels of the M.sm rrnA, M.sm rrnB, and M.tb rrnA operons were found to be in the same range according to the β-gal activities. This is interesting
since the three operons have a different number of promoters, (3, 1, and 2, respectively). The *M.sm rrnA* and *rrnB* operons have previously been reported to have similar expression levels during various stages of growth, as shown by primer extension studies (Gonzalez-y-Merchand *et al*., 1998). In the same study, however it was demonstrated that the expression level of the *M.tb rrnA* operon is between 5 and 15% of the expression level of the combined *M.sm rrn* operons depending on the growth stage and growth conditions (Gonzalez-y-Merchand *et al*., 1998). The discrepancy between the *M.tb rrn* expression levels in the two studies can be attributed to a number of factors. The cultures in the study by Cox and co-workers were not in balanced growth judging by the variations in *rrn* expression and since they reached stationary phase before six generation times had passed. This means that in that particular study there is more than one value for the level of *rrn* expression during exponential growth and therefore the following comparisons should be regarded carefully.

In the current study the *M.tb rrnA* operon was expressed in *M.sm* and not *M.tb*. If trans-acting factors rather than the *rrn* operons *per se* enhance the *rrn* expression levels, then the activity of the *M.tb rrnA* operon expressed in *M.sm* would presumably resemble that of the *M.sm rrnA* operon, also expressed in *M.sm*. This would also be the case if the *in vivo* RNA elongation rates in *M.tb* and *M.sm* were significantly different. The *M.tb* cultures in Gonzalez-y-Merchand *et al.* (1998) were grown under sub-optimal conditions with a doubling time of 72 hrs. *M.tb*
grown under optimal conditions has a doubling time of 15 hrs. Therefore, since
*M. tb* *rrnA* expression increases with increasing growth rate (Verma *et al.*, 1999),
the *rrn* expression in the "slow" cultures was probably much lower than it would
have been in a faster growing culture. Therefore it seems plausible that the three
*rrn* operons have roughly the same capacity for rRNA synthesis at least during
balanced growth and in the same background.

9.3 Feedback regulation of *rrn* operons

The *rrnB* operon of *M. sm* is feedback activated upon a reduction in *rrn* gene
dosage but there is no indication of a similar mechanism in the *M. tb* *rrnA* operon.
On the other hand in *M. sm* the inactivation of the *M. sm* *rrnB* operon has less
impact on growth than inactivation of the *M. sm* *rrnA* operon. This suggests that
the *M. sm* *rrnA* operon is at least as capable of compensating for the reduced gene
dosage as the *rrnB* operon; in other words an indication of feedback activation of
the *M. sm* *rrnA* operon. Does this mean that the *M. sm* *rrnA* but not the *M. tb* *rrnA*
operon is feedback regulated? The primary difference between these two operons
is the presence of the P2 promoter in the *M. sm* *rrnA* operon. A plausible
explanation would therefore be that the P_B and P2 promoters are feedback
activated, whereas the P1 and PCL1 promoters are not. The sequences of the P_B
and P2 promoters resemble each other but the size of the spacer between the –35
and -10 regions is 17 basepairs in PB and 18 basepairs in P2 (Gonzalez-y-Merchand et al., 1997). One of the DNA determinants for feedback activation in *E.coli* rrn P1 promoters is the 16 basepair spacer. Changing this to 17 basepairs abolishes feedback activation of expression upon rrn deletion (Voulgaris et al., 2000). Although the spacer length in mycobacterial *rrn* promoters generally varies (Gonzalez-y-Merchand et al., 1997), the different lengths of the PB and P2 spacers to some extent argues against these two promoters being the only promoters that are feedback activated. An alternative explanation could be that the difference in spacer lengths reflects different σ factor requirements (e.g. SigA vs. SigB) and that the determinants for feedback activation in mycobacterial *rrn* promoters are not related to the spacer. The deletion of one of two *rrn* operons in *M.sm* has a limited effect on cells in balanced growth (Figure 4.4). Experiments with *E.coli* have demonstrated that the effect of *rrn* deletions is much more pronounced upon nutritional up-shifts e.g. from stationary phase to exponential growth than it is in balanced growth, possibly because of a rapid increase in *rrn* expression under these conditions (Condon et al., 1995a). Two of the four *M.sm* *rrn* promoters (PB and P2) display a dramatic increase in activity after a nutritional up-shift (Gonzalez-y-Merchand et al., 1998). This suggests that the *M.sm* *rrn* knockout strains could be severely impaired upon such an up-shift, i.e. when the cells require maximum *rrn* expression.
In conclusion the \textit{M.\textit{tb} rrnA} operon has an expression level comparable to that of each of the two \textit{M.\textit{sm} rrn} operons, and \textit{M.\textit{sm}} steady-state growth is virtually unaffected by the inactivation of one of two \textit{rrn} operons. This suggests that the \textit{M.\textit{tb} rrnA} operon has the capacity to sustain growth rates comparable to \textit{M.\textit{sm}} growth rates. In other words the reason for the slow growth of \textit{M.\textit{tb}} is not the single \textit{rrn} operon.

\textbf{9.4 The mycobacterial minimal and core promoters}

The individual \textit{rrn} minimal promoters display large variations in activities. This may to some extent be attributed to different \(\sigma\) factor requirement/levels. It is worth noting that since the \textit{rrn} promoters are recognised by different \(\sigma\) factors, the promoter strength reported in the present study will also reflect the relative concentrations of various \(\sigma\) factors. Therefore the definition of promoter strength is not as strict as if the promoters were all recognised by the same \(\sigma\) factor.

Three promoters, \textit{M.\textit{tb} P1}, \textit{M.\textit{tb} PCL1}, and \textit{M.\textit{sm} P_{B}} were assayed for activity with and without their native transcription start point, i.e. the activities of the minimal promoters and the core promoters were compared. The results demonstrate that all three promoters benefit from the region between the \(-10\) box and the transcription start point but to very different extents. Thus, the strongest promoter, \textit{M.\textit{tb} PCL1} displays only 10\% increase in activity from minimal
promoter to core promoter. In contrast the weakest promoter, *M.tb* P1 displays approximately 400% increase in activity from minimal promoter to core promoter. In other words there is an inverse correlation between the strength of a minimal promoter and the advantage of including the native transcription start point. It has not been determined whether this relation applies to other promoters or whether this difference is conferred by the σ factor or by the DNA.

The investigated minimal promoters display a 30-fold difference in β-gal activity and these activities appear to be achieved by different means. For example the PCL1 promoters from *M.ch* and *M.ab* (as well as P2-P4 from the same species) have acquired an extended −10 region. The *M.ch/ab* PCL1 minimal promoter is the strongest of the investigated promoters but without the TGN motif this promoter is weak compared to e.g. the *M.tb* PCL1 promoter. Therefore this motif does not increase promoter activity as much as it probably could have, were the −10 region the same as that in the *M.tb* PCL1 promoter. A comparison of the two promoters reveals that the *M.tb* PCL1 promoter has a higher degree of homology to the consensus than the *M.ch/M.ab* PCL1 promoters (Gonzalez-y-Merchand *et al.*, 1997). The binding constant, $K_B$, of a promoter increases with increasing homology to consensus, although both hexamers are involved in multiple steps of initiation (Ellinger *et al.*, 1994a; Record *et al.*, 1996). In contrast the extended −10 region has been shown to increase the rate of open complex formation, $k_f$, without
increasing $K_B$ (Burr et al., 2000) suggesting that the $M.ch/ab$ PCL1 minimal promoters have evolved towards a lower $K_B$ and instead increasing a different parameter in the initiation pathway. The extended $-10$ region is only present in $M.ch$ and $M.ab$, i.e. mycobacterial $rrn$ operons with many promoters (Gonzalez-y-Merchand et al., 1997). In an operon with many promoters a high $K_B$ may cause promoter occlusion due to stalling of the RNAP (Ellinger et al., 1994a). Furthermore it may not be as critical due to the proximity of other (strong) promoters which could possibly result in an increase in the local concentration of RNAP. In such an operon, an increase in the rate of open complex formation ($k_f$) and in the rate of promoter clearance appear to be more appropriate.

9.5 The promoters of the $M.tb$ $rrnA$ operon

The two promoters of the $M.tb$ $rrnA$ operon clearly demonstrate that the strength of minimal promoters, core promoters, and core promoters plus flanking regions do not always correlate. The $M.tb$ PCL1 minimal promoter is a strong promoter and the strength does not increase much when promoter proximal sequences are included in the reporter construct. Thus the difference between minimal promoter and core promoter is approximately 10% and the highest activity of the PCL1 promoter constructs is less than three-fold higher than the activity of the minimal
promoter (Table 6.1). The activation conferred by the region immediately upstream of the promoter is less than five-fold, Table 9.1:

Table 9.1: The activating potential of the \textit{rrn} promoters' upstream regions.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Upstream region*</th>
<th>Activation by upstream region**</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{M.tb} P1</td>
<td>300</td>
<td>185-fold</td>
</tr>
<tr>
<td>\textit{M.tb} PCL1</td>
<td>84</td>
<td>4-fold</td>
</tr>
<tr>
<td>\textit{M.sm} \text{P_B (small)}</td>
<td>80</td>
<td>177-fold</td>
</tr>
<tr>
<td>\textit{M.sm} \text{P_B (large)}</td>
<td>140</td>
<td>388-fold</td>
</tr>
</tbody>
</table>

* Number of basepairs counting from the transcription start point
** The value is the maximum activation conferred by the indicated region

The table indicates the extent of the upstream region included in the reporter constructs and the activation fold conferred by this region. From these results it is clear that all three promoters are activated by sequences upstream of the \textendash-35 region but that the activating potentials of these regions are very different.

An obvious explanation for this weak activation would be the limited size, i.e. 47 basepairs, of the PCL1 upstream region (84 basepairs upstream of the transcription start point). However in comparison the promoter proximal 43 basepairs of the \textit{rrnB} upstream region have an activation potential of almost 200-fold indicating that the size of the PCL1 upstream region is not the sole explanation for its limited activating potential.

The PCL1 promoter has been reported to contribute to 75% of total rRNA synthesis in \textit{M.tb} (Gonzalez-y-Merchand \textit{et al.}, 1998), and the expression level of the entire operon as a \textit{lacZ} reporter gene fusion is 1100 units/mg protein.
Therefore, in theory the expression level of the reporter construct containing the PCL1 promoter with flanking regions should be $0.75 \times 1100 = 825$ units/mg protein. The expression of the PCL1 promoter including upstream region and leader is 174 units/mg protein, i.e. $1/5$ of the expected value (KAM61, Table 6.1). Since the region downstream of the PCL1 promoter suppresses rather than increases $lacZ$ expression (discussed below), it appears that the $M.tb$ PCL1 promoter relies on sequences further upstream, and perhaps even an intact P1 promoter, in order to reach an expression level that can account for its contribution in vivo.

The $M.tb$ P1 promoter has a different constellation of regulatory elements. This promoter has a very low activity as a minimal promoter (the weakest of the wild type minimal promoters) but activity increases dramatically from minimal promoter to core promoter, i.e. the eight basepairs between the $-10$ region and the transcription start point confer a 400% increase in activity. The region downstream of the P1 transcription start point to +41 has a weakly enhancing effect, which increases promoter activity less than two-fold. Interestingly the activation conferred by the P1 downstream region is diminished in the presence of the P1 upstream region. The P1 upstream region is highly activating and enhances P1 promoter activity almost 200-fold (Table 9.1), but as is the case with the P1 downstream region, the P1 upstream region loses some of its potential in the
presence of the downstream region (Figure 6.2). Thus the regions flanking the P1 promoter are both activating but to different extents and dependent on each other. Both regions have a higher activating potential in the absence of the other indicating that they act on the same step(s) of transcription initiation. If calculations similar to the ones above are applied, a number of interesting observations can be made. The expression level of the entire operon is 1100 units/mg protein and the P1 promoter contributes with 25% corresponding to 275 units/mg protein (Gonzalez-y-Merchand et al., 1998). The P1 reporter construct covering the largest region is KAM49 with an activity of 4388 units/mg protein, i.e. 16-fold over the expected value. This difference can to some extent be accounted for by the absence of the rrnA leader region, which in a different context confers a seven-fold reduction in activity. Alternatively or in addition, transcription from the P1 promoter could be inhibited by the presence of the PCL1 promoter in its native context, i.e. promoter occlusion. Therefore the two promoters of the M.tb rrnA appear to influence the expression of each other and could therefore be regarded as a single unit rather than two individual promoters in one operon. The PCL1 promoter requires sequences upstream of the P1 promoter for activity presumably while the P1 promoter itself is active. The region between the two promoters is capable of enhancing the activity of both promoters although it is upstream of one and downstream of the other. Finally the leader region, i.e. the region downstream of both promoters suppresses the
expression from both promoters. Reporter constructs in which the activity of one or the other promoter had been abolished by single point mutations should give an indication of the two promoters' mutual dependence. This approach would obviously eliminate a phenomenon such as promoter occlusion.

9.6 The M.tb rrn leader region

The M.tb rrn leader region, in this study defined as the region between the PCL transcription start point and the mature 16S RNA, displayed some interesting phenomena. Initially it was demonstrated that the entire region confers a moderate reduction in lacZ expression. This is possibly due to the presence of one or more pause sites within the region. The proposed structure of the M.tb rrn leader is shown in Figure 7.2A. Although this remains hypothetical, a number of the indicated stem-loop structures resemble, and could act as, transcriptional pause sites.Pause sites are also found in the rrn leaders of E.coli (Kingston and Chamberlin, 1981). The function of these pause sites is not fully understood but it seems likely that the elongation complex requires pausing in order for the rrn leader to fold into the proper conformation. This is sustained by the finding that mutations in the tL region of E.coli leads to ribosome assembly defects believed to be caused by misfolding (Theissen et al., 1993). Furthermore transcription of E. coli rrn genes with T7 RNAP, elongating five times as fast as the native E. coli
RNAP, leads to misfolding of the rRNA (Lewicki et al., 1993). Therefore the conclusion is that the *M.tuberculosis* *rrn* leader harbours one or more pause sites, analogous to the *Escherichia coli* *rrn* leaders, responsible for the observed reduction in *lacZ* expression. *In vitro* transcription of the *M.tuberculosis* *rrn* leader may resolve this question.

A truncation of the *M.tuberculosis* *rrn* leader severely impairs the expression of all the reporter constructs. It seems likely that this phenotype arises from a different molecular mechanism than the transcriptional pausing described above. There seem to be two possible explanations for this block in *lacZ* expression. The truncated leader transcript possibly mimics a misfolded *rrn* transcript, which could be a target for degradation *in vivo*. In other words the reporter transcripts containing the truncated leader region are extremely unstable. Alternatively the elongation complexes transcribing the truncated leader constructs could be unstable. Since the "normal" elongation complexes, i.e. the ones transcribing the leader-less constructs are stable a modification that confers this instability would be required. The *Escherichia coli* *rrn* leader harbours the *nut* site, which signals a modification of the elongation complex into an anti-terminated (and more stable) elongation complex. The *Escherichia coli* BoxA has been shown to be sufficient for *rrn* anti-termination *in vivo* and *in vitro* (Berg et al., 1989; Squires et al., 1993). The BoxA mediated increase in transcription elongation rate, which according to some groups is tantamount to *rrn* anti-termination, requires NusA, NusB, and NusG (Vogel and Jensen, 1997; Zellars and Squires, 1999). Furthermore the interaction
between BoxA and NusB requires NusE (Nodwell and Greenblatt, 1993). In other words it appears that BoxA is sufficient for the assembly of the entire *E.coli rrn* anti-termination complex. *In vitro* experiments performed in the present study indicate that the situation may be different in *M.tb*. Electrophoretic mobility shift assays with the *M.tb rrn* leader and *M.tb* NusA and NusB demonstrate that the two proteins require different regions of the leader region for a stable interaction. Furthermore the *M.tb* NusB may require the entire *rrn* leader region in order to bind. It is possible that the protein interacts with the proposed *nut* site as well as sequences in the second half of the leader, i.e. a secondary *nut* site. The *M.tb* NusB protein exists as a homodimer in the crystal structure and in solution (Gopal *et al.*, 2000). It is conceivable that if the protein exists as a dimer in the cell each monomer binds to different halves of the *rrn* leader transcript thereby stabilising the interaction. A schematic drawing of this model has been shown in Figure 9.1. If the assembly of the *M.tb* anti-termination complex requires the entire *rrn* leader because NusB is only loosely (or not at all) associated with the first half, then the assembly would be initiated but not finished in the presence of a truncated leader. The incomplete complexes could be intrinsically unstable and dissociate as soon as the first pause site or terminator is encountered. Alternatively an accessory factor could target the incomplete complexes and cause dissociation. The purpose of this mechanism would be analogous to the proposed function of the *tL* region
Figure 9.1: A model of the interaction between *M. tb* NusB and the *rrn* leader.
The *M. tb* NusB protein exists as a dimer in solution but it is still unknown if the protein acts as a monomer or a dimer in the cell (Gopal et al., 2000). The *in vitro* binding assays performed in the present study indicate that NusB interacts weakly with the truncated *rrn* leader region and requires the second half of, or the entire leader region for a stable interaction. A model for the interaction between the NusB dimer and an *rrn* leader transcript with two binding sites is shown in panel A. The NusB dimer has been shown in green and the RNA binding sites are indicated as red cricles. In the case of a full-length leader region the NusB dimer interacts with two binding sites thereby obtaining a stable interaction (panel A). In the case of a truncated leader NusB can only interact with one binding site and the interaction is therefore less stable (panel B).
of *E.coli*, i.e. a discrimination between anti-terminated and not anti-terminated elongation complexes (Zacharias and Wagner, 1987). Both intrinsic and factor mediated dissociation of the elongation complex should be detectable in an *in vitro* transcription assay. Electrophoretic mobility shift assays with the second half of the *rrn* leader and with a mutant NusB protein impaired in dimerisation should establish whether the model in Figure 9.1 is valid.

9.7 The *M.sm* *rrnB* operon

The *M.sm* *rrnB* operon has the advantage of harbouring a single promoter thereby facilitating an investigation of promoter and flanking regions. The P_B minimal promoter is as weak as the P1 promoter but in contrast to the P1 promoter the activity does not increase dramatically by including the native transcription start point. Nor do the additional 9 basepairs to +10 change much. The strength of the *rrnB* promoter lies in the highly activating upstream region, which contains, within approximately 100 basepairs, elements responsible for increasing the P_B minimal promoter activity more than 300-fold. The activation depends on helical phasing and therefore presumably involves the RNAPα subunit, but a sequence-specific interaction between this protein and the *rrnB* UAR could not be detected. The region upstream of −70 activates transcription between two and five-fold and it is unlikely that this activation should be factor independent due to the distance
between this region and the core promoter. The conclusion is therefore that the activation of the \( P_B \) promoters is likely to be factor dependent and that this so far unidentified factor interacts with the RNAP\( \alpha \) subunit. Curiously the \( rrnB \) UAR appears to have an inhibitory effect in \( E.coli \). According to the \( \beta \)-gal activities, the deletion of the UAR increased promoter activity. A closer look at the sequences of the constructs reveals that, in the reporter construct without the \( rrnB \) UAR (i.e. KAM57) the transcriptional terminators of the promoter probe vector are immediately upstream of the \( P_B \) promoter. The promoter proximal terminator consists of a stretch of eight T-residues centred around \(-53\) relative to the transcription start point and this region could perhaps act as a weak UP element in \( E. coli \) (Estrem et al., 1998; Estrem et al., 1999). The conclusion is therefore that the \( rrnB \) UAR is not recognised in \( E.coli \) and the two-fold increase in activity observed by a deletion of this region is an artefact.

The \( rrnB \) leader region appears to suppress activity in a manner analogous to the \( M.tb \ rrnA \) leader. It is, however difficult to draw final conclusions about this region since, in the reporter constructs the leader is truncated 93 basepairs upstream of the mature 16S. The truncation does not block activity as is the case in the \( rrnA \) operon but it may still reduce it relative to a full-length leader. The extremely high \( \beta \)-gal activity, obtained with leader-less \( rrnB \) reporter constructs, indicates a very high potential for rRNA synthesis. On the other hand, the absence
of the leader region would probably generate problems in the native \textit{rrn} gene. The leader region contains the anti-termination motifs. Since the reporter gene is translated, anti-termination is not critical but the native \textit{rrn} transcript is not translated and the lack of anti-termination would presumably result in premature termination of transcription (provided that the system resembles the \textit{E.coli} system). In other words a deletion of the \textit{rrn} leader region in the native gene would probably lead to a severe reduction in activity rather than the observed increase.

9.8 The \textit{rrn} operons of Mycobacteria – a comparison

The two operons investigated in this study turned out to have a number of elements in common in spite of their apparent different structures. Figure 9.2 shows a schematic drawing of a model operon, based on the findings in the \textit{M.tb} \textit{rrnA} and \textit{M.sm} \textit{rrnB} operons. In this model operon the two promoters of the \textit{rrnA} operon have been depicted as a single unit for comparative reasons. Both \textit{rrn} operons have a UAR with approximately the same activating potential, i.e. 2-300-fold, which is comparable to the \textit{E.coli} \textit{rrn} UAR (Condon \textit{et al.}, 1995b). Within this upstream region lies a palindromic sequence (palindrome number I in Figure 6.4), which is perfect in the \textit{rrnA} operon but has a single basepair mismatch in the \textit{rrnB} operon. In the \textit{rrnA} operon a second palindromic sequence is located further
upstream (see Figure 6.4). The significance of these sequences has not been established, but it is possible that they act as binding sites for transcriptional regulators. The sequences are highly conserved between mycobacterial species but this could also be due to the fact that the region lies within the \textit{murA} coding region (Gonzalez-y-Merchand \textit{et al}., 1997).

The region upstream of the mature 16S gene, i.e. the \textit{rrn} leader, appears to have a suppressing effect in both operons possibly due to the presence of one or more pause sites. The \textit{rrnA} reporter construct, which ends at the base of the L1 structure has an activity which is intermediate of the truncated and the full-length leader constructs. The \textit{rrnB} construct does not include the entire L1 structure and yet has an activity comparable to the full-length \textit{rrnA} construct. Either the two leader regions do not contain the same regulatory signals or the activity of the \textit{rrnB} operon with full-length leader has a much higher activity than the one reported in this study. One indication that the two leaders could contain different functions/activities is that the \textit{rrnB} leader does not contain a BoxB element (Ji \textit{et al}., 1994b). Therefore a comparison between the functions of the two leaders, \textit{in vivo} and \textit{in vitro}, should further the understanding of anti-termination in mycobacteria and in general. The \textit{M.tb rrn} leader contains binding sites for both NusA and NusB. The experiments indicated that the two proteins bind to different halves of the \textit{rrn} leader or alternatively that the NusB protein binds to both halves by having two binding sites (Figure 9.1).
Figure 9.2: The mycobacterial model rrn operon.
The figure illustrates a model rrn operon with the features identified in either one or both mycobacterial rrn operons in the present study. The promoter is depicted as a single unit, P, which in M.tb represents two promoters and in M.sm rrnB represents one promoter. Features such as UAR and leader region are flanking the promoter unit and are found in both operons. The UAR enhances promoter activity indicated by a green upwards arrow. The UAR of both operons contain a palindromic sequence (II), which could act as a binding site for a transcription factor, indicated as X. In addition the UAR of M.tb rrnA contains a second palindromic sequence (I). The leader region of both operons suppresses expression (indicated with a red downward arrow) presumably on post-initiation level. The first half of the rrn leaders contain a homologue of the E.coli nut site implicated in rrn anti-termination, although the BoxB element is absent from the rrnB operon (Ji et al., 1994a; Ji et al., 1994b). M.tb NusA binds to this half of the M.tb leader region but the exact binding site has not been identified (indicated by a question mark). M.tb NusB displays a weak interaction with this region and a much stronger interaction with the full-length rrn leader. It has not been determined whether NusB requires the entire leader or just sequences in the second half but a model in which the leader contains two binding sites (nut and nut*) for NusB is a possibility (nut* does not mean that a second nut-like site has been identified in the downstream half of the leader, only that it may bind NusB).
10 Future prospects

The work presented here contains a number of opportunities for further investigations. A primary objective is to obtain a functional *M. tb* in vitro transcription system. Having such a system will further the understanding of the basic mechanisms of transcription in *M. tb*. An in vitro transcription system based on the *M. tb* RNAP would furthermore aid in the characterisation of the *rrnA* and *rrnB* leader regions, identifying pause and termination signals in these parts of the two *rrn* operons. The in vitro requirement for Nus factors in *M. tb* anti-termination could be determined by investigating RNAP read-through at Rho-dependent terminators. Finally an in vitro transcription system should clarify whether the UARs of the two *rrn* operons are factor dependent or factor independent.

A further characterisation of the in vitro interactions between Nus factors and *rrn* leader regions should identify the exact binding sites of NusA and NusB. This could be accomplished by EMSA, toe-printing or by SELEX.

Mutant proteins of NusA and NusB could be used in EMSA to determine the structural requirements of these two proteins in the interaction with the *rrn* leader region. These could include a NusA protein in which one of the two KH domains has been deleted and a monomeric NusB protein.
A number of additional reporter constructs should identify differences and similarities between the two \textit{rrn} leader regions in terms of regulation, whereas the existing reporter strains could be used under different growth conditions to identify the nature of the elements. Together with the already obtained results the suggested experiments should give a comprehensive insight into the regulation of \textit{rrn} expression in mycobacteria.
Acknowledgements

Some of the people who helped me through this project.

First I’d like to thank my boss and supervisor Dr Jo Colston. Jo has unfailingly kept his door open to me giving advice and comments. Thanks for your patience and help and for giving me the chance to obtain a PhD.

Thanks to Dr K G Papavinasasundaram for supervision, comments and company and to Dr B. Gopal for help and proteins and company. Also thanks to Dr Cox for starting off the project and to the EU for financing it.

To the rest of my lab (past and present) Colin, Richard, Sapna, Krishna, Bosco and Cath (who sort-of-belongs to our lab). And thanks to the rest of the division for comments, company, coffee and the occasional cake.

I’d like to thank Dr E. O. Davis for plasmids, pEJ414 and pEJ425, Dr Peter Sander for strains, SMR5, KO14 and KO16 and Dr B. Gopal for Nus proteins.

On the private side I’d like to thank family and friends, in DK and in UK, for patience and understanding (I think) and for still being around.

Last but obviously not least I’d like to thank my best friend and future husband, Finn, for critical comments and excellent emotional assistance: love, patience, understanding, fun, calm…….
References


Grange, J. M. (1996) J appl bact symp supp, 81, 1S-9S.


Appendices
Appendix A: Promoter and leader sequences
Sequence of the promoter and leader region of the *M.tb* *rrnA* operon.
Promoters, P1 and PCL1 are indicated in green, transcription start point as asterisks and anti-termination motifs, BoxB and BoxA are highlighted in yellow. The stop codon of the upstream *murA* gene is indicated in red and the mature portion of the 16S rRNA is shown in bold italics. The grey shaded nucleotides indicated the difference in upstream region between KAM58 and KAM60.
Sequence of the promoter and leader region of the M. sm rRNA operon.

Promoters, P1, PCL1 and P2, are indicated in green, transcription start point as asterisk and anti-termination motifs, BoxB and BoxA are highlighted in yellow. The stop codon of the upstream gene, murA, is indicated in red. The mature portion of the 16S rRNA is shown in bold italics. The reporter construct KAM59 includes the entire sequence shown plus additional basepairs both upstream and downstream of the shown sequence.

M. sm rRNA

GCCATCGGCGAGGACGTATGTCACCACTCCAGGACCTACGGGCTCTGCCCTCAGTAGCGACTCTGCCTGCT
CGCCGCGCGGCCGACGCGACGCGGGCGACGCGACGCGACGCGACGCGACGCGACG
CGAGACGCACGCGACGCGACGCGACGCGACGCGACGCGACGCGACGCGACGCGACGCGACG

ATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG

APGACCGGCAACAACCCCACTCCTGATGGTTTCACGAGGACTGGGAGCACTACGCTGAAGACGACG

TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT

AGGCGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

TAGTCAGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG

©
M. sm rrnB

The promoters, P₉ is indicated in green, transcription start point as asterisk and the anti-termination motif, BoxA is highlighted in yellow. The grey shaded nucleotides indicate regions which were not included in the full-length reporter construct, KAM55. The stop codon of the upstream gene, tyrS, is indicated in red and the mature portion of the 16S rRNA is shown in bold italics.

CGGCAGCGGTGGACCTGCCCAGGGCGGCGTTATGTGAAACACGCGCGTGCAAAGCGGAGTGGGT
ACCACAAGAATCCGATTTCCTGCATGAGGGCTGCTGTGTTGGCGCTGGCGAAGCCACATTCGCCGG
GTGCGCCGCGTGGACCCGCTGACCAGGGAAA TAGCCCTCTGACCTGGGATTTGACTCCCAGTTTCCA
AGGACGTAACTTA TCCAGCTGACTGACACCGCCACTGGGGAAGCAGAAGCAGAATGGCCGAGAGAC
TCCCACTAAGGTGGGGGATCCTCGCTGCACATAGAATGCAAGCGGGTTTTAGCCGCGGATCTTGC
GGCAAGTCGGCGCGTGTTTGTGAAACTCAATAGTGTGTTGTTGTTGTTTGTGTTTGTGTTTGTCC
GCCTCTTTTTCCGTTTAGGGTGATTTTGTGATGCCAGTTTTGGTGTCTTTTTGTAGTCAAGCTTT
TCTCTGATTTGTAATTACCTGCTTTTGGATGGTGTGTTTTGTTGT
**Appendix B: List of reporter plasmids**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pEJ414/425</td>
<td>see Figure 4.1</td>
<td></td>
</tr>
<tr>
<td>pKA34</td>
<td>MP(^a) of <em>M. tb rRNA</em> PCL1</td>
<td>B8264, B8265(^b)</td>
</tr>
<tr>
<td>pKA35</td>
<td>MP of <em>M. tb rRNA</em> P1</td>
<td>B8266, B8267(^b)</td>
</tr>
<tr>
<td>pKA36</td>
<td>MP of P2 consensus promoter</td>
<td>B8268, B8269(^b)</td>
</tr>
<tr>
<td>pKA37</td>
<td>MP of <em>M. sm rRNA</em> P(_B) (−36 to −8)</td>
<td>B8270, B8271(^b)</td>
</tr>
<tr>
<td>pKA38</td>
<td>MP of <em>M. tb rRNA</em> PCL1 (T−7C)</td>
<td>C1890, C1889(^b)</td>
</tr>
<tr>
<td>pKA39</td>
<td>MP of <em>M. tb rRNA</em> PCL1 (T−12C)</td>
<td>C1887, C1888(^b)</td>
</tr>
<tr>
<td>pKA40</td>
<td>MP of <em>M. tb rRNA</em> PCL1 A−10</td>
<td>C1883, C1884(^b)</td>
</tr>
<tr>
<td>pKA41</td>
<td>MP of <em>M. tb rRNA</em> PCL1 A−35</td>
<td>C1885, C1886(^b)</td>
</tr>
<tr>
<td>pKA42</td>
<td>MP of <em>M. ch rRNA</em> PCL1</td>
<td>C3766, C3769(^b)</td>
</tr>
<tr>
<td>pKA43</td>
<td>MP of <em>M. tb rRNA</em> PCL1 (TG−14/−15CC)</td>
<td>C3767, C3768(^b)</td>
</tr>
<tr>
<td>pKA48</td>
<td><em>M. tb rRNA</em> −300(P1), +104(PCL1)(^c)</td>
<td>F3676, F3677(^c)</td>
</tr>
<tr>
<td>pKA49</td>
<td><em>M. tb rRNA</em> −300 to +41(P1)</td>
<td>F3676, F4873(^c)</td>
</tr>
<tr>
<td>pKA50</td>
<td><em>M. tb rRNA</em> −300 to −8(P1)</td>
<td>F3676, F8985(^c)</td>
</tr>
<tr>
<td>pKA51</td>
<td><em>M. tb rRNA</em> −37 to +41(P1)</td>
<td>F4088, F4085(^b)</td>
</tr>
<tr>
<td>pKA52</td>
<td><em>M. tb rRNA</em> −84 to+104(PCL1)</td>
<td>F5028, F3677(^c)</td>
</tr>
<tr>
<td>pKA53</td>
<td><em>M. tb rRNA</em> −84 to−7(PCL1)</td>
<td>F4086, F4087(^b)</td>
</tr>
<tr>
<td>pKA54</td>
<td><em>M. tb rRNA</em> −36 to+104(PCL1)</td>
<td>F8989 , F3677(^c)</td>
</tr>
<tr>
<td>pKA55</td>
<td><em>M. sm rRNA</em> −140 to +207</td>
<td>F10926, F11010(^c)</td>
</tr>
<tr>
<td>pKA55(_h)</td>
<td>As pKA55 but in pEJ425 background</td>
<td>F10926, F11010(^c)</td>
</tr>
<tr>
<td>pKA56</td>
<td><em>M. sm</em> rrnB –140 to +10</td>
<td>F10926, F13746&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>pKA57</td>
<td><em>M. sm</em> rrnB –40 to +207</td>
<td>F13745, F11010&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA58</td>
<td><em>M. tb</em> rrnA –345(P1), +219(PCL1)&lt;sup&gt;g&lt;/sup&gt; (0.7 kb <em>XbaI-HindIII</em> frgm. from pMtbA, App. C)</td>
<td></td>
</tr>
<tr>
<td>pKA58&lt;sub&gt;h&lt;/sub&gt;</td>
<td>As pKA58 but in pEJ425 background</td>
<td></td>
</tr>
<tr>
<td>pKA59</td>
<td><em>M. sm</em> rrnA –345(P1), +229(PCL1)&lt;sup&gt;g&lt;/sup&gt; (0.8 kb <em>XbaI-HindIII</em> frgm. from pMsmA, App. C)</td>
<td></td>
</tr>
<tr>
<td>pKA60</td>
<td><em>M. tb</em> rrnA –300(P1), +219(PCL1)</td>
<td>F3676, F13396&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA61</td>
<td><em>M. tb</em> rrnA –84 to +219(PCL1)</td>
<td>F5028, F13396&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA62</td>
<td><em>M. tb</em> rrnA –36 to +219(PCL1)</td>
<td>F3676, F13396&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA64</td>
<td><em>M. tb</em> rrnA –36 to +127(PCL1)</td>
<td>F3676, F24779&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA67</td>
<td><em>M. tb</em> rrnA –8 to +104(PCL1) <em>HindIII</em> frgm. at position +10 in pKA56</td>
<td>F30406, F3677&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA68</td>
<td><em>M. tb</em> rrnA –8 to +219(PCL1) <em>HindIII</em> frgm. at position +10 in pKA56</td>
<td>F30406, F13396&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA69</td>
<td><em>M. sm</em> rrnB (–140 to –37) <em>XbaI</em> frgm (<em>M. sm</em> rrnB upstream region in pEJ414)</td>
<td>F10926, F26814&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA70</td>
<td><em>M. sm</em> rrnB (–140 to –37) <em>XbaI</em> frgm into pKA37 (8 bp insert between core promoter and UAR)</td>
<td>F10926, F26814&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA71</td>
<td>same as pKA37 without CC insert</td>
<td>F31480, F31481&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA72</td>
<td><em>M. sm</em> rrnB P&lt;sub&gt;b&lt;/sub&gt; (–39 to –8)</td>
<td>F31483, F31482&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA73</td>
<td><em>M. sm</em> rrnB P&lt;sub&gt;b&lt;/sub&gt; (–53 to –8)</td>
<td>F31485, F31484&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA74</td>
<td><em>M. sm</em> rrnB (–140 to –37) <em>XbaI</em> frgm into pKA71 (6 bp insert between core promoter and UAR)</td>
<td>F10926, F26814&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA75</td>
<td><em>M. sm</em> rrnB UAR <em>XbaI</em> frgm and pKA72 both digested with <em>XbaI</em> and end-filled (13 bp insert between core promoter and UAR)</td>
<td></td>
</tr>
<tr>
<td>pKA76</td>
<td>Deletion of two bp in pKA75 by SDM (11 bp insert between core promoter and UAR)</td>
<td>F36686, F36687</td>
</tr>
<tr>
<td>pKA77</td>
<td><em>M. sm</em> rrnB P&lt;sub&gt;b&lt;/sub&gt; (–41 to –8)</td>
<td>F36847, F36850&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA78</td>
<td><em>M.sm rrnB</em> Pₜ (−52 to −8)</td>
<td>F36849, 36848&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>pKA79</td>
<td><em>M.sm rrnB</em> Pₜ (−60 to −8)</td>
<td>F40052, F40053&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA80</td>
<td><em>M.sm rrnB</em> Pₜ (−70 to −8)</td>
<td>F40054, F40055&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA81</td>
<td><em>M.sm rrnB</em> Pₜ (−80 to −8)</td>
<td>F40056, F40057&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA82</td>
<td>Insert of two bp in pKA56 by SDM</td>
<td>F40626, F40627</td>
</tr>
<tr>
<td>pKA83</td>
<td>Insert of one bp in pKA56 by SDM</td>
<td>F40624, F40625</td>
</tr>
<tr>
<td>pKA84</td>
<td><em>M.sm rrnB</em> Pₜ (−36 to +10)</td>
<td>F43876, F43875&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA85</td>
<td><em>M.sm rrnB</em> Pₜ (−80 to +10)</td>
<td>F43878, F43877&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA87</td>
<td><em>M.sm rrnB</em> −140 to −8</td>
<td>F50702, F50705&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA88</td>
<td><em>M.tb rrnA</em> PI core promoter</td>
<td>F59382, F59383&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA89</td>
<td><em>M.tb rrnA</em> PCL1 core promoter</td>
<td>F59380, F59381&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKA90</td>
<td><em>M.sm rrnB</em> Pₜ core promoter</td>
<td>F59378, F59379&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

All inserts were *XbaI-HindIII* fragments in pEJ414 unless otherwise stated.

SDM = site directed mutagenesis

- a) **MP = Minimal Promoter:** the region from 5' end of −35 box to 3' end of −10 box; all MP fragments (except pKA71) were flanked by CC to separate the promoter elements from the restriction site.
- b) Promoter insert generated by annealing of oligonucleotides
- c) Promoter insert generated by PCR amplification
- d) CC-insert between promoter fragment and restriction site
- e) Numbers refer to the position relative to the transcription start point of the following bracketed promoter
- f) Consensus elements and *M.sm* spacer according to (Gonzalez-y-Merchand *et al.*, 1997)
### Appendix C: List of plasmids (other)

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET15b</td>
<td><em>E. coli</em> expression vector (Studier et al., 1990)</td>
<td></td>
</tr>
<tr>
<td>pUC19</td>
<td>(Yanish-Perron et al., 1985)</td>
<td></td>
</tr>
<tr>
<td>pGEM-3Zf(+)</td>
<td>Vector for <em>in vitro</em> transcription (Promega)</td>
<td></td>
</tr>
<tr>
<td>pMtbA</td>
<td>0.7 kb <em>PstI</em>-HindIII frgm of <em>M. tb</em> <em>rrnA</em> in pUC19</td>
<td>F13395, F13396</td>
</tr>
<tr>
<td>pMsmA</td>
<td>0.8 kb <em>PstI</em>-HindIII frgm of <em>M. sm</em> <em>rrnA</em> in pUC19</td>
<td>F13395, F13396</td>
</tr>
<tr>
<td>pusB</td>
<td><em>M. sm</em> <em>rrnB</em> (−140 to −37) XbaI fragment in pUC19</td>
<td>F10926, F26814</td>
</tr>
<tr>
<td>TOPO-β</td>
<td>3.8 kb frgm of <em>M. tb</em> <em>rpoB</em> in pCR-Blunt TOPO II</td>
<td>F47446, F45576</td>
</tr>
<tr>
<td>TOPO-βmut</td>
<td>mutated version of TOPO-β by SDM</td>
<td>F52118, F52119</td>
</tr>
<tr>
<td>TOPO-β’</td>
<td>4.3 kb frgm of <em>M. tb</em> <em>rpoC</em> in pCR-Blunt TOPO II</td>
<td>F46577, F46578</td>
</tr>
<tr>
<td>TOPO-β’mut</td>
<td>mutated version of TOPO-β’ by SDM</td>
<td>F53396, F53397</td>
</tr>
<tr>
<td>TOPO-ω</td>
<td>333 bp frgm of Rv1390 in pCR-blunt TOPO II</td>
<td>F53224, F53225</td>
</tr>
<tr>
<td>TOPO-sigA</td>
<td>1.7 kb frgm of <em>M. tb</em> <em>sigA</em> in pCR-blunt TOPO II</td>
<td>F53395, F40459</td>
</tr>
<tr>
<td>pKAα</td>
<td>1.2 kb <em>NdeI</em>-BamHI frgm in pET15b</td>
<td>F32919, F32920</td>
</tr>
<tr>
<td>pKAβ</td>
<td><em>NcoI</em>-EcoRI frgm from TOPO-β’mut in pET15b</td>
<td></td>
</tr>
<tr>
<td>pKAβ’</td>
<td><em>NdeI</em>-BamHI frgm from TOPO-β’mut in pET15b</td>
<td></td>
</tr>
<tr>
<td>pKAω</td>
<td><em>NdeI</em>-EcoRI frgm from TOPO-ω in pET15b</td>
<td></td>
</tr>
<tr>
<td>pSigA</td>
<td><em>AflIII</em>-EcoRI frgm from TOPO-sigA in pET15b (NcoI/EcoRI)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D: Oligonucleotides

B8264:
CTAGACCTTGACTCCATTGCCGGATTTGTATTAGACTCCA

B8265:
AGCTTGGAGTCTAATACAAATCCGGCAATGGAGTCAAGGT

B8266:
CTAGACCGTCTCGGTGCCGAGATCGAACGGGTATGCTCCA

B8267:
AGCTTGGAGCTACCCGTCGGATCTCGGCACCGAGACGGT

B8268:
CTAGACCTTGACTCCATTTTCCACGTCGGCAGACGGT

B8269:
AGCTTGGAGAAATAATACGAGGTTCGTCGCGAAGTCAAGGT

B8270:
CTAGACCTTGACTCCCAGTTTCCAAGGACGTAAACTTCCA

B8271:
AGCTTGGGATTTACGTCCTTGGAAACTGGGAGTCAAGGT

C1883:
CTAGACCTTGACTCCATTGCCGGATTTGTATCCCA

C1884:
AGCTTGGATACAAATCCGGCAATGGAGTCAAGGT

C1885:
CTAGACCCCATTTGCGGATTGTATTAGACTCCA

C1886:
AGCTTGGAGTCTAATACAAATCCGGCAATGGGAGT
C1887:
CTAGACCTTGACTCCATTGCCGGATTTGTATCAGACTCCA

C1888:
AGCTTTGGAGTCTGATACAAATCCGGCAATGGAGTCAAGGT

C1890:
CTAGACCTTGACTCCATTGCCGGATTTGTATTAGACCCCA

C1889:
AGCTTTGGGCTCTAATACAAATCCGGCAATGGAGTCAAGGT

C3766:
CTAGACCTTGACTCAAGTTCATTGGACTTGGTACAGTCCA

C3767:
CTAGACCTTGACTCAAGTTCATTGGACTCCGTACAGTCCA

C3768:
AGCTTTGGGACTGTACACGGAGTCTCCAATGAACCTTGAGTCAAGGT

C3769:
AGCTTTGGGACTGTACCAAGTCCAATGAACTTGAGTCAAGGT

F3676:
GCAGTCTAGAATCGATGATCACGGAGAAC

F3677:
ATACAAGCTTTATCAAGAGCATGGCCAAAAATA

F4085:
AGCTTTGACCAGGTTCCGTTCAATCCATAGATATCCATAGGGTAGGACCAGTCG
CCTAACAGCATACCCGGATGGCCGACTCGACCGAGCGGT

F4086:
CTAGAGTATAAGGGACGGTACCTATGATATCTATGGATGACCCGAAACC
TTGGTCTTGACTCCATTGGGATAATGATAGACTCCA

F4087:
AGCTTTGGAGTCTAATACAAATCCGGCAATGGAGTCAAGACCAGTCGGT
CATCCATAGATATCCCATAGGTGACCGTCGCTAACACT
F4088:
CTAGACCGTCCTCGGTGCCGAGATCGAACGGGTATGCTGTTAGGCGACG
GTCACCTTATGGATATCTATGGATGACCGAACCTGGTCA

F4873:
CGATAAGCTTACCAGGTTGGGTCTCATCCATAGAT

F5028:
GCAGTCTAGAGTTAGGCGACGGTCACCT

F8989:
AGCGTCAGATTGACTCCATTGCCGGATTTGTATTAG

F8985:
GCAGAAGCTTAGCATACCCGTTCCGATCTCGGCA

F10926:
AGCTTCTAGACTGCATGAGGGCTGGCTGGTG

F11010:
AGCTAAGCTTAAACGGGAAAAAGAGGCAGGACAAA

F13395:
ACGTTGCGACCTGCGTTCC

F13396:
AGCTAAGCTTCTGAGCCAGGATTCAACTCTCCA

F13745:
AGCTTCTAGAGGATTTGACTCCCAGTTTCCAAGGAC

F13746:
AGCTAAGCTTTCTGCTGACCTGGAATAAGTTACGCTCTT

F24779:
AGATAAGCTTTGGGGACGCCCCCGACAA

F26814:
AGGTTCTAGAATCCCAAGGTCAGAGG
F30406: AGCTAAGCTTCTGGCAGGGTTGCCCCGA

F31482: AGCTTAAGTTACGTCTTGGAAACTGGGAGTCAAATCT

F31483: CTAGAGATTTGACCTCCCCAGTTTTCCAAGGACGTAACTTA

F31484: AGCTTAAGTTACGTCTTGGAAACTGGGAGTCAAATCCCCAGGTCAGA GGGT

F31485: CTAGACCCCTCTGACCTGGGGATTGGACTCCCATTTCCAAGGACGTAA CTTA

F32919: TCGAATCGCATATGCTGATCTCACAG

F32920: ATGAGGATCCCCCTTCGGCTCGGTCGTCGTG

F36686: CCTGGGGATTCTAGCTAGAGTTGACTCCCAGTTTTCCAAGGACG

F36687: CGTCCTTGGAAACTGGGAGTCAACTCTAGCTGAATCCCAAGGACG

F36847: CTAGAGGGATTTGTAGCTCCCATTTCCAAGGACGTAACTTA

F36848: AGCTTAAGTTACGTCTTGGAAACTGGGAGTCAAATCCCCAGGTCAGA GGT

F36849: CTAGACCTCTGACCTGGGGATTGGACTCCCATTTCCAAGGACGTAAAC TTA
F36850:
AGCTTAAAGTTACGTCCCTTGAAACTGGGAGTCAAATCCCT

F40052:
CTAGAAAAATAGCCCTCTGACCTGGGATTTGACTCCCAAGTTTCCAAG
GACGTAACCTTA

F40053:
AGCTTAAAGTTACGTCCCTTGAAACTGGGAGTCAAATCCCCAGGTCAGA
GGGCTATTATT

F40054:
CTAGACTGACCAGGGAAATAGCCCTCTGACCTGGGATTTGACTCCC
AGTTTCCAAGGACGTAACCTTA

F40055:
AGCTTAAAGTTACGTCCCTTGAAACTGGGAGTCAAATCCCCAGGTCAGA
GGGCTATTATTCCCTGGTCAGT

F40056:
CTAGAGTGACCGCGTCTGACCAGGGAAATAGCCCTCTGACCTGGGA
TTTGACTCCCAAGTTTCCAAGGACGTAACCTTA

F40057:
AGCTTAAAGTTACGTCCCTTGAAACTGGGAGTCAAATCCCCAGGTCAGG
GGGCTATTATTCCCTGGTCAGACGCGGTCACT

F40459:
AACGGGATCCCTCGGCGGGCGCTCTCAGTCCAGGTAGTC

F40624:
GCCCTCTGACCTGGGATTTTGACTCCCAGTTTCC

F40625:
GGAAACTGGGAGTCAAAATCCCCAGGTCAGAGGGC

F40626:
GCCCTCTGACCTGGGATTTTGACTCCCAGTTTCC
F43875:
AGCTTTTCGCTCTGACCTGGAATAAGTTACGTCCCTTGAAAACCTGGGAGC
AAT

F43876:
CTAGATTGACTCCCAGTTTTCCAAGGACGTAACTTATTCCAGGTGAGAG
CGAA

F43877:
AGCTTTTCGCTCTGACCTGGAATAAGTTACGTCCCTTGAAAACCTGGGAGC
AATCCCAAGGTCAGAGGCTATTTTCCCTGGGTCAGACGGCGGTCACT

F43878:
CTAGAGGTGCACCGGCTCTGACCAGGGAAAATAGCCCTCTGACCTGGGGA
TTTGACTCCCAGTTTTCCAAGGACGTAACTTATTCCAGGTGAGACGAA

F46576:
GCCGCAGCGCGTCGAGATGAT

F47446:
GCAGATTCCCCAAGAGCAA

F50702:
CTAGACTGCATGAGGGCTGGCTGTGTTGCAGGCAGTGCAAGGGCAAGCGCCACA
TTGCGGGGCGTCGGCGCCGATTGACGACTGACCCAGGAAATAGCCCT
CTGACCTGGGAGATTGACTCCCAGTTTTCCAAGGACGTAACTTAA

F50705:
AGCTTTAAGTTACGTCCCTTGAAAATGCGGAGTCAAATCCCCAGGTGAGA
GGGCTATTTTTCCCTGTCAGACCGGCTCAACCGCGCCGCAACCACCCCGCAT
GTGGCCTTCGACCGCCGAAACACGCGGAGGCGCCACTCAGTGCAGT

F52118:
CCGCCAGTGTCGTGCGGCAATTCCGCAATGGATCCCGCCAGAGC

F52119:
GCTCTGCGGGAAATCCTGCGGGAATTCCAGCAGCACACTGCGGG

F53224:
CATATGAGTATTCGCGAGTCCGAGGCGC
F53225: AGGCCTTGCCCTGCTACTCG

F53395: TCTAACATGTCAGCGACCAAAGCAACGCGAC

F53396: GTTAGGGGAAGGGAGTCTATGCTCGACGTCACCTTTCTTCG

F53397: CGAAGTGGAGTCAGACGTCACTCCCTTTCCCCTAAC

F59378: CTAGATTGACTCCAGTGGGACTAACTTATTCCAGGA

F59379: AGCTTCCCTGAATAAGTACGTCTTTGGAAACTGGAGTCAAT

F59380: CTAGATTGACTCCATTGCCGGATTTGTATTAGACTCCAGGA

F59381: AGCTTCCCTGCAATACGCTAACATCCGGGAATGAGTCAAT

F59382: CTAGAGTCTCGGTGCCGAGATCGAACGGGTATGCTGTTAGGCGA

F59383: AGCTTCCCTGCAATACGCTAACATCCGGGAATGAGTCAAT

Pmimt: ACGAGGGGCATTCACACCAGATTG

lac518R: TTTCCAGTCAAGCGTTTGTTAAA
Appendix E: Media, buffers, gels

Growth of bacteria

L-Broth (1000 ml)

10 g Bacto Tryptone, 5 g Yeast Extract, 1 g NaCl

Modified Dubos Medium (1000 ml)

1 g KH₂PO₄, 6.25 g Na₂HPO₄·12 H₂O, 1.25 g Na₃-citrate, 0.6 g MgSO₄·7 H₂O

2 g Asparagine, 2 g Casamino acids, 5 ml 10% Tween 80, 20 ml 10% Glycerol

Adjust to pH 7.2

Mix with 40 ml Dubos Medium Albumin (Difco) after autoclaving
**SDS-PAGE: buffers and gels**

**SDS sample buffer (Laemmli)**

10% β-mercaptoethanol (v/v); 6% SDS (w/v); 0.2 mg/ml Bromophenolblue; 6% glycerol; 0.25M TrisHCl pH 6.8

**4 x Stacking Buffer (1000 ml; pH 6.8)**

60.5 g Trizma base, 4 g SDS

**4 x Resolving Buffer (200 ml; pH 8.8)**

36.4 g Trizma base, 0.8 g SDS

**10% (12%) Resolving Gel**

5 (6) ml Protogel (Bio-Rad, 30% acrylamide/0.8% bisacrylamide);

3.75 ml 4 x Resolving buffer; 6.25 (5.25) ml water; 30 µl APS (25%)

10 µl TEMED

**4% Stacking Gel**

1.33 ml Protogel; 2.5 ml Stacking buffer; 6.17 ml water; 20 µl APS (25%)

10 µl TEMED
Electrophoretic mobility shift assays

8% non-denaturing acrylamide gel (1 x TBE)

10 ml Protogel (30% acrylamide/0.8% bisacrylamide); 3.75 ml 10 x TBE;
23.8 ml water; 56 µl APS (25%); 56 µl TEMED

For native gels with 1/2 x TBE use 1.88 ml 10 x TBE and 25.68 ml water.

5 x DNA-EMSA buffer

100 mM HEPES pH 7.5; 1.0 mM EDTA; 50 mM (NH₄)₂SO₄; 5 mM DTT;
75 mM KCl; 75 mM MgCl₂; 250 µg/ml BSA (PIERCE); 10 µg/ml Poly dI:dC (Roche)

2 x RNA-EMSA buffer

40 mM HEPES pH 7.5; 0.4 mM EDTA; 20 mM (NH₄)₂SO₄; 2 mM DTT
30 mM MgCl₂; 400 mM KCl; 4 µg/ml tRNA (Roche); 100 µg/ml BSA (PIERCE)
0.5% Triton x-100 (SIGMA)
Other buffers and gels

Z buffer (1000 ml, pH 7.0)

16.1 g Na₂HPO₄·7H₂O, 5.5 g Na₂HPO₄·H₂O, 0.246 g MgSO₄·7H₂O,
0.75 g KCl

Adjust pH to 7.0 with NaOH

Add β-mercaptoethanol freshly to a final concentration of 38 mM

10 x TBE (1000 ml, pH 7.0)

12.11 g Trizma base; 61.8 g Boric Acid; 18.6 g EDTA

6% Acrylamide sequencing gel

40.4 g urea; 8.0 ml 10 x TBE; 9.6 ml Long Ranger (50% stock solution, JT Baker)

34 ml water; 280 μl APS (25%)

The ingredients are mixed thoroughly to dissolve the urea and filtered before
adding 25 μl TEMED and pouring the gel.
Appendix F: Bacteria

E. coli DH5α (Sambrook et al., 1989; Gibco BRL)

E. coli BL21 DE3 pLysS (Studier et al., 1990; Novagen)

M. smegmatis mc²155 (Snapper et al., 1990)

M. smegmatis SMR5 (mc²155, rpsL (streptomycin resistance; Sander et al., 1996)

M. smegmatis KO14 (rpsL⁺, rrnA; Sander et al., 1996)

M. smegmatis KO16 (rpsL⁺, rrnB; Sander et al., 1996)

Reporter strains were all derivatives of M. smegmatis mc²155 numbered according to the integrated plasmid