Antitermination in *Mycobacterium tuberculosis*

Shaun Kevern Twyman Cochrane

Thesis submitted in partial fulfilment of the requirements of the University College of London for the degree of Doctor of Philosophy

April 2004

Divisions of Mycobacterial Research and Protein Structure
National Institute for Medical Research
Mill Hill
London, U.K.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Mycobacterium tuberculosis **is the leading cause of death from a single infectious agent. The varying efficacy of the BCG vaccine and the emergence of multi-drug resistant strains of *M. tuberculosis* have made it essential that novel drug and vaccine targets are identified. The antitermination mechanism, probably used in transcriptional regulation of the single *rrn* operon, is one such target. Antitermination is a mechanism by which RNA polymerase is able to transcribe through both Rho-dependent and –independent terminators. Antitermination in association with Nus (N-utilising) factors was initially discovered within the λ-phage where it regulates the transcription of early and late genes. Subsequent investigations in *Escherichia coli* demonstrated antitermination and the Nus proteins were involved in transcriptional regulation of the seven *rrn* operons present in the genome.**

The presence of Nus homologs and *Nut* sequences (the assembly sites for the antitermination complex) within the *M. tuberculosis* genome indicate that antitermination may occur during *rrn* operon transcription. The aim of this project has thus been an understanding of the characteristics, functions and interactions of the Nus proteins within the *M. tuberculosis* antitermination complex.

A definite, yet likely to be weak, interaction between NusB and NusE is shown. The stoichiometry of this interaction (NusB monomer or dimer bound to NusE monomer) was investigated and results may indicate a heterodimer. The dissociation equilibrium constant for the NusB dimer was estimated to be 8 × 10^{-8} M. The NusG protein was characterised and shown to be monomeric in solution with a highly elongated shape with approximately 20 % α-helical secondary structure. A C-terminal domain, containing the KOW *rRNA* binding motif, was identified by limited proteolysis. Lastly the Rho termination factor was over-expressed and purified and interactions with NusA and NusG investigated. No interactions have yet been detected.

Acknowledgements

First and foremost I would like to thank my fiancée, Vicky Tsoni, for her love, her understanding, her remarkable tolerance and patience, her endless encouragement and support, her wonderful smile and her laughter and for her ability to make sure there were more good times to remember than bad times over the last four years and, even then, keeping me smiling through the bad times.

I would also like to thank my family for their love, a listening ear at the end of a phone line and encouragement of my studies and especially my "Nana" who, throughout my life, has shown me love and laughter, had an amazing calming influence on me as well as an incredible interest in every aspect of my studies.

I would like to thank Ian Taylor for invaluable help and advice in completing the PhD and to Steve Smerdon for his supervision and both for having to spend hours reading numerous drafts of the thesis.

I would also like to thank Val Ennis-Adeniran for her never-ending encouragement, support and friendship and her ability to raise my self-esteem and self-belief during the lowest of times.

Lastly, I would like to thank the late Jo Colston for his continued support and encouragement during the first three years of my PhD. His achievements in mycobacterial research and his incredible determination in and contribution to the worlds fight against tuberculosis and leprosy have had an enormous impact on my belief in biomedical science and its necessity for the improvement of the life of millions affected by the world's diseases.

Table of Contents

Table of Figures

Chapter 1 : Introduction

Chapter 3 : Biochemical and Biophysical Techniques

Chapter 4 : Results – The Antitermination Complex and NusB and NusE.

Chapter 5 : Results – NusG and the Rho Termination Factor

Chapter 6 : Discussion

List of Tables

Abbreviations

°C	Degrees Centigrade
μg	Micrograms (10^{-6} grams)
μl	Microlitres (10^{-6} litres)
Å	Angstroms
AIDS	Acquired Immuno-Deficiency Syndrome
AMP	Adenosine-Mono-Phosphate
BCG	Bacille-Calmette-Guérin
cAMP	Cyclic Adenosine-Mono-Phosphate
CAP	Catabolite Activator Protein
CD	Circular Dichroism
CFE	Cell Free Extract
DTT	Dithiothreitol
DLS	Dynamic Light Scattering
EDTA	Ethylenediamine-tetraacetic Acid
ELISA	Enzyme-linked Immunosorbant Assay
ESI MS	Electrospray Ionisation Mass Spectroscopy
F.T.	Flow Through
GST	Glutathione-S-Transferase
HIV	Human Immunodeficiency Virus
IPTG	Isopropyl β-D-thiogalctopyranoside
K_a	Association Constant
K_d	Dissociation Constant
kb	Kilobases
kDa	Kilodaltons
MALDI MS	Matrix Assisted Laser Desorption/ Ionisation Mass Spectroscopy
mg	Milligrams (10^{-3} grams)
ml	Millilitres (10^{-3} litres)
mM	Millimolar (10^{-3} molar)
mRNA	Messenger RNA
MWCO	Molecular Weight Cut Off

N(A/G) MP	Nucleoside (Adenine/Guanine) Mono-Phosphate
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
pI	Isoelectric Point
pmol	Picomoles
PVDF	Polyvinylidene Flouride
RNAP	RNA Polymerase
RNAPα	RNA Polymerase α-subunit
rpm	Revolutions per minute
rrn	ribosomal RNA operon
rRNA	ribosomal RNA
S	Svedberg
SAP	Shrimp-Alkaline-Phosphatase
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	SDS Polyacrylamide Gel Electrophoresis
SLS	Static Light Scattering
TB	Tuberculosis
tRNA	transfer RNA
U	Units
UV	Ultra-violet
λ	Lambda Phage

Chapter 1

Introduction

1.1. *Mycobacterium tuberculosis* – history and biology.

1.1.1. A brief history of *Mycobacterium tuberculosis*.

Mycobacterium tuberculosis, the causative agent of 'Tuberculosis' has been present in the human population for thousands of years and is now one of the leading causes of death in the 20th Century, claiming close to 3 million lives every year (Chan and Iseman, 2002; www.statenj.us/health/cd/tbhistry.htm).

Tuberculosis was first documented around 460 B.C. (then known as phthisis or consumption) where it was acknowledged as the most widespread disease of the time. It was however only in the seventeenth century that disease symptoms and progression were first accurately identified and described. Sylvius, in 1679, showed the consistent presence of tubercles in the lungs which subsequently went on to cause cavities and abscesses. It was only in 1720 that minute microorganisms may cause "consumption" was first hypothesised by the physician, Benjamin Martin. He was also the first to propose mechanisms of transmission and made important epidemiological recommendations that played a significant role in slowing down "consumption" infection. The first dedicated means of treating consumption was provided by Hermann Brehmer (himself a cured sufferer) in the form of the sanatorium. The sanatoria were always found in clean mountainous regions with an abundance of fresh air and patients were well provided with the necessary nutrients.

It was during the mid 19th century that major insight into the disease was made and rapid progression subsequently followed. Jean-Antoine Villimen was the first to show cross species transmission, from human to cattle and cattle to rabbits. However it was Robert

Koch that, arguably, made the most important discovery in the fight against consumption. He developed a stain that allowed him to see *Mycobacterium tuberculosis*, and it was only then that the true cause of consumption was recognized. Unfortunately this discovery had little effect on treatment of the sick and the sanatoria were still the primary means of treatment. In addition artificial pneumothorax (lung collapse) was shown to increase the chances of cure and was introduced as therapy against consumption.

Research on *M. tuberculosis* continued and an important breakthrough was made by Calmettte and Guerin when they passaged the cattle form of tuberculosis through specific culture media resulting in a decrease in virulence of the bacterium. This was the basis of the BCG vaccine against the human form of tuberculosis and is still in extensive use today.

During the Second World War a fundamental breakthrough needed to fully combat the disease was made. Chemotherapy became the customary means of antibacterial treatment using antibiotics such as sulfonamide and penicillin. Both of these molecules were however ineffective against *M. tuberculosis* and it was only in 1943, with the isolation of streptomycin from *Streptomyces griseus* that chemotherapy against *M. tuberculosis* could be carried out with maximum efficiency and minimal toxicity to the host. It was, in fact, applied to the first human TB patient in 1944, who made a dramatic and full recovery. Additional anti-TB drugs were developed soon afterwards and were important in preventing streptomycin-resistant strains from increasing in number. Important anti-TB drugs developed during this era include ρ-aminosalycilic acid (1949), isoniazid (1952), pyrazinamide (1954), cycloserine (1955), ethambutol (1962) and

rifampin (1963). All of these antibiotics are used in combination today to treat and prevent the emergence of *M. tuberculosis* drug resistant strains (htttp://www.statenj.us/health/cd/tbhistry.htm).

1.1.2. Tuberculosis drugs, treatment and relapse.

Drug treatment has, for the last forty years, relied on a combination of drugs with a treatment period lasting from six months to two years, primarily to prevent antibiotic resistance from emerging. Multiple drug therapy was first proposed in 1962 by the Medical Research Council. Streptomycin, para-aminosalycilic acid (PAS) and isoniazid given over a period of two years were recommended. This treatment could also be given on an out-patient basis freeing up hospital beds. The subsequent use and discovery of rifampicin shortened the therapy period to 9 months and subsequent use of pyrazinamide shortened the treatment course even further (Dormandy, 1999).

1.1.3. The modern tuberculosis epidemic.

The use of these drugs resulted in a dramatic decrease in the number of cases in developed countries after the Second World War. During the 1960's there was widespread belief that tuberculosis had been conquered (it was even predicted that tuberculosis could be eradicated in the USA by 2005 (Dormandy, 1999). Unfortunately this opinion was short lived as the drop in incidence began to level out (or even increase) in the mid 1980's. In the USA, the increase in mortality increased by less then 1% in the 1980's but in Eastern Europe and South America the rise was approximately 5%, in South Eastern Asia by about 200% – 300% and in sub-Saharan Africa the rise in mortality was close to 500%. Once again tuberculosis was the leading cause of death

amongst infectious organisms resulting in 7% of all deaths and 26% of avoidable deaths (Dormandy, 1999).

The major reason for this dramatic increase in incidence was the HIV pandemic. Latent tuberculosis infection frequently develops into active tuberculosis as a result of HIV infection and the subsequent deterioration in the host immune system. HIV also increases susceptibility to primary tuberculosis infection and facilitates progression of the disease (Fatkenheuer *et al.*, 1999). Tuberculosis accounts for 11% of AIDS deaths worldwide (www.who.int). It has been suggested that HIV positive individuals were 40 times more likely to contract tuberculosis than healthy individuals and 20 times more likely to die from it (Dormandy, 1999). In a similar fashion tuberculosis stimulates the release of cytokines and decreases the number of CD4 cells thereby speeding up the progression of HIV infection to AIDS (Fatkenheuer *et al.*, 1999).

The close relation between HIV and tuberculosis did not mean that the rise in incidence only occurred in HIV positive individuals. Poorly managed tuberculosis treatment and prevention programmes have led to a marked increase in drug resistant and multiple drug resistant strains. There are currently strains of *M. tuberculosis* that are resistant to at least one of the antibiotics in every country surveyed by the WHO and strains resistant to all of the main antibiotics used to treat tuberculosis have been documented (Chan and Iseman, 2002). These multiple drug resistant strains have emerged as a result of patients who have failed to comply with a drug treatment regime, doctors failing to prescribe the correct drugs or an unreliable drug supply (www.who.int). Lastly increased numbers and movement of immigrants into developed countries is an important factor that has led to a rise in incidence in these countries. In the USA approximately 40% of

tuberculosis cases are in foreign born individuals. Crowded refugee camps are also prime sites for the easy and quick spread of the bacillus.

1.1.4. The mycobacterial family.

The two most medically important members of this family are *Mycobacterium leprae* (the causative agent of leprosy, discovered in 1868) and *Mycobacterium tuberculosis.* The genus Mycobacterium has numerous members isolated from all corners of the environment. Ultimately they can be divided into those that are pathogenic to humans and those that are not, as well as fast growers and slow growers (see Table 1).

Table 1 : Simplified classification of the Mycobacteria (Kanai, 1991)

	Pathogenic	Non Pathogenic
Slow Growth	*M. tuberculosis*	*M. microti*
	M. leprae	*M. lepraemurium*
	M. bovis	*M. gordonae*
	M. africanum	*M. farcinogens*
	M. kansasii	*M. gastri*
	M. marinum	*M. nonchromogenicum*
	M. simiae	*M. terrae*
	M. asiaticum	*M. triviale*
	M. scrofulaceum	*M. paratuberculosis*
	M. szulgai	
	M. avium	
	M. intracellulare	
	M. xenopi	
	M. malmoense	
	M. haemophilum	
	M. ulcerans	
Fast Growth	*M. fortuitum*	*M. smegmatis*
	M. chelonei subsp. Chelonei	*M. phlei*
	M. chelonei subsp absessus	*M. chitae*
		M. flavescens
		M. parafofortuitum
		M. thermoresistible
		M. aurum
		M. duvalli
		M. neoaurum
		M. gilvum
		M. vaccae
		M. komossense
		M. senegalense

1.1.5. *Mycobacterium tuberculosis.*

The *M. tuberculosis* complex of organisms (including *M. tuberculosis*, *M. bovis*, *M. bovis* BCG, *M. africanum* and *M. microti*) are likely to have originated from a *M. tuberculosis* like human pathogen (Brosch *et al.*, 2002) as opposed to the originally believed bovine (*M. bovis*) origins (Cole *et al.*, 1998).

Figure 1 : Electron micrograph of the rod shaped *Mycobacterium tuberculosis* (http://www.wadsworth.org/databank/mycotubr.htm).

M. tuberculosis is a straight or slightly curved rod between 1 and 4 μm in length and 0.3 to 0.6 μm in width (see Figure 1). The bacterium is non-spore forming and non-motile. They are Gram-positive and identified based on their acid-fast properties (they are not decolourised by acids after staining). Optimum growth temperature is 37°C with an optimum growth pH of 6.4 to 7.0. Doubling time is in the region of 14 to 20 hours in culture. *M. tuberculosis* requires aerobic conditions for growth although it is capable of adapting to microaerophilic conditions. In general, mycobacterial genomes contain a high guanine/cytosine content of between 64.0 and 66.4 % (Kanai, 1991).

The entire genome of *M. tuberculosis* H37Rv has now been sequenced (Cole *et al.*., 1998). It is 4400 kilobases (kb) in length and contains 3294 Open Reading Frames (ORF's). Approximately 70 % of the genes have been identified with the remaining coding for proteins of unknown function. The H37Rv genome contains sixteen copies of the insertion sequence IS6110 and six copies of the more stable IS1081 insertion sequence (Cole *et al.*, 1998). The genome also contains all those genes needed for essential amino-acid, vitamin and enzyme co-factor expression. However, some of these pathways may differ from other bacteria (Cole *et al.*, 1998). Thirteen putative sigma factors have been identified along with one hundred additional proteins involved in transcriptional regulation. This is not surprising given the range of environments in which the organism survives. The genome also encodes numerous drug modifying enzymes such as β-lactamases, aminoglycoside acetyl transferases and drug efflux systems. These along with the thick cell wall give *M. tuberculosis* a natural resistance to a wide range of antibiotics (Cole *et al.*, 1998).

The cell wall of *M. tuberculosis* is unusual in that it contains an additional layer above the peptidoglycan cell wall. It is rich in unusual lipids, glycolipids and carbohydrates. The organism also carries out unique biosynthetic pathways to generate the components of the cell wall, such as mycolic acid, mycocerosic acid, phenolthiocerol, lipoarabinomannan and arabinogalactan (Cole *et al.*, 1998). The thick hydrophobic nature allows the wall to act as a highly efficient barrier. This explains the bacilli's resistance to a number of antimicrobial agents but is also vital since it protects the bacteria from oxidative stresses experienced during intracellular infection (Brennan and Nikiado, 1995; www.uct.ac.za/depts/mmi). It may also be responsible for the hosts'

inflammatory reactions and may play a role in pathogenicity (Cole *et al.*, 1998). The thick cell wall is also important during extracellular transmission of the organism as it prevents drying and allows the bacteria to survive for long periods of time in the environment (Mims *et al.*, 1993). Because of the thick, impenetrable cell wall, porins found in the membrane are the principal routes for entry of hydrophilic molecules into the cell.

1.2. Transcription : initiation, elongation, termination.

1.2.1. DNA, RNA and proteins.

DNA found in most viruses and all prokaryotes and eukaryotes contains all the information (in the form of genes) needed for the synthesis of the proteins and enzymes required for the survival, replication and ultimately death of each cell in that organism. The RNA formed during transcription may then have any one of a number of vital roles. The RNA may be used in translation. It may have structural roles and may be used as part of the ribosome itself, or form transfer RNA (tRNA). Alternatively RNA may be used as the genome of viruses or may be involved in regulation itself. The roles transcription and translation play in the cell means they are the most effective steps at which protein and enzyme synthesis can be regulated and thus control the rate at which the organism ultimately grows and divides.

1.3. Prokaryotic RNA polymerase and transcription.

DNA-dependent RNA polymerase (RNAP) is the enzyme used for the conversion of DNA into a complementary RNA sequence; all RNAP's carry out the following reaction:

$$NTP + (NMP)_n \xrightarrow[DNA]{Mg^{2+}} (NMP)_{n+1} + PP_i$$

RNA polymerases are multi-subunit enzymes with a variety of functions, including promoter binding, DNA melting and RNA chain initiation, elongation and termination.

The *E. coli* RNA polymerase has been well characterised and the core RNAP has been shown to be made up of three types of protein subunit, 2 subunits of alpha (α - 36.5 kDa each), a single beta subunit (β - 150.6 kDa) and a single beta prime subunit (β′ - 155.2 kDa). Core RNAP is capable of elongation and termination. A fifth subunit, omega (ω - 10.5 kDa) has been shown to be part of the core enzyme but it is non essential for viability and core enzyme containing the remaining four subunits is capable of *in vitro* transcription (Gentry *et al.*, 1991). Studies have however shown that deletion of ω results in no change in phenotype but that ω may be responsible for RNAP stabilisation (Mukherjee and Chatterji, 1997). The holoenzyme contains an additional subunit, σ^{70} (70.2 kDa) and is capable of specific promoter binding and transcription initiation (Opalka *et al.*, 2000). Sigma^{70} binds very tightly to the core RNAP with a dissociation equilibrium constant of approximately 10^{-9} M (Gill *et al.*, 1991) .

The transcription cycle involves binding, initiation, elongation and termination (reviewed in Record *et al.*, 1996 and Richardson and Greenblat, 1996) . The RNAP binds specific DNA sequences known as promoters found upstream of the 5′ end of RNA coding regions (operons). The promoter sequence determines how tightly the RNAP binds, the frequency of initiation and subsequent elongation. The promoter region may also contain binding sites for transcriptional regulatory proteins. The RNAP binds to approximately 60 nucleotides covering the region -40 (40 nucleotides upstream

of the transcription start site) to +20 (20 nucleotides downstream of the transcription start site). The sequences of nucleotides found at the –35 and –10 regions are specific recognition sites for the RNAP and are found in the majority of prokaryotic promoters. RNAP is capable of non-specific binding to DNA but is then thought to slide along the DNA until it forms a stable complex with the promoter regions, most likely at the –35 and –10 regions simultaneously. This is termed a "closed promoter complex."

Transcription initiation then commences with the unwinding or "melting" of 11 residues of the double stranded DNA in the –9 to + 2 region. This is termed an "open promoter complex." Initiation then continues rapidly with the first nucleoside-5′-triphosphate (usually ATP or GTP) binding to RNAP. Binding of the NTP to RNAP is determined by the DNA complementary base found at the +1 position of the unwound DNA. A second NTP (determined by the nucleotide found at position +2 on the template strand of the unwound DNA) then binds RNAP and a phosphodiester bond between the first and second NTPs is formed by nucleophillic attack of the 3′-hydroxyl group of the first nucleotide on the 5′-α-phosphorous of the second nucleotide.

Elongation then continues with sequential binding of complementary NTPs to the RNAP, phosphodiester bond formation and translocation of the RNAP in the 5′ direction of the template strand. Once an RNA strand 10 nucleotides in length has been produced the σ^{70} subunit dissociates from the holoenzyme leaving the core enzyme to elongate and eventually terminate RNA synthesis. This allows the σ^{70} subunit to bind free core enzyme at additional promoter sites. Messenger RNA chains elongate at approximately

40 nucleotides per second whereas *rRNA* is transcribed at a rate of >80 nucleotides per second.

Termination requires a sequence of DNA that stops the RNAP and releases both the RNA transcript and the enzyme. Two types of terminators may be found at the end of RNA transcripts. Both usually contain a stem-loop structure at the 3′ end of the transcript but they differ in that one has six uridine residues immediately downstream of the stem-loop structure whereas the second type of terminator does not. Weak interactions between the 3′ RNA uridine residues and adenine residues on the DNA template are sufficient to cause destabilisation of the interaction between RNAP and the DNA and to allow the release of the RNAP and the RNA transcript (Yarnell and Roberts, 1999). The terminator lacking the uridine residues relies on protein terminator factors such as Rho. Rho-dependent terminators are frequently C-rich, unstructured and show little sequence conservation (Henkin, 1996).

Interestingly, in a small number of cases, prokaryotic holo-enzymes make use of additional σ factors. These additional σ factors bind to the core polymerase allowing it to recognise different promoters. For example, the σ^{32} subunit allows for the expression of those genes involved in heat-shock reactions (Yura *et al.*, 1993), σ^{54} is required for nitrogen metabolism (Kustu *et al.*, 1991), σ^{F} is required for the regulation of sporulation in *Bacillus subtilis* (Haldenwang, 1995) and those enzymes needed for chemotaxis and σ^{S} is needed during carbon starvation (Loewen and Hengge-Aronis, 1994) .

The recently solved crystal structure of the *Thermus aquaticus* RNAP shows a "crab claw" shaped molecule with a 27 Å wide channel capable of binding double stranded

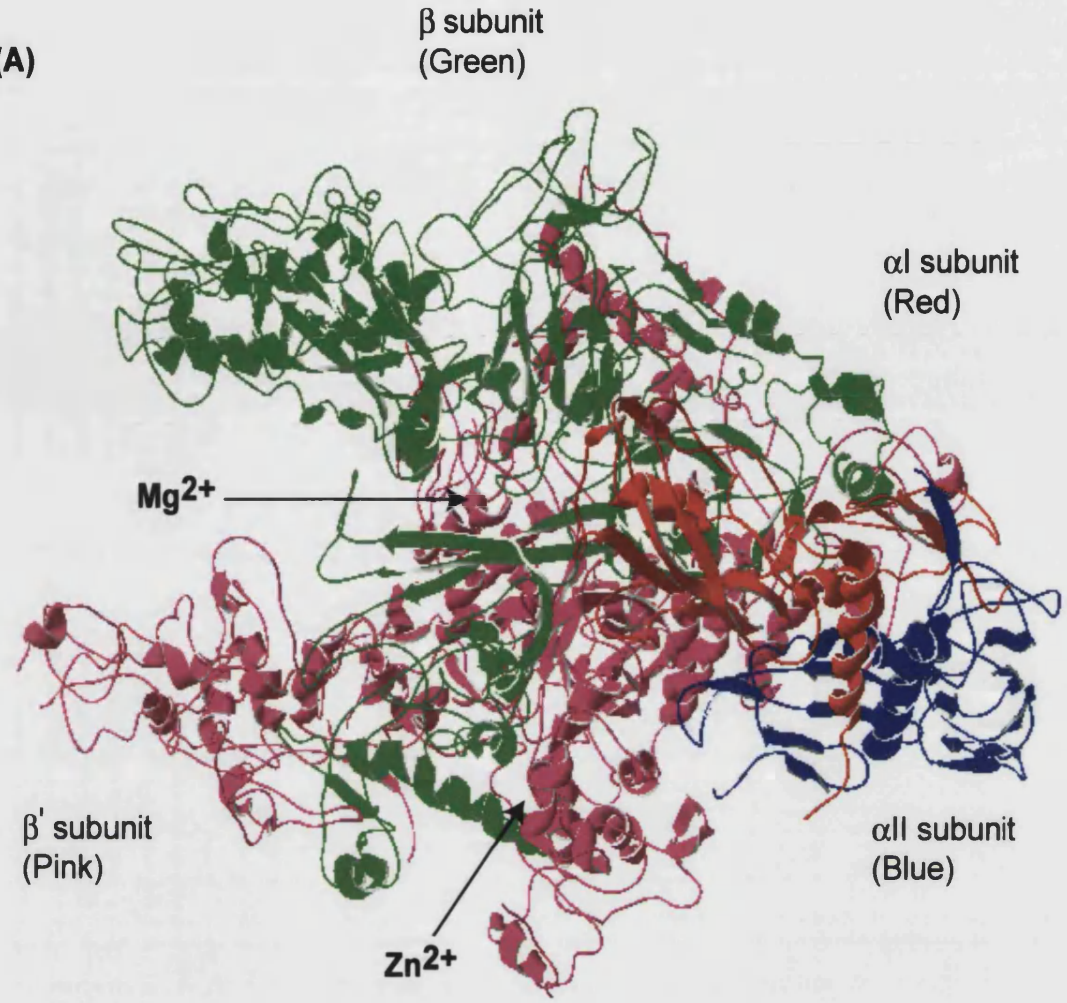
DNA. An Mg^{2+} ion found on the channel back wall is responsible for catalytic activity. The ion is held in place by an amino acid motif conserved through all prokaryotic and eukaryotic RNAPs (Zhang *et al.*, 1999). One arm of RNAP is the β subunit and the other arm is essentially the β′ subunit. The overall size of the enzyme is 150 Å long, 115 Å tall and 110 Å wide. See Figure 2.

The sites for double-stranded DNA, hybrid DNA/RNA and single stranded RNA binding have all been identified. Double-stranded DNA-binding occurs about nine base-pairs from the DNA fork and shows strong non-ionic interactions with the RNAP. Seven-base pairs are involved in this interaction and make contacts with a zinc-finger found at the N-terminal of the β′ subunit. The hybrid DNA/RNA binding site is found close to the Mg^{2+} catalytic site and is characterised by weak ionic interactions (Nudler *et al.*, 1996). A 43 amino-acid region in the C-terminal domain of the β subunit is responsible for this interaction. The single stranded RNA binding site is directly adjacent to the hybrid binding site and covers the RNA region –8 to –14 (Nudler *et al.*, 1996; Zhang *et al.*, 1999). See Figure 2.

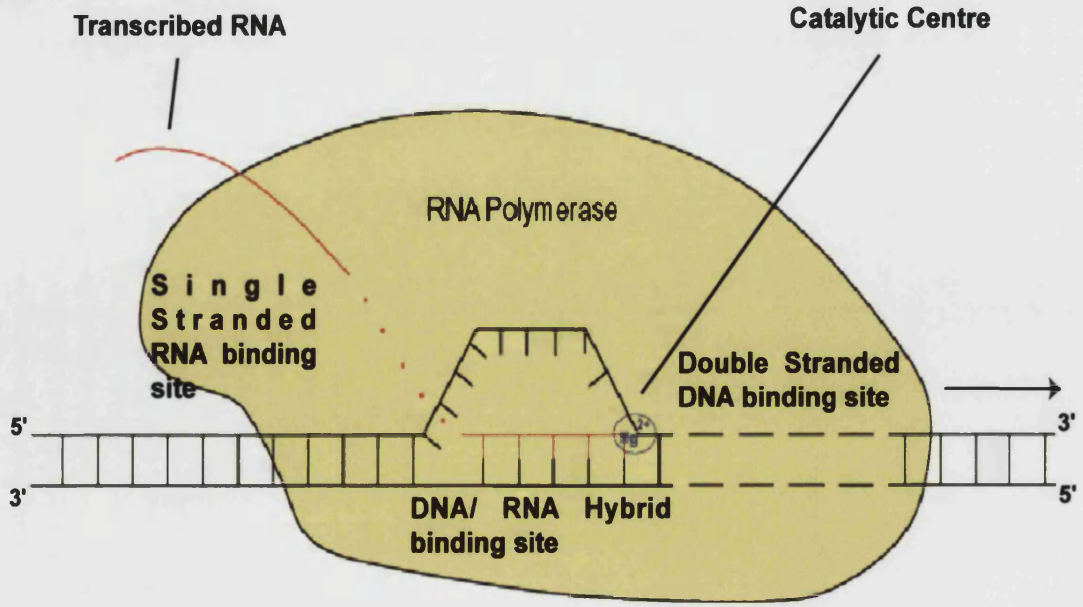
Figure 2 : Prokaryote RNA polymerase structure and elongation complex.

(A) The crystal structure of *T. aquaticus* core RNAP is shown. The structure was detected at a resolution of 3.3 Å. The location of the Mg^{2+}, found at the catalytic centre, is labelled as is the Zn^{2+} atom found in the β' subunit. Subunits are coloured αI - Red, αII - Blue, β - Green and β' - Pink (Steiner et al., 1999). **(B)** The diagram shows the RNA polymerase transcription elongation mechanism. Single-stranded RNA, DNA/RNA hybrid and double-stranded DNA binding sites are all indicated. Complexed proteins are not shown. The direction of elongation is indicated by the arrow. Adapted from Nudler, 1999.

(A)


β subunit
(Green)
αI subunit
(Red)
Mg2+
β' subunit
(Pink)
αII subunit
(Blue)
Zn2+

(B)


Transcribed RNA
Catalytic Centre
RNA Polymerase
Single Stranded RNA binding site
Double Stranded DNA binding site
5'
3'
3'
5'
DNA/ RNA Hybrid binding site

1.4. Regulation of prokaryotic gene expression.

"The level of transcription for any particular gene usually results from a complex series of control elements organized into a hierarchy that coordinates the metabolic activities of the cell" (Zubay, 1993).

Most of our understanding of gene regulation comes from important work carried out in the *E. coli* and *B. subtilis* catabolic, amino-acid and pyrimidine biosynthetic operons. They have shown the extraordinary variety of mechanisms used to simply increase or decrease levels of transcription, as well as sense levels of molecules within cells and adjust the level of transcription of the relevant operons accordingly. A few prokaryotic regulatory systems are detailed below.

1.4.1. Regulation of transcription initiation.

1.4.1.1. Activation and repression : The *lac* operon.

The first and most important site for transcription control is the point of initiation. The first mechanism for regulation is simply found in the sequence of nucleotides in the promoter region of any given gene. The closer the -35 and -10 sequences are to the consensus -35 and -10 sequences, the more rapidly and stably RNA polymerase binds. Negative supercoiling is an additional means of DNA structural regulation. Negative supercoiling facilitates the unwinding and melting of the DNA double helix allowing for the formation of the open-promoter-complex.

Activator and repressor proteins are probably the most effective means of regulation. Activators increase the affinity of the RNA polymerase for the promoter whereas repressors have the opposing effect. One of the best studied systems for activation and

repression is the *lac* operon in *E. coli.* The operon allows *E. coli* to carry out diauxic growth where glucose is utilized first followed by disaccharides such as lactose and also allows the bacteria to use alternative carbon sources in the absence of glucose. The *lac* operon contains the genes for β-galactosidase, a permease and thiogalactaside transacetylase, all required for β-galactoside hydrolysis. Expression of this operon is stimulated by the apoactivator Catabolite Activator Protein (CAP) which, in the presence of the co-activator cyclic-AMP (cAMP), greatly enhances the affinity of the RNA polymerase for the promoter. cAMP levels increase in the absence of glucose and consequently induce a structural change in CAP. This results in CAP binding tightly to the region upstream of the promoter, thus attracting the polymerase to the promoter binding site. cAMP-CAP activation increases the transcription rate by 20 to 50 fold over transcription with RNA polymerase alone. cAMP-CAP activation is also found in the galactose operon, required for galactose catabolism, and the arabinose operon, needed for arabinose utilization (Busby and Ebright, 1999; Ebright, 1993; Record *et al.*, 1996; Zubay 1993).

Repression (inhibition of transcription) in the *lac* operon is carried out by a repressor found immediately upstream of the operon. The repressor binding site (found in the -8 to +28 region) shows dyad symmetry thus allowing the repressor (consisting of four identical subunits) to bind at two symmetrical sites. The repressor binding site overlaps with the RNA polymerase binding site thus preventing transcription initiation (Friedman *et al.*, 1995; Nick and Gilbert, 1985).

A third means of regulation occurs in the presence of an inducer. These are molecules that bind to repressor and prevent it from binding to the promoter region, allowing

transcription of the *lac* operon to occur (Riggs *et al.*, 1970). Inducers of the *lac* operon all contain an intact, unsubstituted galactosidic residue such as allolactose and the gratuitous inducer isopropyl-β-D-thiogalactoside (IPTG).

1.4.1.2. Initiation regulation in amino-acid biosynthesis : the *trp* operon.

A second system in which transcription is regulated at the level of initiation is in the biosynthesis of amino-acids. Those genes involved in tryptophan biosynthesis are best understood as a result of the work done by Charles Yanofsky and co-workers. L-tryptophan binds an aporepressor, encoded by the *trpR* gene, resulting in a structural change, generating repressor which is able to bind to the promoter region. This prevents polymerase binding and hence the genes required for tryptophan synthesis are not expressed. Again the repressor binding site shows dyad symmetry with the repressor subunits showing similar symmetry (Gunsalus and Yanofsky, 1980; Otwinowski *et al.*, 1988; Squires *et al.*, 1975,).

1.4.2. Regulation of transcription termination and terminator read-through.

Transcriptional initiation does not always mean automatic synthesis of the RNA encoded by that particular gene. Even after transcription initiation there are still opportunities for transcription to be regulated. Many genes and operons contain intrinsic terminators within regions upstream of translation start sites meaning that transcription can be effectively stopped before functional regions of a gene are transcribed. Transcription termination and antitermination (the ability of RNA polymerase to read through a terminator) provide additional mechanisms for transcriptional regulation.

1.4.2.1. Regulation through leader peptide translation : the *E. coli trp* operon.

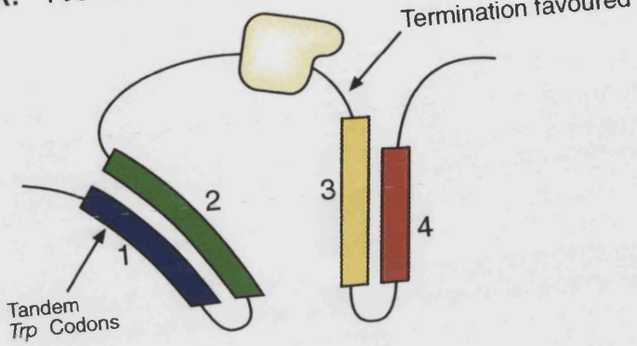
The tryptophan operon does not show positive regulation in the same way that the catabolic genes show cAMP-CAP upregulation. The *trp* operon does however exploit another ingenious mechanism to ensure tryptophan synthesis in times of starvation (See Figure 3). The leader region (approximately 200 residues downstream of the transcription start site) of the operon contains a translatable region containing two adjacent tryptophan codons. In addition, the mRNA is capable of forming three possible secondary stem-loop structures, one of which is an intrinsic terminator.

The first structure (the 1:2 stem-loop) results in pausing of the polymerase. This allows a ribosome to initiate translation of the leader region, thereby coupling transcription and translation. The binding of the ribosome disrupts the pause signal and transcription resumes with the ribosome following shortly behind it. In conditions of excess tryptophan the cell contains high concentrations of trp-charged tRNAs. This means that the ribosome translates easily through the leader region containing the two tandem trp codons, preventing the formation of the 2:3 stem-loop. This in turn allows the formation of the 3:4 stem-loop structure, a typical transcription terminator. In this way the mRNA-polymerase complex is disrupted and transcription is terminated.

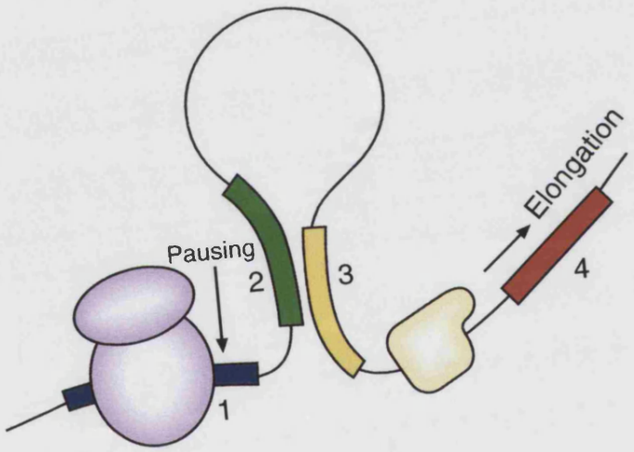
Figure 3 : Regulation of the *E. coli trp* operon by leader peptide translation.

The *trp* operon leader contains two tryptophan codons and four regions all capable of forming stable stem loop structures (labelled 1 – 4 in the diagram). **(A)** shows the naturally occurring stem loop structures in the absence of translation. **(B)** shows the formation of the antiterminating 2:3 stem loop during tryptophan starvation. The stem loop forms as a result of the stalled polymerase at the 1:2 stem loop and the stalled ribosome at the tandem *trp* codon location. The presence of the ribosome at the tandem *trp* codons, disrupts the 1:2 stem loop which in turn releases the polymerase to progress through the 2:3 stem loop and downstream into the operon. **(C)** shows the terminating 3:4 stem loop structure which forms during excess tryptophan conditions. The ribosome does not stall at the tandem *trp* codons preventing the formation of the 2:3 stem loop. The 3:4 stem loop however remains intact, preventing further polymerase elongation. The positions of the ribosome (purple) and the RNAP (yellow) are shown in each scenario. Adapted from Zubay *et al.* (1993).

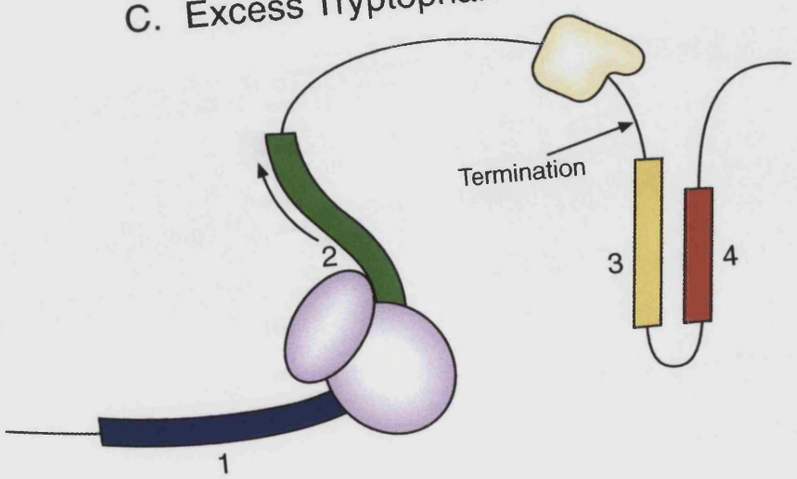
A. No Translation



B. Tryptophan starvation



C. Excess Tryptophan



In times of tryptophan starvation, a process termed antitermination occurs. During translation of the leader peptide the ribosome stalls at the tandem Trp codons. This allows sufficient time for the 2:3 stem-loop structure to form preventing formation of the overlapping 3:4 terminator stem-loop. In this case the polymerase continues transcription through the terminator. (Gollnick and Babitzke, 2002; Henkin, 1996; Yanofsky, 1981).

The *his*, *phe* and *leu* operons all show regulation in a similar manner, as does the pyrimidine biosynthesis operon *pyrBI*. In the case of UTP biosynthesis there is a pause signal followed by a sequence of U residues and a terminator signal, but there is no anti-terminator structure in the leader region of the operon. Transcription through the terminator occurs as a result of the coupling of transcription and translation through the pausing of the polymerase at the poly(U) sequence. The close proximity of the ribosome to the polymerase prevents the formation of the terminator signal. Excess UTP means the polymerase does not pause sufficiently long enough for the ribosome to trail the polymerase in close proximity. This results in the formation of the terminator structure (Gollnick and Babitzke, 2002; Roland *et al.*, 1985).

1.5. The ribosome, *rRNA* and antitermination.

A selection of the wide variety of mechanisms used for transcriptional regulation in prokaryotes has been discussed in detail above. This thesis has focused on the characterisation of the "N utilising (Nus)" proteins in *M. tuberculosis* and their mechanistic roles in antitermination during ribosomal RNA operon expression.

Antitermination is the ability of the RNA polymerase to read through intrinsic Rho-dependent and Rho-independent terminators (Richardson and Greenblat, 1996). In addition to the regulation of expression of catabolic and biosynthetic genes, antitermination regulates ribosomal RNA (*rRNA*) expression using the Nus proteins. The Nus proteins are believed to be involved in the read-through of RNA polymerase in the transcription of the *rRNA* operon in *M. tuberculosis*. Ribosomes are the catalytic tools responsible for the synthesis (translation) of proteins within prokaryotic and eukaryotic cells and *rRNA* forms an essential component of both the prokaryotic and eukaryotic ribosomes. The prokaryotic ribosome is made up of three *rRNA* molecules, two in the 50S ribosomal subunit (the 5S and 23S *rRNA*) and one, the 16S *rRNA* in the 30S ribosomal subunit. The three *rRNA* molecules make up approximately 50% of the mass of the ribosome and are believed to be responsible for its catalytic activity (Ban *et al.*, 2000; Nissen *et al.*, 2000; Noller, 1991). The regulation of *rRNA* synthesis is important in defining the growth rates of bacteria and allowing the organisms to adapt to changing environments. For example, cells in logarithmic phase contain far more ribosomes than cells in lag or stationary phase. As ribosomal activity is constant throughout all growth rates (Sorensen *et al.*, 1994), a cell must increase its ribosomal production in order to increase its protein production and hence its growth rate (Keener and Nomura, 1996; Condon *et al.*, 1995). Antitermination plays an important role in this regulation as it allows for the "transcription of the long, highly complex and untranslated *rRNA*s required for increased ribosome production" (Keener and Nomura, 1996)

1.6. The history of antitermination and the "N-utilising (Nus)" proteins.

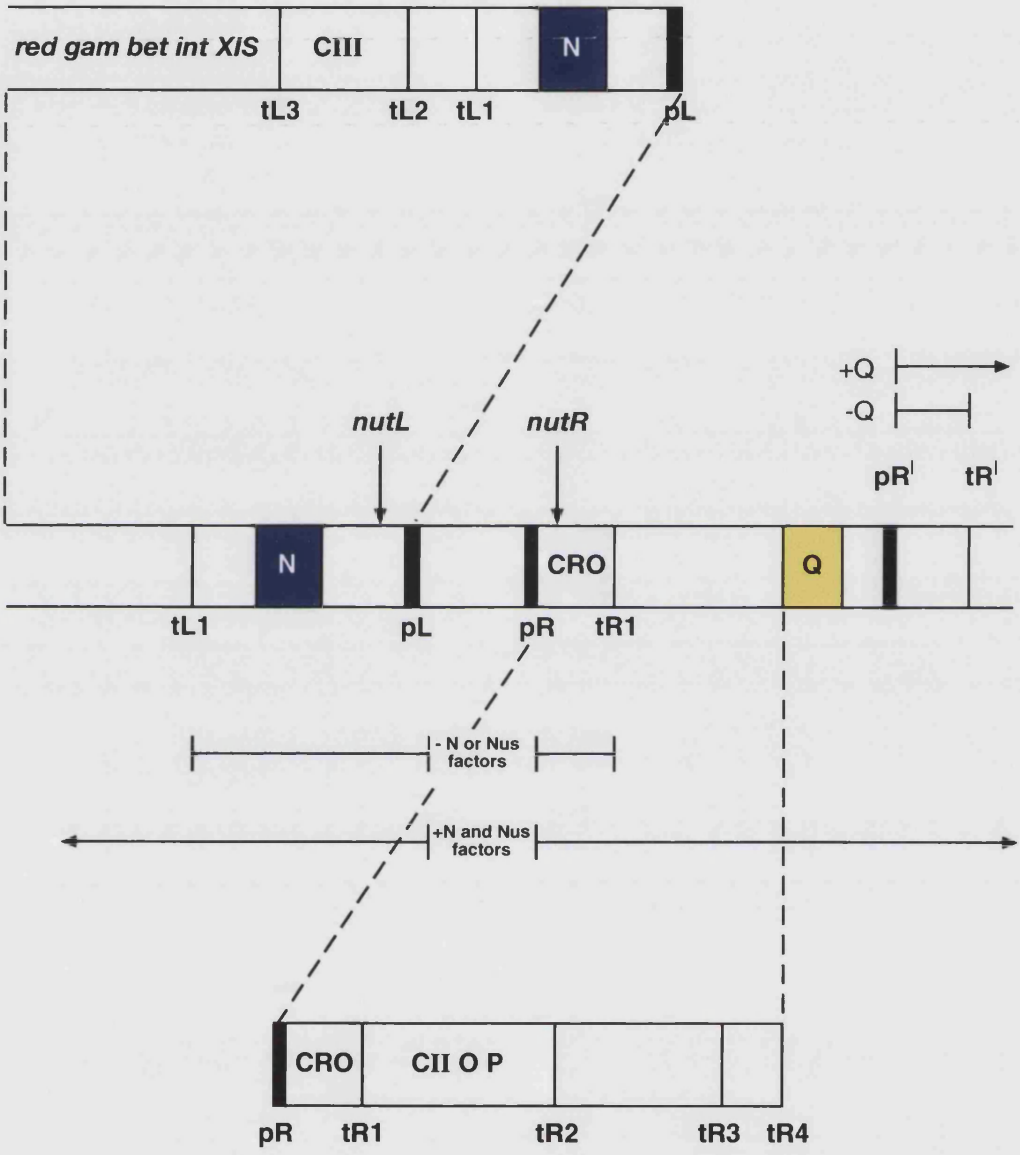
1.6.1. Bacteriophage lambda.

Antitermination utilising the Nus proteins was first identified and studied in bacteriophage Lambda (λ) replication. Roberts (1969) predicted that the N product of λ allowed the transcription machinery to transcribe through Rho-dependent terminators found downstream of the early promoter sites.

Early gene transcription in the λ prophage state initiates at two promoters, termed pL (left) and pR (right), see Figure 4. Transcription from pL transcribes the N gene whereas the *cro* gene is transcribed from pR. Rho-dependent terminators are found immediately downstream of N and *cro*, tL1 and tR1 respectively. Downstream of *cro* are genes O and P (required for λ replication) which are, again, followed by three Rho-dependent terminators tR2, tR3 and tR4. As concentrations of the N protein increase within the bacterial cell, transcripts initiating at pR are able to transcribe through tR1, tR2 and tR3 allowing for expression of the downstream genes O and P (required for λ DNA replication) and transcripts initiating from pL are able to transcribe through tL1, tL2 and tL3 allowing for expression of the downstream genes *red*, *gam*, *bet*, *int*, and *xis* (required for integration of the λ genome into the bacterial chromosome). The protein *c*II is also produced after transcription from pR, the levels of which determine whether the phage enters lysogeny or the lytic cycle (Friedman and Court, 1995; Richardson and Greenblatt, 1996; Zubay, 1993).

Figure 4 : Bacteriophage lambda genome structure and gene expression.

A simplified diagram of part of the λ genome is shown with the antiterminator proteins N and Q highlighted in blue and yellow. Early transcription terminates at tR1 and tL1 resulting in the synthesis of the protein N. N in combination with the host Nus factors allows for the transcription of genes required for the establishment of lysogeny, phage genome integration and a second antiterminator Q, required for entry into the lytic cycle. The regions of transcription in the presence or absence of N and the host Nus factors are shown (Adapted from Friedman and Court, 1995).


red gam bet int XIS
CIII
N
tL3
tL2
tL1
pL
+Q
-Q
nutL
nutR
pR'
tR'
N
CRO
Q
tL1
pL
pR
tR1
- N or Nus factors
+N and Nus factors
CRO
CII O P
pR
tR1
tR2
tR3
tR4

Studies on *E. coli* mutants that were unable to support λ growth, and *in vitro* antitermination systems allowed for the identification of host proteins involved in λ phage antitermination, termed the N-utilising (Nus)" proteins. The Nus proteins include NusA, NusB, NusE (identical to the ribosomal protein S10) and NusG and are believed to interact with N and modify the RNA polymerase into a termination-resistant form (Li *et al.*, 1992; Mason and Greenblatt, 1991). All the Nus proteins are required to form a stable antitermination complex. Cis-acting sequences, termed *Nut* sites, found between pL1 and tL1 and pR1 and tR1 were also found to be required for λ antitermination. The *Nut* sites act at the RNA level (Nodwell and Greenblat, 1991). NusA binds to sequences termed boxA and boxB found in the *Nut* site. NusA, NusE and NusG are all capable of interacting directly with the RNA polymerase and NusB interacts with NusE. The λ N protein binds to the *Nut* site and in turn interacts with the RNA polymerase-NusA complex. (Greenblatt and Li, 1981b; Horwitz *et al.*, 1987; Mason and Greenblatt, 1991; Mason *et al.*, 1992a,b; Mogridge *et al.*, 1995). DeVito and Das (1994) proposed that NusA facilitates the binding of N to the RNA polymerase and that the remaining Nus factors stabilise the complex.

N and the Nus proteins assemble at *Nut* sites on the RNA transcripts (found downstream of *cro* and upstream of N) (Salstrom and Szybalski, 1978; Rosenberg *et al.*, 1978). These *Nut* sites are made up of three, component sequences, boxA, boxB and boxC separated by spacer regions (Friedman and Gottesman, 1983; Olson *et al.*, 1982, Olson *et al.*, 1984; Salstrom and Szybalski, 1978). A comparison of five lamboid *nut* sites has allowed for definition of boxA and boxB consensus sequences (Friedman and Gottesman, 1983) and mutational studies on boxA have shown its importance in λ

antitermination (Olson *et al.*, 1982). The λ *Nut* sites are found far upstream of the terminators the N-Nus antitermination complex acts on. It is thus believed that looping of the RNA, such that the complex is continuously in contact with the polymerase, maintains a termination-resistant polymerase (Richardson and Greenblatt, 1996). Interestingly boxA sequences are found in the ribosomal (*rrn*) operons of many prokaryotes (Berg *et al.*, 1989). *In vitro* studies have shown that prokaryotic antitermination complex assembly takes place on the transcribed RNA *Nut* site (Mason and Greenblatt, 1991; Nodwell and Greenblat, 1991).

A second round of antitermination occurs immediately preceding late gene transcription in λ. It is at this stage that entrance into the lysogenic or lytic cycle is decided. During transcription of the early genes, the gene Q is also transcribed. Q is the key factor required for late lytic gene expression (those genes required for cell lysis and maturation of virus particles) and, like N, binds a specific sequence, *Qut,* found between the -35 and -10 promoter sites. *Qut* contains both boxA and boxB but not boxC sequences (Friedman and Gottesman, 1983) but is a non-transcribed DNA sequence (found between the -10 and -35 promoter sites) as opposed to the RNA *Nut* sequence (Friedman and Court, 1995). NusA has been shown to be the only host Nus factor involved in Q antitermination (Barik and Das, 1990).

The lysogenic cycle is followed if *c*II levels, produced during early gene expression, are high. In this case the genes *c*I and *int* are transcribed from the pRE and pI promoters. The production of the Int protein allows λ genome integration into the *E. coli* genome where as *c*I prevents transcription from the pL and pR promoters and hence the

production of Q. The mechanism by which *c*II levels are determined is still not understood (Richardson and Greenblat, 1996).

1.6.2. The Nus proteins and antitermination in *E. coli.*

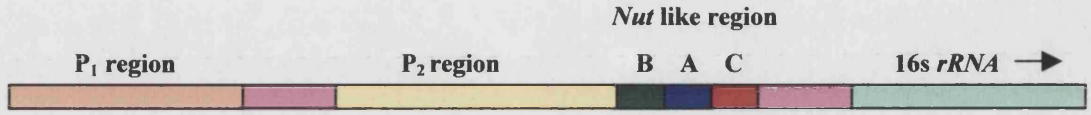
Antitermination has been identified as a mechanism of regulation of gene expression in *E. coli* (Holben and Morgan, 1984, Aksoy *et al.*, 1984). Expression of *E. coli rrn* operons do not show polarity, defined as "the decreased expression of a distal gene (or sequences) of an operon resulting from specific genetic signals in proximal sequences" (Aksoy *et al.*, 1984). It was therefore proposed that the *E. coli rrn* operons must exhibit some sort of antitermination mechanism. *Nut* sequences were first identified in the *rrn* operons by Berg *et al.* (1989) and Li *et al* (1984). Sharrock *et al.* (1985) were the first to implicate some of the Nus factors in *rrn* antitermination. A great many subsequent experiments showed that *rrn* antitermination was highly analogous to λ N-mediated antitermination.

1.6.2.1. *E. coli rrn* operon structure.

E. coli contains seven *rrn* operons (*rrnA –rrnG*) coding for the three ribosomal RNA fragments. The operons are located asymmetrically on one half of the *E. coli* chromosome and differ in that different tRNA genes are found in the 16S-23S *rRNA* spacer regions and at the ends of the operons (Bachman, 1990; Komine *et al.*, 1990). They contain two promoters, P_1 and P_2, followed by the 16S *rRNA* coding region, a 4S tRNA coding region, the 23S and lastly the 5S *rRNA* coding regions . The entire operon is transcribed initially as a large 30S fragment which is then processed by RNase activity into the four molecules (Lund and Dahlberg, 1977; Perry, 1976). Study of the

rrnB operon shows a highly conserved boxA sequence downstream of P_2 and in the 16S – 23S spacer regions (Li *et al.*, 1984) indicating a λ phage like antitermination mechanism, utilising Nus proteins. Berg *et al.* (1989) noted that all seven *rrn* operons have identical boxA and boxC sequences and all show dyad symmetry at boxB allowing for the formation of a stable stem-loop structure (See Figure 5). BoxA sequences are also highly conserved in the *rrn* leader and spacer regions of a diverse range from the *Bacteria* domain and some from the *Archea* species. See Table 2.

(A)



(B)

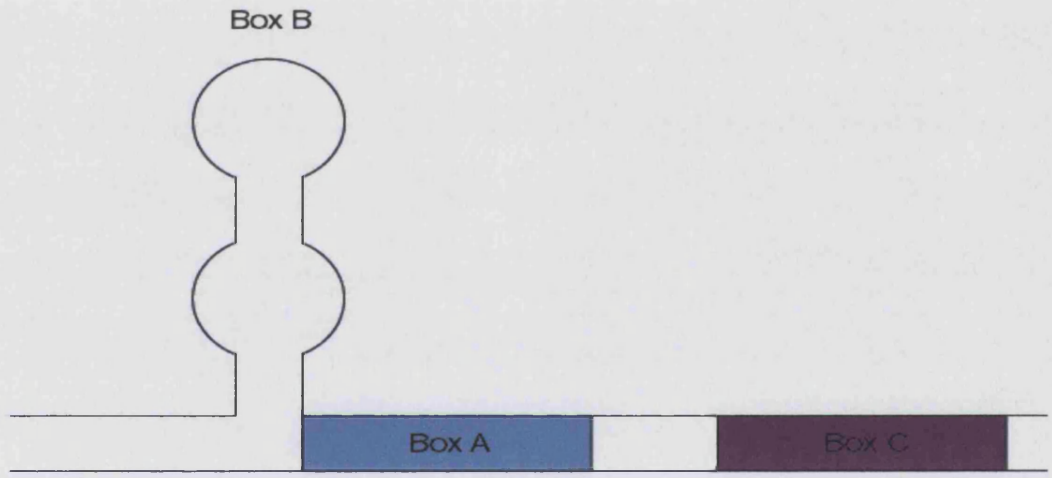


Figure 5 : *E. coli rrnG* leader region.
(A) Schematic showing the positions of P_1, P_2, boxB, boxA, boxC and the beginning of the 16s *rRNA* coding region (Condon *et al.*, 1995; Keener and Nomura, 1996; Li *et al.*, 1984). **(B)** shows the RNA secondary structure of the transcribed leader region of the *rrnG* operon in *E. coli*. The boxB hairpin is labelled and boxA and boxC shown in blue and purple respectively. Adapted from Squires *et al.* (1993).

Table 2 : BoxA Sequences.
BoxA sequences for a variety of organisms and phages (Berg *et al.*, 1989; Olson *et al.*, 1982) show high homology and thus their likely role in transcriptional antitermination.

Organism	BoxA Sequence
Escherichia coli (All *rrn* operons)	TGCTCTTT
Bacillus subtilis (All *rrn* operons)	(A/T)G(A/T)CTTT
Thermus thermophilus (*rrn* 16S and *rrn* 23S)	GGATCTTG
Caulobacter crescentus (All *rrn* operons) α-purple eubacterium	GG(G/C)TCTTT
Pseudomonas aeruginosa	TGCTCTTT
Mycobacteriums tuberculosis	TGTTGTTTG (Leader region) TGTTCTTTG (Spacer region)
Lambda Phage *nutR*	CGCTCTTA
Lambda Phage *nutL*	CGCTCTTA
Lambda *qut*	CGCTCGTT
Phage 21 *nutR*	GCTCTTTA
Phage P22 *nutL*	CGCTCTTTA

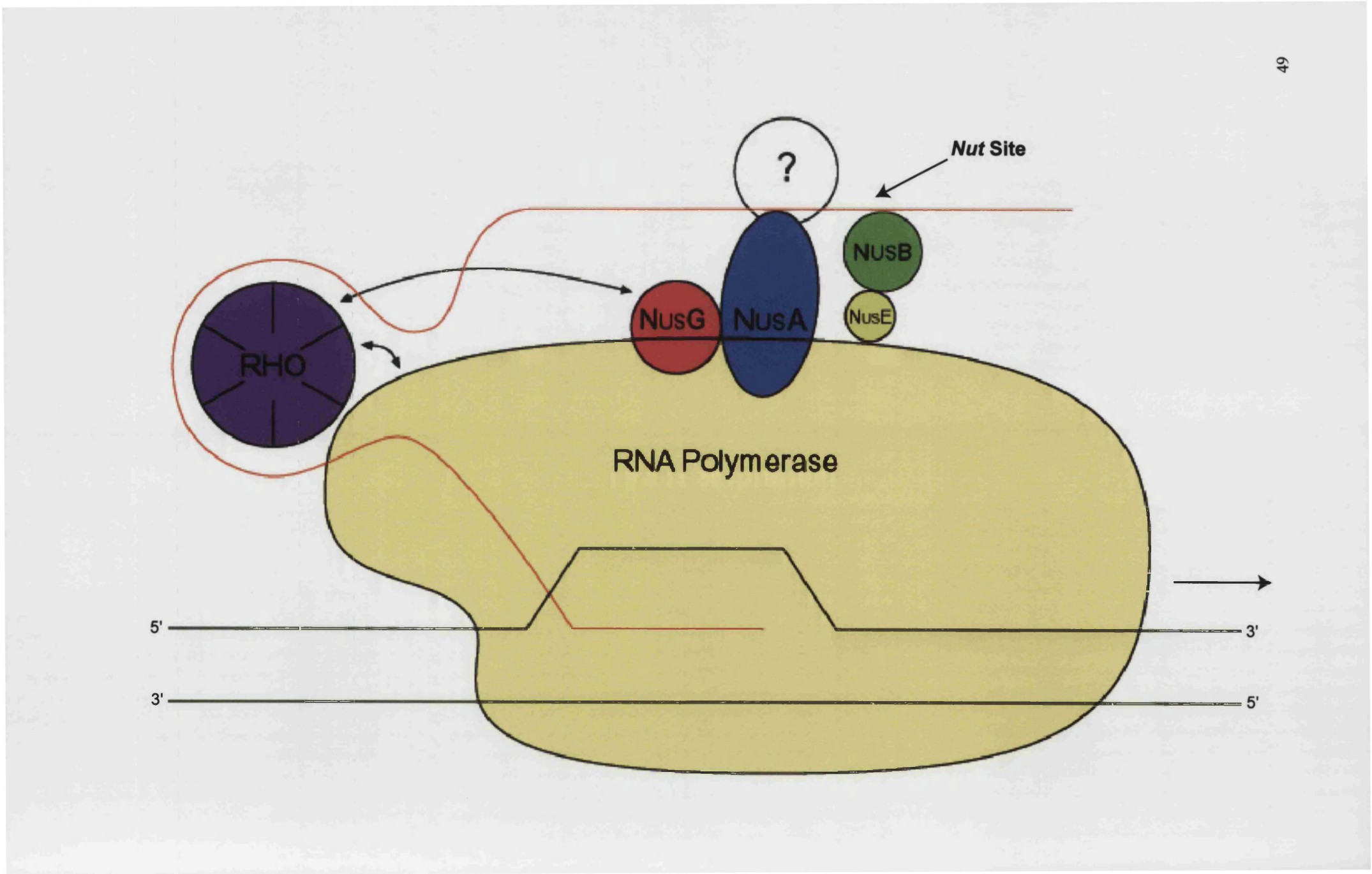
The first evidence that the Nus proteins were involved in *E. coli* transcriptional regulation was the observation that *rRNA* transcription was inhibited in a NusB mutant *E. coli* strain (Sharrock *et al.*, 1985). This mutant (*nusB5*) is also cold-sensitive probably as a result of defective *rRNA* production (Taura *et al.*, 1992). Antitermination and expression of downstream reporter genes was achieved *in vitro* using a DNA template with a promoter, anti-terminator, terminator structure, purified RNA polymerase, termination factor Rho, NusA, NusB, NusE, NusG and cell extract depleted of NusB.

Efficient expression was only achieved with the cell extract indicating that as yet unidentified components are required for antitermination (Squires *et al.*, 1993). NusB and NusE form a heterodimer (Luttgen *et al.*, 2002; Mason *et al.*, 1992a) which was shown to interact directly with *rrn* boxA (Luttgen *et al.*, 2002; Nodwell and Greenblatt, 1993). In addition NusE interacts with the RNA polymerase (Mason and Greenblatt, 1991). NusA, NusE and NusG are all known to interact with RNA polymerase directly and probably stabilise the antitermination complex (Greenblatt and Li, 1981a; Li *et al.*, 1992; Mason and Greenblatt, 1991). NusA is believed to replace the RNA polymerase σ subunit after transcription initiation (Greenblatt and Li, 1981a). The terminator Rho has been shown to interact with RNA polymerase (Darlix *et al.*, 1971), NusG (Li *et al.*, 1993; Pasman and von Hippel, 2000) and NusA (Schmidt and Chamberlin, 1984) and may thus be sequestered into the antitermination complex, altering its interaction with the RNA polymerase (Condon *et al.*, 1995). Alternatively the formation of the antitermination complex may increase the rate of the transcription machinery thereby uncoupling Rho from the RNA polymerase and reducing termination (Jin *et al.*, 1992).

A model has been proposed (Luttgen *et al.*, 2002) where the antitermination complex assembles on the rrn *nut* sites in the leader region. The complex then interacts with RNA polymerase and remains attached to the *nut* site as the polymerase transcribes in the 5′ to 3′ direction. This results in looping of the RNA and modification of the polymerase, allowing it to transcribe through terminators found in the leader and spacer regions of the *rrn* operons (See Figure 6).

Figure 6 : The antitermination complex in *E. coli*.

The antitermination complex in *E. coli* is shown. Interactions occur between the RNA polymerase and NusA, NusE and NusG with NusA replacing the RNA polymerase σ subunit. NusB and NusE form a heterodimer. NusG interacts with Rho and the ? indicates unknown factors required for *in vitro* antitermination. The *nut* site and direction of elongation are shown. Adapted from Luttgen *et al.*, 2002.


?
Nut Site
NUSB
NUSE
NUSG
NUSA
RHO
RNA Polymerase
5'
3'
3'
5'

1.7. Antitermination in *Mycobacterium tuberculosis*.

In contrast to the λ and *E. coli* systems, little is known about antitermination in *M. tuberculosis*. A major difference from other prokaryotes is that there is only a single *rrn* operon present in the entire genome (Cole *et al.*, 1998). The operon shows the classical *rrn* structure observed in other bacterial systems (Gonzalex-Y-Merchand *et al.*, 1996; Kempsell *et al.*, 1992), containing two promoters, p1 and pCL1 (with -35 and -10 elements). The leader region contains boxA, boxB and boxC followed by coding regions for 16S *rRNA*, 23S *rRNA* and 5S *rRNA*. The spacer region between the 16s *rRNA* and 23s *rRNA* region contains only boxA and boxB (Gonzalez-y-Merchand, 1996; Kempsell *et al.*, 1992). BoxA and boxC sequences show homology to both λ and *E. coli* sequences. The boxA sequence is TGTTGTTTG in the leader region and TGTTCTTTG in the spacer region. The boxC sequence is AGTGTGTTT. BoxB possesses dyad symmetry with the ability to form a stable stem-loop structure (Kempsell *et al.*, 1992). See Figure 7.

In addition to the RNA elements required for antitermination in the ribosomal operon, homologous genes to all the Nus proteins are found in the *M. tuberculosis* genome (Cole *et al.*, 1998) indicating *rRNA* transcription may be regulated by a similar antitermination mechanism to that seen in the λ phage and in *E. coli*.

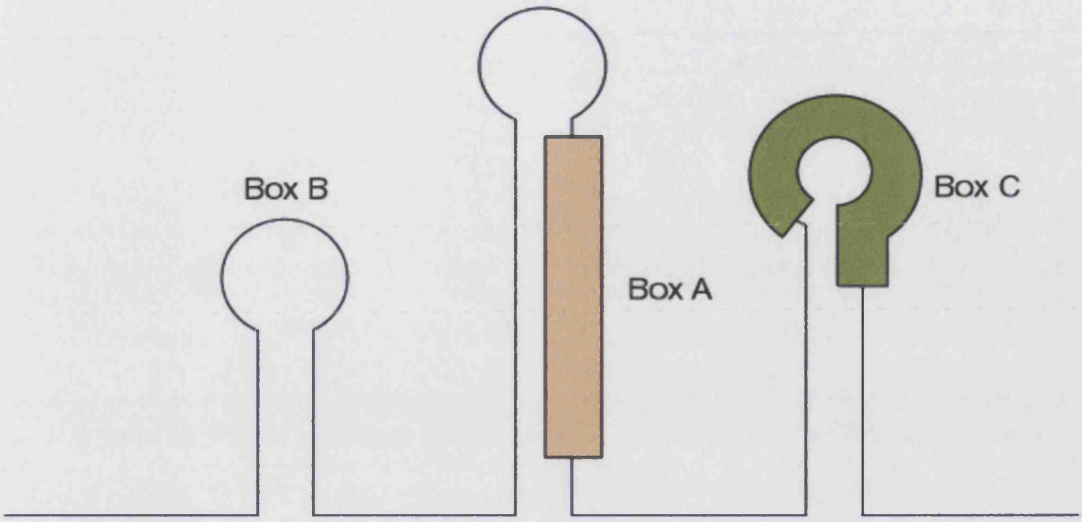


Figure 7 : *M. tuberculosis rrn* leader region secondary structure.
The secondary structure and positions of boxA, the stem-loop boxB and boxC in the *M. tuberculosis rrn* operon leader region are shown. BoxA (orange) and boxC (green) sequences, homologous to the λ and *E. coli* boxA and boxC sequences are highlighted.

1.8. The Nus Proteins.

1.8.1. NusA.

NusA is 56 kDa protein in *E. coli* and is essential for the survival of the organism. Its key function within the cell is regulation of transcription. It is responsible for slowing down the rate of elongation by stimulating the RNA polymerase to pause at certain natural pause sites (Landick and Yanofsky, 1987) and it increases termination efficiency at intrinsic terminators (such as λtR2), probably through stem-loop structure stabilisation. NusA promotes release of the nascent transcript by preventing the RNA polymerase from binding the stem-loop structure (Artsimovitch and Landick, 1998; Schmidt and Chamberlin, 1987). NusA however also decreases termination efficiency at other terminators, such as the *E. coli* terminator *trpoBa* (Linn and Greenblatt, 1992) and the λ Rho-dependent terminator λtR1 (Lau *et al.*, 1982). The regulation of this

paradoxical action may be brought about by NusA's interaction with different components of the transcription and translation systems. Carlomagno and Nappo (2003) have hypothesised that the interaction of NusA with *mRNA* signals, other Nus proteins or ribosomal subunits may enhance transcriptional termination where as the interaction of NusA with ribosomes or specific ribosomal proteins found on nascent transcripts may enhance elongation.

Its role in antitermination was also revealed through an *E. coli* mutant strain carrying the *nus*A1 mutation that failed to support λ growth (Friedman *et al.*, 1973). A role for NusA in antitermination was also predicted from studies of another NusA mutant, the cold sensitive *nus*A*cs*10. *Nus*A*cs*10 mutant decreases the rate of elongation of RNA polymerase resulting in considerably less read-through of terminators. NusA may thus be responsible for increased RNA polymerase elongation rates during antitermination (Vogel and Jensen, 1997).

NusA is able to make direct contacts with the RNA polymerase shortly after transcription initiation and once the σ^{70} subunit has escaped the polymerase. It makes direct contacts with the C-terminal domain of the α-subunit as well as with both the β and β′ through contacts in both the NusA N- and C-terminal domains (Liu *et al.*, 1996; Ito and Nakamura, 1996; Mah *et al.*, 1999). However, only the N-terminal domain, containing one of the RNA polymerase binding regions and the RNA binding region, is needed to enhance both termination and antitermination (Mah *et al.* 1999). NusA also binds the N-terminal region of N in λ phage antitermination via a C-terminal region (Mogridge *et al.*, 1998; Mah et al., 1999). N appears to be unstructured in solution but

on interaction with RNA and NusA it folds into separate domains. The interaction of N with NusA is also independent of RNA polymerase (Van Gilst and von Hippel, 1997).

Sequence and structural homology indicates the presence of three RNA binding domains, an S1 and two KH domains (Gopal *et al.*, 2001a). NusA is believed to interact with the boxA and boxB RNA *nut* sites but does not cause a mobility shift in the absence of N. It is therefore likely that N mediates NusA RNA binding in the λ antitermination mechanism (Mogridge *et al.*, 1995). Mah *et al.* (1999) showed that only NusA constructs containing a wild-type S1 domain are capable of binding an N-*nut* complex and that this domain is therefore essential for antitermination functions. Zhou *et al.* (2002) subsequently showed that wild-type KH domains are also essential for RNA binding in that mutations in either KH domain abolish RNA binding even in the presence of N. The N-terminal RNA polymerase binding region appears to be essential for both termination and antitermination by the λ N protein where as the C-terminal RNAP-binding domain is not required for either of these processes (Mah *et al.*, 1999).

The crystal structure of *M. tuberculosis* NusA shows it is made up of two distinct regions, an N-terminal domain similar to the RNA polymerase binding domain of SigA, and a C-terminal region containing three RNA binding domains, an S1 domain and two KH domains (Gopal *et al.*, 2001a). The N and C-terminal segments are attached by a highly flexible linker. The *M. tuberculosis* NusA crystal structure shows similarities with that of *T. maritima* (Worbs *et al.*, 2001), although there are slight changes in the positioning of the domains and the linker region in *T. maritima* is helical in structure. Nevertheless the overall architectures are similar.

1.8.2. NusB.

In *E. coli* NusB is a monomeric 16 kDa protein and was first identified when a *nus*B5 mutant was isolated that was unable to support the growth of λ phage and resulted in premature transcription termination within *rRNA* operons *in vitro* (Sharrock et al., 1985). *E. coli* strains with this mutation are also cold sensitive, again indicating a possible role in *rRNA* transcription regulation (Taura *et al*, 1992).

The structures of NusB from both *E. coli* and *M. tuberculosis* have been solved. *E. coli* NusB is an all helical protein with a novel fold. The majority of conserved residues in NusB homologs are responsible for stabilising the core of the protein. Two conserved amino-acids (Lys 82 and Arg 86) are found on the surface of NusB and form a small, positively charged cavity (Altieri *et al.*, 2000). This feature is also conserved in the *M. tuberculosis* NusB structure. The two NusB homologs are 34 % identical and show good (r.m.s. deviations between Cα atoms of 2.0 Å) structural similarity (Altieri *et al.*, 2000; Gopal *et al.*, 2000). A potential RNA binding site was also proposed by Gopal *et al.* (2000) based upon the N-terminal region of *M. tuberculosis* NusB containing a phosphate binding site (defined by the presence of two conserved arginines, Arg 10 and Arg 14).

A major difference between the two proteins is that the *E. coli* NusB is a monomer as shown by sedimentation equilibrium ultracentrifugation whereas *M. tuberculosis* NusB is dimeric (Figure 8) as seen in the crystal lattice and also shown by sedimentation equilibrium ultracentrifugation (Altieri *et al.*, 2000; Gopal *et al.*, 2000). The dimer is formed as a result of numerous polar and non-polar interactions. This is the first

indication that *rRNA* transcription regulation and antitermination in *M. tuberculosis* may have a different mechanism to that seen in λ and *E. coli.*

E. coli NusB and NusE form a heterodimeric complex with a dissociation equilibrium constant of 10^{-7} M. Complex formation is mediated via non-ionic interactions (Mason *et al.*, 1992a). This complex then interacts specifically with boxA *nut* sequences (Nodwell and Greenblatt, 1993). The heterodimer has a higher affinity for boxA RNA than NusB alone indicating that it is the complex that interacts with boxA RNA and not allosteric NusE activation of NusB only RNA binding (Luttgen *et al.*, 2002). NusE is unable to bind boxA in the absence of NusB (Nodwell and Greenblatt, 1993). In contrast to the *E. coli* NusB-NusE complex, Gopal *et al.* (2001b) tentatively showed only very weak interactions between *M. tuberculosis* NusB and NusE.

NusB has been shown to increase Rho termination efficiency by stimulating the release of RNA polymerase at sub-optimal Rho-dependent terminators (Carlomagno and Nappo, 2001). It has also been speculated that NusB may be involved in translation as the *nus*B5 mutation resulted in decreased peptide elongation rates (Shiba *et al.*, 1986).

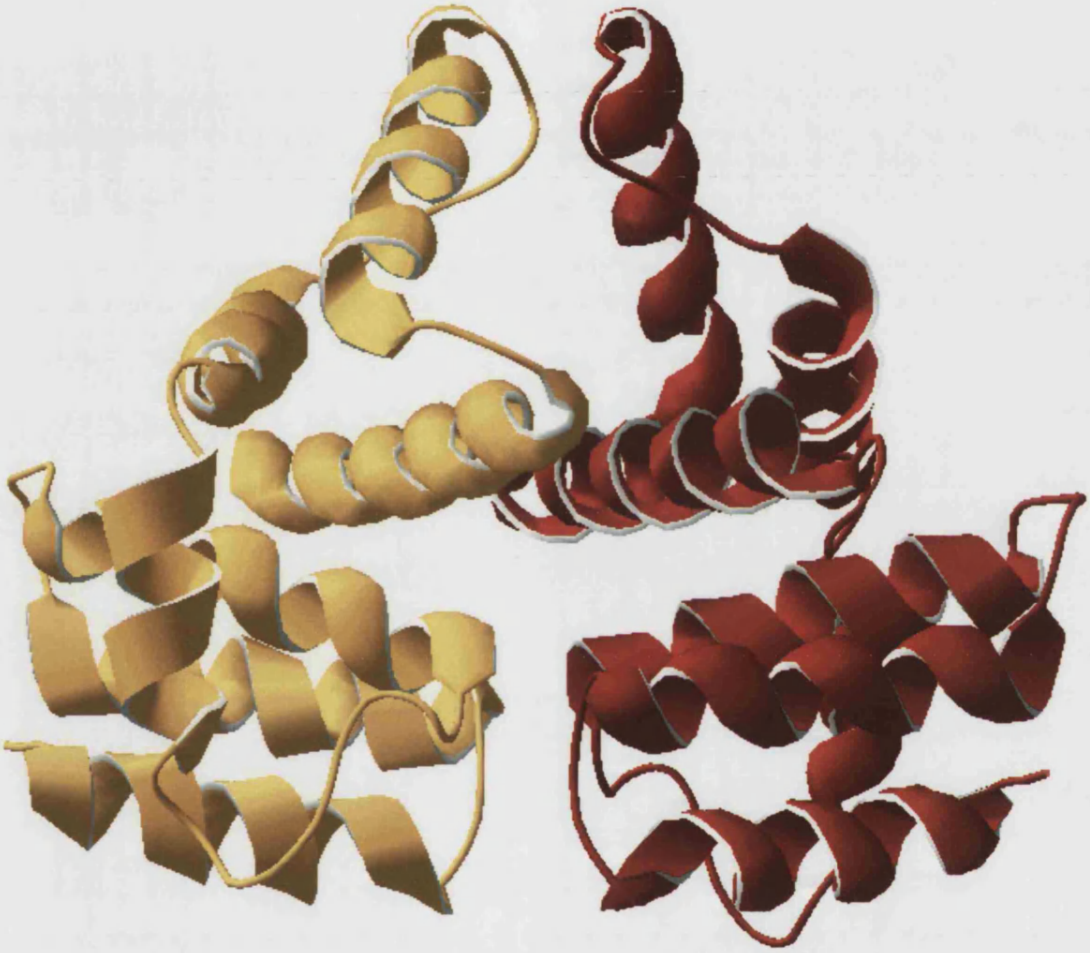


Figure 8 : *M. tuberculosis* NusB dimer structure.
A ribbon representation of the NusB dimer. Dimerisation occurs through a number of polar and non-polar interactions resulting in a triangular shaped molecule. This is in contrast to *E. coli* NusB which is monomeric (Gopal *et al*, 2000; Guex and Peisch, 1997; Schwede *et al*, 2003).

1.8.3. NusE.

NusE is identical to the ribosomal protein S10 and is essential for cell survival. It is found in the 30S ribosomal subunit where it is one of the last proteins to be added to the 30S complex before a functional 30S subunit is made (Squires and Zaporojets, 2000).

NusE's involvement in antitermination was discovered when a *nus*E71 mutation inhibited the growth of λ phage as N was inactive (Friedman *et al.,* 1981) and the addition of NusE to *in vitro* antitermination experiments resulted in increased read-through (Squires *et al.*, 1993). Warren and Das (1984) showed that NusE was involved in N-mediated antitermination but was not responsible for coupling antitermination and translation. It has however not been possible to determine whether NusE participates in antitermination as part of the 30S ribosomal subunit or as a soluble, independent form (Das, 1993, Warren and Das, 1984).

NusE has also been shown to make direct interactions with NusB (described previously) as well as with the RNA polymerase with an association equilibrium constant of 10^6 M^{-1} (Mason and Greenblatt, 1991). NusE also increases the affinity of NusB for both the λ *nutR* boxA and the *rrn* boxA by a minimum of 25-fold (Luttgen *et al.*, 2002).

M. tuberculosis NusE has been isolated. The free protein is monomeric, largely unstructured, with a helical content of 12 – 16 % at 30°C. (Gopal *et al.*, 2001b).

1.8.4. NusG.

NusG, the last Nus factor to be identified, is also essential to *E. coli* and, similarly to NusA, appears to be multi-functional with respect to gene expression regulation.

Although both proteins promote read-through in the antitermination complex, they also exhibit antagonistic functions. NusG decreases pause times at attenuator structures while NusA increases pause times, NusG increases the rate of Rho-dependent termination while NusA appears to decrease the rate (Burova *et al.*, 1995; Burns *et al.*, 1998; Schmidt and Chamberlin, 1987; Pasman and von Hippel, 2000). NusG's role in expression regulation and antitermination has been suggested by its ability to bind both RNA polymerase and Rho. Weak binding of NusG to RNA polymerase has been demonstrated (Li *et al.*, 1992) while NusG binds tightly to the hexameric Rho with a dissociation equilibrium constant of 1.5×10^{-8} M (Pasman and von Hippel, 2000).

The crystal structure of *Aquifex aeolicus* NusG has recently been solved (Steiner *et al.*, 2002; See Figure 9). It is made up of three domains and has overall dimensions of approximately 75 × 70 × 38 Å. Domain I has homology to the ribonucleoprotein (RNP) motif thus indicating it may be involved in nucleic-acid binding. Domain I also shows homology with the S6 ribosomal protein (which makes contacts with S18 in the ribosome (Ban *et al.*, 2000)) perhaps indicating it may also be involved in protein – protein interactions. Domain III contains a KOW sequence motif (Kyrpides *et al.*, 1996). These are involved in *rRNA* binding in prokaryotic L24 ribosomal protein and in the eukaryotic *e*L26 and *e*L27 ribosomal proteins. The KOW domain has also been implicated in protein binding through its structural homology to the tudor domain in the human SMN protein (Selenko *et al.*, 2001). The protein binding site in the KOW domain is believed to be different to its nucleic acid binding site, suggesting that *rRNA* and protein binding can occur simultaneously (Steiner *et al.*, 2002).

Domain II shows some similarity to an immunoglobulin fold (which are often involved in eukaryotic protein – protein interactions) however the NusG domain II structure could not be matched with a specific class of immunoglobulin fold. There is also no sequence homology and it has been impossible to reliably predict the function of domain II which appears as a loop in the homology model of the *E. coli* NusG (Steiner *et al.*, 2002). It may simply be involved in the stability of the protein in the thermophilic organism.

NusG also has links to translation and lowered amounts of NusG lead to a decreased translation rate (Burova *et al.*, 1995; Zellars and Squires, 1999). As the KOW motif is predominantly found in ribosomal proteins, perhaps this is not entirely surprising (Squires and Zaporojets, 2000). Little is known about the *M. tuberculosis* NusG. It carries an additional forty one residues at its N-terminus and does not contain domain II found in *A. aeolicus* NusG (Steiner *et al.*, 2002).

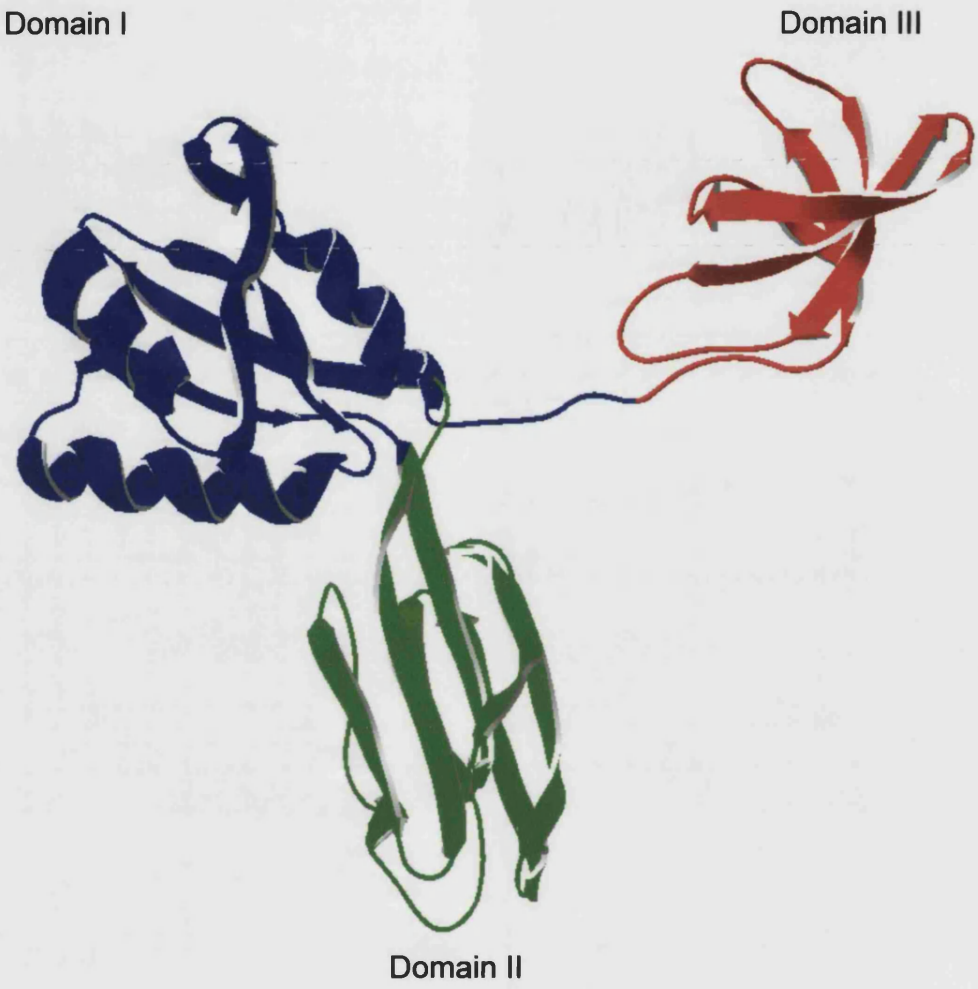


Figure 9 : *A. aeolicus* NusG structure.
A ribbon representation diagram of the *A. aeolicus* NusG structure. The three domains (domain I - blue; domain II - green; domain III - red) are indicated (Guex and Peisch, 1997; Schwede *et al*, 2003; Steiner *et al*, 2002).

1.9. Rho termination factor.

The Rho protein is responsible for termination of gene or operon expression in bacteria. It utilises the energy from ATP hydrolysis to stimulate nascent transcript release. Rho is a hexamer of identical subunits forming a ring shaped-structure (Yu *et al.*, 2000). Each subunit contains an RNA-binding domain at the N-terminus (Allison *et al.*, 1998; Modrak and Richardson, 1994; Skordalakes and Berger, 2003) and an ATPase domain making up the C-terminal region (Dolan *et al.*, 1990; Dombroski *et al.*, 1988a,b; Skordalakes and Berger, 2003).

Rho-dependent terminators are made up of two parts, the *rut* region (Rho binding site) along with the site of transcript release (Richardson and Greenblat, 1996). The terminator region can span up to 150 nucleotides in length. Termination and transcript release is brought about by pause sites downstream of the *rut* sites when nucleotide addition by the polymerase is at its slowest rate (Artsimovitch and Landick, 2000; Morgan *et al.*, 1983; Richardson, 2002).

The structure of *E. coli* Rho has been solved by Skordalakes and Berger (2003). The Rho hexamer was shown to contain a 12 Å gap between two of the subunits presumably allowing for the loading of RNA into the central cavity. The model for Rho action proposed by Skordalakes and Berger involves the nascent transcript *rut* site binding the inwardly pointing, RNA-binding domains of the Rho subunits. This automatically forces the 3′ end of the transcript into the cavity, either through the gap present in the ring or until the gap opens within the ring. The presence of RNA in the central cavity may then lead to the closing of the gap. This in turn activates the ATPase function of Rho

allowing it to translocate along the RNA. From the structure they were however not able to establish whether translocation causes looping of the nascent RNA, as a result of *rut* sites remaining bound to Rho, or whether these interactions are disrupted and Rho simply translocates down the transcript with no RNA looping. The transcript must then be released by means of a Rho interaction with the RNAP. Richardson (2003) proposes that this interaction may cause Rho to pull the transcript from the RNAP catalytic site and that this may be coupled with the forward movement of RNAP (without transcript elongation) along the DNA template, as shown by Yarnell and Roberts (1999).

The exact role of Rho in antitermination has not been defined. Two theories have been proposed; the first involves the sequestering of Rho by NusG thus preventing access to the *rut* site or RNA polymerase. The second involves the acceleration of the elongation complex to such an extent that Rho is unable to catch up with and dissociate the RNA polymerase from the template. The second theory seems to be the more plausible given the high elongation rates of *E. coli rrn* operons when compared with *mRNA* elongation (› 80 nucleotides s^{-1} compared with 40 nucleotides s^{-1}) which may, in turn, be due to the antitermination mechanism (Vogel and Jensen, 1994a,b; Vogel and Jensen, 1995).

1.10. Identification of protein – protein interactions in the λ and *E. coli* antitermination mechanisms.

Numerous techniques were used in the study of the λ and *E. coli* antitermination mechanisms to determine protein – protein interactions. These are vital in resolving the components of the complex. Subsequent to the identification of the likely components of the antitermination complexes, through defective λ growth or *rRNA* synthesis in *E. coli*

mutants, *in vitro* transcription assays were used to determine the ability of the proposed complexes to transcribe through Rho-dependent terminators. In λ, the minimum components were identified as the host Nus factors, the λ N protein and the λ *nut* site (Mason and Greenblat, 1991). Squires *et al.* (1993) and Torres *et al.* (2001) showed the requirements of the Nus factors, RNAP, *rRNA nut* sites and S4, in addition to unknown factors, for *E. coli rRNA* antitermination-dependent terminator readthrough. The inter-protein interactions within these complexes were subsequently established using techniques such as pulldown assays and sedimentation equilibrium ultracentrifugation. An interaction was established between NusB and NusE using immobilized GST-NusE and purified NusB (Mason *et al.*, 1992a) and the 1 : 1 stoichiometry of this interaction resolved using sedimentation equilibrium ultracentrifugation (Luttgen *et al*, 2002). The interaction between NusG and Rho was also established using immobilzed NusG and Rho being "pulled" from *E. coli* cell free extracts (Li *et al.*, 1993) and the 1 : 6 (NusG : Rho) stoichiometry again determined using sedimentation equilibrium ultracentrifugation (Pasman and von Hippel, 2000). Interactions between NusA and the λ N protein and the *E. coli* RNAP subunits were established using pulldown assays and immobilized GST-N or His-NusA as shown by Mah *et al.* (1999) and Liu *et al.* (1996). Alternative techniques used to demonstrate protein – protein interactions include the use of size exclusion chromatography to show the co-elution of NusE with RNAP (Mason and Greenblat, 1991) and NMR spectroscopy used to tentatively show a weak interaction between *M. tuberculosis* NusB and NusE (Gopal *et al.*, 2001b). The use of pulldown assays is thus a well established technique for protein – protein interaction

determination and sedimentation equilibrium ultracentrifugation important in the establishment of the subsequent stoichiometrys.

1.11. Project aims.

The presence of only a single *rrn* operon in *M. tuberculosis* means that the regulation and expression of the operon is vital to survival of the organism. As antitermination is important in the regulation of *rrn* operon expression, antitermination and the components of the antitermination complex are likely to be critical to *M. tuberculosis* viability. An understanding of the antitermination mechanism will thus provide even greater comprehension of the biology of the bacterium. This is clearly fundamental for the development of new drug and antibiotic targets. Indeed, the components of this mechanism may even make useful targets themselves.

The aim of the project has been an understanding of the characteristics, functions and interactions of the *M. tuberculosis* Nus proteins within the antitermination complex. This thesis presents biochemical and biophysical data used to try and achieve this aim and resolve some of the structural and functional differences evident between the *M. tuberculosis* and λ*E. coli* models.

Chapter 2

Materials and Methods

2.1. Bacterial strains and growth media.

E. coli strain XL-1 Blue (Stratagene) was used for site-directed-mutagenesis.

E. coli strain DH5α (Invitrogen) was used for all plasmid vector amplification and cloning procedures.

E. coli strain BL21(DE3)p*Lys*S (Novagen) was used for all protein expression and protein purification.

E. coli strains were grown in Luria Broth (10 g Bacto-Tryptone, 5 g Yeast Extract, 10 g NaCl in 1 L Distilled Water) or Terrific Broth (10 g bacto-tryptone, 24 g yeast extract, 4 ml glycerol, 12.54 g K_2HPO_4, KH_2PO_4 in 1 L distilled H2O).

2.2. Plasmid vectors.

Plasmids pET15b – NusA, NusB, NusE, NusE C50S, NusG and RNAPα were supplied by Dr. B. Papavinasundarum

Plasmid pET15b (Novagen) was used as an expression vector to enable the production and purification of NusA, NusB, NusB FE22.23AA, NusE, NusE C50S, NusG and RNA polymerase-α-subunit (RNAPα). The plasmid contains an ampicillin cassette allowing for antibiotic selection. It also contains a coding sequence for a six-histidine tag. Cloning of protein genes in the *Nde*1 and *Bam*H1 restriction sites results in an in-frame his_6-tag, occurring on the N-terminus of the translated protein. A diagram of the pET15b-NusG plasmid is shown in Appendix 2C.

Plasmid pET-22b (Novagen) was used for the cloning and over expression of Rho termination factor. Cloning of Rho into the *Nde*1 and *Xho*1 sites allowed for a C-

terminal His-tag. The plasmid contains an ampicillin cassette allowing for antibiotic selection.

Plasmid pGEX-4T-1 (Amersham Pharmacia Biotech) was used for the cloning and over expression of NusB. Cloning NusB into the *Bam*H1 and *Eco*R1 sites resulted in an N-terminus GST-tag which could subsequently be removed by thrombin digestion. The plasmid contains an ampicillin cassette allowing for antibiotic selection.

Plasmid pGEX-6P-1 (Amersham Pharmacia Biotech) was used for the cloning and co-expression of NusB and NusE. A NusB-RBS-NusE-His_6 construct was cloned into the *BamH*1 site allowing for an N-terminal GST-tag on NusB and a C-terminal His-tag on NusE. The GST-tag on NusB could subsequently be removed by "PreScission Protease" (Amersham Pharmacia Biotech) cleavage. The plasmid contains an ampicillin cassette allowing for antibiotic selection.

2.3. Site-directed-mutagenesis of wild-type NusB to NusB FE22.23AA.

2.3.1. Primer design.

The phenylalanine and glutamic acid, at positions 22 and 23 respectively, of the *M. tuberculosis* NusB are found at the centre of the dimer interface and are likely to be involved in dimer formation in the protein (Gopal et al., 2000). It was thus decided to mutate Phe22 (codon : TTC) and Glu23 (codon : GAG) to Ala (codons GCC and GCT). Primers required for site-directed-mutagenesis must contain approximately 25 – 35 nucleotides and must contain the mutated codon(s). Primers, 31 nucleotides in length, were designed with the double alanine codons in the centre. Sequences of the two

primers designed and used in the mutation of FE22.23AA in NusB are shown in Appendix 1 (Primers **A** and **B**). The Ala codons are highlighted.

2.3.2. Amino acid mutation using site-directed mutagenesis.

Site-directed-mutagenesis was carried out using the Quickchange kit (Stratagene). The reaction mixture was made up of :

5 µl 10X Reaction Buffer
58 ng NusB Template (Wild-type *M. tuberculosis* NusB in pET15b)
229 ng Upper Primer **A** (See Appendix 1)
218 ng Lower Primer **B** (See Appendix 1)
1 µl dNTP mix
40 µl Distilled H_2O
2.5 U *Pfu Turbo* DNA polymerase

Temperature cycling was then carried out, using a Peltier thermal cycler (MJ Research), using the following parameters :

Step 1. 95°C for 30 seconds
Step 2. 95°C for 30 seconds
Step 3. 55°C for 1 minute
Step 4. 68°C for 15 minutes
Steps 2 – 4 were then repeated 16 times

1 µl of restriction enzyme *Dpn*1 was added to the mutation reaction mixture in order to digest the parental (non-mutated) supercoiled dsDNA. Digestion was carried out at 37°C for 1 hour.

1 µl of the mutated DNA (*Dpn*1 treated) was then added to 50 µl *E. coli* XL-1 Blue supercompetent cells and transformation carried out by incubating the cell/plasmid mixture on ice for 45 minutes. The sample was then heat-shocked for 1 minute at 42°C and immediately placed on ice for a further 2 minutes. 1 ml of Luria broth was then added to the cell/plasmid mixture which was subsequently incubated at 37°C for 45 minutes. 100 µl of the mixture was then plated on Luria agar plates (10 g bacto-tryptone,

5 g yeast extract, 10 g NaCl, 15 g difco Agar in 1 L distilled water) containing 0.01% w/v ampicillin and incubated overnight at 37°C. The remaining 900 μl was pelleted at 4000 rpm for 5 minutes and then resuspended in 100 μl Luria broth and plated on a second Luria agar plate containing 0.01% w/v ampicillin.

The pET15b plasmid containing mutated NusB (FE22.23AA NusB) was purified from single colonies using the Qiaprep Quickspin mini-prep kit (Qiagen). The presence of the mutation was determined by DNA sequencing (Cytomyx, Cambridge).

2.4. Cloning of NusB into pGEX-4T-1.

PCR, using KOD DNA polymerase (Novagen), was used to amplify the NusB sequence from *M. tuberculosis* genomic DNA. The following reaction conditions were used for the NusB PCR:

Reaction mixture
26 μl H_2O
5 μl 10 × PCR Reaction Buffer (Novagen)
5 μl 1 mM dNTPs (Novagen)
2 μl 25 mM $MgSO_4$ (Novagen)
5 μl of 20 ng/ μl *M. tuberculosis* genomic DNA
3 μl of 5 pmol/ μl forward primer **C** (See Appendix 1)
3 μl of 5 pmol/ μl reverse primer **D** (See Appendix 1)
1 μl (1U) KOD DNA polymerase

Temperature cycles
Step 1. 2 min. 94°C
Step 2. 15 s 94°C
Step 3. 30 s 55°C
Step 4. 1 min 68°C
Repeat steps 2 - 4 30 ×

*Bam*H1/*Eco*R1 restriction enzyme digestion was then carried out using standard reaction conditions. The restriction digest reaction mixture was incubated at 37°C for 3 hours. Restriction enzymes, polymerase, oligonucleotides and buffers were then removed from the sample using a Qiaquick PCR Purification kit (Qiagen). *Bam*H1/*Eco*R1 digestion of pGEX-4T-1 plasmid was carried out so as to allow for the insertion of the NusB PCR product. The plasmid was cut using standard reaction conditions. The reaction mixture was incubated at 37°C for 3 hours. Digestion was confirmed by running 1μl of the

product on a 1 % agarose gel. Ligation of the amplified and digested NusB and the cut pEX-4T-1 plasmid was then performed. Reaction conditions were as per manufacturer instructions (New England Biolabs). The ligation reaction mixture was incubated at 16°C overnight.

4 μl of the sample was then transformed into *E. coli* DH5α cells (as described for *E. coli* XL-1 Blue cells). Single colonies were then picked from the plates and grown in 5 ml Luria broth, containing 0.01 % ampicillin, overnight. Plasmids were then purified from the overnight cultures using a Qiaprep Spin Miniprep kit (Qiagen). Positive clones were identified using a *Bam*H1/*Eco*R1 restriction enzyme digest using standard reaction conditions. The restriction digest products were then run on a 1 % agarose gel, with a 1 kb DNA ladder marker (Promega), and positive clones were identified by insert release at the correct size. Positive clones were then sequenced (Cytomyx, Cambridge). A diagram of the plasmid construct is shown in Appendix 2a.

2.5. Cloning of GST-NusB – NusE-His into pGEX-6P-1.

The pGEX-4T-1 – NusB plasmid was first digested with *Eco*R1 allowing for the insertion of amplified NusE. Digestion was carried out for 3 hours at 37°C using standard reaction conditions.

Removal of the 5′ phosphates, on the overhangs, by dephosphorylation was then performed using shrimp alkaline phosphatase (SAP – Roche). This prevents recircularisation of the cut plasmid. Dephosphorylation was carried out as per manufacturer instructions (Roche).

SAP was inactivated by incubation at 65°C for 15 minutes. All enzymes, oligonucleotides and buffers were removed using a Qiaquick PCR purification kit.

NusE, containing a synthetic N-terminus ribosome binding site (RBS), was amplified using PCR and the reaction conditions set out below :

Reaction mixture
26 μl H_2O
5 μl 10 × Reaction Buffer
5 μl 1 mM dNTPs
2μl 25 mM $MgSO_4$
5 μl 20 ng/ μl *M. tuberculosis* genomic DNA
3 μl 5 pmol/ μl forward primer **E** (See Appendix 1)
3 μl 5 pmol/ μl reverse primer **F** (See Appendix 1)
1 μl (1U) KOD DNA polymerase

Temperature cycles
Step 1. 2 min 94°C
Step 2. 15 sec 94°C
Step 3. 30 sec 55°C
Step 4. 1 min 68°C
Repeat steps 2 – 4 30 ×

The primers used for NusE (with an N-terminus RBS) are shown in Appendix 1. The NusE forward primer contains a RBS sequence obtained from pGEX-4T-1.

The entire RBS-NusE PCR product was run on a 1 % agarose gel and the band then removed from the gel using a Qiaquick gel extraction kit (Qiagen). This was then digested with *Eco*R1, for 3 hours at 37°C, using standard reaction conditions.

The restriction enzyme was then inactivated at 65°C for 20 minutes and all enzymes, oligonucleotides and buffers removed using a Qiaquick PCR purification kit.

Ligation of *Eco*R1 digested RBS-NusE and pGEX-4T-1 – NusB was carried out using standard reaction conditions. Positive clones were then identified by insert release after *EcoR*1 restriction digestion and verified by DNA sequencing (Cytomyx, Cambridge).

This plasmid construct was then used as a PCR template for amplification of NusB-RBS-NusE and the synthetic addition of a His_6-tag to the C-terminus of NusE. The PCR

reaction conditions are shown below and the primer sequences **G** and **H** shown in Appendix 1:

Reaction mixture
30.6 μl H_2O
5 μl 10 × Reaction Buffer
5 μl 1 mM dNTPs
2 μl 25 mM $MgSO_4$
0.4 μl (0.36 μg) pGEX-4T-1 – NusB – RBS – NusE Template DNA
3 μl 5 pmol/ μl forward primer **G** (See Appendix 1)
3 μl 5 pmol/ μl reverse primer **H** (See Appendix 1)
1 μl (1U) KOD DNA polymerase

Temperature cycles
Step 1. 94°C 2 min
Step 2. 94°C 15 s
Step 3. 55°C 30 s
Step 4. 68°C 1 min
Repeat steps 2 – 4 30 ×

The entire PCR sample was then run on a 1 % agarose gel and extracted from the agarose using a Qiaquick gel extraction kit before *Bam*H1 digestion of the amplified product. The digest was carried out at 37°C for 3 hours. Oligonucleotides, enzymes and buffer were then removed from the PCR preparation using a Qiaquick PCR purification kit (Qiagen).

Plasmid pGEX-6P-1 was digested with *Bam*H1 restriction enzyme digestion using standard reaction conditions.

The 5′ ends were then dephosphorylated using SAP to prevent recircularisation of the plasmid. Dephosphorylation was carried out as per manufacturer instructions (Roche).

SAP was inactivated by treatment at 65°C for 15 minutes. Enzymes, oligonucleotides and buffers were then removed from the linearized plasmid preparation using a Qiaquick PCR purification kit.

Ligation of the amplified NusB-RBS-NusE-His_6 in pGEX-6P-1 was lastly carried out using manufacturer instructions (New England Biolabs). Positive clones were then identified by insert release after *BamH*1 restriction digestion and verified by DNA

sequencing (Cytomyx, Cambridge). A diagram of the plasmid construct is shown in Appendix 2b.

2.6. Cloning of *M. tuberculosis* Rho termination factor.

The gene for the *M. tuberculosis* Rho termination factor was amplified from genomic DNA using PWO DNA polymerase (Eurogentec) and using the following reaction conditions and temperature cycles :

Reaction mixture
37 µl H_2O
2.5 µl 20 ng/ µl *M. tuberculosis* genomic DNA
1 µl 5 mM dNTPs (Eurogentec)
5 µl 25 mM $MgSO_4$ (Eurogentec)
2 µl 25 pmol/ µl Forward Primer **I** (See Appendix 1)
2 µl 25 pmol/ µl Reverse Primer **J** (See Appendix 1)
1 µl (2.5U) PWO DNA polymerase (Eurogentec)

Temperature cycles
Step 1. 94°C 3 min
Step 2. 94°c 1 min
Step 3. 58°C 30 s
Step 4. 72°C 1 min 30 s
Step 5. 72°C for 5 min
Repeat steps 2 – 4 30 ×

Oligonucleotides, enzymes and buffer were removed from the PCR preparation using a Qiaquick PCR purification kit (Qiagen) and the sample then precipitated using 5 µl 5M NaCl, 1 µl 0.5M EDTA and 350 µl cold 100 % ethanol. The ethanol-precipitation mixture was then incubated overnight at -20°C before centrifugation at 13000 rpm for 15 minutes. The pelleted DNA was allowed to dry and resuspended in 42.5 µl H_2O. *Xho*1/*Nde*1 restriction digestion was then carried out as per manufacturer instructions (New England Biolabs). The restriction digestion mixture was incubated at 37°C for 3 hours. *Xho*1 and *Nde*1 were then heat inactivated at 65°C for 20 minutes before enzymes, oligonucleotides and buffers were removed using a Qiaquick PCR purification kit (Qiagen). The Rho PCR product was then run on a 1 % agarose gel and the product cut from the gel using a Qiaquick gel extraction kit (Qiagen).

The Rho PCR product was ligated into *Xho1/ Nde*1 cut pET22b as described previously. Positive clones were identified by insert release after *Xho*1/*Nde*1 restriction digestion and 1 % agarose gel electrophoresis. DNA sequencing was carried out to verify correct Rho sequence replication (Cytomyx, Cambridge). A diagram of the plasmid construct is shown in Appendix 2d.

2.7. Protein expression transformations.

All plasmid constructs used for protein expression and purification were transformed into *E. coli* BL21(DE3)pLysS competent cells. 100 μl of the mixture was then plated on Luria agar plates containing 0.01% w/v ampicillin and 0.002 % w/v chloramphenicol incubated overnight at 37°C. The remaining 900 μl was pelleted at 4000 rpm for 5 minutes and then resuspended in Luria broth and plated on a second Luria agar plate. Single colonies were then used for protein expression and purification.

2.8. Protein Purification.

2.8.1. Purification of NusB, NusB FE22.23AA, NusE and NusE C50S.

After transformation a colony was picked from the plate and used to check for over expression of the relevant protein. The colony was used for the inoculation of 3 ml of Luria Broth containing 0.01% w/v ampicillin and 0.002 % w/v chloramphenicol . The culture was allowed to grow for 3.5 hours. 300 μl of the culture was used to inoculate another 3 ml of Luria broth (containing 0.01% w/v ampicillin and 0.002% chloramphenical) and allowed to grow for 30 minutes. Expression from the T7 lac promoter was induced by the addition of 3 μl 1M IPTG and the culture allowed to grow for 3 hours. The cells were then pelleted by centrifugation at 3500 rpm for 20 minutes.

The supernatant was discarded, the cells resuspended in 20 µl PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.0), 7.5 µl Nupage 4× LDS sample buffer (Invitrogen) and 3 µl 10× sample reducing agent (Invitrogen) and then heated at 70°C for 10 minutes. 10 µl was then loaded onto a 10 % Nupage pre-cast Bis-Tris polyacrylamide gel (Invitrogen) and run at 200 volts (V), 200 mA. A negative control with no IPTG induction was also used.

If over expression was observed, large scale purification was carried out by adding 500 µl of a 3 ml culture grown for 3.5 hours, to 50 ml of Luria broth containing 0.01% w/v ampicillin and 0.002 % w/v chloramphenicol. This was grown overnight at 37°C. 10 ml of this culture was subsequently used to inoculate 750 ml of terrific broth containing 0.01% w/v ampicillin and 0.002 % w/v chloramphenicol . This culture was then grown until the $O.D._{600}$ reading was between 0.8 and 1.0. Expression was then induced by addition of 750 µl of 1M IPTG and the cells allowed to grow at 37°C for 5 hours. The cells were then pelleted by centrifugation at 4000 rpm for 20 minutes, washed with 180 mM NaH_2PO_4/ Na_2HPO_4, 130 mM NaCl and pelleted at 4000 rpm for 20 minutes.

Pelleted cells were then resuspended in 100 ml of lysis buffer (50 mM NaH_2PO_4, 10 mM Tris, 8 M Urea, 100 mM NaCl, 1 complete EDTA free protease inhibitor cocktail tablet, pH 6.0) and sonicated for 5, 1 minute bursts with 3 minute intervals on ice. Cell debris was removed by centrifugation at 18000 rpm at 4°C for 60 minutes and the supernatant incubated for 1 hour with 10 ml of pre-equilibrated Talon Resin (Clontech) at 4°C. The resin was then loaded into a gravity flow column (Sigma) and washed with 20 ml lysis buffer. The bound protein was then eluted using elution buffer (50 mM

NaH_2PO_4, 20 mM PIPES, 8 M Urea, 100 mM NaCl, pH 8.0). Stepwise dialysis, to renature the proteins, was carried out using 3.5 kDa MWCO Snakeskin Pleated dialysis tubing (Pierce), through 7 M, 5 M, 3 M, 1.5 M and 0M Urea, 200 mM NaCl, 50 mM Tris, pH 7.5). The protein was concentrated to 5 ml using a 10 kDa MWCO centrifugal filter tubes (Vivaspin) and loaded onto an XK (26/70) column packed with Superdex 75 (Amersham Pharmacia Biotech) equilibrated in 50 mM Tris, 300 mM NaCl, 1mM EDTA, 1mM DTT pH7.5. Fractions containing NusB, NusB FE22.23AA, NusE or NusE C50S were concentrated using 3.5 kDa MWCO centrifugal filter tubes.

Protein concentrations were calculated using the following absorbance/concentration ratios and extinction coefficients: NusB = 12660 $M^{-1}cm^{-1}$ (0.67 $O.D._{280}$ = 1 mg/ml); NusB FE22.23AA = 12660 $M^{-1}cm^{-1}$ (0.68 $O.D._{280}$ = 1 mg/ml); NusE = 3840 $M^{-1}cm^{-1}$ (0.28 $O.D._{280}$ = 1mg/ml) and NusE C50S = 3840 $M^{-1}cm^{-1}$ (0.28 $O.D._{280}$ = 1 mg/ml).

2.8.2. Purification of GST-NusB.

Over expression of GST-NusB was assayed before commencing purification and cells were induced, grown, washed and pelleted as described before.

Pelleted cells were resuspended in 100 ml lysis buffer (300 mM NaCl, 50 mM Tris, 1 mM EDTA, 1 mM DTT, 1 complete EDTA free protease inhibitor cocktail tablet, pH 7.5) and sonicated 5 × 1 minute bursts with 3 minute intervals on ice. Cell debris was removed by centrifugation at 18000 rpm, 4°C for 60 minutes. The supernatant was then added to 10 ml of glutathione sepharose 4B fastflow resin (Amersham Pharmacia Biotech) in a gravity flow column (Bio-Rad) and the resin was washed with 5 × 20 ml of lysis buffer. GST-NusB was then eluted with 20 × 1.5 ml elution buffer (lysis buffer +

20 mM glutathione (Sigma)) and pooled samples concentrated to 5 ml using a 30 kDa MWCO centrifugal filter tubes (Vivaspin).

An XK (26/70) column packed with Superdex 200 (Amersham Pharmacia Biotech), attached to an AKTA prime pump and fraction collector (Amersham Pharmacia Biotech), was then equilibrated with GF buffer (300 mM NaCl, 50 mM Tris, 1 mM EDTA, 1 mM DTT, pH 7.5) before loading of 5 ml of GST-NusB. GST NusB was then eluted in 4.5 ml fractions using GF buffer running at 1.5 ml/ min.

Pooled peak elution fractions were concentrated using 30 kDa MWCO centrifugal filter tubes (Vivaspin) and the concentration estimated using the absorbance/concentration ratios and extinction coefficient 53580 $M^{-1}cm^{-1}$(1.245 $O.D._{280}$ = 1mg/ml).

2.8.3. Purification of co-expressed GST-NusB and NusE-His_6.

Over expression of GST-NusB and NusE-His_6 was checked as before. Cells were induced, grown, washed and pelleted as above.

Pelleted cells were resuspended in 100 ml lysis buffer (300 mM NaCl, 50 mM Tris, 1 complete EDTA free protease inhibitor cocktail tablet, pH7.5) and sonicated for 5, 1 minute bursts with 3 minute intervals on ice. Cell debris was removed by centrifugation at 18000 rpm, 4°C for 60 minutes. The supernatant was then incubated (rolling) with 10 ml of Ni-NTA His•Bind Superflow resin (Novagen) for 1 hour at 4°C, allowing for the selection of His-tagged NusE. The resin/ supernatant mixture was then loaded onto a gravity-flow column (Sigma) and flow through discarded. The resin was washed with 20 ml lysis buffer and then with 5 × 10 ml lysis buffer + 20 mM imidazole. Protein(s) were eluted with 20 × 1.5 ml lysis buffer + 500 mM imidazole. Pooled fractions were

collected and dialysed (using Snakeskin Pleated Dialysis Tubing, Pierce) into a Tris/ NaCl buffer (300 mM NaCl, 50 mM Tris, 1 mM EDTA, 3 mM DTT, pH7.5) overnight at 4°C.

The dialysed protein solution was then circulated through 2 ml of Glutathione Sepharose 4B resin (loaded on a gravity-flow column (Bio-Rad)), using a P1-pump (Amersham Pharmacia Biotech), for 4 hours at 4°C. This allowed for the selection of GST-tagged NusB.

GST-tag cleavage was then carried out by incubating the GST-NusB bound resin with 75 μl (75U) PreScission protease in 1ml 50 mM Tris, 150 mM NaCl, 5mM DTT, 5mM EDTA buffer for 1 hour at room temperature. This fraction was collected and the resin then washed with 4 × 1ml Tris, NaCl, DTT, EDTA buffer to elute all traces of protein(s). Elution fractions should therefore contain GST-free NusB and His-tagged NusE.

The elution fractions were pooled and loaded onto an equilibrated XK (26/70) column packed with Superdex 200. Proteins were eluted using GF buffer (300 mM NaCl, 50 mM Tris, 1 mM EDTA, 3 mM DTT, pH7.5), peak elution fractions were pooled and concentrated using a 5 kDa MWCO centrifugal filter tubes.

2.8.4. Purification of Rho termination factor.

Over expression of Rho was assayed by small scale culture and cells then induced, grown, washed and pelleted as before.

His-tag purification of Rho was carried out using Ni-NTA His•Bind Superflow resin. Pooled elution samples were then dialysed into a Tris/ NaCl buffer (50 mM NaCl, 50 mM Tris, 1 mM EDTA, 3 mM DTT, pH7.5) overnight at 4°C.

The His-affinity purified protein was loaded onto a 5 ml HiTrap Heparin HP column (Amersham Pharmacia Biotech) using a P1-pump. This utilises the nucleic acid binding properties of Rho. An AKTA Prime pump and fraction collector were then used to pump a linear 50 mM – 1M NaCl gradient in 30 minutes through the Heparin column. Rho was eluted at approximately 600 mM NaCl. Peak elution samples were pooled and concentrated to 5 ml, using 30 kDa MWCO centrifugal filter tubes, before loading onto an XK (26/70) Superdex 200 column. Gel filtration was carried out using GF buffer (300 mM NaCl, 50 mM Tris, 1 mM EDTA, 3 mM DTT, pH7.5) and peak elution samples pooled and concentrated using 30 kDa MWCO centrifugal filter tubes.

The concentration of purified Rho was determined using the absorbance/concentration ratio and extinction coefficient 8960 $M^{-1}cm^{-1}$ (0.135 O.D. = 1 mg/ml)

2.8.5. NusA, NusG and RNAPα.

Purified NusA, RNAPα and NusG were supplied by Dr Ian Taylor and Dr B. Beuth. (NIMR, London) Protein concentrations were determined using the following absorbance/concentration ratios and extinction coefficients:

NusA = 17900 M^{-1} cm^{-1} (0.45 $O.D._{280}$ = 1mg/ml), NusG = 19060 M^{-1} cm^{-1} (0.69 $O.D._{280}$ = 1 mg/ml) and RNAPα = 17210 M^{-1} cm^{-1} (0.43 $O.D._{280}$ = 1 mg/ml)

2.9. Protein storage.

All proteins were stored in 50 % glycerol at -20°C

2.10. Electrospray ionisation mass spectroscopy of purified proteins.

All proteins submitted to electrospray ionisation mass spectroscopy (ESI MS) were diluted to a concentration of 2 μM in 0.1 % formic acid, 10 % acetonitrile (v/v). ESI MS (using a Platform Electrospray Mass Spectrometer, Micromass, Manchester, UK) was then carried out by Dr. S Howell and Dr. L Haire (NIMR, London).

2.11. Pulldown assays.

2.11.1. His-affinity pulldown.

250 μl of Talon Resin (Clontech) was washed twice with a PBS wash/binding buffer, pH 7.0. 50 μl of purified protein (NusA (98 μM), NusB (659 μM), NusE (163 μM) or RNA polymerase α-subunit (157 μM)) and 50 μl wash/binding buffer was added to the resin and allowed to incubate on a roller at 4°C for 1 hour. The resin was then washed twice with the wash/binding buffer before adding 50 μl of *M. tuberculosis* cell free extract and allowed to incubate for 1 hour at 4°C. Next the resin was washed twice with 100 μl wash/binding buffer. The resin was then washed four times with 100 μl elution buffer (PBS, 250 mM Imidazole, pH 7.5).

Samples of all supernatant's from the above steps were run on a 10% Bis-Tris pre-cast gels (Invitrogen). Supernatants were obtained by centrifugation at 13 000 rpm for 1 minute. Supernatants were then carefully removed.

2.11.2. Western blot analysis of pulldown assays.

Wash and elution samples were then probed with NusA, NusB, NusE or RNAPα antibodies. Wash and elution samples from each respective pulldown were run on 10% Bis-Tris pre-cast Gels at 200 V, 200 mA. Gels were then equilibrated in semi-dry transfer buffer (39 mM Glycine, 48 mM Tris Base, 0.037% SDS, 20% Methanol) for 15 minutes and placed on PVDF Immobilon-P blotting membrane (Millipore), treated with methanol then semi-dry transfer buffer. Three layers of 3M filter paper (Millipore) were placed above and below the gel and membrane and placed on a semi-dry blotting unit (Biometra). Transfer was then run at 60 mA per gel for 1 hour. Once transfer was complete the membranes were air dried and transfer checked by staining with Ponceau S stain (Sigma). Membranes were then incubated in 0.1% TTBS (20 mM Tris, 0.5 M NaCl, 0.1 % Tween 20) containing 10% milk powder and 2 mM EDTA. Membranes were then washed in 0.1% TTBS (without milk powder) for 15 minutes, washed in water and incubated in a 1/250 dilution of primary antibody in 0.1% TTBS containing 3% milk powder and 2 mM EDTA for 1 hour. Membranes were subsequently washed three times in 0.1% TTBS before incubating in a 1/250 dilution of anti-mouse secondary antibody, conjugated with horseradish peroxidase (HRP), (Dako) for 1 hour. Membranes were then washed four times in 0.1% TTBS, twice in 1X TBS (20 mM Tris, 500 mM NaCl, pH 7.5) and once in PBS pH 7.2 – 7.4. Blots were lastly developed using (1 mM Di-aminobenzidine, 3 mM NiCl, 0.005% H_2O_2), rinsed in water and allowed to air dry.

2.12. GST-NusB pulldowns and NusA, NusE and RNAPα antibody western-blotting.

2.12.1. GST-NusB pulldowns.

200 μl of glutathione sepharose 4B was equilibrated with 2 ml wash buffer (300 mM NaCl, 50 mM Tris, pH7.5). 100 μl of 3.2 mg/ml GST-NusB was incubated with the resin for 30 minutes at 4°C. 100 μl wash buffer was used as control. The resin was then centrifuged at 13000 rpm, 1 min, 4°C, the supernatant removed (flow through (F.T.)) and the resin washed 3 × 1 ml with wash buffer. 100 μl of a 4.5 mg/ml *M. tuberculosis* CFE was then added to the resin and incubated for 30 minutes, 4°C. The resin was centrifuged at 13000 rpm, 1 min, 4°C, the supernatant removed and the resin washed 5 × 1 ml wash buffer (Wash). 100 μl of elution buffer (wash buffer + 20 mM glutathione) was then added to the resin and incubated for 30 minutes, 4°C. The resin was subsequently pelleted (13000 rpm, 1 min, 4°C) and the elution fraction removed. This was repeated 3 times (with 5 minute incubations) and 20 μl of the F.T, wash and elution samples run on 10 % Bis-Tris pre-cast gels, at 200V, 200 mA, with wide-range ECL molecular weight markers (Amersham Pharmacia Biotech) and stained with Simply-Blue Safe stain (Invitrogen).

2.12.2. Western blots using NusA, NusE and RNAPα antibodies.

Blotting of pulldown samples was carried out as before. Blots were, however, developed using the ECL Plus chemiluminescent detection reagents (Amersham Pharmacia Biotech). 1:1000 and 1:5000 primary and secondary antibody dilutions respectively were used.

After the last PBS wash the membrane was developed as per ECL Plus instructions (Amersham Pharmacia Biotech). The detection solution was drained off, the membrane wrapped in Saran wrap and placed in an X-ray film cassette. Autoradiography film (Kodak MXB NIF100) was placed on top of the membrane and exposed for 30 seconds. The film was then developed using an AGFA developer.

Preliminary development was done using ECL protein molecular weight markers (Amersham Pharmacia Biotech) and conjugating streptavidin-HRP to the markers using ECL Streptavidin-HRP conjugate (Amersham Pharmacia Biotech). After probing of the membrane with primary and secondary antibodies and subsequent 0.1 % TTBS washes (as above), the membrane was incubated in a 1:10000 dilution of Streptavidin HRP-conjugate for 1 hour at room temperature. The membrane was then washed 4 × 5 min 0.1 % TTBS, 2 × 5 min 1 × TBS and lastly 2 × 15 min PBS. Membrane development was then carried out using the ECL Plus Detection kit.

2.13. Preparation of *M. tuberculosis* cell free extracts.

0.5 ml of an *M. tuberculosis* culture in logarithmic phase was used to inoculate 175 ml of Dubos broth (Difco) containing 0.2 % glycerol and 10 ml Dubos medium albumin (Difco). The culture was grown for 7 days at 37°C reaching an approximate $O.D._{600}$ of 1.3.

Cells were harvested by centrifugation at 13000 rpm for 30 min. Cells were washed once with 150 ml PBS (137 mM Nacl, 2.7 mM KCL, 10 mM Na_2HPO_4, pH 7.0) and centrifuged at 10000 rpm for 30 min. The supernatent was discarded and cells washed twice with 2 ml PBS followed by centrifugation at 10000 rpm for 30 min. The pelleted

cells were resuspended in 500 µl PBS containing 1 mM Pefabloc (Roche) and an equivalent volume of 150 – 212 µm glass beads (Sigma) added. Cells were disrupted by beadbeating, using a ribolyser (Hybaid), 3 × 1 min with 1 min intervals on ice, at the 4200 setting. Centrifugation was carried out at 13000 rpm for 15 min, the supernatent removed and centrifuged for a further 5 min at 13000 rpm. Lastly the supernatent was filtered through a 0.22 µm centrifugal filter (Amicon) by centrifugation at 13000 rpm for 15 min.

Protein concentration was then estimated using the BCA protein assay kit (Pierce) as per manufacturer instructions.

2.14. Co-refolding and His-affinity pulldowns of NusB/ NusE and FE22.23AA/ NusE mixtures.

2.14.1. Purification and refolding of NusB/ FE22.23AA and NusE.

His-tagged NusB and NusB FE22.23AA were purified as described in 2.8.1. The N-terminal His-tags were removed by thrombin digestion. 12.7 µg (48U) of thrombin (Haemotologic Technologies, Inc.) was added per mg of protein and incubated overnight at 4°C. The reaction was stopped by the addition of 10 µl PPack inhibitor (Calbiochem) per 40U of thrombin used and left at room temperature for 30 minutes. The protein solution was dialysed into 300 mM NaCl/ 50 mM Tris, pH 7.5 and the free his-tag then removed by passing the protein solution through 125 µl of Talon his-affinity resin (Clontech).

NusE was affinity purified, as above, but not renatured. 30 µM NusB and 10 µM NusE were placed in 3.5 kDa MWCO Slide-a-lyser cassettes (Pierce) and refolded by stepwise

dialysis through 8M, 7M, 5M, 3M, 1.5M Urea, 300mM NaCl, 50 mM Tris, pH 7.5 and final dialysis into 300mM NaCl, 50 mM Tris pH 7.5. Each dialysis step was carried out for two hours at 4°C.

His-tagged NusE was then selected for by incubating the NusB or FE22.23AA and NusE mixture with Talon his-affinity resin for 30 min, 4°C. The supernatant was removed and the resin washed 10 × 1 ml 300mM NaCl, 50 mM Tris, 20 mM imidazole, pH 7.5 (Wash). Finally NusE was eluted with 3 × 100 μl elution buffer (300mM NaCl, 50 mM Tris, 500 mM imidazole, pH 7.5).

Wash and elution samples were separated on 16 % Tris-glycine pre-cast gels (Invitrogen) and stained with Simply-Blue safe stain.

2.15. Dynamic light scattering.

Dynamic light scattering was carried out using a DynaPro-801 dynamic light scattering machine (DynaPro). Protein samples were diluted to 1 mg/ml and measurements carried out using a 10 mm path-length quartz cuvet.

2.16. Circular dichroism of purified proteins.

Proteins were dialysed into 20 mM Tris, 300 mM NaCl, pH 7.5 and diluted to 9 μM for far-UV circular dichroism (CD) and to 1 mg/ml for near-UV CD. Spectra were recorded using a Jasco-J715 Spectropolarimeter. Far-UV CD was carried out in a quartz-cuvet with a path-length of 1 mm and near-UV CD in a quartz cuvet with a path-length of 10 mm. Wavelengths 260 nm – 190 nm were used for far-UV CD and 340 nm – 255 nm for near-UV CD. The scan rate was set at 100 nm/min, time constant 0.25 s and the spectral

bandwidth 2 nm. Final spectra are the average of fifteen scans. Data analysis was carried out using Specproc (Dr. S. Martin, NIMR, London).

2.17. Sedimentation equilibrium ultracentrifugation .

Sedimentation equilibrium ultracentrifugation was carried out using a Beckman Optima XL-A analytical ultracentrifuge and an AN-60 Ti rotor. Sample concentrations ranging from 0.25 mg/ml to 1.5 mg/ml were prepared and loaded into a 6-channel centre piece and centrifugation carried out using the following speed, delay and absorbance scan settings:

NusB and NusB FE22.23AA			**NusG**		
18000 rpm	24 hours	1 scan 280 nm	14000 rpm	20 hours	1 scan 280 nm
18000 rpm	2 hours	1 scan 280 nm	14000 rpm	2 hours	1 scan 280 nm
18000 rpm	2 hours	1 scan 280 nm	14000 rpm	2 hours	1 scan 280 nm
22000 rpm	15 hours	1 scan 280 nm	17000 rpm	16 hours	1 scan 280 nm
22000 rpm	2 hours	1 scan 280 nm	17000 rpm	2 hours	1 scan 280 nm
22000 rpm	2 hours	1 scan 280 nm	17000 rpm	2 hours	1 scan 280 nm
32000 rpm	15 hours	1 scan 280 nm	21000 rpm	16 hours	1 scan 280 nm
32000 rpm	2 hours	1 scan 280 nm	21000 rpm	2 hours	1 scan 280 nm
32000 rpm	2 hours	1 scan 280 nm	21000 rpm	2 hours	1 scan 280 nm
			42000 rpm	18 hours	1 scan 280 nm

2.18. Analytical gel filtration.

Analytical gel filtration was carried out using a Jasco PU-1580 HPLC pump, a Jasco UV-1575 UV/Vis detector and a Jasco UV-970M 4λ UV detector. A calibrated Superdex 200 HR 10/30 pre-packed column (Amersham Pharmacia Biotech) was used for NusB and NusB FE22.23AA gel filtration and a Superose 6 HR 10/30 pre-packed column (Amersham Pharmacia Biotech) was used for Rho, NusA and NusG gel filtration. Sample volumes of 100 μl were loaded onto columns using concentrations of 18.5 μM Rho, 6.6 μM NusA, 15.4 μM NusG, 27 μM NusB FE22.23AA and

concentrations of 0.26 μM to 27 μM for NusB. A flow rate of 0.4 ml/min was used for all separations. All analysis was done using Borwin Chromatography software.

2.19. Limited proteolysis of NusG.

Trypsin and Chymotrypsin (Promega) were used for limited proteolysis of NusG. Trypsin or chymotrypsin (92 ng/ml) was added to 50 μl of 2.3 μg/μl of NusG to give a 1:250 dilution of the enzyme. Proteolysis was carried out at room temperature and 8 μl of reaction mixture removed at time points 0, 2, 5, 10, 20, 40 and 70 minutes. Cleavage was stopped with the addition of 1 μl 20 mM Pefabloc (Roche) and 1 μl of Protector (Roche). Proteolytic products were then run on 10 % Bis-Tris pre-cast gels with low range molecular weight markers (Sigma) and stained with Simply Blue safe stain.

2.20. N-terminal sequencing of proteolytic products.

Protein samples for N-terminal sequencing were resolved on 10 % pre-cast gels and then blotted onto PVDF membrane paper (ProBlot). The membrane was initially treated with 100 % methanol, followed by incubation with blotting buffer (10 mM CAPS, pH 11.0, 10 % methanol). The gel was also incubated in blotting buffer for five minutes prior to assembly of the blotting sandwich. Blotting was then carried using a Biometra blotter at 60 mA/ gel, 5 – 10 V for 1 hour. Membranes were then stained with 0.1 % Coomassie Blue R-250, 40 % methanol, 1 % acetic acid until bands appeared and then destained using 50 % methanol. Band sequencing was then carried out by Babraham Biosciences Technologies, Babraham, U.K.

2.21. Amino acid sequence alignment of NusG.

Amino acid sequence alignment of *M. tuberculosis*, *A. aeolicus* and *E. coli* NusG was carried out using the T-coffee program (www.chembnet.org/software/Tcoffee.html - Notredame *et al.*, 2000).

2.22. *M. tuberculosis* NusG secondary structure prediction.

M. tuberculosis NusG secondary structure prediction was carried out using the PSI-Predict program (http://bioinf.cs.ucl.ac.uk/psipred - Jones, 1999).

Chapter 3

Biochemical and Biophysical Techniques used in this Study

3.1. Sedimentation equilibrium ultracentrifugation.

Analytical ultracentrifugation is an important technique for the study of the oligomerisation state of macromolecules. It allows for molecular weight determination, homogeneity analysis of a solution, complex size determination and the calculation of the strength of interactions within these complexes. This makes the method invaluable in initial characterisation of macromoleculer complexes and the components of these complexes. Unlike dynamic light scattering and size exclusion chromatography (other methods commonly used for molecular weight size determination), the measurement of molecular weight by sedimentation equilibrium ultracentrifugation is independent of the shape of the molecule. Molecular shape is unaccounted for in dynamic light scattering and size exclusion chromatography as the parameter measured by these techniques is related to the translational diffusion coefficient (D_T) rather than molecular weight itself. D_T is correlated to molecular weight however it is directly related to the frictional coefficient, f through equation 1.

$$D_T = RT/N_A f \quad (1)$$

where R = Gas constant,
T = temperature (Kelvin)
N_A = Avogadro's number

The frictional coefficient (f) is indicative of the ability of the molecule to undergo translational motion, affected by both molecular weight and shape and as D_T is directly dependent on f, it is obvious as to why molecular weight determination by dynamic light scattering and size exclusion chromatography are shape-dependent (Cantor and Schimmel, 1980).

Sedimentation equilibrium ultracentrifugation is therefore used for molecular weight determination independent of molecule shape. During centrifugation two processes occur, the first being sedimentation towards the bottom of the cell as a result of a gravitational force (dependent on the buoyant molecular weight of the molecule) and the second being diffusion opposing the establishment of the concentration gradient. Both of these forces are accounted for in the concept of flux (the rate of flow of molecules across a given surface) and shown in equation 2.

$$J_i = -L_i\left(\partial U_i / \partial r\right) \quad (2)$$

where J_i = flux for component i
$\delta U_i/\delta r$ is a generalised gradient of chemical potential,
L_i is conductivity (and is related to f and contributes to both flow due to sedimentation and diffusion)
r = radial distance from the centre of rotation

The movement of molecules within an ultracentrifuge cell can then be shown by equation 3:

$$J = \left[\frac{M(1-v\rho)}{N_A f}\right]\omega^2 rC - \left[\frac{RT}{N_A f}\right]\left(\frac{dC}{dr}\right) \quad (3)$$

where M = molecular weight
v = partial specific volume
ρ is the solute density
ω is the angular velocity

$[M(1\text{-}v\rho)/N_A f]$ is the term for the sedimentation coefficient (S) and $[RT/N_A f]$ corresponds to the translational diffusion coefficient (D_T) and U_i has been replaced by the concentration, C. At low rotor speeds (as used in sedimentation equilibrium ultracentrifugation), back diffusion, as a result of the chemical potential gradient, balances movement due to the gravitational field until equilibrium is reached. This

results in an exponential increase in concentration with increasing radial distance (See Figure 10). At equilibrium net movement within the cell becomes zero at all points (i.e. $J = 0$).

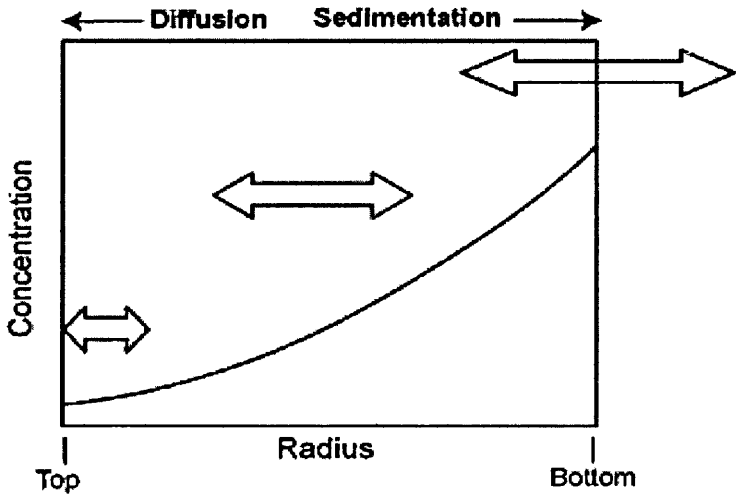


Figure 10 : Diffusion and sedimentation forces occurring within an ultracentrifugation cell.
The diagram shows the relative strength and direction of balanced diffusion and sedimentation forces occurring during sedimentation equilibrium ultracentrifugation. At equilibrium, the concentration distribution of the solute is a squared exponential. Adapted from (G. Ralston, Introduction to Analytical Ultracentrifugation, Beckman).

Equation 3 can thus be written as :

$$\left[\frac{M(1-v\rho)}{N_A f}\right]\omega^2 rC = \left[\frac{RT}{N_A f}\right]\left(\frac{dC}{dr}\right) \quad (4)$$

meaning that the $N_A f$ (relating to molecular shape) is removed and the concentration gradient of the molecule is simply related to molecular weight and is completely independent of shape. Equation 4 is then rearranged to form equation 5 and integration carried out between C_0 and C_x and r_0 and r_x giving equations 6a and 6b:

$$\int_{C_0}^{C_x}\left(\frac{1}{C}\right)dC = \frac{M\omega^2(1-v\rho)}{RT}\int_{r_0}^{r_x} r dr \quad (5)$$

$$\ln\left(\frac{C_x}{C_0}\right) = \frac{M\omega^2(1-v\rho)}{2RT}\left(r_x^2 - r_0^2\right) \text{ (6a) or}$$

$$C_x = C_0 e^{\frac{M\omega^2(1-v\rho)}{2RT}(r_x^2 - r_0^2)} \text{ (6b)}$$

The spectrophotometer of the Beckman XL-A is able to measure O.D. at a set wavelength at numerous points in the cell between r_0 and r_x. Absorbance can then replace C in equation 6b based on the Beer-Lambert law :

$$A_x^{\lambda} = \varepsilon^{\lambda} C_x l \quad (7)$$

The molecular weight is determined by least-squares fitting of equation 6b between r_0 and r_x and solving for M and A_0. Data analysis is carried out using Origin Optima XL-A/ XL-I Data Analysis software with partial specific volumes (calculated from the amino acid content of the protein) and solvent densities determined using Sednterp software. Ideally offset values (non-sedimenting material causing a no-zero baseline) included in data analysis should be as close to zero as possible. In the case of polymers, data will not ideally fit a monomer model and dissociation constants can thus be determined for monomer – polymer equilibrium (Cantor and Schimmel, 1980; G. Ralston, Introduction to Analytical Ultracentrifugation, Beckman; Reviewed in Laue and Stafford, 1999; Schuster and Toedt, 1996).

3.2. Dynamic light scattering.

Dynamic Light scattering (DLS) is useful in the characterisation of molecular weight or conformation of a protein as well as the interaction or aggregation of protein molecules. It is based on the principle of constructive and destructive interference of transmitted light as a result of the Brownian motion of the molecules. The technique utilises a laser beam (monochromatic, single beam of light) directed through an aqueous protein preparation.

In static light scattering (SLS), the absolute intensity of the scattered light is used to estimate the molecular weights of particles. Scattering by a single molecule will result in reradiation of the beam of light in all directions and the intensity of this light, in planes normal to the plane of polarization, can be predicted by the Rayleigh scattering relation:

$$I_{S,\theta} = \frac{4\pi^2 M^2 \sin^2\theta (dn/dc)^2 I_0}{N_A{}^2 \lambda^2 r^2} \quad (8)$$

where θ = the angle between the incident light and the direction of scatter
M = molecular mass
dn/dc = refractive index increment
I_0 = intensity of the incident light
N_A= Avogadro's Number
λ = wavelength of the incident light
R = distance of molecule to the point of detection

When numerous particles are illuminated by a laser beam, the intensity of the scattered light at the detector is given by the square of the vector sum of the light reradiated from each particle.

This method is useful for the calculation of the molecular weight of a given molecule or solution of molecules as well as the radius of gyration of a molecule provided the

particle is big enough. It, however, requires information such as *dn/dc* and a very accurate value for the concentration of the solution.

DLS negates the need for such information through the use of the movement of the molecules in solution, by Brownian motion. DLS utilises the resultant fluctuations in intensity and the time taken for these fluctuations to occur, to estimate the diffusion constant (D_T) of the molecule. During DLS, the scattered light (photon count) is detected by an avalanche photo diode (APD) and information regarding the time dependence of the intensity fluctuations is determined. Time dependence of these fluctuations is determined using a mathematical procedure known as autocorrelation. Autocorrelation involves the multiplication of the signal from the APD by itself, but delayed by a delay time τ, and summed and averaged until a statistically relevant number of intensity fluctuations have been obtained. This gives rise to the intensity autocorrelation function, $G^2(\tau)$ as shown in equation 9 :

$$G^2(\tau) = \int I(t)I(t+\tau)dt / N \quad (9)$$

where $I(t)$ = Intensity at time t
$I(t+\tau)$ = Intensity at time t + delay time τ
N = background photon count (measurement of photon count of solvent only)

It can be been shown that $G(\tau)$ decays exponentially with τ if a single molecular species is responsible for the scattering.

The intensity autocorrelation coefficient can then be converted to an amplitude autocorrelation coefficient using equation 10:

$$G^2(t,t+\tau) = 1 + \left|G^1(t,t+\tau)\right|^2 \quad (10)$$

Molecules undergoing simple Brownian motion will then give equation (11) :

$$G^1(\tau) = 1 + \exp(-2D_T q^2 \tau) \quad (11)$$

where $G(\tau)$ = autocorrelation function
D_T = translational diffusion coefficient
τ = decay time
And where q is defined by the wave vector:

$$q = \frac{4\pi n_s \sin(\theta/2)}{\lambda} \quad (12)$$

where θ = the scattering angle
n_s = the refractive index of the solvent
λ = the wavelength of the incident radiation
and the translational diffusion coefficient of the molecule is given by Stokes-Einstein equation :

$$D_T = \frac{k_B T}{6\pi\eta R_H} \quad (13)$$

where k_B = Boltzmann constant
T = temperature (Kelvin)
η = the viscosity of the solvent
R_H = the hydrodynamic radius of the scattering molecule

DLS thus has the advantage over SLS in that solution concentrations and refractive index increments do not have to be determined when estimating the hydrodynamic radius of the experimental molecule.

The calculated correlation functions can then be used to fit the following function :

$$\ln\left|G^1(\tau)\right| = -2D_T q^2 \tau \quad (14)$$

and a plot of lnG (τ) vs. τ made and D_T estimated from the slope of this graph. As D_T is now known, R_H can be calculated based on equation 13, and molecular weights

estimated using relationships between R_H's and molecular weights of well characterised proteins. If the slope is not a straight line it is an indication of polydispersity (a mixture of different sized particles).

Dynamic light scattering is therefore useful for the estimation of molecular weights of monomeric, polymeric or aggregated protein molecules. As the shape of the protein molecules is incorporated in D_T, dynamic light scattering also allows for the prediction of compact or loosely packed proteins as the calculated molecular weights will not match the actual molecular weights. For example a tightly packed protein will have a lower than actual predicted molecular weight whereas a loosely packed or extended protein will have a higher than actual predicted molecular weight. The shape dependence of D_T is shown by the relation of the frictional coefficient (f) with the diffusion coefficient seen in equation 1:

$$D_T = RT/Nf \quad (1)$$

where $f = 6\pi\eta R_H$ (15) and R_H is the hydrodynamic radius of the molecule

A ratio of f/f_0 (frictional ratio), where $f_0 = 6\pi\eta R_0$ (R_0 is the radius of a sphere), is thus the difference in shape of the experimentally determined molecule from that of a regular sphere. These can however only be determined if the shape and radius of the molecule are known (DynaPro MS and MSTC : Theory and Data Interpretation, 1998; Coligan *et al.*, 2003; van Holde *et al.*, 1998).

3.3. Circular dichroism.

Circular dichroism is an important technique useful for the study of protein secondary structure, conformational changes to proteins in changing extrinsic environments, protein denaturation and protein-ligand interactions. Circular dichroism (CD) utilises circular polarised light (light where the magnitude of the electric vector remains constant while its direction changes). Circular polarised light can be produced by the superimposition of two plane-polarised beams that have the same wavelengths and amplitudes but differ in phase by a quarter of a wavelength and by 90° in the plane of polarisation. The electric vector thus takes on the shape of a helix which can either be left or right handed. In linear polarized light, the direction of the electric vector is constant where as the magnitude varies.

Optical rotary dispersion spectroscopy (ORD) was the polarimetry technique which initially used plane polarised light. ORD measures the ability of an asymmetric or chiral molecule to cause a change in the plane of the polarised light due to changes in the differential refractive indices of the left and right-handed circular polarised light without absorption being required. The angle of rotation (θ_λ) of the resultant beam is measured. ORD is useful for the determination of the chirality of molecules.

Circular dichroism then became the technique of choice as more information could be obtained from the effects of chiral molecules on circular polarised light. CD is an investigation of the difference in absorption of the left and right-handed circular polarised light. Proteins are made up of highly ordered structures, such as α-helices and β-sheets and the asymmetric orientation of the peptide bonds in these ordered structures

leads to differences in absorption of the right and left-handed polarised light. This occurs mainly in the far-UV (260 – 190 nm). The amino acids tyrosine, tryptophan and phenylalanine are the major chromophores resulting in changes in circular polarised light absorption in the near-UV (340 – 255 nm) range. These differences in absorption cause the resultant beam (a recombination of both left and right-handed emergent beams) to form an ellipse. CD spectropolarimeters are capable of measuring this difference in absorption and hence the ellipticity (θ – ratio of the magnitudes of the major and minor axes of the ellipse) of the resultant elliptically polarised light.

As CD is an absorbance measurement, it is based on the Beer-Lambert law:

$$A = \varepsilon_M . C l \quad (16)$$

where A = absorbance
ε_M = molar extinction coefficient (M^{-1} cm^{-1})
C = protein concentration (M)
l = path length (cm)

The CD version of this equation is shown below:

$$\Delta\varepsilon_M = {}^{\Delta A}\!/_{C l} \quad (17)$$

where $\Delta\varepsilon_M = (\varepsilon_{Left} - \varepsilon_{Right})$ = the molar CD extinction coefficient

The difference in absorption is frequently presented as θ by commercial spectropolarimeters which is related to circular dichroism by equation 18 :

$$\Delta A = \theta / 3298 \quad (18)$$

CD of proteins is measured in two regions, the near and the far UV regions. Near-UV CD (Wavelengths 340 nm – 255 nm) is useful for determining whether a protein folds

into a 3-dimensional structure or simply retains its secondary structure. Near-UV CD is determined by the amino-acids Phe, Tyr, Trp and Cys and all four give rise to specific peaks at defined wavelengths. Phe gives rise to peaks in the 255 – 270 nm region, Tyr gives rise to peaks in the 275 – 282 nm region and Trp gives rise to peaks above 280. The intensity of the peaks is determined by the immobility or interaction of these residues with neighbouring residues with both immobility and interaction resulting in higher intensities. This thus allows for the prediction of the presence of tertiary structure

Far-UV CD is an important technique for the determination of a proteins secondary structure content. The spectra determined by peptide bond absorption are purely dependent on the presence of α-helices, β-sheets and random coils or unstructured regions. The spectra of all helical proteins show intense negative peaks at positions 208 nm and 222 nm and an intense positive peak between 191 nm and 193 nm. β-sheet containing proteins usually show a single negative peak in the 210 nm – 225 nm region and a positive peak in the 190 nm – 200 nm region however the intensities are lower than those shown by α-helices. β-sheets are in fact better defined at very short wavelengths (below 180 nm). Unstructured or denatured peptides have a strong negative peak in the 195 nm – 200 nm region and a weaker band between 215 nm and 230 nm. Proteins containing both α-helices and β-sheets have spectra dominated by the α-helical components and show peaks in the 190 nm – 195 nm, 208 nm and 222 nm regions. A broad peak extending from 210 nm to 220 nm may also be seen as a result of the combined effects of the α-helices and β-sheets (Martin and Bayley, 2002; van Holde *et al.*, 1998; Wilson and Walker, 1994).

3.4. Electrospray mass spectroscopy.

Electrospray ionisation mass spectroscopy is a useful technique for verification of molecular masses of purified proteins. The technique involves the transfer of the protein molecule into gas phase via electrospray ionization (ESI). The sample is first dissolved in a polar solvent and then forced through a narrow capillary, the end of which is under a strong electric field. This results in the production of highly charged droplets. As the droplets move towards the detector, they undergo solvent evaporation resulting in diminishing droplet size and subsequent charge density increase. Once all solvent has been removed, the intact, charged proteins pass through a sampling cone, into an intermediate vacuum region and onto the mass analyser under high vacuum. ESI MS results in proteins gaining multiple positive charges on ionisation. The majority of the nitrogen ions in the protein backbone can be protonated as can the amine groups in lysine and arginine. As each protein molecule differs in the number of added protons, the function of the mass analyser is to resolve the mass/ charge (m/Z) and measure the intensities for each set of differently charged protein molecules.

The data is then plotted in the form of intensity *vs.* m/Z. Molecular mass (M) can then be determined using the equation:

$$m/Z = \frac{(M + nH^{\pm})}{n} \quad (19)$$

where m/Z = mass/ charge
n = the integer number of charges on the ions
H = the mass of a proton (1.008 Da)

As n is not usually known, simultaneous equations can be used to determine the number of charges by assuming the peaks (m/Z of differently charged protein molecules) are separated by a single charge. *n* can thus be defined by the equation:

$$n = \frac{m/Z(n+1) - H^{\pm}}{m/Z(n) - m/Z(n+1)} \quad (20)$$

A number average can thus be determined for *n* by solving for all peaks and the molecular mass of the protein then calculated by substitution of *n* back into equation 19.

(Wilm, 2000, www.astbury.leeds.ac.uk/Facil/MStut/mstutorial.htm).

3.5. N-terminal sequencing.

Edman sequencing (or degradation) is used for the sequencing of peptides and proteins from their N-terminal end. It requires the modification of the N-terminal amino acid by phenylisothiocyanate (PITC) which is subsequently removed by weak acid hydrolysis. The released 2-anilino-5-thiazolinone amino acid derivative is then converted to a phenylthiohydantoin amino acid and HPLC analysis used for amino acid identification. The single amino acid shortened polypeptide now contains a free N-terminus which can then be modified by PITC and the cyclical sequencing reaction can then continue (Coligan *et al.*, 2003).

Chapter 4

Results -

The Antitermination Complex and NusB and NusE

4.1. The antitermination complex in *M. tuberculosis*.

4.1.1. Purification of hexa-his-N-terminally tagged NusA, NusB, NusE and αRNAP.

All proteins were purified to homogeneity as described in the materials and methods (See appendix 2 for general characteristics of cloned and purified proteins).

4.1.2. His-affinity pulldowns.

Studies in both the λ phage and *E. coli* have successfully revealed some of the components of antitermination complexes. Mason and Greenblatt (1991) reconstituted transcription reactions using purified RNA polymerase holoenzyme, the λ N protein, NusA, NusB, NusE, NusG and the *nut* containing RNA and all were subsequently isolated as a single, stable complex using High Performance Liquid Chromatography (HPLC) gel filtration. A model for the order of assembly was also proposed where S10 may initially be interacting weakly with the RNA polymerase. As σ^{70} is released early on in chain elongation NusA binds tightly to the polymerase and NusG may also interact weakly with the polymerase. After transcription of the *nut* site, N binds NusA, NusA, NusE and NusG bind the polymerase tightly and NusE and NusB interact. This results in a stable, antitermination complex on the surface of the RNA polymerase (Mason and Greenblatt, 1991). However, an *E. coli rrn* antitermination complex consisting of RNAP, Rho, NusA, NusB, NusE and NusG was not able to transcribe through a Rho-dependent terminator indicating the complex is not fully functional. Readthrough could only be achieved when the above complex was allowed to form in the presence of *E. coli* cell free extracts. This indicated the need for additional unidentified factors for

accurate and efficient antitermination (Squires *et al.*, 1993). Four such factors may have been identified, the S4, L3, L4 and L13 ribosomal proteins (Torres *et al.*, 2001). The addition of these proteins to *in vitro* antitermination complexes, described above, showed high rates of read-through with the majority of the effect as a result of S4. The S4 protein does in fact show properties very similar to that of NusA and binds tightly to the RNA polymerase (Torres *et al.*, 2001). This is thus a strong indication that ribosomal protein S4 is an additional antitermination factor.

With these results in mind, experiments were set up to isolate and identify components of the antitermination complex from *M. tuberculosis*. It has not yet been established whether antitermination is a regulatory mechanism used in *M. tuberculosis*, and if so, whether the same factors are involved as in λ and *E. coli* antitermination and whether they are capable of forming similar complexes to those found in λ and *E. coli*. Initial aims of the project were thus the characterisation of the components of the antitermination complex in *M. tuberculosis* using affinity pulldown experiments.

Affinity pulldown assays are techniques used to identify protein – protein interactions by pulling proteins or complexes of proteins from a cell free extract (CFE) via a tagged protein tightly bound to an affinity resin. This is commonly done using a nickel or cobalt based resin in the case of His-tagged proteins or glutathione sepharose in the case of GST-tagged proteins. The tagged protein (defined as the "bait" protein) is attached to the affinity resin and CFE from the relevant organism then passed through the "bait" protein primed affinity column. Proteins in the CFE interacting with the "bait" protein remain on the column until elution of the "bait" protein. The presence of interacting

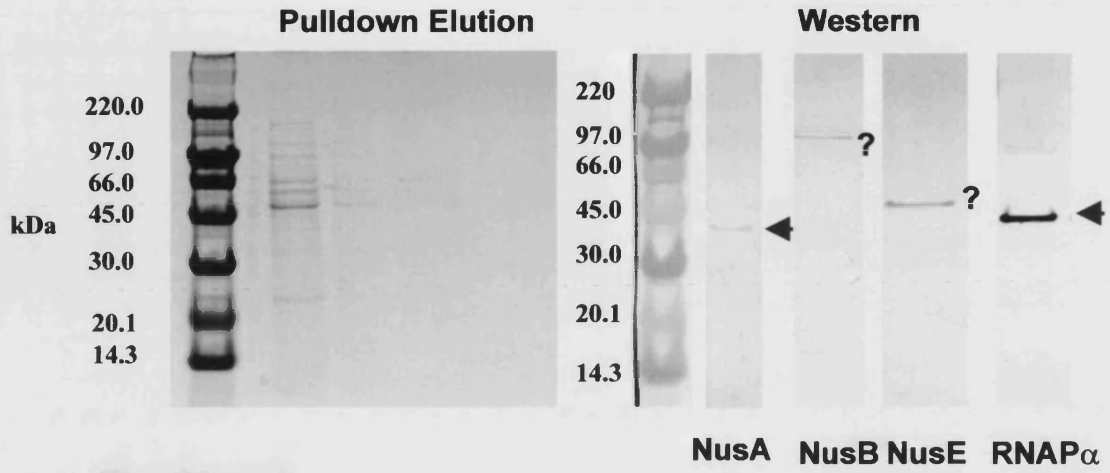
proteins can then be detected by a visualization technique after electrophoretic separation on an SDS-polyacrylamide gel.

Affinity pulldown experiments were carried out using N-terminally His_6-tagged NusA, NusB, NusE, and the α-subunit of RNA polymerase (αRNAP). These five proteins were individually attached to a His-affinity resin, and cell free extract (CFE) then applied to the Nus or αRNAP primed columns. Those factors present in the CFE capable of binding any of the Nus factors or αRNAP should have subsequently been retained on the column and then eluted. Samples are applied to denaturing SDS-acrylamide gels, allowing for separation of isolated protein complexes, and the "pulled out" proteins identified using MALDI (Matrix Assisted Laser Desorption/ Ionisation) MS or western blotting. Negative control experiments are run in parallel where no tagged protein is added to the column prior to the loading of CFE. This is thus indicative of those proteins with affinities for the resin only.

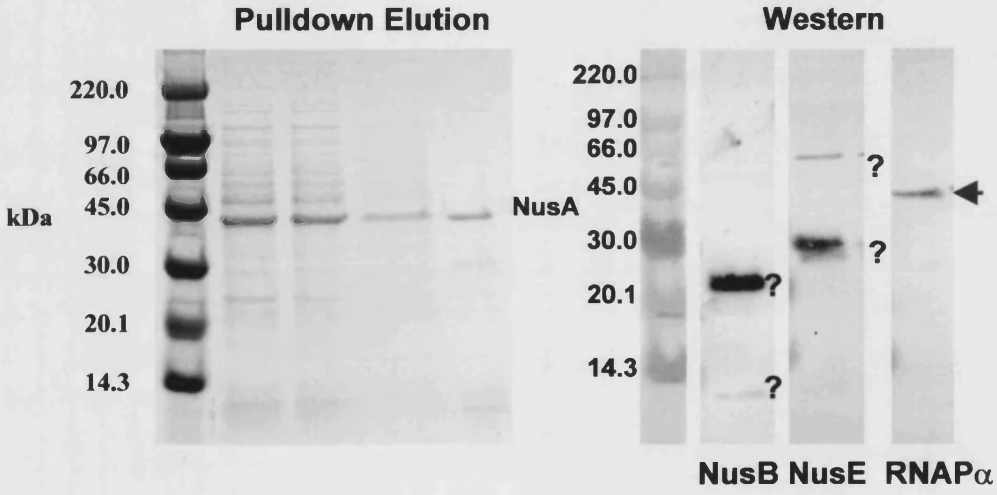
The results of these experiments are shown in Figure 11. Unfortunately both the experimental and negative control gels immediately indicate the non-specificity of His-affinity resins. A considerable number of proteins within the *M. tuberculosis* CFE have affinities for the resin and thus are retained on the column regardless of the "bait" protein. They are then eluted along with the His-tagged Nus or αRNAP protein along with proteins that may have made biologically relevant contacts with the Nus factors or RNAPα. Initially it was hoped that the technique would provide sufficient and specific resolution of "pulled out" complexes and that identifications could be made using MALDI MS. However due to the high number of bands present in elution samples, this was not possible (See Figure 11).

Antibodies to NusA, NusB, NusE, and αRNAP, raised in mice (Mr B. Butler, NIMR, London), were then used to probe Nus or αRNAP primed column eluants for any of the Nus proteins or αRNAP. This would at least provide verification of interactions between the predicted components of the *M. tuberculosis* antitermination complex. This however could also not be achieved due to the low specificity or cross-reactivity of the antibodies used (See Figure 11). Some bands were seen in the negative control indicating the binding of antibodies to proteins with affinities for the resin. Both NusA and RNAPα seem to bind to the resin as they are also present in all other pulldowns. It is also likely that the β′ RNAP subunit would bind the resin due to the presence of a Zn^{2+} binding site. The NusB and NusE antibodies also appear to cross react with proteins which have affinities for the resin. The NusE antibody cross-reacts with a ≈ 45 kDa protein in all pulldowns (including the negative control) and a ≈ 28 kDa protein in the NusA pulldown. Interestingly the NusE pulldown pulls out a protein corresponding to the NusB size (indicted by NusB? In Figure 11). However the band is not recognised by the NusB antibody in the western blot, perhaps indicating a lack of specificity of the antibody. The NusB antibody binds a high molecular weight protein in the negative control and the NusE pulldown but seems to show non-specificity in the NusA and RNAPα pulldowns.

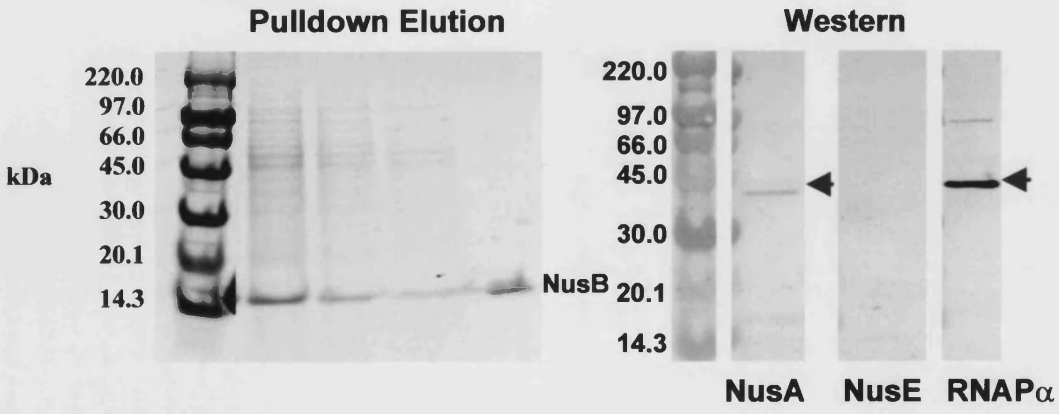
A : Negative Control



B : NusA



C : NusB



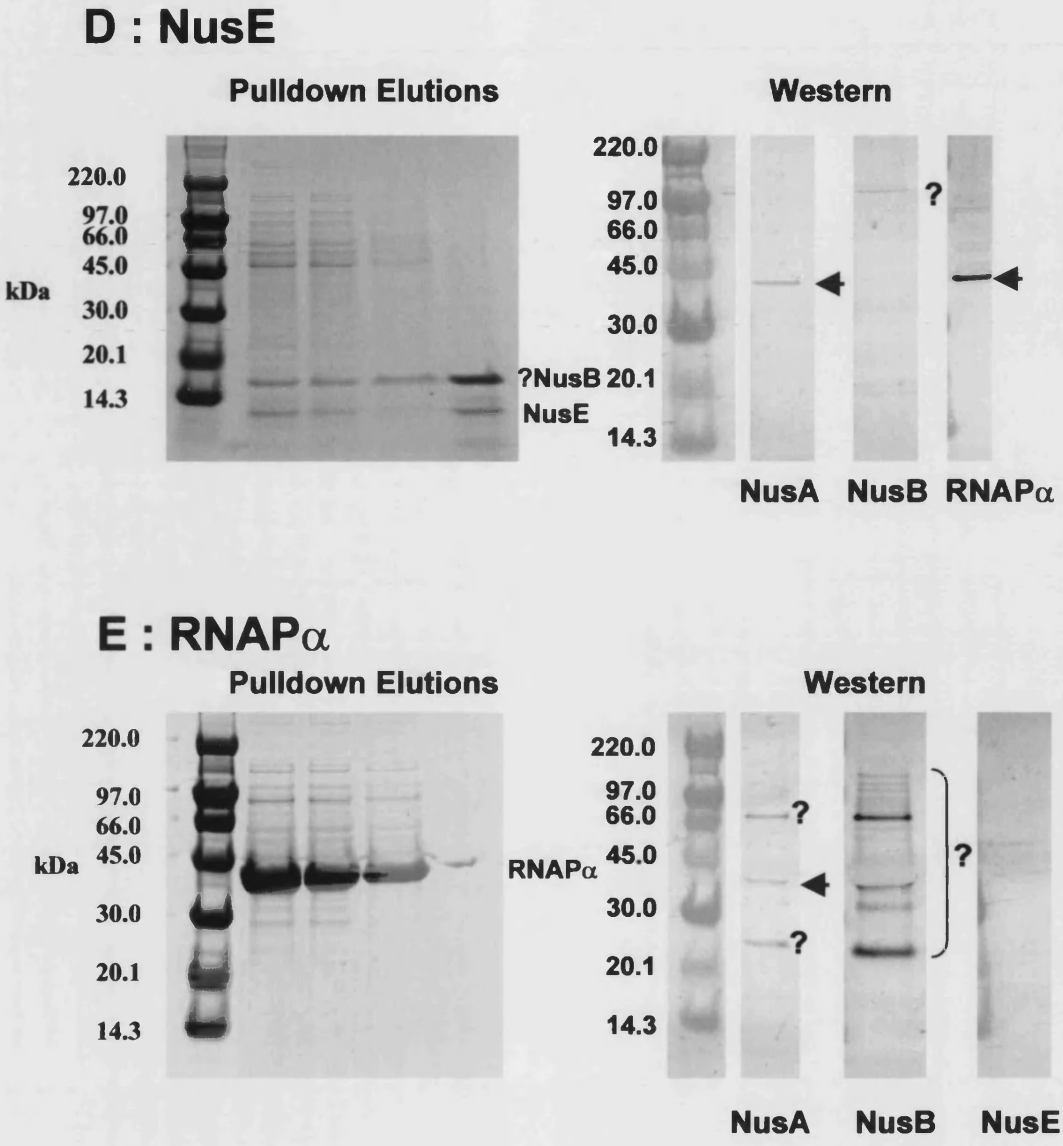


Figure 11 : His-affinity Pulldowns
N-terminally His tagged affinity pulldowns using NusA, NusB, NusE and RNAPα and Western Blots using mouse derived antibodies against NusA, NusB, NusE and RNAPα. **(A)** Negative Control Pulldown. **(B)** NusA Pulldown. **(C)** NusB Pulldown. **(D)** NusE Pulldown. **(E)** RNAPα Pulldown. Both NusA and RNAPα appear to stick to the resin as bands correspond to their sizes can be seen in all elution lanes. They are indicated by a ←. Bands found at unusual sizes are indicated by a **?**. The band corresponding to the NusB size in the NusE pulldown is indicated by **?NusB**.

4.1.3. Cloning, expression and purification of GST N-terminally tagged NusB.

As His affinity pulldowns proved unsuccessful, it was decided to N-terminally tag NusB with glutathione-S-transferase (GST) and carry out similar pulldown experiments using a glutathione-sepharose resin. Glutathione-sepharose is more frequently employed in pulldown experiments.

NusB was cloned into a GST containing vector, pGEX-4T-1 resulting in an N-terminally tagged NusB. The resulting construct was then transformed into *E. coli* BL21(DE3)*pLysS* cells. GST-NusB over expressed very well and proved to be soluble. This is in contrast to the N-terminally His-tagged form of NusB which was highly insoluble. The over expressed protein was then easily purified using a 2 step procedure. A glutathione-sepharose affinity column was used to select for the GST-tagged NusB. This gave a high level of purity (see Figure 12) however size exclusion chromatography was required to remove a few remaining contaminants. After purification concentrations in the region of 7 mg/ml (estimated using an extinction coefficient of 53580 M^{-1} cm^{-1} at an absorbance of 280 nm) were obtained and a yield of 30 mg/1.5 L of culture. This was suitable for the affinity pulldowns to be carried out.

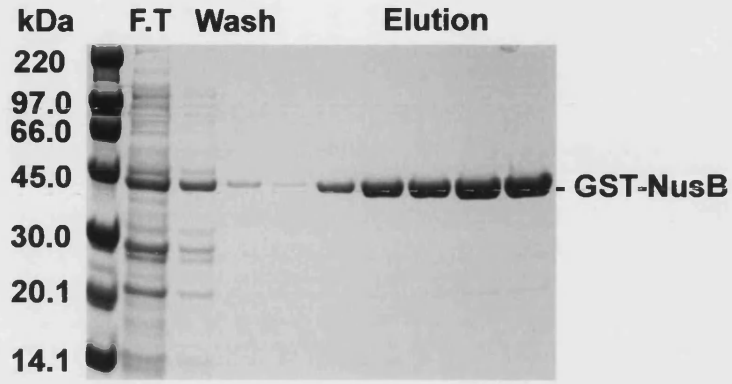
4.1.4. Glutathione sepharose affinity pulldowns.

Purified GST-NusB was used in glutathione sepharose affinity pulldowns. The technique is identical to that of His-affinity pulldowns except GST-tagged protein is immobilized on a column of glutathione-sepharose. *M. tuberculosis* CFE is then passed through the column and the GST-tagged protein along with bound proteins are eluted

with glutathione. Thus any proteins in the *M. tuberculosis* CFE making contacts with NusB should be eluted with the tagged NusB.

In this case no bands can be seen in the elution lanes of either the experimental or negative control pulldown experiments (See Figure 13). This may be due to the fact that small volumes of *M. tuberculosis* culture are used for producing CFE (between 100 and 400 ml). The resulting CFE would thus contain low concentrations of proteins and would not be seen after Coomassie staining. It was thus not possible to identify proteins pulled out by the NusB using MALDI MS as previously hoped. Elution samples from the GST-NusB and negative control pulldown experiments were analysed by western blotting with antibodies against NusA, NusE and αRNAP. This was again carried out in an attempt to verify interactions known to occur in the λ and *E. coli* systems. No bands were seen with neither the NusA antibodies nor the αRNAP antibodies indicating there are no significant interactions between NusB and these two proteins. However, NusE antibodies showed a definite band at the size corresponding to NusE thus indicating that NusE had been pulled from the CFE by NusB. No other bands could be seen indicating the specificity of NusB for NusE. The negative control (where no GST-NusB was loaded onto the glutathione-sepharose column) showed no band confirming the interaction of NusB with NusE. No cross-reactivity of the NusE antibody was seen (as with the His-affinity pulldowns). This is the first time that this has been shown using *M. tuberculosis* NusB and NusE and corresponds to the results seen in λ and *E. coli* (Mason *et al.*, 1992a; Luttgen *et al.*, 2002).

(A) Glutathione-sepharose purification



(B) Size exclusion chromatography

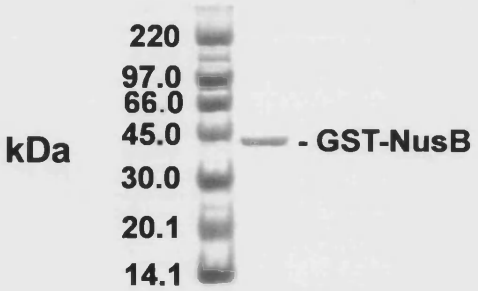


Figure 12 : Purification of GST-NusB
(A) Glutathione-sepharose affinity chromatography was initially used to purify GST-NusB.
(B) Size exclusion chromatography (Superdex 200) was then used to achieve homogeniety of the protein. Concentrations of 7 mg/ml (170 μM) were obtained.

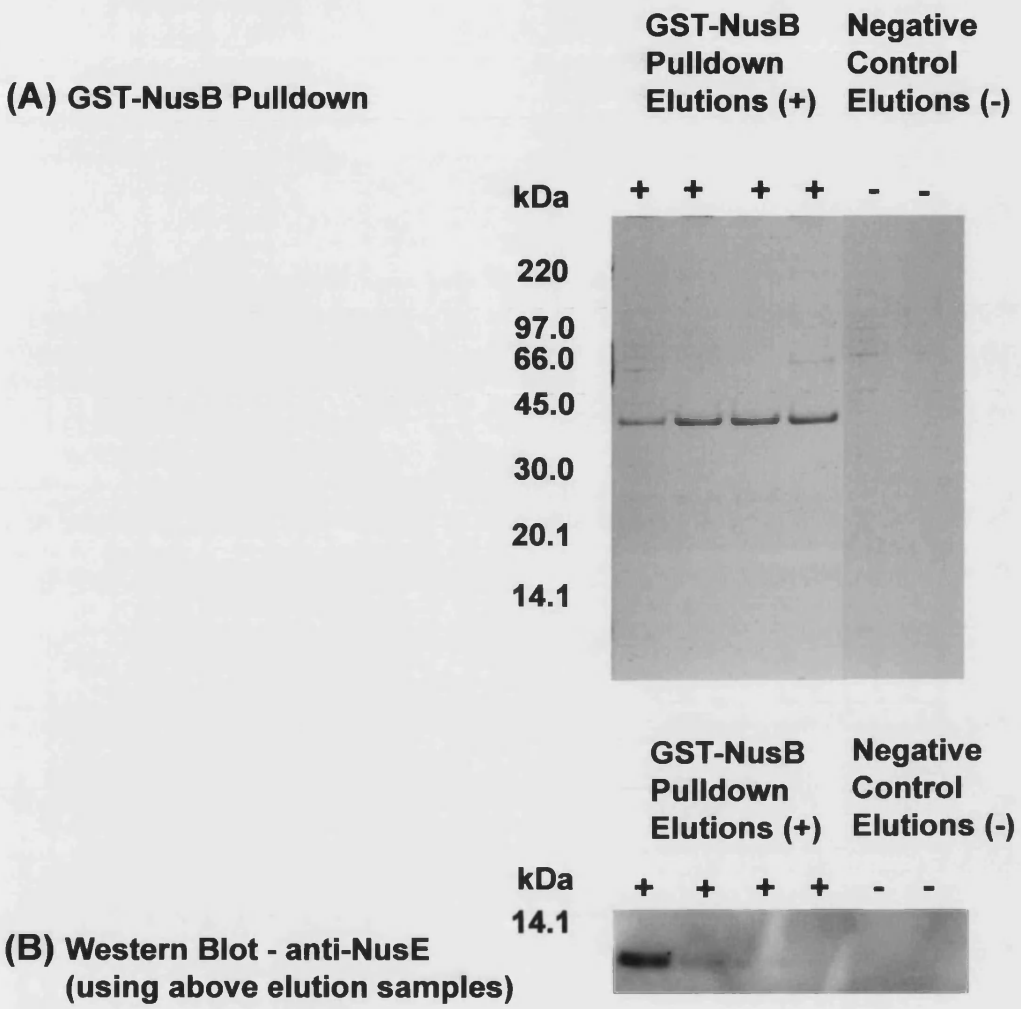


Figure 13 : GST-NusB Pulldowns and anti-NusE Western Blots.
(A) GST-NusB Pulldowns showing elution samples (+). The first two elution lanes of the negative control pulldowns (-) are also shown During the negative control pulldowns, no GST-NusB was attached to the glutathione-sepharose resin but CFE is still passed through the cloumn. This allows for the determination of non-specific protein binding to the resin.
(B) Anti-NusE Western blots on GST-NusB and negative control elution samples. A band can clearly be seen in the GST-NusB elution lanes. NusE is thus pulled from the CFE by NusB indicating an interaction between these two proteins.

4.2. The NusB dimer.

The role of the *M. tuberculosis* NusB dimer in antitermination has proved elusive and this has been further complicated by the fact that in the well characterised *E. coli* system, the protein is monomeric (Altieri *et al.*, 2000; Gopal *et al.*, 2000). In *E. coli* it is believed that NusB functions as a monomer (Mason and Greenblatt, 1991) and it was thus necessary to establish the importance of the *M. tuberculosis* dimeric NusB.

Gopal *et al.* crystallised and solved the structure of the *M. tuberculosis* NusB (shown in Figure 8) and revealed the homodimeric nature of the *M. tuberculosis* protein in contrast to the monomeric *E. coli* form. Despite the dimeric properties of the *M. tuberculosis* NusB, there are still structural similarities to the *E. coli* NusB structure with six of the seven α-helices in each molecule superimposing (Altieri *et al.*, 2000; Gopal *et al.*, 2000).

Interactions between the two NusB molecules in the dimer are both polar and non-polar with eleven water molecules buried within the dimer interface. The non-polar interface is made up of residues Val 18, Ala 19, Phe 22, Val 26, Ala 34, Val 88, Val 97, and Leu 122. The charged residues Lys 15, Arg 16 and Glu 23 all make contacts with the non-polar residues Val 26 and Ala 19 (Gopal *et al.*, 2000). Lastly Phe 22 forms an aromatic stack at the core of the dimer interface. The phenylalanine at position 22 is interesting as it corresponds to the tyrosine at position 18 in *E. coli*, which when mutated to aspartic acid results in the cold sensitive phenotype (the NusB5 strain – Court *et al.*, 1995; Friedman *et al.*, 1976)

Analysis of the dimer interface indicated that mutations of Phe 22 and Glu 23 to Ala were likely to have a large impact on the interface and the most likely to result in monomer formation (See Figure 14). The mutation of F22A would result in a large cavity within the hydrophobic interface while the E23A change would disrupt some of the hydrogen bond and network polar interactions at the dimer interface.

4.2.1. Monomerising the NusB dimer.

Site-directed-mutagenesis was used to generate F22A and E23A double mutation. The success of the mutation was verified using DNA sequencing and the resulting FE22.23AA mutant was over expressed and purified. As with the wild-type, the protein was insoluble and purified by denaturation and subsequent refolding. The mutation was reconfirmed using ESI MS (See Figure 15). A second species (with molecular weight 18820 ±1.55; 178 da greater than the expected molecular weight) is present in the preparation (species B in the ESI MS spectra) and is likely to be a result of α-N-gluconylation of the N-terminus of NusB FE22.23AA as described by Geoghegan *et al.* (1999).

It was then necessary to determine whether the resulting mutations had in fact caused dissociation of the NusB dimer. The FE22.23AA mutant was analysed using sedimentation equilibrium ultracentrifugation, in order to accurately define its solution molecular weight. The solution molecular weight of FE22.23AA NusB obtained by sedimentation equilibrium ultracentrifugation was 17 ± 3 kDa. This corresponds to the expected monomeric molecular weight (See Figure 16). The wild-type NusB protein gave a molecular weight of 36 ± 4 kDa under the same conditions. Thus, the double

mutation in NusB disrupts the dimer interface sufficiently to prevent any significant dimer formation in solution. Further analysis of the data also showed there was no apparent concentration dependence (at 80 μM, 53.5 μM and 26 μM) of either the FE22.23AA or the wild-type proteins on the molecular weights determined by sedimentation equilibrium ultracentrifugation indicating a strong wildtype dimer and very weak dimerisation affinities of the mutant monomer.

Before the monomeric NusB mutant could be used in any further studies, it was also necessary to ensure that the secondary structure of the mutant was still essentially identical to that of the wild-type. Changes in structure may affect functionality and stability of the mutant thus rendering functional and interactive studies difficult to interpret and compromising the comparison with the wildtype.

Far-UV circular dichroism (the 260 to 190 nm range) is an ideal technique for establishing secondary structure of a protein and can be used to estimate the number of helical residues within a protein. CD spectra were recorded for both the wild-type and FE22.23AA mutant NusB proteins at a concentration of 9 μM and at temperatures of 20 °C. The results are shown in Figure 17. The results indicate the secondary structure content of wild-type and mutant NusB is almost identical as the far-UV CD spectra almost overlap. This means the two mutations introduced into NusB are unlikely to have had a large effect on the secondary structure of the protein. The monomeric NusB is therefore likely to fold similarly to the wild-type protein, maintain functionality and would still be able to make interactions with *in vivo* partner proteins if dimerisation was not required for its biological activity.

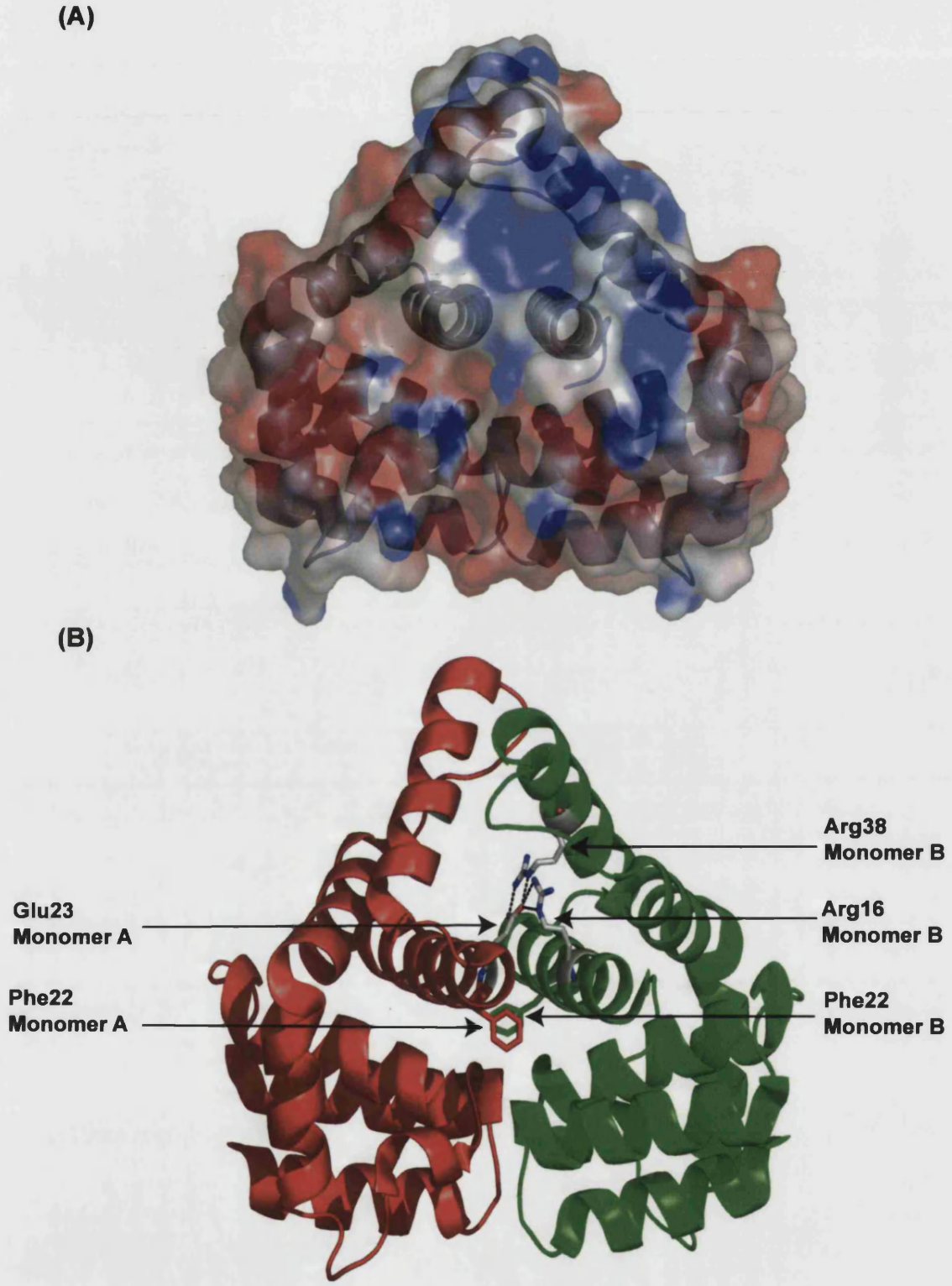


Figure 14 : Electrostatic representation of the NusB dimer surface and select interactions between the two monomers.
(A) shows the electrostatic potential on the dimer surface with blue shading indicating positive charge, red shading indicative of negative charge and white indicating neutral charge. **(B)** shows hydrogen bond interactions between the Glu23 on monomer A (red) and the Arg16 and Arg38 of monomer B (Green). The phenylalanine stack at the centre of the interface is also shown (Gopal *et al.*, 2000).

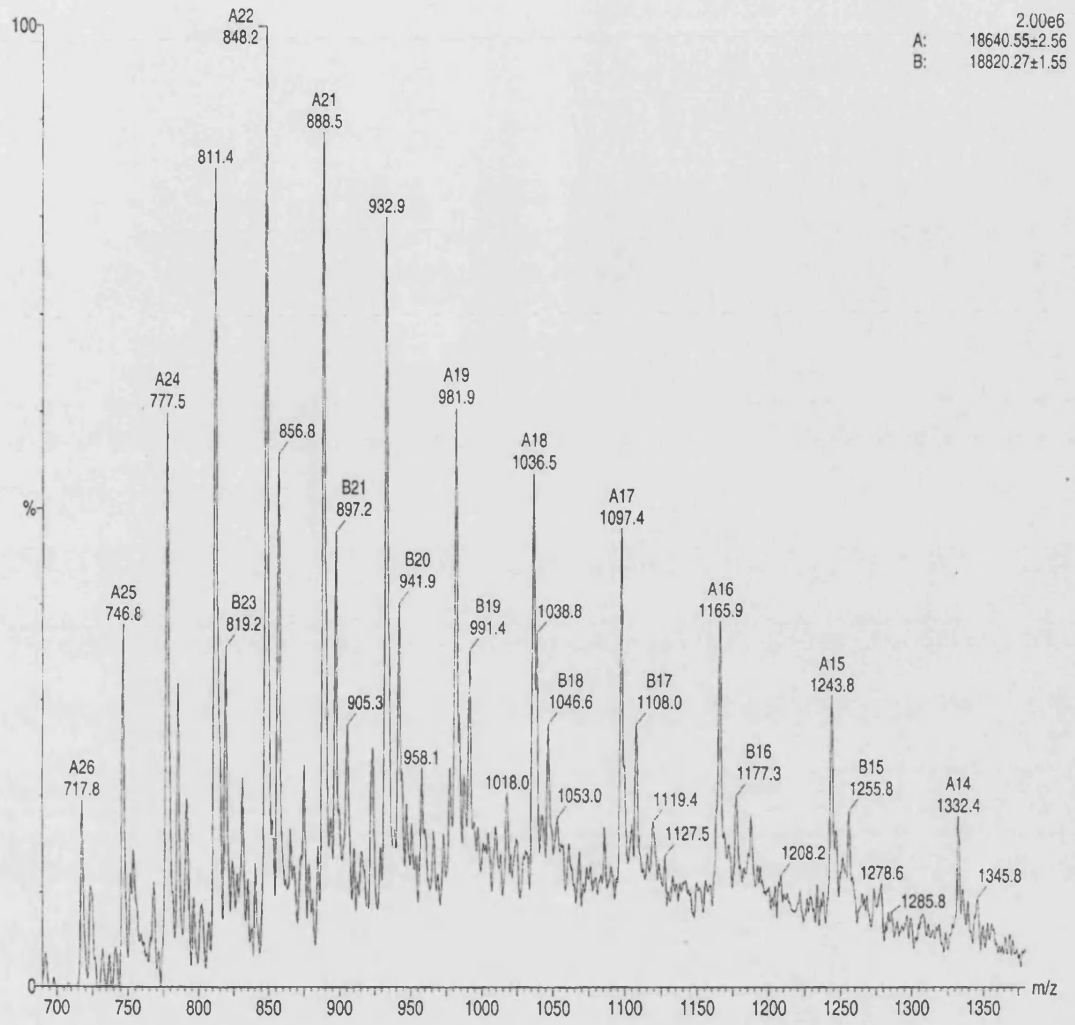
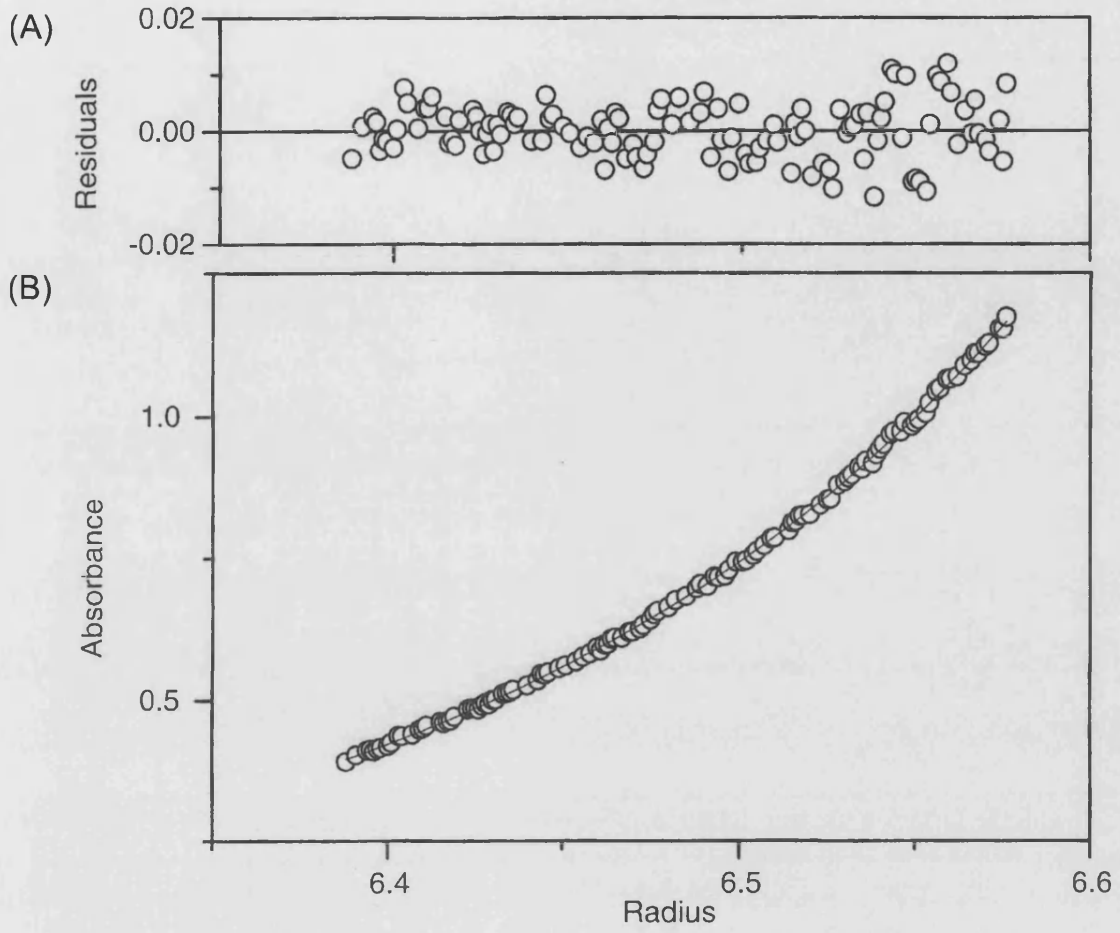


Figure 15 : Confirmation of FE22.23AA Molecular Weight using ESI MS. The molecular weight, and hence the presence of the amino acid mutations, was determined using ESI MS. The molecular weight was shown to be 18640 kDa (Species A) indicating the N-terminal methionine is missing. The calculated molecular weight of the mutant is 18638 kDa without the N-terminal methione. ESI MS thus confirmed the presence of the mutations. Species B (18820 kDa) is the α-N-gluconylated form.



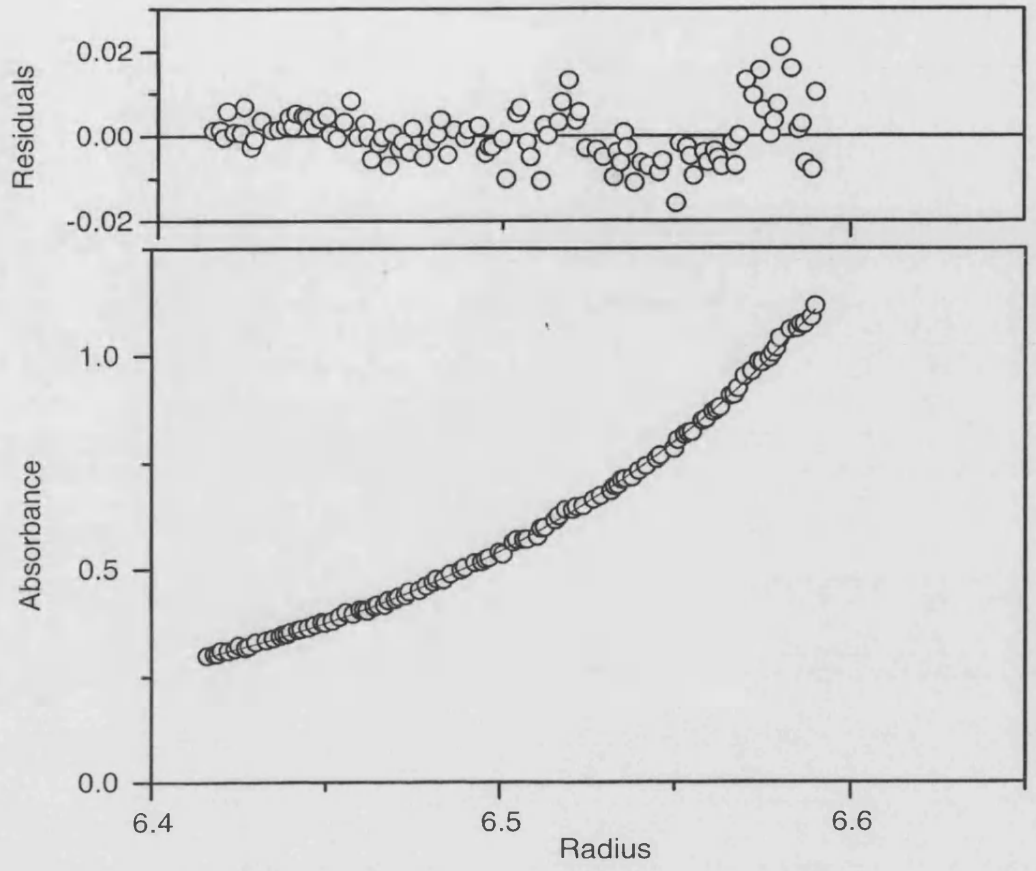
DOF = 891 Variance = 2.97215E-5
Fitted Parameters:

Co	Offset
0.600	0.039
0.371	0.038
0.187	-0.004
0.598	0.050
0.355	0.040
0.184	-0.005
0.578	0.050
0.374	0.055
0.177	-0.004

M = 17092 B = 0

Speed = 22000
Time = 167680
Temp = 20
V-bar = 0.738
Rho = 1.011

Figure 16a : Sedimentation equilibrium ultracentrifugation of FE22.23AA NusB. Equilibrium sedimentation ultracentrifugation was carried out to determine the effect of mutation at the dimer interface in NusB. The data shows a molecular weight of 17 ± 3 kDa for the mutant indicating a monomeric species. NusB wildtype showed a molecular weight of 36 ± 4 kDa using equilibrium sedimentation ultracentrifugation therefore indicating a dimeric species (seen in Figure 15b). **(A)** shows the random distribution of of residuals around a non-linear fit. **(B)** shows the radial distribution of the sample once equilibrium has been reached. The plot is a global fit of nine parameters (three concentrations recored at three time points).



DOF = 865 Variance = 5.49596E-5

Fitted Parameters:

Co	Offset
0.406	0.002
0.243	0.018
0.104	0.036
0.409	0.024
0.223	0.056
0.096	0.045
0.439	0.046
0.228	0.070
0.101	0.051

M = 36149 B = 0

Speed = 18000
Time = 101890
Temp = 20
V-bar = 0.738
Rho = 1.011

Figure 16b : Sedimentation equilibrium ultracentrifugation of NusB wildtype protein.

The molecular weight of NusB wildtype was determined using equilibrium sedimentation ultracentrifugation. The molecular weight was 36 ± 4 kDa indicating a dimeric species. The plot is a global fit of nine parameters (three concentrations recorded at three time points).

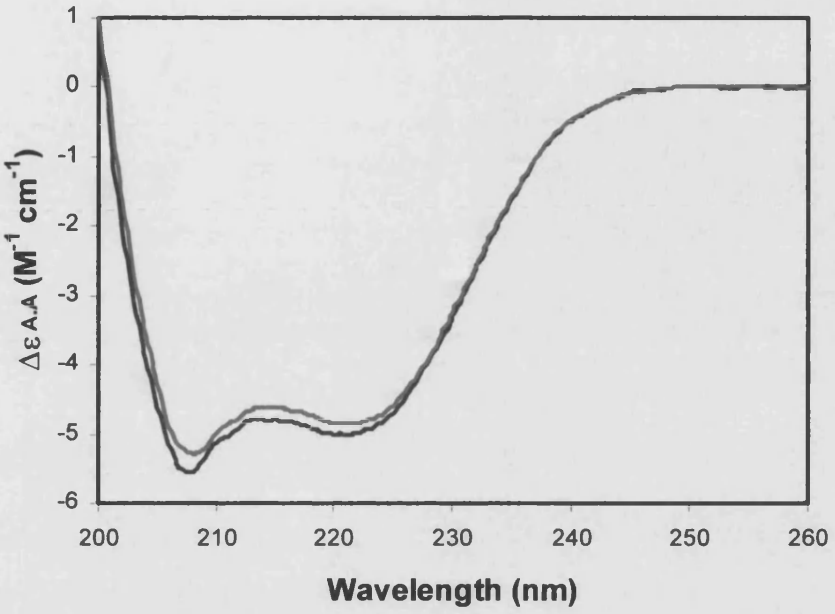


Figure 17 : Circular dichroism of NusB wildtype (-) and the FE22.23AA mutant (-).

The CD spectra of the wildtype and mutant proteins are very similar indicating the mutations introduced into NusB have had little effect on the secondary structure of the protein.

4.2.2. Interactions of wild-type and FE22.23AA mutant NusB with NusE.

The occurrence of an interaction between NusB and NusE in *M. tuberculosis* has only been hypothesised based on studies in *E. coli*. The *M. tuberculosis* NusB dimer indicates a different model of NusB – NusE interaction compared to that of the *E. coli* model. This may in turn have an effect on the whole antitermination mechanism in *M. tuberculosis*. The complex may be trimeric consisting of the NusB homodimer and a single NusE molecule, it may be heterodimeric (a single NusB and NusE molecule) or the interaction may not even occur at all.

The last possibility seems unlikely as the GST-NusB pulldown experiments have demonstrated an interaction between NusB and NusE (see section 4.1.4). Experiments were therefore set up to firstly confirm the NusB and NusE interactions *in vitro* and secondly to determine the stoichiometry of the interaction.

Studies were thus carried out in order to characterise the interaction between the wild-type and mutant NusB proteins and NusE. Luttgen *et al.* (2002) showed that *E. coli* NusE was only soluble when expressed in the presence of *E. coli* NusB. Attempts to express and purify NusE in the absence of NusB resulted in insoluble protein and it proved impossible to refold the protein. Similarly both *M. tuberculosis* NusB and NusE were insoluble when expressed separately. It was possible to refold NusB and obtain crystals, however NusE showed very little stability or structure when refolded (Gopal *et al.*, 2000; Gopal *et al.*, 2001b).

In one set of experiments denatured *M. tuberculosis* untagged-NusB and His_6-NusE were refolded together and then applied to a His-affinity column. The idea being that His_6-NusE would be immobilized on the column and theoretically interact with a stochiometric amount of refolded NusB. The His_6-NusE- NusB complex could then be eluted from the column and analysed by SDS-PAGE. This would allow the estimation of the NusB-NusE stoichiometry.

The results of these experiments can be seen in Figure 18. Encouragingly the results immediately show an interaction between NusB and NusE as well as FE22.23AA and NusE. This can be seen by the increased amounts of NusB and FE22.23AA present in the elution fractions after the elution of NusE when compared with the wash fractions.

This means that during refolding, NusB or FE22.23AA is binding NusE and remains attached to NusE when the refolded mixture is passed through a His-affinity column. On elution, both NusB or FE22.23AA and NusE are present in the fractions indicating binding. This therefore reconfirms, *in vitro*, those results obtained using GST-NusB pulldowns and anti-NusE western blots.

Unfortunately it was not possible to elucidate the stoichiometry of NusB/FE22.23AA – NusE interaction. From the gels it can be seen that there are higher concentrations of NusE than NusB or FE22.23AA. This is unexpected as the stoichiometry was expected to be either 1 : 1 (heterodimer of NusB or FE22.23AA with NusE) or 2 : 1 (heterotrimer of the NusB dimer with NusE). However assuming that FE22.23AA-NusE interactions are heterodimeric (a monomeric FE22.23AA would be expected to interact with a monomer of NusE assuming the *E. coli* model), the similarities between intensities of the bands seen in the NusB-NusE interactions and the FE22.23AA-NusE interactions may suggest a 1 : 1 stoichiometry. As FE22.23AA NusB is monomeric all unbound FE22.23AA would have flowed through the column and thus the bands seen on the gels would represent a single FE22.23AA NusB interacting with a single NusE. If wild-type NusB was interacting with NusE as a dimer (resulting in a heterotrimer) one would expect to see higher band intensities of NusB than those seen with FE22.23AA as two molecules of wild-type NusB would remain attached to NusE compared to the single molecule of FE22.23AA. The band intensities in the two experiments are however similar indicating that a single wild-type NusB molecule may bind NusE as does a single FE22.23AA mutant NusB thereby predicting a 1 : 1 stoichiometry of NusB : NusE.

The increased amounts of NusE (present in the elution lanes) may be due to the fact that NusE never becomes saturated with the loaded NusB wild-type or FE22.23AA and thus equimolar concentrations of NusE and NusB wild-type or FE22.23AA are never eluted. This may be indicative of a weak interaction between NusB wild-type and FE22.23AA with NusE. The NusE may be competing with the NusB wildtype dimer for binding at the dimer interface resulting in a weak interaction and subsequent differences in the NusE – NusB band intensities. It would then, however, be expected that NusE interactions with the FE22.23AA monomer would be stronger (as there is no competition) resulting in more intense FE22.23AA band intensities compared with those seen with NusB during NusB – NusE binding. This is however not the case; similar band intensities are seen with both FE22.23AA during NusE – FE22.23AA binding and NusB during NusE – NusB binding. This may be explained by a weak interaction between NusE and FE22.23AA as the FE22.23AA mutations in NusB (found within the dimer interface) may have disrupted the NusE binding site resulting in a weaker interaction. This again points towards a 1 : 1 stoichiometry for NusB – NusE binding in the *M. tuberculosis* antitermination complex.

These results need to be interpreted with some degree of caution as band intensities are dependent on such factors as staining efficiency. The results are thus only comparative and need to be proved quantitatively.

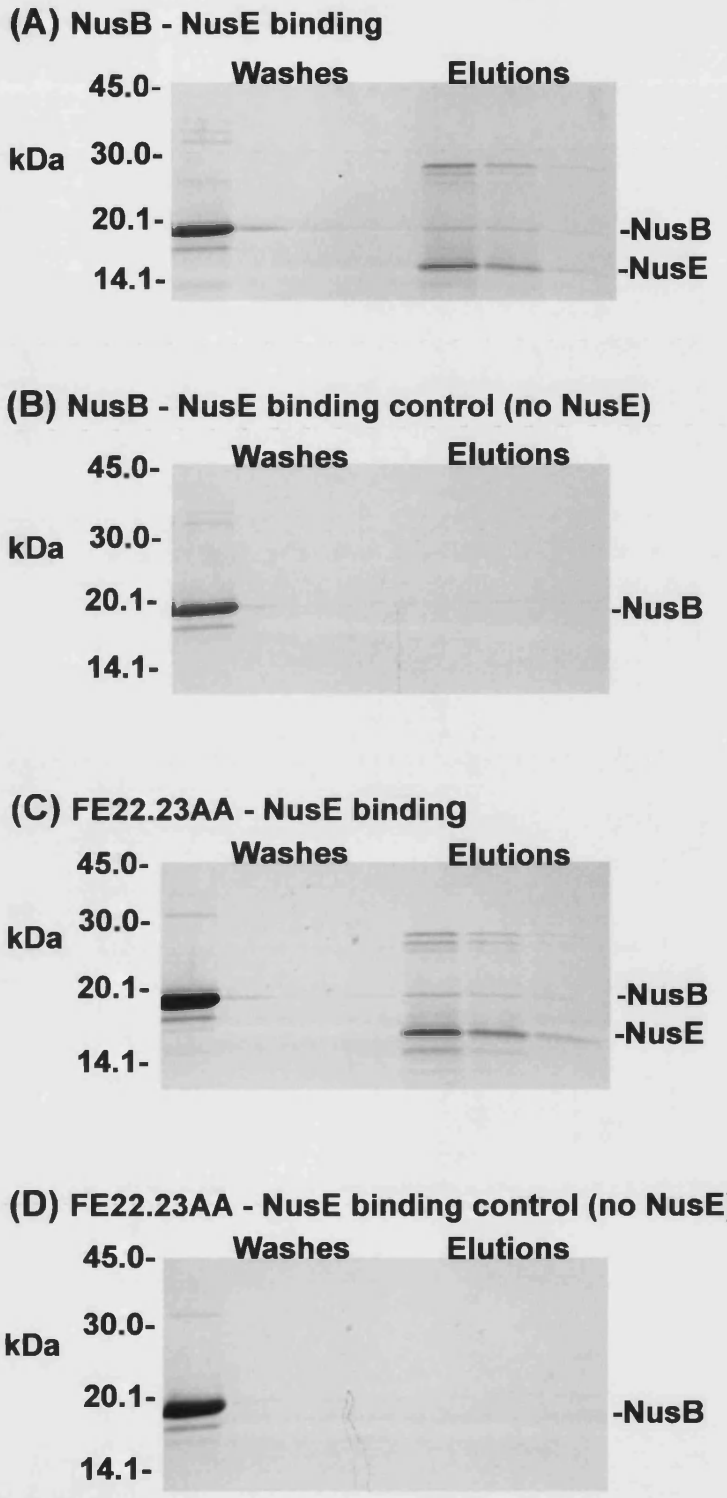


Figure 18 : Refolding and binding of NusB and FE22.23AA with NusE.

(A) Refolding and binding of NusB and NusE.

(B) Control refolding and binding of NusB and NusE. No NusE loaded on His-affinity column.

(C) Refolding and Binding of FE22.23AA with NusE.

(D) Control refolding and binding of FE22.23AA and NusE. No NusE loaded on His-affinity column.

4.2.3. Co-expression of *M. tuberculosis* NusB and NusE.

On the basis of the refolding/binding experiments, it was decided to attempt co-expression of NusB and NusE. NusB was thus cloned into pGEX-6P1 resulting in an N-terminal GST-tag. NusE was then cloned 3′ to NusB and a synthetic Shine-Dalgarno sequence inserted just upstream of NusE and a synthetic hexa-His tag inserted immediately downstream of NusE. This resulted in a GST N-terminally tagged NusB and a C-terminally His_6 tagged NusE which would be expressed simultaneously from the same *tac* promoter found upstream of the GST-tag. Positive clones were verified by DNA sequencing and restriction analysis and then transformed into *E. coli* and the proteins over expressed and purified.

GST-NusB was again found to be soluble and, importantly, so was NusE-His_6. This was a significant step forward in the study of NusE as it had never been purified in a soluble manner. This is similar to what was reported by Luttgen *et al.* (2002) with the *E. coli* NusE. It seems that soluble NusE can only be achieved when expressed in the presence of NusB.

NusB and NusE were then purified using tandem affinity columns in order to first select for NusE and then NusB. A His-affinity column was used to pull the NusE-His_6 from the lysed cell extract. As it has already been shown that *M. tuberculosis* NusB and NusE associate, the likelihood is that NusB should be pulled from the extract through its interaction with NusE. This can clearly be seen in Figure 19a. A GST-affinity column was then used to select for GST-NusB. The GST-tag is then removed from NusB using the "PreScission Protease" allowing for the elution of NusB. Again, as a result of the

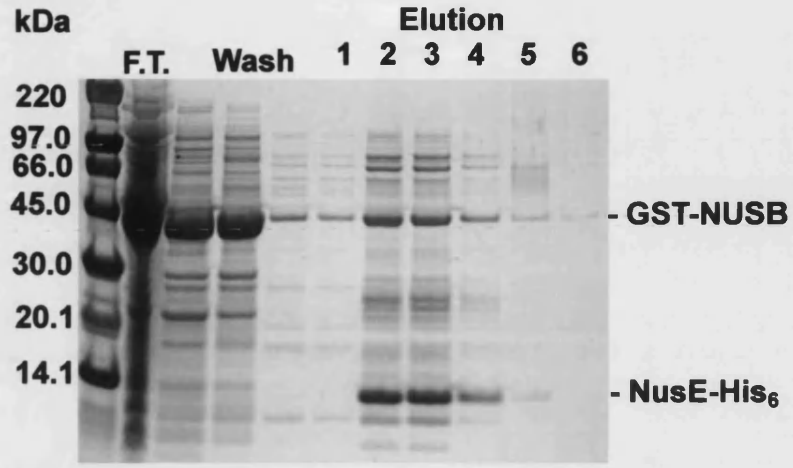
NusB-NusE interactions it would be expected that the two proteins would be eluted together through this method of purification. This was in fact the case, as can be seen in Figure 19a.

The co-elution of NusB and NusE again confirms the interaction between *M. tuberculosis* NusB and NusE as seen in the GST-affinity pulldown from *M. tuberculosis* CFE and the co-refolding and His-affinity pulldown experiments. Thus, there seems no doubt that *M. tuberculosis* NusB and NusE are able to associate. It has not however been possible to determine the NusB-NusE stoichiometry of this complex. Again the band intensities of NusB and NusE, when co-eluted from the glutathione-sepharose column, are similar (see Figure 19a) possibly indicating comparable concentrations of each protein suggesting a heterodimer (single NusB molecule interacting with a single NusE molecule). Attempts were made to try and confirm this by analysing the NusB-NusE complex (eluted from glutathione-sepharose) by size-exclusion chromatography. Unfortunately the NusB-NusE complex appeared to dissociate during size-exclusion chromatography (See Figure 19b). Only NusB is seen in the elution fractions as the dissociation of the complex may result in instability of NusE and subsequent precipitation of the protein. This may be explained by the fact that the interaction between NusB and NusE is relatively weak and during gel filtration the concentration of NusB and NusE falls to well below the K_d value for the complex (as a result of the approximately 10-fold dilution occurring when carrying out gel-filtration experiments). It was also hoped that the NusB-NusE complex could be analysed by analytical gel filtration or sedimentation equilibrium ultracentrifugation allowing for the calculation of

the dissociation constant. However due to the dissociation of the complex during gel filtration, this was not possible.

Co-expression of NusB and NusE therefore clearly shows an interaction between NusB and NusE indicated by the solubility of NusE in the presence of NusB and the co-elution of NusB and NusE through two affinity chromatography steps. The results hint at a 1 : 1 stoichiometry as band intensities are similar after the second round of affinity chromatography but the interaction between the two proteins appears to be weak as shown by the apparent dissociation of the complex during size exclusion gel filtration. Again this weak interaction may be explained by the competition of NusE for the binding of NusB at the dimer interface. Again, results based on band intensity must be treated with caution because of possible differences in staining efficiency.

(A) His-affinity purification of co-expressed NusB-NusE



(B) GST-affinity purification of co-expressed NusB-NusE

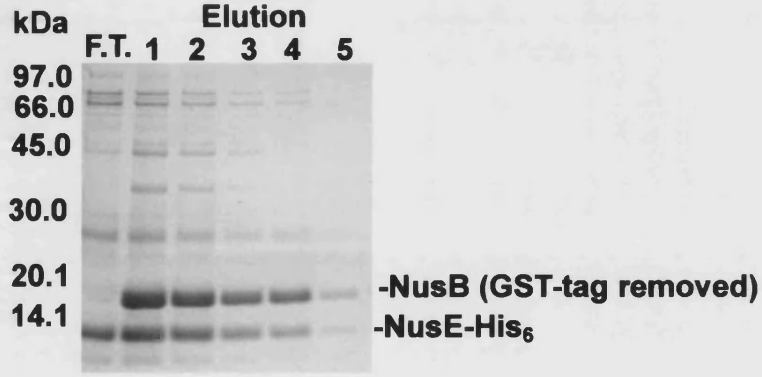


Figure 19a : NusB-NusE co-expression and purification.
(A) GST-NusB and NusE-His_6 were first purified using a His-affinity column and selecting for the NusE-His_6 protein. NusB co-elutes with NusE as seen in the increased band intensity of NusB on NusE Elution (elution lanes 2 - 5).
(B) GST-NusB and NusE-His_6 were then applied to Glutathione-Sepharose. The two proteins appear to be co-eluted at similar concentrations (shown by similar band intenstities in elution lanes 1 - 4) indicating the proteins may form a heterodimer.

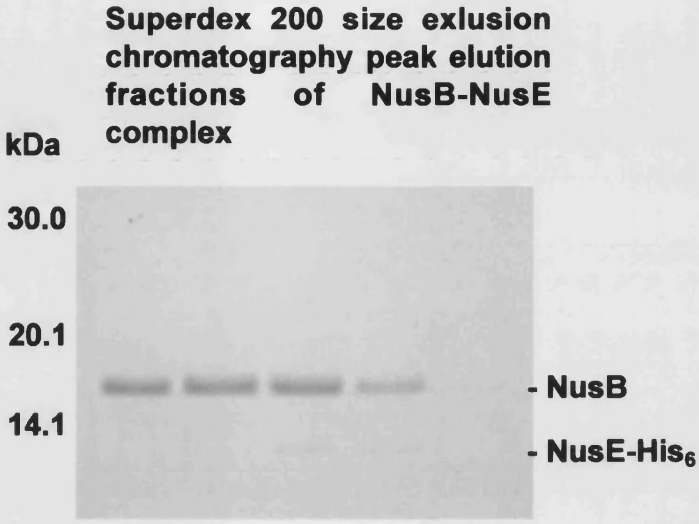


Figure 19b : Size exclusion chromatography of the NusB-NusE complex.
The gel shows selected peak elution fractions after size-exclusion chromatography. NusB can be clearly seen however only traces of NusE are seen meaning the complex is no longer co-eluting (in contrast to co-elution after GST-affinity chromatography -see Figure 19a).

4.2.4. The dissociation equilibrium constant of the NusB dimer.

It was not possible to determine the stoichiometry of the NusB – NusE complex during the refolding and co-expression experiments. Both experiments demonstrated weak binding between NusB and NusE. This is in contrast to the *E. coli* model, where the dissociation equilibrium constant is in the 10^{-7} M range (Mason *et al.*, 1992a). The low affinity of NusE for NusB may be due to the requirement of NusE to compete with NusB self-association. We therefore thought it useful to determine the dissociation equilibrium constant for NusB self-association.

High sensitivity HPLC analytical gel filtration was used to analyse dissociation of the NusB dimer. Analytical gel filtration is useful for complex composition determination and qualitative interactive studies. In this case a calibrated (for molecular weight) Superdex 200 (Amersham Pharmacia Biotech) column was used and column elution's monitored at wavelengths of 220 nm and 280 nm allowing for the observation of the large O.D. range required.

NusB wild-type and the FE22.23AA NusB mutant were used to establish retention times for the 100 % dimer and monomer respectively. The retention times were 38.7 min. for 27 μM loading concentration of the NusB wild-type and 41.4 min for 27 μM of the FE22.23AA mutant. These correspond to molecular weights of 27 203 kDa and 15 981 kDa respectively using a standard curve for the calibrated column. These are therefore the upper and lower limits for the retention times for the dissociating dimeric complex, shown in Figure 20. The differences between the calculated molecular weights using amino-acid composition and those determined using analytical gel filtration may be as a

result of the shapes of the proteins. Compact proteins move slower through gel filtration columns resulting in higher retention times. Statistical error during the fitting of the calibration curve may lead to discrepancies in the molecular weight values determined, from retention times, and formula molecular weights.

Concentrations ranging from 106 μM to 0.26 μM of wild-type NusB were then loaded on the column and the retention times measured for each concentration. There was no significant decrease in retention times between 26 μM and 106 μM; this concurs with sedimentation equilibrium data where no concentration dependency was seen between 26 μM and 80 μM loading concentration (section 4.2.1). All retention times were determined using wavelengths of 280 nm. As the concentrations of wild-type NusB decreased below 26 μM (loading concentration) the retention times increased (moved towards the NusB monomer retention time) indicating a dissociating complex. See Table 3. The retention times were then converted to molecular weights using a standard curve for the calibrated column. The calculated average molecular weights were then used to determine the fraction of monomer present in the NusB wild-type samples loaded onto the column using the following equation :

$$\text{Fraction Monomer} = \left[\frac{\text{Dimer}M_w - \text{Calc}M_w}{M_w\text{Dimer} - M_w\text{Monomer}} \right] \quad (21)$$

A plot of fraction monomer *vs.* total protein eluted (M – estimated as a 10-fold dilution of the loading concentration) then allows an estimation of the K_d of the NusB dimer (Figure 21). This is determined at the total concentration of monomer equal to 0.5 (i.e.

50 % of the total protein is in monomeric form). A fraction monomer value of 0.5 occurs at an X-value of approximately 8×10^{-8} M.

$$K_d = {[M]^2}/{[D]} \quad (22)$$

where $[M]$ = concentration monomer and $[D]$ = concentration dimer

$$\text{but } F_M = {[M]}/{[M_T]} \quad (23) \text{ and } [M_T] = [M] + 2[D] \quad (24)$$

where F_M is fraction monomer, $[M_T]$ is total protein eluted, $[M]$ is concentration monomers present and $[D]$ is concentration dimers present.

$$\text{From (23) } [M] = F_M [M_T] \quad (25)$$

$$\text{and from (24) } [D] = \frac{[M_T] - [M]}{2} \quad (26)$$

$$\text{Substituting (26) into (22) gives } K_d = \frac{2[M]^2}{[M_T] - [M]} \quad (27)$$

$$\text{And substituting (25) into (27) gives } K_d = \frac{2[F_M M_T]^2}{[M_T] - F_M [M_T]} \quad (28)$$

$$\text{Thus when the fraction monomer is 50\% } (F_M = 0.5) \; K_d = \frac{2 \times 0.5^2 \times M_T^{\,2}}{M_T - 0.5 M_T} \quad (29)$$

$$K_d = \frac{0.5 M_T^{\,2}}{0.5 M_T} \quad (30)$$

Thus from (30) when $F_M = 0.5$, $K_d = M_T$ or K_d is given by the total protein concentration eluted at which 50 % is monomeric. This is of course only an estimation of the K_d value for the NusB dimer. Limitations of the method are that it was not possible to achieve the full range of dimer to monomer dissociation (a full sigmoidal curve could not be plotted) and that the dissociation was not in equilibrium. The technique has however given a good indication as to the possible K_d value.

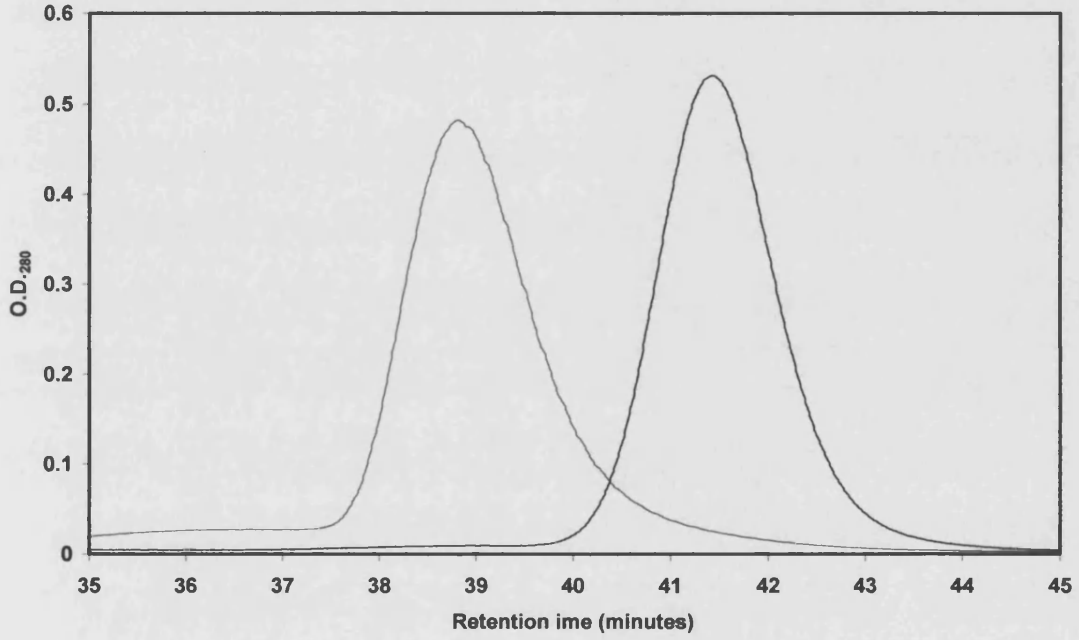


Figure 20 : Analytical gel filtration showing retention times of NusB dimer (-) and FE22.23AA NusB monomer (-) on a Superdex 200 Column.
The retention times of the NusB dimer and NusB mutant monomer were determined using gel filtration HPLC. There is a clear difference between the NusB wildtype and mutant at identical concentrations again indicating the mutations have led to dimer disruption. Retention times of 38.8 (dimer) and 41.4 (monomer) correspond to molecular weights of 27 203 kDa and 15 981 kDa respectively determined from the standard curve for the molecular weight calibrated column.

Table 3 : Analytical gel filtration of the NusB dimer – determination of the K_d.
Determination of the average molecular weight of decreasing concentrations of NusB dimer allowing for the determination of fractions of monomer present in the NusB wildtype loaded on the column. A plot of Fraction Monomer Loaded vs. Total Protein Concentration Eluted allows for the subsequent calculation of the dissociation equilibrium constant of the NusB wildtype dimer.

NusB Wildtype Concentration (μM)	Retention Time (Min)	Calculated Molecular Weight (kDa)	Fraction of Monomeric NusB Loaded	Total Protein Concentration Eluted (μM)
106 μM	38.7 min	27 765	-	-
53.0 μM	38.8 min	27 203	0.05	5.3 μM
27.0 μM	38.8 min	27 203	0.05	2.7 μM
13.3 μM	38.9 min	26 652	0.09	1.3 μM
5.30 μM	39.0 min	26 113	0.14	0.53 μM
2.70 μM	39.2 min	25 066	0.23	0.27 μM
0.53 μM	40.1 min	20 850	0.59	0.053 μM
0.26 μM	40.7 min	18 442	0.79	0.026 μM
27 μM FE22.23AA	41.4 min	15 981	1.00	-

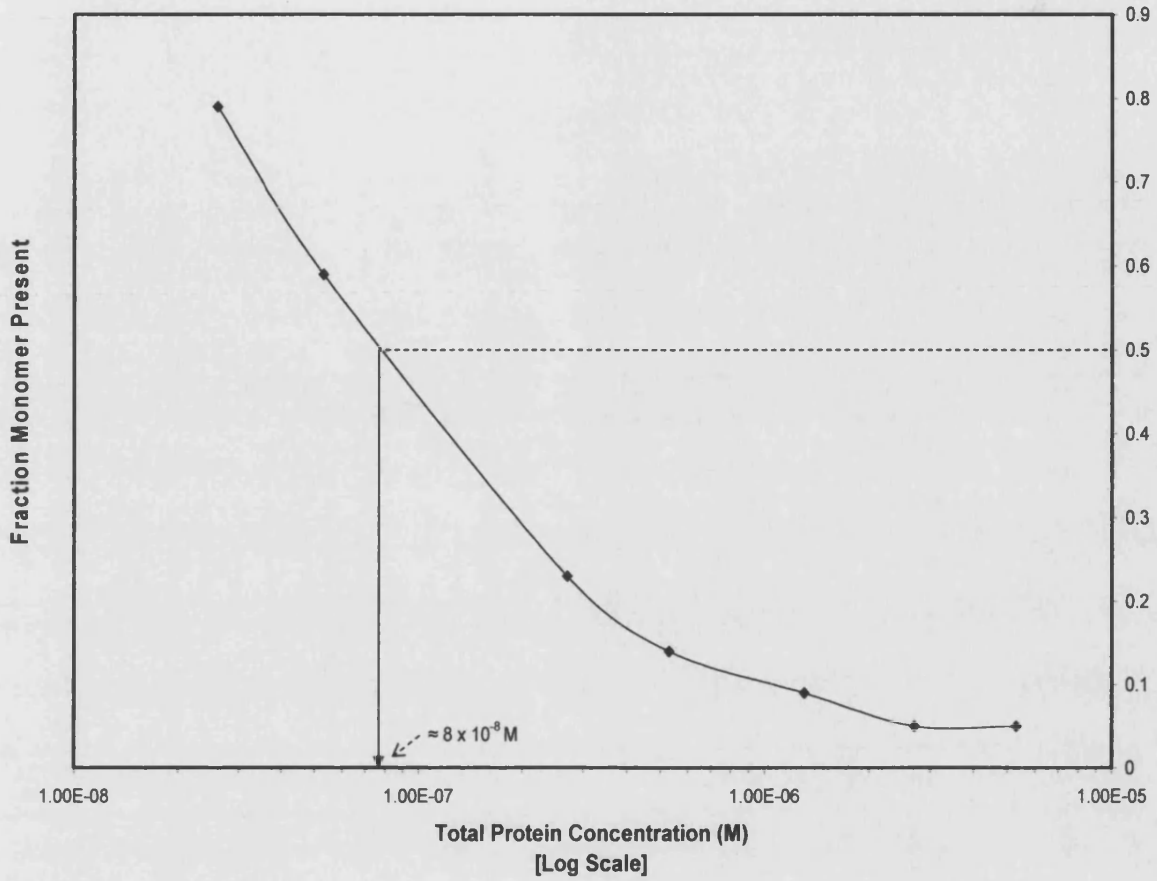


Figure 21 : Plot of Fraction Monomer Present *vs.* Total Protein Concentration (log scale).

Using Table 3, a plot of the fraction monomer present in the total protein eluted *vs.* total protein concentration eluted from the column allows for the calculation of the dissociation equilibrium constant of NusB (i.e. when 50 % of the protein is dimerised and 50 % is in monomeric form). Thus at a fraction monomer present of 0.5 (50 % of the protein is in monomeric form) the X-value is approximately 8×10^{-8} M indicating a K_d of 80 nM.

Chapter 5

Results –

NusG and the Rho Termination Factor

The NusB – NusE interaction was investigated in the previous chapter. A second interaction between NusG and Rho (Pasman and von Hippel, 2000) is also believed important in antitermination and both proteins were thus studied.

5.1. Characterisation of the *M. tuberculosis* NusG protein

In comparison to NusG from other organisms, little is known of the *M. tuberculosis* NusG protein. Interaction studies have been carried out on *E. coli* NusG and it has been shown to bind to both RNA polymerase and Rho. It has also been implicated in translation (Zellars and Squires, 1999). The crystal structure of *A. aeolicus* NusG revealed it to be a three domain protein. Sequence homology modelling then allowed for the prediction of the *E. coli* NusG structure. Two of the three domains have nucleic acid, *rRNA* and protein – protein interaction functional characteristics however a role for the third domain could not be determined (Steiner *et al.*, 2002).

None of these properties have however yet been shown in NusG from *M. tuberculosis*. Sequence analysis along with biochemical and biophysical studies have been employed to investigate the structural and domain organisation of *M. tuberculosis* NusG.

5.1.1. Amino-acid sequence analysis of *M. tuberculosis* NusG.

NusG is a 238 amino-acid, 26 kDa protein (http://genolist.pasteur.fr/TubercuList/). The T-Coffee alignment program (Notredame *et al.*, 2000) was used to align NusG from the organisms *M. tuberculosis*, *A. aeolicus* and *E. coli*. This can be seen in Figure 24. The alignment immediately shows the presence of an N-terminal extension found on the *M. tuberculosis* NusG when compared with that from both *A. aeolicus* and *E. coli*. The

function of this extension is unknown and deletion studies would be very useful in the determination of the function or necessity of these forty one amino-acids. Immediately following the N-terminal extension is a region of high homology with both *A. aeolicus* and *E. coli* (amino-acids 42 – 99). This corresponds to part of domain I in the *A. aeolicus* structure and is thus likely to be involved in both nucleic acid and protein – protein interactions. A region present in *A. aeolicus* but not in either of *E. coli* or *M. tuberculosis* corresponds to domain II in the *A. aeolicus* structure and is of unknown function. It may have a role in structural stability of the protein in this thermophilic organism. Amino-acids 100 to 155 then form the rest of domain I (with respect to the *A. aeolicus* structure). A region linking domain I and a second domain (corresponding to domain III of *A. aeolicus)* is made up of amino-acids 156 to 181. The second domain (amino-acids 182 to 238) shows strong homology with both *E. coli* and *A. aeolicus* and corresponds to domain III of *A. aeolicus* NusG. Again it is likely to be involved in *rRNA* binding and protein – protein interactions. The KOW motif present in domain III of *A. aeolicus* NusG is also present in *M. tuberculosis* NusG (amino-acids 184 to 211 – sequence highlighted in Figure 22) thus confirming the proteins *rRNA* binding and protein – protein binding roles (Kyrpides *et al.*, 1996; Steiner *et al.*, 2002).

1-VTTFDGDTSAGEAVDLTEANAFQDAAAPAEEVDPAAALKA
ELRSKPGDWYVVHSYAGYENKVKANLETRVQNLDVGDYIFQV
EVPTEEVTEIKNGQRKQVNRKVLPGYILVRMDLTDDSWAAVR
NTPGVTGFVGATSRPSALALDDVVKFLLPRGSTRKAAKGAAS
TAAAAEAGGLERPVVEVDY**EVGESVTVMDGPFATLPATISEV**
NAEQQKLKVLVSIFGRETPVELTFGQVSKI**-238**

Figure 22 : Amino-acid sequence of NusG with the KOW homologous sequence highlighted

From amino acid sequence analysis (see figures 23 and 24), it is probable that *M. tuberculosis* NusG is made up of two domains connected by a linker region. Based on sequence homology, the N-terminal (domain I) and C-terminal domains (domain III) probably share the nucleic acid and protein – protein binding functions as their equivalents in *A.aeolicus* NusG.

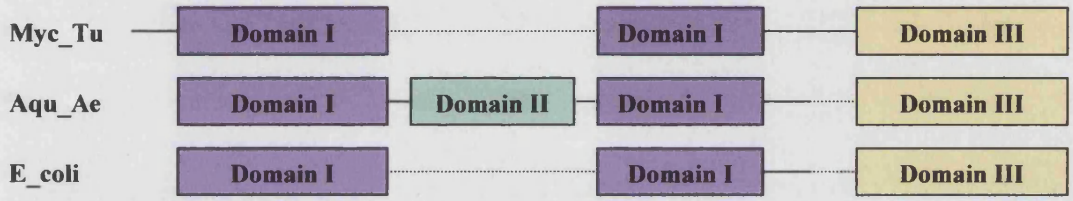


Figure 23: Domain homology of *M. tuberculosis*, *A. aeolicus* and *E. coli* NusG.
Sequence homology analysis (see Figure 23) allowed for the representation of domain homology of *M. tuberculosis* (Myc_Tu) and *E. coli* (E_coli) NusG with respect to the known *A. aeolicus* (Aqu_Ae) NusG domain organisation. Dotted lines indicate regions of *A. aeolicus* NusG not present in *M. tuberculosis* or *E. coli* NusG and regions of *M. tuberculosis* NusG not present in *A. aeolicus* or *E. coli* NusG. The N-terminal extension found on *M. tuberculosis* NusG is shown (Steiner *et al.*, 2002).

A prediction for *M. tuberculosis* NusG secondary structure is shown in Figure 25. The alignment predicts an α-helical content of 26 %, a β-sheet content of 27 % and non-structured content of 47 % for NusG.

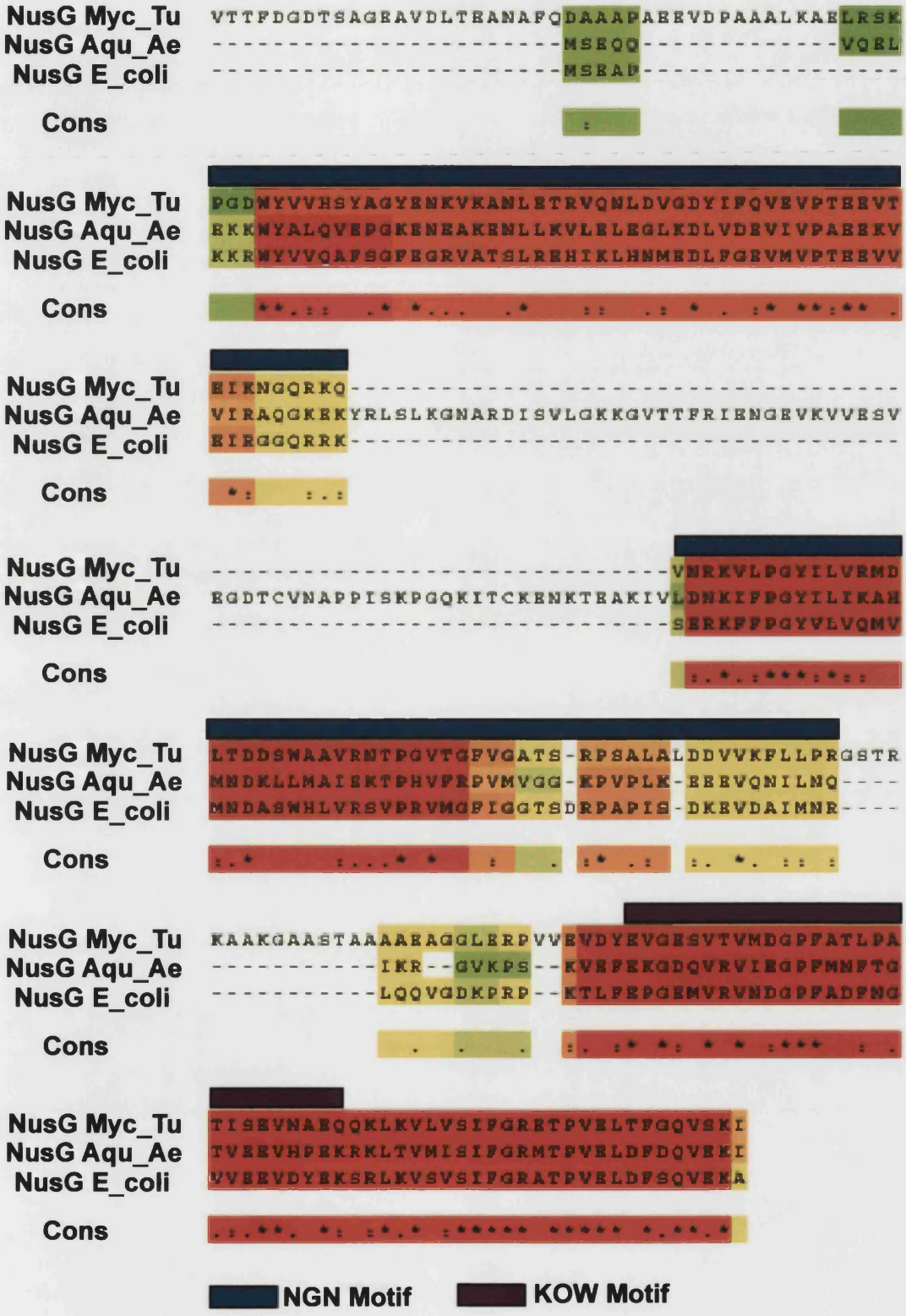
5.1.2. Expression and purification of NusG.

NusG was cloned into pET15b (Novagen) and over expressed with an N-terminal His-tag. Purification was carried out using a His-affinity resin followed by ion-exchange using Source Q (Amersham Pharmacia) and size-exclusion chromatography (Superdex

75). Concentrations of upwards of 2 mg/ml (yield of 10 mg/L of culture) were achieved and were thus suitable for carrying out biochemical and biophysical characterisation. The purified protein was analysed using ESI MS in order to confirm the molecular weight and hence confirm the identity of the protein; shown in Figure 26. ESI MS gave a molecular weight of 27479.89 ± 1.39 Da. The molecular weight of NusG (including the N-terminal His-tag) is 27577.8 Da so ESI MS indicates the N-terminal methionine is removed. The ESI MS also shows a second species is present in the protein sample with a molecular weight of 27655.00 ± 1.27 Da. This corresponds to α-*N*-gluconylation of the N-terminus. This common modification of the His-tag sequence has been shown by Geoghegan *et al.*, where there is a resultant 178 Da increase in molecular weight in N-terminally His-tagged, recombinant proteins. A 258 Da modification is also occasionally seen as a result of 6-phosphogluconylation of the N-terminus however it was not seen during the purification of NusG.

Figure 24 : Amino Acid Sequence Alignment of NusG from *M. tuberculosis, A. aeolicus* and *E. coli.*

The T-Coffee alignment program was used to show alignment between the NusG amino acid sequences from *M. tuberculosis* (Myc_Tu), *A. aeolicus* (Aqu_Ae) and *E. coli* (E_coli). The alignment shows homology throughout most of the *M. tuberculosis* NusG length. When compared with sequence, structural and biochemical information from *A. aeolicus*, *M. tuberculosis* NusG appears to contain two domains both involved in nucleic acid and rRNA binding as well as protein – protein interactions. Good levels of homology are defined by red shading, above average homology defined by orange shading, below average by yellow shading and poor homology by green shading. The KOW motif and NGN motif are also shown.


NusG Myc_Tu
NusG Aqu_Ae
NusG E_coli
Cons
NusG Myc_Tu
NusG Aqu_Ae
NusG E_coli
Cons
NusG Myc_Tu
NusG Aqu_Ae
NusG E_coli
Cons
NusG Myc_Tu
NusG Aqu_Ae
NusG E_coli
Cons
NusG Myc_Tu
NusG Aqu_Ae
NusG E_coli
Cons
NusG Myc_Tu
NusG Aqu_Ae
NusG E_coli
Cons
NGN Motif
KOW Motif

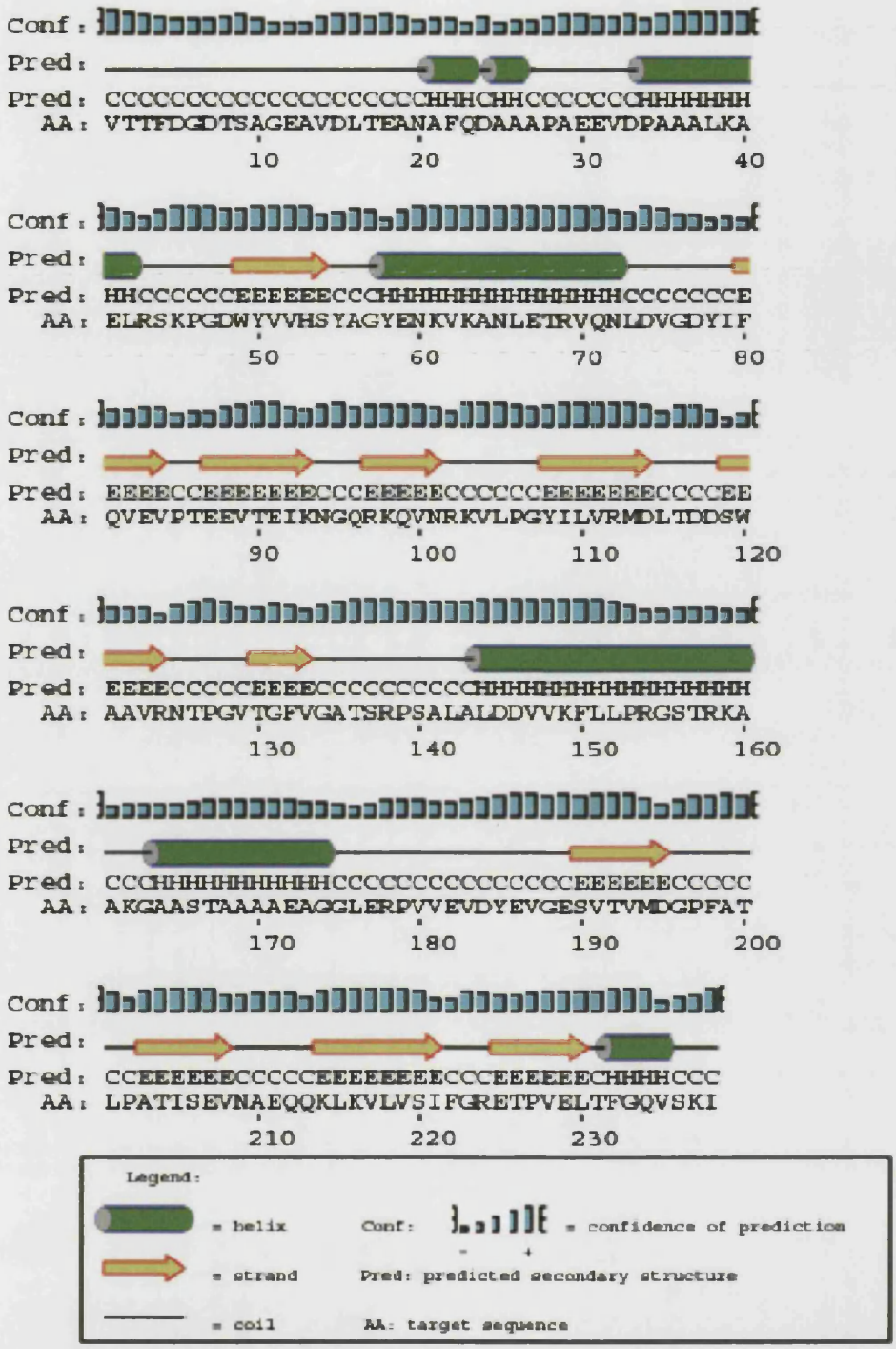


Figure 25 : Secondary structure prediction for *M. tuberculosis* NusG using PSI-Predict.

The predicted secondary structure of NusG contains both a-helices (approximately 26 %)and ß-sheets (27 %). Just under half (47 %) of the protein is predicted to be unstructured.

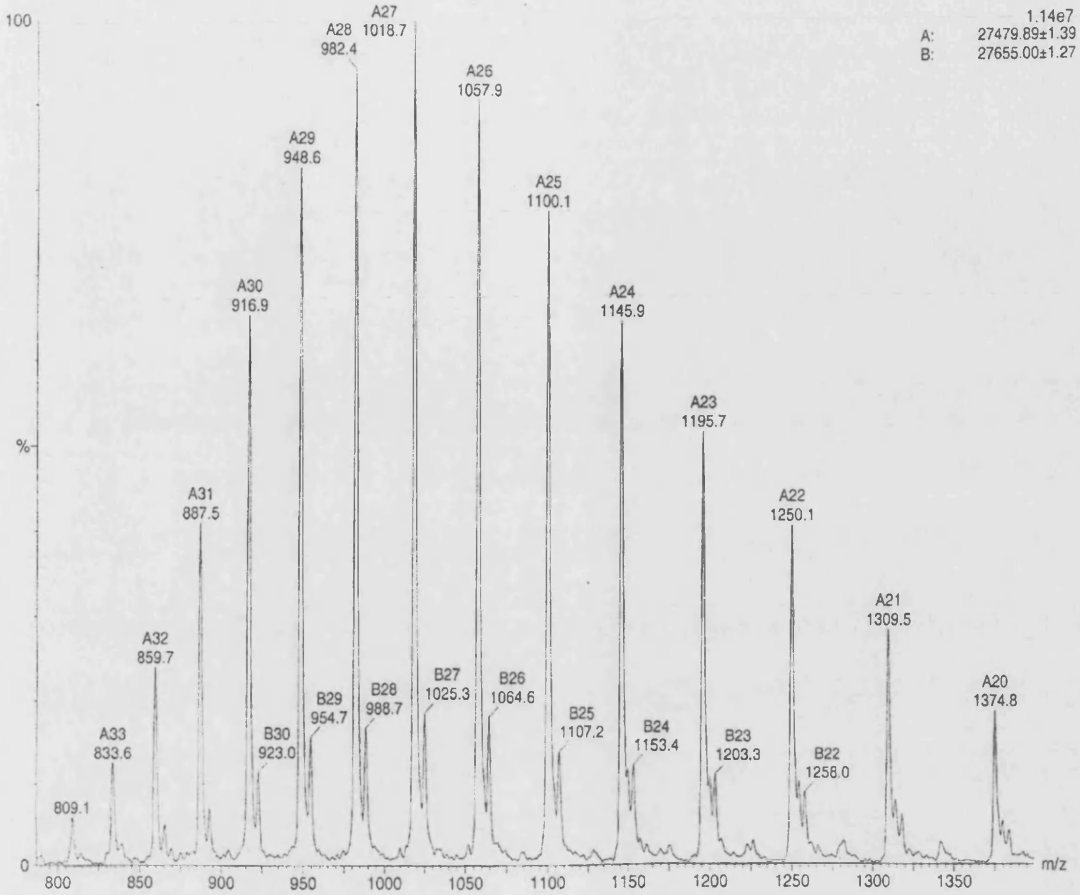


Figure 26 : ESI MS of *M. tuberculosis* NusG

The correct molecular weight for NusG is 27479.89 kDa (Calculated M.W. : 27478.7 kDa) taking into account methionine processing. A second species can also be seen with an adddtional mass of 176 Da corresponding to N-terminal α-N-gluconylation (Geoghegan *et al,* 1999).

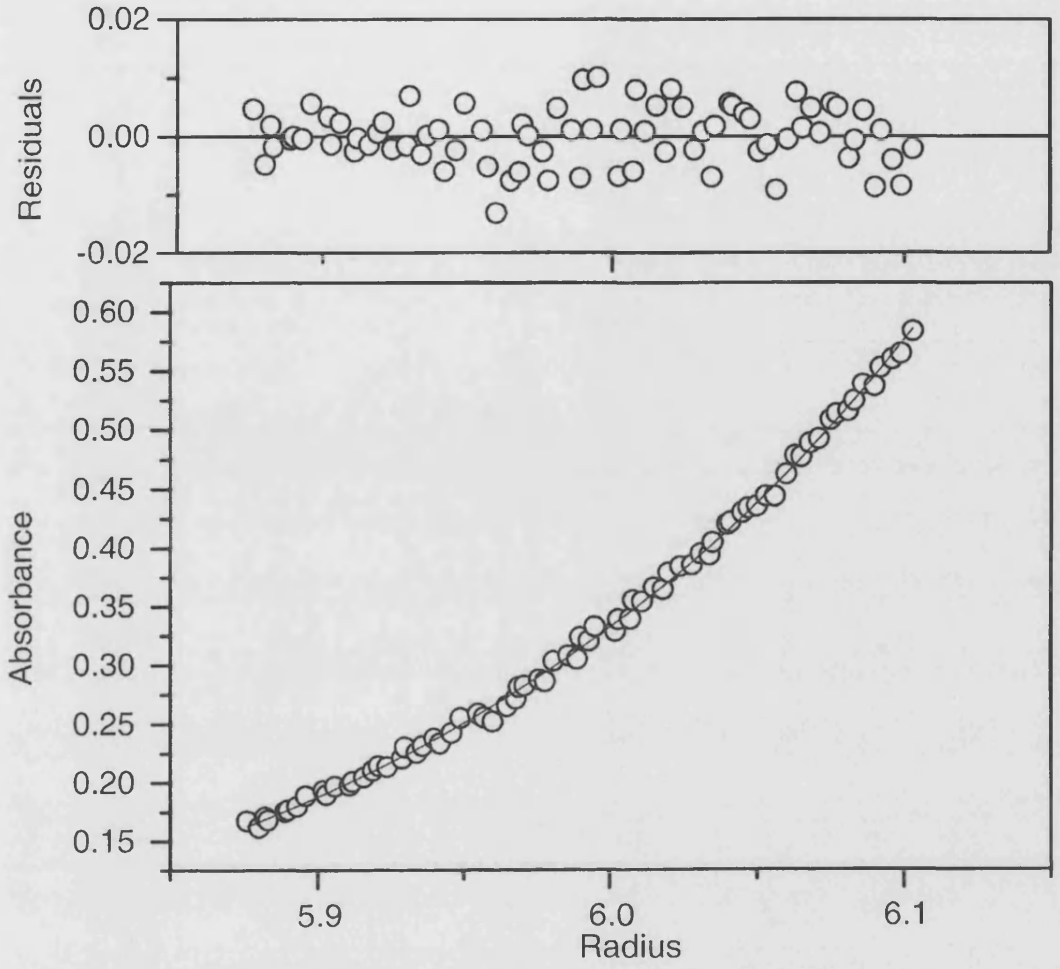
5.1.3. Sedimentation equilibrium ultracentrifugation of NusG.

The solution oligomeric state of NusG was analysed using sedimentation equilibrium ultracentrifugation as detailed earlier. NusG at a concentration of 0.4 mg/ml (15 μM) gave a molecular weight of 25.4 ± 1 kDa. This therefore indicates a monomeric form of the protein in solution. See Figure 27.

5.1.4. Dynamic light scattering of NusG.

Dynamic light scattering was carried out on a 1 mg/ml (37 μM) NusG solution (DynaPro). The results are shown in Table 4. The measured translational diffusion coefficient is the molecular parameter that governs the translational motion of NusG in solution. A hydrodynamic radius (R_H) prediction can be determined using Stokes equation (equation 13), D_W^{20} and D_T. The polydispersity is a measurement of the standard deviation of the distribution that best fits the autocorrelation function decay. A polydispersity value less than 25 % of the value of R_H indicates a monodisperse sample. The polydispersity of NusG is 10 % of that of the R_H indicating the NusG preparation is highly monodisperse.

The results gave a predicted molecular weight of 54.27 kDa, a much higher value than the monomeric formula molecular weight of NusG. The predicted molecular weight is in fact double that of the formula molecular weight and may thus indicate a NusG dimer in solution. As sedimentation equilibrium ultracentrifugation demonstrated a monomeric protein, the dynamic light scattering data reflects the fact that NusG contains a substantial amount of extended structure. The elongated properties of NusG can be confirmed by calculation of f/f_0.



DOF = 624 Variance = 3.67844E-5

Fitted Parameters:

Co	Offset
0.247	-0.038
0.252	-0.035
0.264	-0.041
0.193	-0.031
0.185	-0.027
0.192	-0.030
0.127	-0.022
0.119	-0.016
0.126	-0.018

M = 25431 B = 0

Speed = 17000
Time = 149848
Temp = 20
V-bar = 0.734
Rho = 1.011

Figure 27 : Sedimentation equilibrium ultracentrifugation of *M. tuberculosis* NusG

Equilibrium sedimentation ultracentrifugation shows an overall molecular weight of 25.4 ± 1 kDa for NusG. The plot is a global fit of a single concentration at nine time points over three speeds.

Table 4 : Dynamic light scattering data for *M. tuberculosis* NusG.

Translational Diffusion Coefficient	6.0×10^{-7} cm^2/sec
Hydrodynamic Radius	3.28 nm
Polydispersity	0.358 nm
Molecular Weight	54.3 kDa
Temperature	17.7 °C
Baseline	1.000

The frictional ratio (f/f_0) is an indication of the deviation in molecular shape of a protein from a hard sphere and was calculated using the equations 31 – 33.

$$V = \frac{M\bar{v}}{N_0} \quad (31)$$

$$r_0 = \sqrt[3]{\frac{3V}{4\pi}} \quad (32)$$

$$\frac{f}{f_0} = \frac{6\pi\eta r}{6\pi\eta r_0} \quad (33)$$

where V=volume
ν = partial specific volume = 0.7430
r = hydrodynamic radius from light scattering data
r_0 = radius of a hard sphere
η = viscosity (D_W^{20} of 0.0102)

f was calculated using the hydrodynamic radius (see table 4) and was found to be 6.1×10^{-8}. f_0 was calculated as 3.7×10^{-8} giving an f/f_0 value of 1.65. This indicates NusG is not a sphere (f/f_0 of 1) and is thus likely to be highly elongated. These data correspond

with secondary structure predictions where approximately 47 % of the protein is unfolded.

5.1.5. Circular dichroism of NusG.

Circular dichroism was carried out to examine the secondary and tertiary structural content of NusG. Both near and far-UV CD spectra were recorded and are shown in Figure 28.

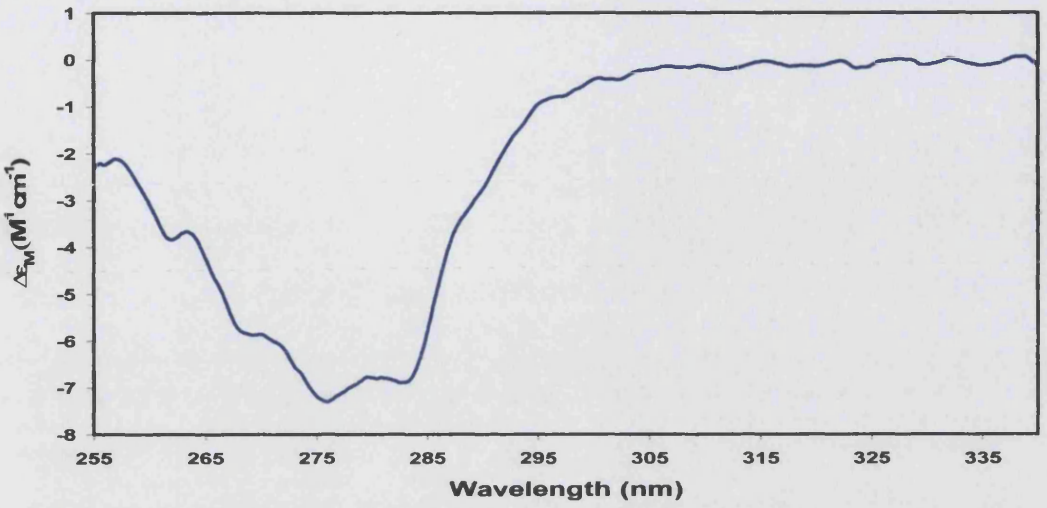
Analysis of the near-UV spectra shows weak peaks in the 255 – 270 nm region indicating that some of the Phe residues are contributing to tertiary structure. Phe residues at positions 22, 80, 132, 150 and 232 are all found in α-helical or β-sheet regions according to secondary structure predictions. The presence of intense peaks in the 275 – 282 region indicate that at least some of the six Tyr are ordered and are involved in tertiary structure formation. According to sequence homology (Figure 24), the tyrosines at positions 50, 56, 58, 78 and 108 may be found in an N-terminal domain (corresponding to domain I of the *A. aeolicus* NusG structure) and the tyrosine at position 185 found in a C-terminal domain (corresponding to domain III of *A. aeolicus* NusG). Secondary structure predictions (Figure 25) show the tyrosines at positions 50 and 108 to be involved in α-helical structure and tyr58 to be involved in β-sheet structure. A declining signal is found above 280 nm meaning the two Trps are contributing little to the overall signal.

The near-UV CD thus indicates there is certainly a degree of tertiary structure within NusG with some of the phenylalanines and tyrosines found within these regions of tertiary structure. The tryptophans (found at positions 49 and 120) do not however seem

to contribute to the NusG tertiary structure possibly indicating that regions in the vicinity of positions 49 and 120 may be unfolded. The secondary structure prediction shows positions 49 and 120 to be found at the beginning of six-residue long β-sheets. The secondary structure prediction may thus be slightly inaccurate and these residues are, in fact, not present in secondary structure folds or that these particular secondary structure regions do not contribute to tertiary structure.

Interpretation of the far-UV CD spectra for NusG seen in Figure 28 indicates the presence of both α-helices and β-sheets. The Δε per residue value at 222 nm is -2.37 $M^{-1}cm^{-1}$ indicating the presence of α-helices. The α-helical content is thus approximately 22 % which corresponds to approximately 52 helical residues. This is similar to that seen in the secondary structure prediction (26 %). A broad plateau is seen between wavelengths 210 nm and 220 nm probably resulting from a combined α/β effect, and hence indicating the presence, of both α-helices and β-sheets within the protein.

(A)



(B)

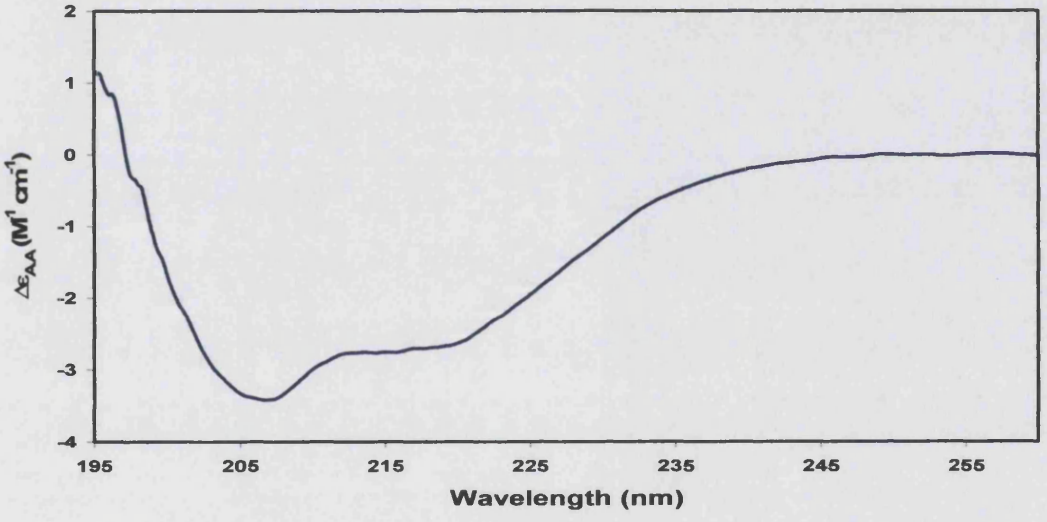


Figure 28 : CD of *M. tuberculosis* NusG (1 mg/ml and 10 µM – 20°C)
(A) The near-UV CD of *M. tuberculosis* NusG indicates the presence of some tertiary structure as a result of peaks in the 255 nm – 270 nm and 275 – 282 nm regions of the spectra. These are as a result of phenylalanine and tyrosine residues respectively which are either immobilised or interacting with neighbouring residues. **(B)** The far-UV CD indicates the presence of both α-helical and β-sheet structure in NusG. The peaks at the 208 nm and 222 nm wavelength indicate α-helical structure and equate to approximately 22 % total α-helical structure in NusG. The plateau between wavelengths 210 nm and 220 nm shows the co-effects of α-helices and β-sheets indicating both are present in NusG. The data correlates with that determined from the secondary structure prediction program where NusG was shown to be 26 % helical and 27 % β-sheet.

5.1.6. Limited proteolysis of *M. tuberculosis* NusG.

The domain organisation within NusG was investigated using limited proteolysis with trypsin and chymotrypsin. Proteolytic enzymes, such as trypsin and chymostrypsin, frequently cleave rapidly at sites within proteins that are unfolded or exposed. Sites found within three dimensional domains are often protected from the actions of the enzymes. The treatment of a protein with trypsin (cleaves the C-terminal sides of Arg and Lys) or chymotrypsin (cleaves C-terminal to Tyr, Phe, Trp and Leu) thus gives an indication of domain structure within a protein (Wilkinson, 2003). N-terminal sequencing of major cleavage products then allows for the placement of domains at specific locations within the protein. Limited proteolysis is thus an important tool for further characterisation of a protein and gives additional information as to possible secondary and tertiary structural features.

A time-course (70 minutes) trypsin and chymotrypsin proteolytic digest was carried out in order to look at the location of domains within NusG. Cleaved protein was removed after 0, 2, 5, 10, 20, 40 and 70 minutes of protease treatment and the products separated on SDS-polyacrylamide gels (See Figure 29)

Trypsin digestion of NusG was rapid and cleavage sites were abundant within the protein. Two stable products were apparent at approximately 3 kDa and 8 kDa. The remainder of the protein was rapidly hydrolysed by the trypsin. Chymotrypsin treatment was less rapid. Three stable products were observed; the 3 and 8 kDa polypeptides seen after trypsin digestion and an additional polypeptide found 2 – 3 kDa below undigested NusG. A small amount of NusG remained uncleaved after chymotrypsin treatment.

Three polypeptide products (the 8 and 34 kDa chymotrypsin and 8 kDa trypsin products) were N-terminally sequenced in order to determine the position of the polypeptides within NusG. The results of the N-terminal sequencing and the positions of the polypeptides within NusG are shown in Table 5 and Figure 30.

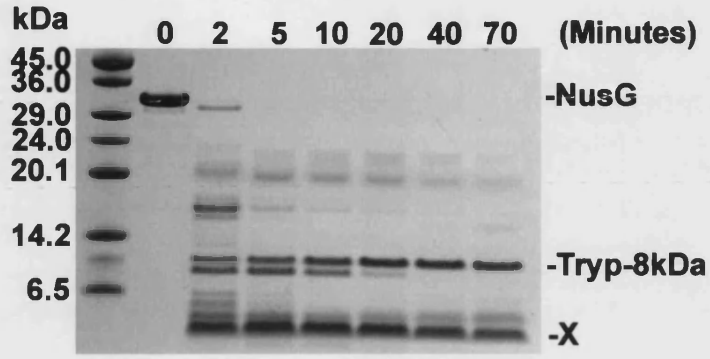
Results from both trypsin and chymotrypsin cleavage indicate a single stable domain at the C-terminal end. Both the trypsin and chymotrypsin 8 kDa stable polypeptides contain the KOW homologous sequence and thus indicate that the C-terminal region of NusG is likely to be structured and important functionally. It is likely to correspond to domain III found in *A. aeolicus* NusG. However when compared with the secondary structure prediction both the trypsin and chymotrypsin cleavage sites are within a predicted helical region. This may mean that this region is helical yet not folded into tertiary structure or that, more likely, the secondary structure prediction is inaccurate, at least in this region.

Based on amino acid sequence homology, an N-terminal domain, corresponding to domain I in *A. aeolicus* NusG, would have been expected to have been identified. No stable polypeptide products were however identified within the N-terminal region possibly indicating the N-terminal domain shows little tertiary structure. It would still be expected that a functional N-terminal domain does exist.

M. tuberculosis is thus likely to be made up of a loosely folded N-terminal domain and a structured C-terminal domain corresponding to domains I and III from *A. aeolicus* NusG respectively. This correlates with results from light scattering and circular dichroism implying that the overall structure of *M. tuberculosis* NusG is likely to be elongated or

largely unfolded. A schematic showing the positions of trypsin and chymotrypsin cleavage sites, relative to the expected positions of the N-terminal and C-terminal domains (based on amino acid sequence homology) is shown in Figure 30b.

(A) Trypsin digest of *M. tuberculosis* NusG



(B) Chymotrypsin digest of *M. tuberculosis* NusG

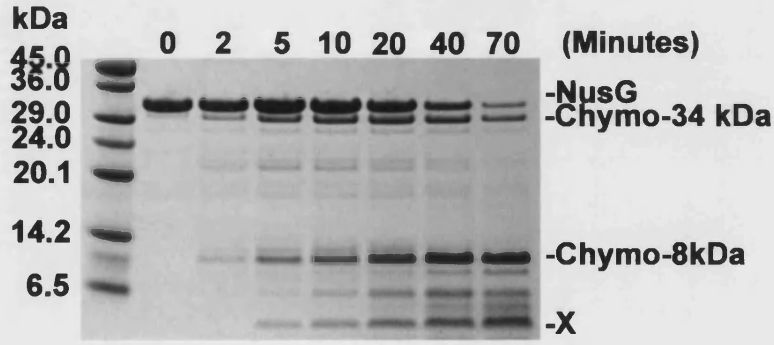


Figure 29 : Limited proteolysis of *M. tuberculosis* NusG using trypsin and chymotrypsin.
(A) Proteolytic digestion of NusG by trypsin (0.004 w/w) shows two stable polypeptides at 3 kDa (**X**) and 8 kDa (**Tryp-8kDa**). Tryp-8kDa was submitted to N-terminal sequencing (See Table 5). The 3 kDa polypeptide was regarded as too small a fraction of NusG and thus N-terminal sequencing was felt to be unneccesary. **(B)** Chymotrypsin digestion (0.004 w/w) gave rise to a 3 kDa polypeptide (**X**), an 8 kDa polypeptide (**Chymo-8kDa**) and a 34 kDa polypeptide (**Chymo-34-kDa**). Chymo-8kDa and Chymo-34kDa were submitted to N-terminal sequencing. The 3 kDa fragment was again regarded as too small and is likely to be the cleaved N-terminus His-tag. Note - NusG is a 25 kDa protein but during SDS-PAGE, NusG moves as a 35 kDa protein.

Table 5 : N-terminal sequences of stable NusG polypeptides after trypsin and chymotrypsin digestion.

BAND	N-TERMINAL SEQUENCE
Tryp-8kDa	GAA(S/R)T
Chymo-8kDa	(L/G)(S/A)PRG
Chymo-34kDa	DGDTS

Tryp-8kDa = trypsin 8 kDa band; **Chymo-8kDa** = chymotrypsin 8 kDa band; **Chymo-34kDa** = chymotrypsin 34 kDa band; the 3 kDa bands from both trypsin and chymotrypsin digests were not sequenced due to the small size of the polypeptide.

(A)

1-VTTF**DGDTS**AGEAVDLTEANAFQDAAAPAEEVDPAAALKA
ELRSKPGDWYVVHSYAGYENKVKANLETRVQNLDVGDYIFQV
EVPTEEVTEIKNGQRKQVNRKVLPGYILVRMDLTDDSWAAVR
NTPGVTGFVGATSRPSALALDDVVKFL**LPRG**STRKAAK**GAAS**
TAAAAEAGGLERPVVEVDYEVGESVTVMDGPFATLPATISEV
NAEQQKLKVLVSIFGRETPVELTFGQVSKI-238

(B)

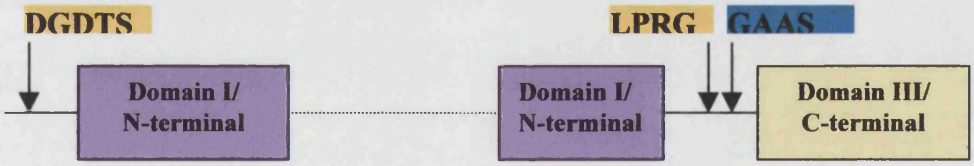


Figure 30 : NusG trypsin and chymotrypsin cleavage sites.
(A) N-terminal sequences of trypsin resistant polypeptides are shown in **Blue**. The polypeptide indicates the presence of a C-terminal domain (containing the KOW motif). N-terminal sequences of chymotrypsin resistant polypeptides are highlighted in **Yellow**. The larger polypeptide (beginning DGDTS) corresponds to the bulk of NusG and signifies that NusG shows a degree of resistance to chymotrypsin digestion. The 8 kDa polypeptide (beginning LPRG) again indicates the presence of a C-terminal domain. **(B)** shows the positions of cleavage sites relative to the position of the *M. tuberculosis* NusG N- and C-terminal domains predicted from amino acid sequence homology (see Figure 24) The dotted line represents a region present in *A. aeolicus* NusG but not in that of *M. tuberculosis*.

5.2. The *M. tuberculosis* Rho terminator.

M. tuberculosis contains a homolog very similar to the Rho termination proteins of numerous prokaryotes (http://genolist.pasteur.fr/TubercuList). *M. tuberculosis* Rho is a 602 amino-acid protein with a molecular weight of 65 kDa. *E. coli* Rho functions as a hexamer with each subunit containing an N-terminal RNA binding domain and large C-terminal ATPase domain (Modrak and Richardson, 1994; Dombroski *et al.*, 1988a,b; Yu *et al.*, 2000). By comparison, it is likely the *M. tuberculosis* Rho will also assemble into a hexamer, with a molecular weight of 390 kDa. The role of *E. coli* Rho in antitermination has not yet been fully resolved although it must, in some way, be prevented access to the transcription elongation complex.

Rho-dependent termination has not been observed within *M. tuberculosis* itself and nor have specific Rho RNA binding sites been found. This makes it difficult to predict the function of *M. tuberculosis* Rho-termination factor in *M. tuberculosis rRNA* antitermination. Rho was therefore cloned, over expressed and purified and interaction studies with NusA and NusG carried out in an attempt to understand any involvement of Rho in *M. tuberculosis rRNA* antitermination.

5.2.1. Cloning, expression and purification of Rho.

The Rho DNA sequence was amplified directly from *M. tuberculosis* genomic DNA using PCR. The product was then cloned into the *Xho*1 and *Nde*1 sites in the pET22b (Novagen) plasmid vector. This results in a C-terminal His-tag on the expressed Rho protein. The recombinant plasmid was then transformed into BL21(DE3)p*LysS E. coli* cells and C-terminally His-tagged *M. tuberculosis* Rho over expressed. Rho was then

purified using a His-affinity resin selecting for the C-terminal His-tag (See Figure 31a). This resulted in relatively large amounts of Rho however the protein was not sufficiently pure. The nucleic acid binding properties of Rho allowed for further purification using a Heparin column (Amersham Pharmacia). Rho eluted at a NaCl concentration of ≈ 0.6 M which correlates with the observations of Sharp *et al.* (1983) who observed elution of *E. coli* Rho from Heparin at about 0.5 M. This increased the purity of Rho dramatically (See Figure 31b). Finally Rho was purified by size-exclusion gel filtration in an attempt to acquire a homogenous preparation (See Figure 31c). Eluant from size-exclusion chromatography was pure and contained Rho. A yield of 15 mg/1.5L of culture and final concentrations of 1.23 mg/ml (18.5 μM) of Rho were achieved.

The peptide molecular weight of purified Rho was determined by ESI MS. A value of 66098.16 ± 4.54 Da was obtained shown in Figure 32. The formula molecular weight for Rho is 66198.5 Da and with methionine processing is 66067.4 Da. ESI MS thus gave a molecular weight between the two, probably indicating a proportion of recombinant Rho contains the N-terminal methionine while the remainder of the Rho molecules have undergone methionine processing. This apparent intermediate mass is probably due to partial processing of the N-terminal methionine, a process dependent on the nature of the second residue. Hirel *et al.* (1989) and (Dalbøge *et al.* (1990) both showed that 10 % of proteins with threonine found at position two retain the N-terminal methionine with the remaining 90 % having undergone methionine processing. This thus appeared to have happened in purified *M. tuberculosis* Rho.

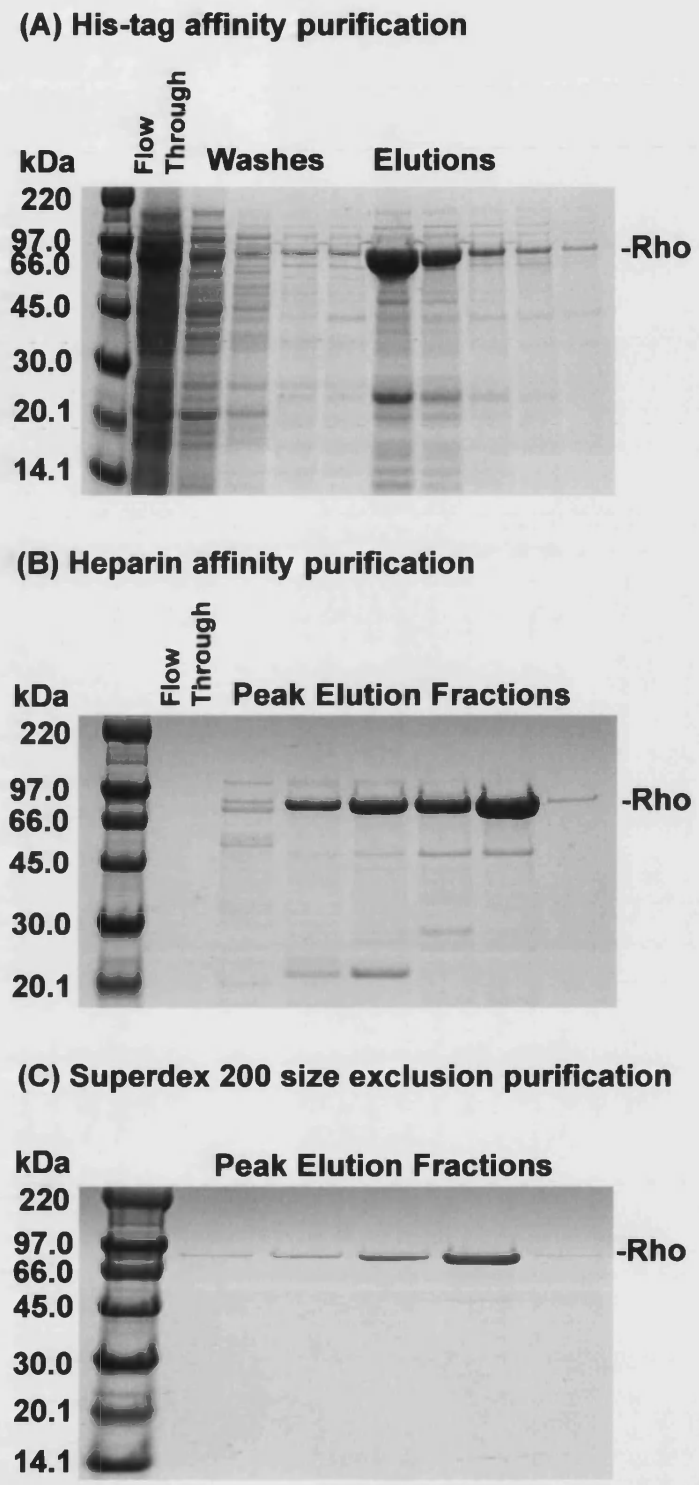


Figure 31 : Purification steps for recombinant *M. tuberculosis* Rho termination factor.

(A) His-tag affinity purification resulted in large amounts of Rho but insufficiently pure.

(B) Heparin-affinity purification resulted in a purer preparation but contaminants still appear to be present.

(C) Superdex 200 size-exclusion chromatography results in a pure preparation.

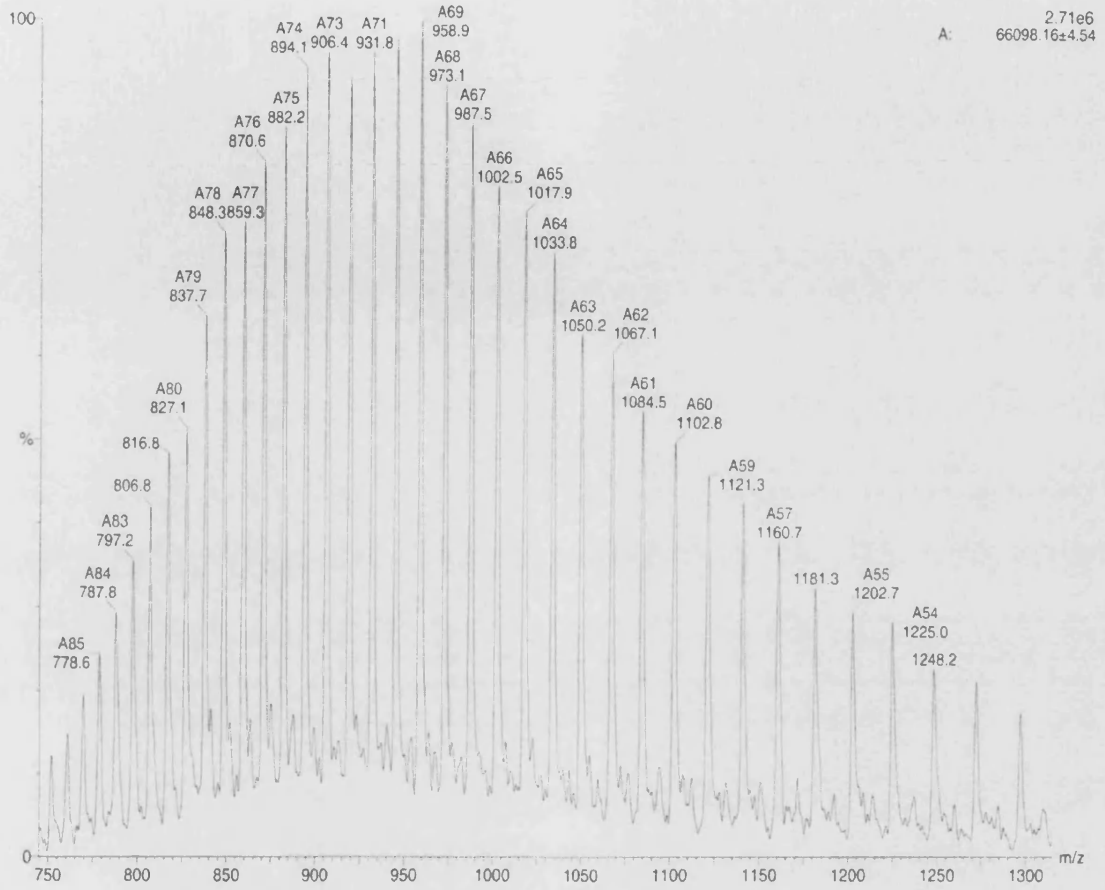


Figure 32 : ESI MS of purified *M. tuberculosis* Rho.
The Mass spectra of Rho shows a moleculer weight of 66098.16 Da. This does not correspond with an N-terminal methionine containing Rho (66198.5 Da) nor with a methionine processed Rho (66067.4 Da). The ESI MS thus indicates both species (methionine containing and methionine processed Rho) are present in the preparation and the calculated molecular weight is determined depending on the ratios of the two species present.

5.2.2. Interaction of Rho with NusA and NusG.

Initial studies on the interaction of Rho with NusA and NusG were carried out using gel filtration HPLC. The technique is based on the decrease in absorbance (or even complete disappearance) of the NusA or NusG peaks as a result of binding to Rho. This technique is sensitive only to strong interactions between proteins and in such cases an increase in Rho absorbance intensity may be seen. The results (seen in Figure 33) show that Rho resolves as two peaks during gel filtration. This is likely to be as a result of degradation over time and the peaks seen may indicate incorrectly assembled Rho hexamers. The binding of *M. tuberculosis* Rho with NusA and NusG were also not observed in these initial experiments. No decrease in NusA or NusG peaks was detected indicating the lack of binding between Rho and NusA and NusG (Figure 33). This may be as a result of Rho degradation and subsequent incorrect assembly of the hexameric protein. These studies are however ongoing and attempts are being made to stabilise the Rho preparation through improved purification procedures

Figure 33 : Elution traces from size-exclusion gel filtration of Rho and Rho in combination with NusA and NusG.

All graphs are plotted as absorbance *vs.* time (minutes). Peaks are labelled by (time, absorbance).

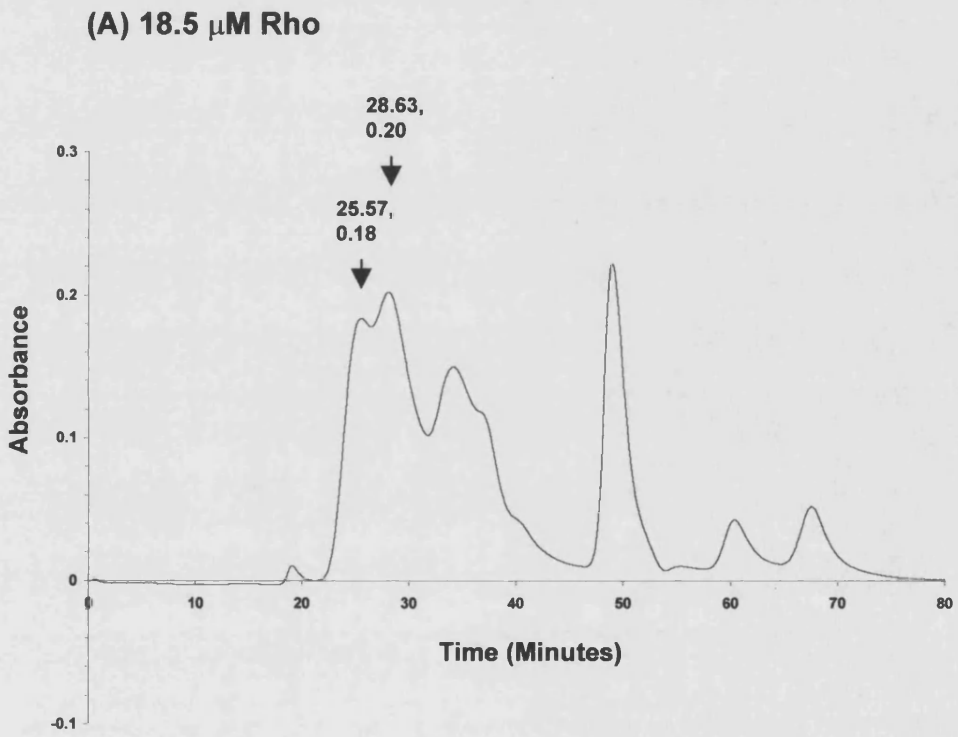
(A) 18.5 μM Rho gave rise to two peaks at 25.57 and 28.17 minutes closest to the expected position for hexameric Rho. Other peaks seen are likely to be additional Rho degradation products.

(B) 6.6 μM NusA gave rise to a single peak at 42.06 minutes.

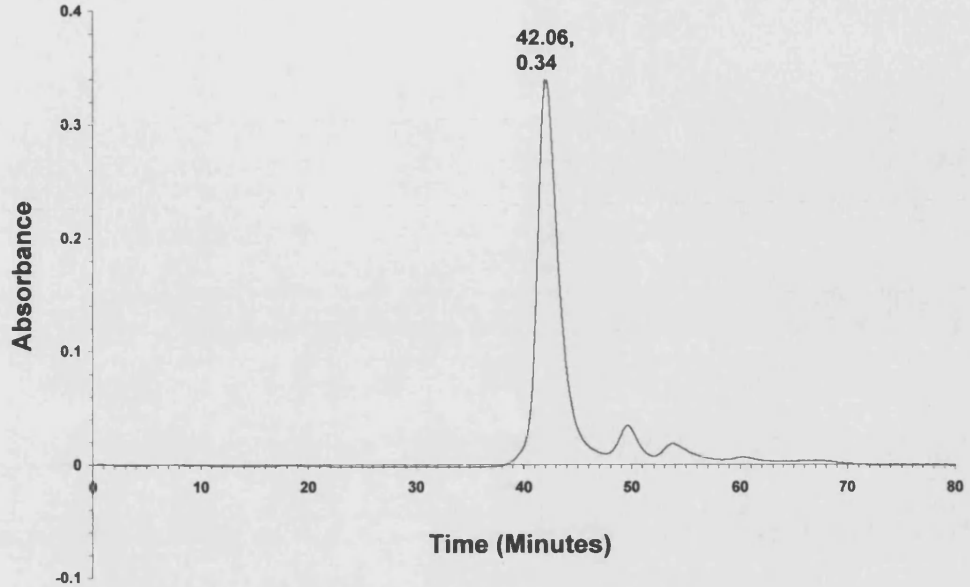
(C) 17.6 μM Rho and 6.6 μM NusA. Two peaks are seen for Rho (25.06 and 28.04 minutes) and a single peak for NusA (42.05 minutes) is seen.

(D) A single peak is seen for 15.4 μM NusG

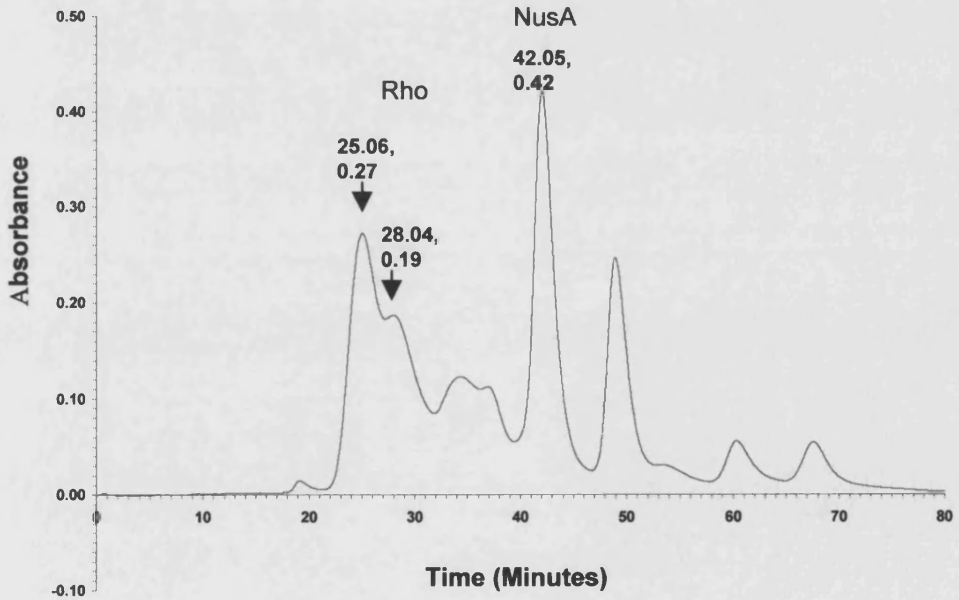
(E) 17.6 mM Rho and 15.4 mM NusG. Two peaks are seen for Rho (25.35 and 28.63 minutes) and a single peak for NusG (43.41 minutes) is seen.


(A) 18.5 μM Rho
28.63,
0.20
25.57,
0.18
Absorbance
0.3
0.2
0.1
0
-0.1
0
10
20
30
40
50
60
70
80
Time (Minutes)

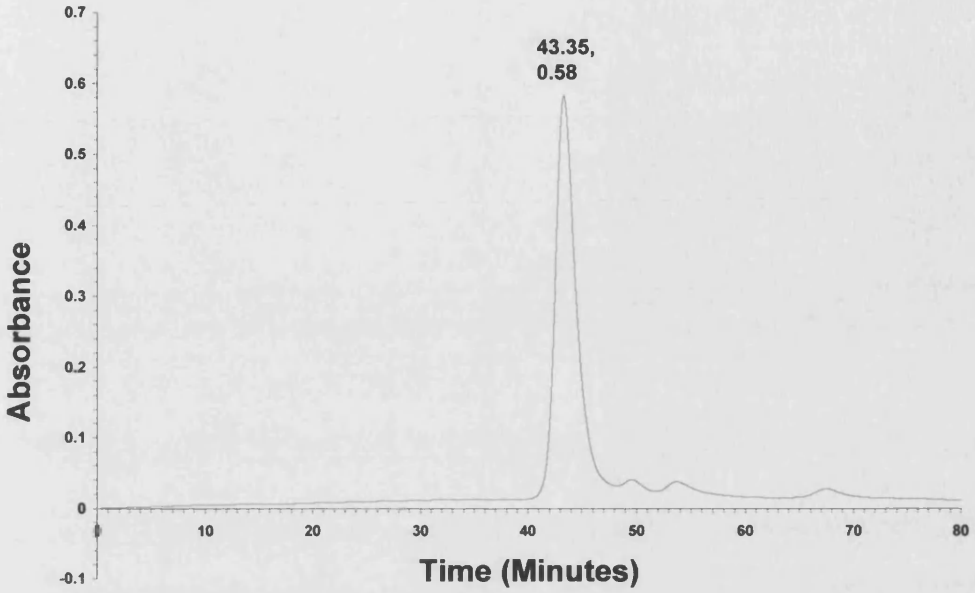
(B) 6.6 μM NusA



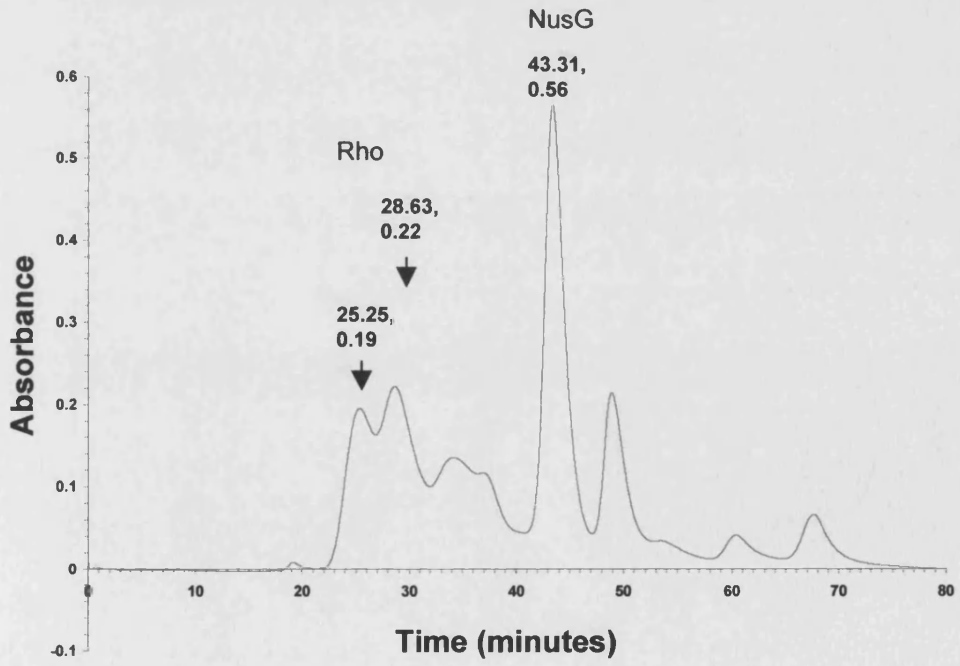
(C) 17.6 μM Rho + 6.6 μM NusA



(D) NusG (15.4 μM)



(E) Rho (17.6 μM) + NusG (15.4 μM)



Chapter 6

Discussion

The antitermination complex in *M. tuberculosis* represents a potentially important target for anti-mycobacterial drug development because of its crucial role in ribosomal RNA synthesis. *M. tuberculosis* contains a single *rrn* operon and accurate regulation and transcription of the operon is essential to the survival of the bacilli. The overall goal of work carried out on antitermination in *M. tuberculosis*, is the determination of the composition of the complex, identification and an understanding of the interactions between components of the complex, identification of RNA contacts and those proteins that make the contacts and an understanding of conditions under which antitermination occurs.

6.1. The *M. tuberculosis* antitermination complex.

Both the λ and the *E. coli* antitermination complexes have been well characterised and models have been proposed for the read-through of terminators in the λ genome and *E. coli rrn* operons (Luttgen *et al.*, 2002; Mason and Greenblatt, 1991). Antitermination in *M. tuberculosis* has not been observed and it is not known whether antitermination is, in fact, a regulatory mechanism for the transcription of the single *rrn* tuberculosis operon. Homologous sequences to the *E. coli* antitermination proteins NusA, NusB, NusE and NusG are present in the *M. tuberculosis* genome as are sequences homologous to the λ *nut* RNA regulatory sequences boxA, boxB and boxC (Cole *et al.*, 1998; Gonzalez-y-Merchand *et al.*, 1996, Kempsell *et al.*, 1992) This therefore suggests that the λ and *E. coli* antitermination models are good models for *M. tuberculosis rrn* transcriptional regulation.

Aims of this study were thus the confirmation or otherwise of the involvement of the various Nus factors in antitermination in *M. tuberculosis* and the identification of additional factors that may be required in the antitermination complex. In *E. coli* it has been shown that NusA interacts with RNA polymerase (Greenblatt and Li, 1981a; Liu *et al.*, 1996) and Rho (Schmidt and Chamberlin, 1984), NusB binds NusE (Mason *et al.*, 1992a), NusE binds RNA polymerase (Mason and Greenblatt, 1991) and NusG binds RNA polymerase (Li *et al.*, 1992) and Rho (Pasman and von Hippel, 2000). The use of His-tagged NusA, NusB, NusE and RNAPα and *M. tuberculosis* CFE's, in His-affinity pulldowns, detected none of these complex-forming interactions. The problems associated with this experiment were primarily due to non-specific interactions of some *M. tuberculosis* CFE proteins with the His-affinity resin. This resulted in high protein background levels during pulldown elutions. NusA and RNAPα (and more than likely RNAP β') also appear to bind to the His-affinity resin, as they are seen in the elution lanes of all His-affinity pulldown experiments. This renders analysis of interactions by this method difficult as it would not be possible to determine whether the interaction was with the tagged "bait" protein, NusA or RNAPα.

GST-affinity pulldowns were then used in an attempt to circumvent the problems seen with His-affinity pulldowns. NusB was N-terminally GST-tagged and pulldown assays carried out. Examination of bound fractions resolved on gels, showed no bands after Coomassie blue staining, firstly indicating the lowered affinity of *M. tuberculosis* soluble proteins for glutathione-sepharose resin, but, also suggesting no strong interactions with NusB. Probing of the bound samples with antibodies against NusE did,

however, show that NusE was present in the elution lanes of GST-NusB pulldown assays, identifying the interaction between *M. tuberculosis* NusB and NusE.

The failure of the NusB His-affinity pulldowns may also be explained by the detection method employed in the western blots, as no NusE band was seen in the His-NusB pulldown. The ECL Plus chemiluminescent detection protocol (Amersham Pharmacia Biotech), used for the GST-NusB pulldown, is up to 1000× more sensitive than the DAB (Di-Aminobenzidine) chromagenic protocol, used for the NusB-His affinity pulldowns. NusE may have been present but simply not detected in the His-affinity pulldowns.

The fact that the NusE band was not visualised by Coomassie blue staining after GST-NusB pulldowns, suggests either the presence of low concentrations of NusE in the CFE and/or weak binding between NusB and NusE. These data compare with the results of Gopal *et al.* (2001b) who reported weak interactions between *M. tuberculosis* NusB and NusE but are in contrast to *E. coli* NusB and NusE, which bind with a dissociation equilibrium constant of 10^{-7} M (Mason *et al.*, 1992a). The NusB – NusE complex may thus be stabilised within the mycobacterial antitermination complex by another, currently unidentified, interaction between either of NusB or NusE and one of the other Nus factors, RNAP or an unknown protein. In *E. coli* the NusB – NusE heterodimer is held in place by an interaction between NusE and the RNAP (Mason and Greenblatt, 1991). A possible difference between the *E. coli* and *M. tuberculosis* systems is that the RNAP may play more of a role in complex stabilisation in *M. tuberculosis* than in *E. coli*. Alternatively there may be an additional, unidentified factor which may interact

with both NusB and NusE thereby stabilising the weak interaction between these two Nus factors.

The combination of the use of GST-tags and high-sensitivity blotting protocols has proved successful in the identification of the NusB and NusE interaction *in vivo* for the first time. These methods can now be used to identify interactions between NusA, NusG, RNAP and, perhaps, unknown complex components.

6.2. The stoichiometry of the NusB – NusE interaction.

The observation of NusB as a dimer in *M. tuberculosis* (Gopal *et al.*, 2000) was somewhat surprising and an indication of mechanistic differences between the *E. coli* and *M. tuberculosis* antitermination systems. The *M. tuberculosis* molecule is dimeric both in the crystal lattice and in solution. *E. coli* NusB is monomeric and the NusB – NusE complex is heterodimeric (Altieri *et al.*, 2000; Luttgen *et al.*, 2002). This difference in NusB is the only disparity that is currently known regarding the Nus factors in *E. coli* and *M. tuberculosis* or the antitermination complex as a whole.

Studies were thus carried out with the aim of elucidating the stoichiometry of the *M. tuberculosis* NusB – NusE complex. Initially a monomeric NusB mutant (FE22.23AA) was constructed that allowed comparisons of the interactions between NusE and the NusB dimer or NusB monomer to be examined. The disruption of the NusB dimer was achieved by a double mutation of Phe22 and Glu23 located at the dimer interface identified from the crystal structure. The resulting monomer was confirmed by analytical ultracentrifugation and shown to have a similar secondary structure to the dimeric form by circular dichroism. It was necessary to confirm this as changes in

secondary structure may obviously lead to disruptions in tertiary structure and subsequent interference in NusE binding. Nooren and Thornton (2003), have observed that transient protein – protein interactions are characterised by the nature of the binding interface. Stable homodimers have large dimer interfaces (> 1000 $Å^2$), which are generally non-polar, less planar and frequently undergo conformational change after dissociation. Transient homodimers are characterised by smaller interfaces (500 $Å^2$ to 1000 $Å^2$) that are more polar and often planar. The NusB dimer lies somewhere between these two extremes with a large dimer interface (2100 $Å^2$) but interactions between the two subunits that are made up of both polar and non-polar interactions including 11 water molecules buried at the dimer interface. In addition, the monomeric mutant also showed little loss of secondary structure possibly indicating little conformational change on dimer dissociation. Overall, NusB thus shows characteristics associated more with weak homodimers, suggesting that antitermination in *M. tuberculosis* may thus require the dissociation of a weak NusB dimer by NusE binding.

In order to further examine this idea, pulldown experiments were carried out using tagged NusE which was refolded in the presence of either untagged NusB or FE22.23AA. This procedure was inspired by the observations of Luttgen *et al.* (2002) who obtained soluble NusB and soluble, stable NusE through co-expression of the two proteins. These refolding experiments (section 4.2.2.) again showed an interaction between NusE and NusB as well as between NusE and the monomeric NusB FE22.23AA. Unfortunately stoichiometry could not be determined from the resulting gels. The NusE band was more intense than the NusB or FE22.23AA bands signifying the presence of increased free amounts of NusE in the NusB/NusE and

FE22.23AA/NusE complex. As explained in section 4.2.2, the co-refolding results may suggest a 1:1 stochiometery as the band intensities seen in the NusB – NusE experiments were very similar to those seen with the FE22.23AA – NusE experiments. This may show that the mass of FE22.23AA (NusB monomer) binding NusE is similar to that of NusB wild-type binding NusE and hence NusB wild-type is interacting with NusE as a monomer and not as a dimer. FE22.23AA – NusE interactions appear weak (as seen with the increased amounts of NusE eluted) possibly because the mutated amino acids at the dimer interface may be involved in NusB – NusE binding. Thus equimolar elution of NusE and NusB wild-type may only occur when NusB wild-type is loaded in very high excess (much higher than the 3 times excess used in this experiment) to ensure saturation.

These experiments, although again revealing the presence of an interaction between NusB and NusE, were not conclusive regarding the NusB : NusE stoichiometry. NusE appears to be highly unstable after purification. Both Gopal *et al.* (2001b) and Luttgen *et al.* (2002) suffered similar problems with *M. tuberculosis* and *E. coli* purified NusE respectively. Luttgen *et al.* (2002) solved the problem through co-expression of NusB and NusE, whereupon NusE was found to be soluble. Similarly when *M. tuberculosis* NusE was co-expressed with NusB both proteins were found to be in the soluble fraction, in contrast to individual expression when both proteins were highly insoluble.

Two-step purification (glutathione sepharose and his-affinity columns) of co-expressed NusB and NusE again demonstrated an interaction between the two proteins. Similar NusB and NusE band intensities after the second round of affinity purification are again suggestive of a 1:1 stoichiometry. However, subsequent gel filtration purification

resulted in dissociation of the NusB – NusE complex possibly as a result of the dilution that occurs during gel filtration and the apparently weak interaction between NusB and NusE.

Taken together, these data suggest that NusB is probably a weak homodimer and may either bind in a 1 : 1 or 2 : 2 stoichiometry with NusE. The *E. coli* NusB – NusE interaction is heterodimeric (Luttgen *et al.*, 2002) suggesting that the *M. tuberculosis* NusB – NusE interaction is also heterodimeric. The observation of an apparent weak NusB – NusE interaction contrasts that found for *E. coli* (Mason *et al.*, 1992a). This may be explained by the fact NusE must compete with the NusB dimer for binding at the NusB dimer interface. Although the interaction of FE22.23AA with NusE showed similar characteristics to that of the NusB dimer – NusE interaction, it is possible that NusE binding partly utilises the amino acids at the dimer interface which were mutated to disrupt the NusB wildtype dimer.

Future work should involve the development of improved NusB – NusE co-expression constructs (allowing for pure and high concentrations of the NusB – NusE complex). The stoichiometry of the resultant NusB – NusE complex can then be determined and, equally importantly, leader region *rRNA* binding assays can be carried out. The NusB – NusE heterodimer binds at the boxA site in the *E. coli rrn* leader regions and the heterodimer is required for this interaction (Nodwell and Greenblatt, 1993). It would thus be interesting to compare this with the *M. tuberculosis* system.

6.3. The *M. tuberculosis* NusB dimer.

The determination of the dissociation equilibrium constant of the NusB dimer was important in establishing the importance of the dimer in antitermination. A K_d of 80 nM (8×10^{-8} M) was estimated for NusB self-association within the limitations of the gel filtration experiment.

The *E. coli* NusB monomer and its interaction with NusE have been well characterised. *E. coli* NusB shows no propensity to dimerise even at the low concentrations (1 mM) required for NMR structure determination (Altieri *et al.*, 2000) and the NusB – NusE interaction is heterodimeric (Mason *et al.*, 1992a). NusB from *E. coli* and *M. tuberculosis*, however, show structural homology (see Figure 34). Analysis of the *E. coli* NusB5 mutation (Y18D), which corresponds with the *M. tuberculosis* Phe22 mutated in this study, has shown the mutation to map to a cluster of aromatic residues that are partially solvent exposed in *E. coli* but are buried in the *M. tuberculosis* dimer interface (Altieri *et al.*, 2000; Friedman *et al.*, 1976; Gopal *et al.*, 2000). The effect of this mutation on NusE has not yet been determined but, should binding be affected, it may be as a result of the Y18D mutation causing a change in overall *E. coli* NusB structure or it may simply perturb the NusE binding interface. The binding of *E. coli* NusB – NusE also appears to be tighter than that seen in *M. tuberculosis*.

Results in this study again show that *M. tuberculosis* NusB – NusE binding may have a stoichiometry of either 1 : 1 (or 2 : 2) and that binding is likely to be weaker than that seen from *E. coli*. A model for *M. tuberculosis* NusB – NusE association would thus be that a component of overall NusE binding may involve a surface that is partially

engaged in NusB : NusB interactions. This would however suggest that the monomerised NusB would bind more tightly to NusE (see refolding experiments, section 4.2.2) however since the mutations are, by definition, at the dimerisation surface, it may be difficult to determine the relative contributions of each of the dimerisation *vs* competition *vs* NusE binding effects to those effects observed in the experiments. Thus the similarities seen between the affinities of the NusB dimer and NusE and the affinities of the NusB monomer and NusE may be due to the contribution of two opposing but equal effects (such as dimerisation and competition).

In the light of the above discussion, a very useful experiment would be the determination of cellular concentrations of NusB in *M. tuberculosis* giving clues as to the concentration of monomeric NusB within the cell. An additional future experiment involves the determination of whether *M. tuberculosis* NusB is capable of complementing an *E. coli* NusB mutant, the cold sensitive *NusB5* *E. coli* strain (Taura *et al.*, 1992). The results would thus demonstrate whether the two NusB proteins are functionally homologous and would be a measure of similarity in antitermination mechanisms between the two microorganisms.

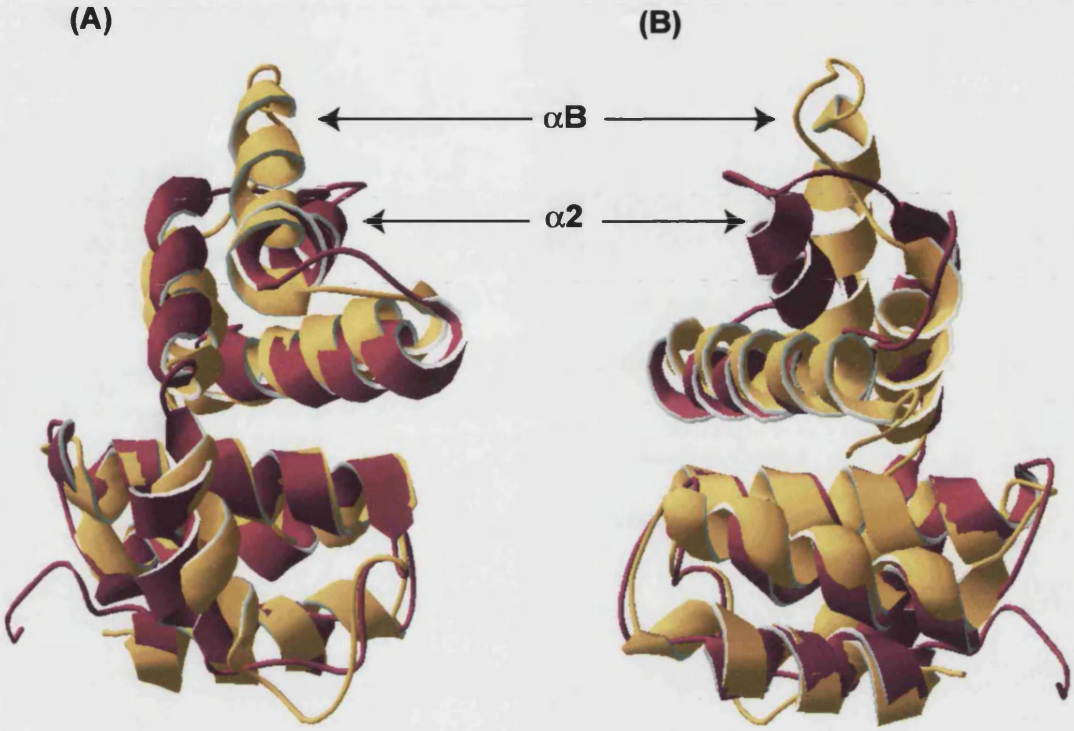


Figure 34 : Structural homology between the NusB molecules of *M. tuberculosis* and *E. coli*.
The overlapping *E. coli* (purple) and *M. tuberculosis* NusB (gold) structures are shown. Homology can be seen in six of the seven α-helices. Both NusB molecules are α-helical proteins made up of seven α-helices each. The two structures show an r.m.s deviation of 2 Å. **(A)** shows the two proteins from a front on view (as seen in the NusB dimer structure in Figure 8) **(B)** shows the homologous structures from the *M. tuberculosis* NusB dimer interface plane (Alteiri *et al.*, 2000 Gopal *et al.*, 2000, Guex and Peitsch, 1997; Schwede *et al.*, 2003).

6.4. *M. tuberculosis* NusG

In comparison with NusG from *E. coli*, little is known about the mycobacterial counterpart. NusG plays a vital role in transcriptional regulation, both in termination and antitermination. NusG is believed to increase the transcriptional elongation rate by approximately 20% when present, yet is required for termination at certain Rho dependent terminators (Burova *et al.*, 1995; Li *et al.*, 1993). These functions of NusG are facilitated by its ability to bind tightly to Rho and weakly to RNA polymerase (Li *et al.*, 1992; Pasman and Von Hippel, 2000). Its role in antitermination is not fully understood but it is known to be required for the increased stability of the antitermination complex (Mason and Greenblatt, 1991). Furthermore the ability of NusG to increase the rate of RNA polymerase elongation may be important in the read-through of terminators. The switch between NusG's role in termination and antitermination may be dependent on the presence of boxA in the transcript (Zellars and Squires, 1999). The recently determined X-ray structure of NusG from *A. aeolicus* has been important in defining the domains involved in protein – protein and RNA interactions and alludes to the ability of NusG to act as a bridge between nucleic acids and other proteins within the complex (Steiner *et al.*, 2002).

In vitro and *in silico* characterisation of the *M. tuberculosis* NusG protein was undertaken in order to shed some light on its role in antitermination. Biophysical and biochemical techniques were used to investigate the structural characteristics of *M. tuberculosis* NusG allowing a comparison with the *A. aeolicus* structure and the *E. coli* homologous model proposed by Steiner *et al.* (2002).

M. tuberculosis NusG was shown to be a monomer using sedimentation equilibrium ultracentrifugation and is likely to be an elongated protein as shown with dynamic light scattering experiments. Circular dichroism experiments showed some secondary structure (approximately 22 % α-helical content and some β-sheet content) and a degree of tertiary structure. Proteolytic experiments identified a stable domain found at the C-terminal end (amino acids 151 – 238). *M. tuberculosis* is therefore likely to be an elongated protein with a structured C-terminal domain potentially involved in interactions with RNAP, Rho or *rRNA*. An N-terminal domain was not identified but is likely to occur. A small amount of NusG remained uncleaved after chymotrypsin treatment and this may indicate the presence of some structure or domain organisation at the N-terminal end as well.

Further sequence analysis of *M. tuberculosis* NusG and comparisons with *E. coli* and *A. aeolicus* NusG revealed 41 non-conserved residues at the N-terminal end of the *M. tuberculosis* NusG. The function of these residues is unclear. The region from position 46 through to the end of the protein is highly conserved and contains two sub regions; the NusG N-terminal (NGN) homology domain (positions 46 – 154) and the KOW motif (positions 184 – 211). The NGN domain appears to be found in all Spt5p and NusG homologues in prokaryotes and eukaryotes (Ponting, 2002). The NGN motif is nearly always associated with a KOW motif (found C-terminal to the NGN domain). *E. coli* TraB and ActX proteins, the *Salmonella typhimurium* TraB and *Serratia entomophilia* AnfA1 proteins are exceptions to this in that no KOW motif is associated with the NGN region. In all of these cases there is a C-terminal sequence present that may contain a cryptic KOW domain or a KOW-like domain. Both *A. aeolicus* and *E.*

coli NusG contain the NGN domain, a gap of approximately 10 residues followed by the KOW domain. *M. tuberculosis* NusG however contains a gap of 30 residues between the NGN and KOW domains. There may thus be a possible role for this region within the *M. tuberculosis* NusG which does not occur in either *A. aeolicus* or *E. coli* NusG. This gapped arrangement of the NGN and KOW domains does occur in a few other proteins (such as NusG from *Mycoplasma gallisepticum* and *Corynebacterium glutamicum*; sequence analysis done using SMART; http://smart.embl-heidelberg.de/; Letunic *et al.*, 2002; Ponting *et al.*, 1999; Schultz *et al.*, 1998; Schultz *et al*; 2000).

Further comparisons of the three NusG molecules were made using structure prediction and the *A. aeolicus* NusG structure and *E. coli* and *M. tuberculosis* NusG sequence homology (using Swiss-Model; Guex and Peitsch, 1997; Schwede *et al.*, 2003). Sequence homology of the *E. coli* NusG with the *A. aeolicus* NusG structure predicts it to be an elongated protein made up of two domains joined by a linker region that is likely to be highly flexible (Steiner *et al.*, 2002). The two domains correspond to domains I and III in the *A. aeolicus* NusG (Figure 35a). The N-terminal domain (corresponding to domain I) is likely to contain the NGN motif and the C-terminal domain (corresponding to domain III) the KOW motif. The C-terminus sequence of *M. tuberculosis* has a high degree of predicted structural homology with domain III of *A. aeolicus* NusG. As this region contains the conserved KOW motif, it is likely to be a domain important in both RNA and protein – protein interaction, hence the high degree of conservation. It is however probable that a domain also exists at the N-terminus of *M. tuberculosis* NusG (containing the NGN motif) with similar functions to those seen in domain I of *A. aeolicus* NusG. Overall, the structure of *M. tuberculosis* NusG is likely to

be similar to that of the predicted *E. coli* structure with two domains joined by a flexible linker region. The topology of the *E. coli* and *M. tuberculosis* structure predictions, based on sequence homology, are shown in Figures 35b and 35c.

An understanding of the structure, properties and functions of NusG will greatly enhance the understanding of both termination and antitermination. The ability of NusG to directly affect the rate of RNA polymerase elongation and its strong interaction with the Rho termination factor indicates that it has important functions in both termination and antitermination. Future work should involve the determination of the structure of *M. tuberculosis* NusG which would give vital clues as to possible functions. In addition, it is important to investigate the interactions NusG may make with Rho and with RNA polymerase and to explore interactions with other proteins which may occur in the antitermination complex.

Figure 35a : Structural homology between *A. aeolicus* NusG and NusG from *E. coli* and *M. tuberculosis*.
(A) The structure of *A. aeolicus* NusG is shown in red with structural homologous regions of *E. coli* NusG shown in blue. *E.coli* NusG appears to contain 2 domains with homology to domains I and III of the *A. aeolicus* NusG. The region showing homology with domain II is likely to loop back into the region homologous with domain I.
(B) The structure of *A. aeolicus* NusG is shown in red with structural homologous regions of *M. tuberculosis* NusG shown in blue. Only the C-terminal region of *M. tuberculosis* NusG shows structural homology with domian III of the *A. aeolicus* NusG.
(Guex and Peisch, 1997; Schwede *et al*, 2003; Steiner *et al*, 2002).

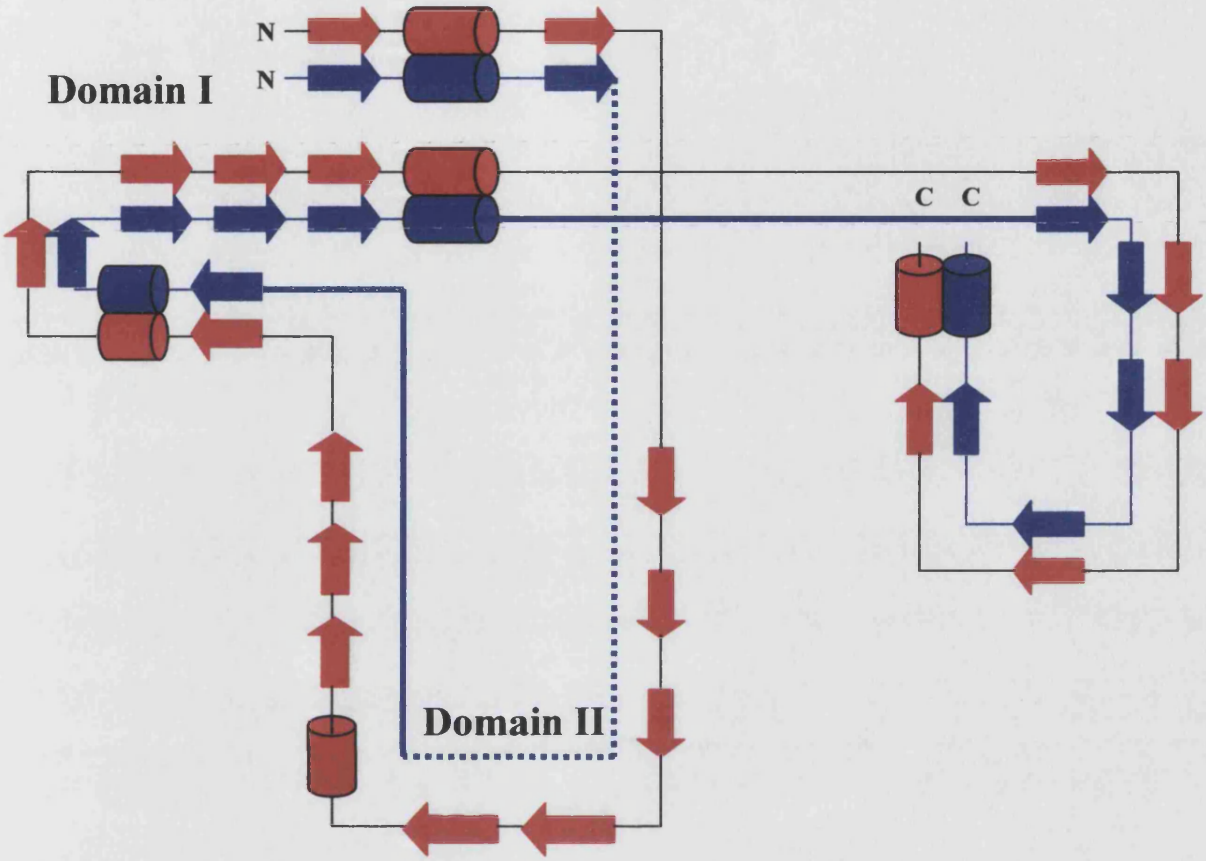


Figure 35b : Topology of the *A. aeolicus* NusG structure compared with that predicted for NusG from *E. coli*.
The *A. aeolicus* topology is shown in red and the *E. coli* in blue (as in Figure 35a). *E. coli* NusG is made up of two domains homologous to domains I and III of *A. aeolicus* NusG. The dotted line is a region that did not map to the *A. aeolicus* NusG structure but is likely to simply loop back to domain I as predicted by Steiner *et al.* (2002)

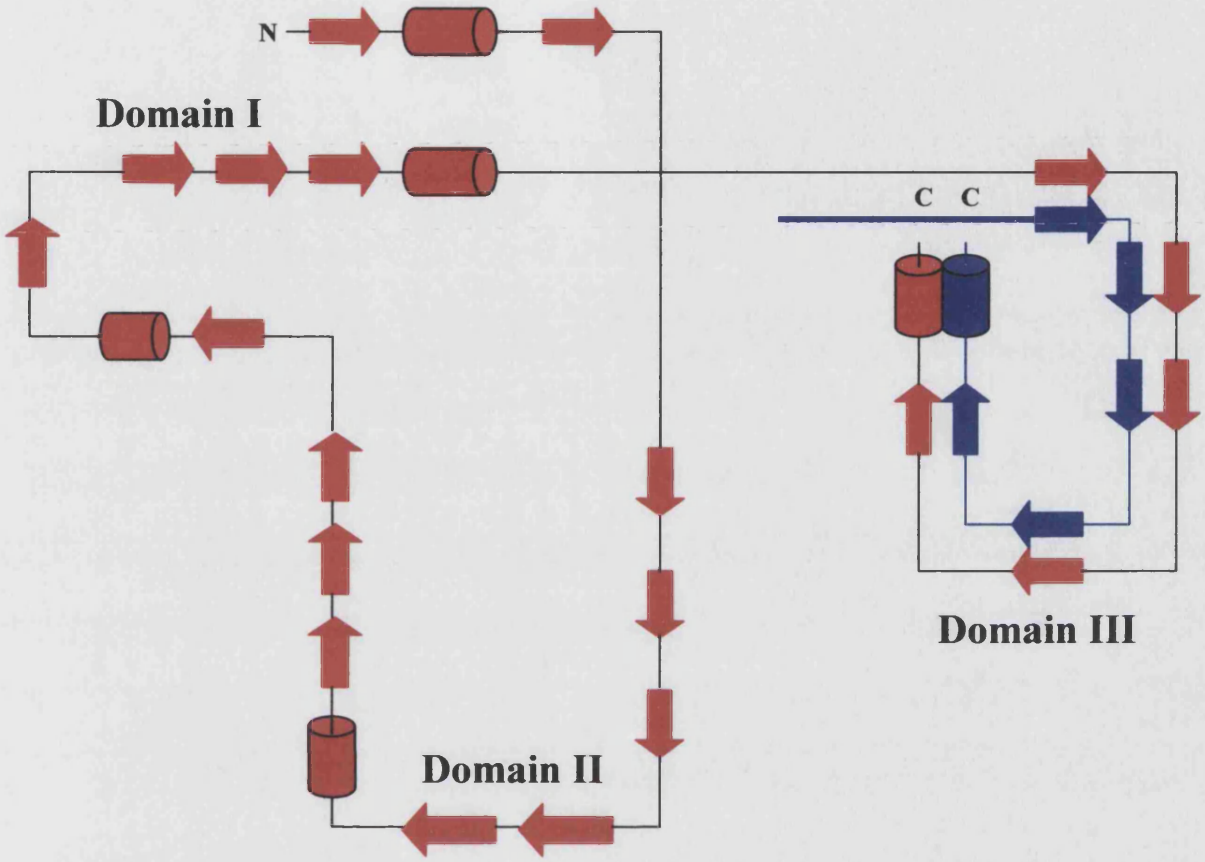


Figure 35c : Topology of the *A. aeolicus* NusG structure compared with that predicted for NusG from M. *tuberculosis*.
The *A. aeolicus* topology is shown in red and the *M. tuberculosis* in blue (as in Figure 35a). The *M. tuberculosis* NusG sequence only showed structural homology with domain III of the *A. aeolicus* NusG structure. An N-terminal domain is, however, likely to occur but does not necessarily show structural homology with that of domain I of *A. aeolicus* NusG.

6.5. An emerging model for *M. tuberculosis* antitermination.

Early studies of the *M. tuberculosis* antitermination complex have not yet yielded enough data to make a detailed comparison with the well documented *E. coli* model. Although *M. tuberculosis* is not closely related to *E. coli* (Fu and Fu-Liu, 2002), work carried out in this thesis does point towards a similar antitermination mechanism but with some distinct differences. The NusB dimer is probably disrupted and is likely to bind to NusE to form a heterodimer, corresponding with the *E. coli* model. The interaction between NusB and NusE appears to be weak, in contrast to the *E. coli* model. Additional factors (such as RNAP, other Nus factors or unknown factors) may therefore be involved in stabilising the NusB – NusE heterodimer. NusG seems to share structural characteristics with that of *E. coli* NusG and they are therefore likely to show functional similarities.

The knowledge obtained from work carried out in this thesis combined with the knowledge of the NusB structure and NusA structure and function (Gopal *et al.*, 2001a) does, however, allow for the tentative postulation of a model for *M. tuberculosis* antitermination (See Figure 35). NusA, NusB, NusE and NusG are all highly likely to be involved in an *rrn* antitermination mechanism occurring within *M. tuberculosis.* A NusB – NusE heterodimer and NusA are likely to make contacts with the *rRNA* transcript leader region while NusA and NusG probably make contacts with the RNA polymerase. This may be all that is necessary to convert the RNA polymerase into a terminator resistant form however the presence of Rho and the requirement for additional factors in *E. coli* antitermination suggest that they are very likely to be necessary for *M. tuberculosis* antitermination.

Confirmation of the proposed *M. tuberculosis rRNA* antitermination complex composition would initially involve the cloning of NusA, NusE, NusG and RNAPα into GST expression vectors and pulldown assays carried out as shown in this thesis using GST-NusB. The use of purified proteins may be used to confirm interactions seen in *E. coli* and *M. tuberculosis* cell free extracts used to identify additional factors required for complex formation. An alternative technique to pulldown assays would be the use of ELISA, where individual Nus factors or RNAP are crosslinked to a microtitre plate and interactions between the attached protein and the remaining individual factors and combinations of the remaining factors established by probing with antibodies to the added factors. Once interactions had been established, sedimentation equilibrium ultracentrifugation would be essential for stoichiometric predictions for all interactions.

rRNA nut site binding assays should also be carried out on all individual Nus factors and combinations of Nus factors in order to establish the RNA binding sites of individual proteins and the complex as a whole.

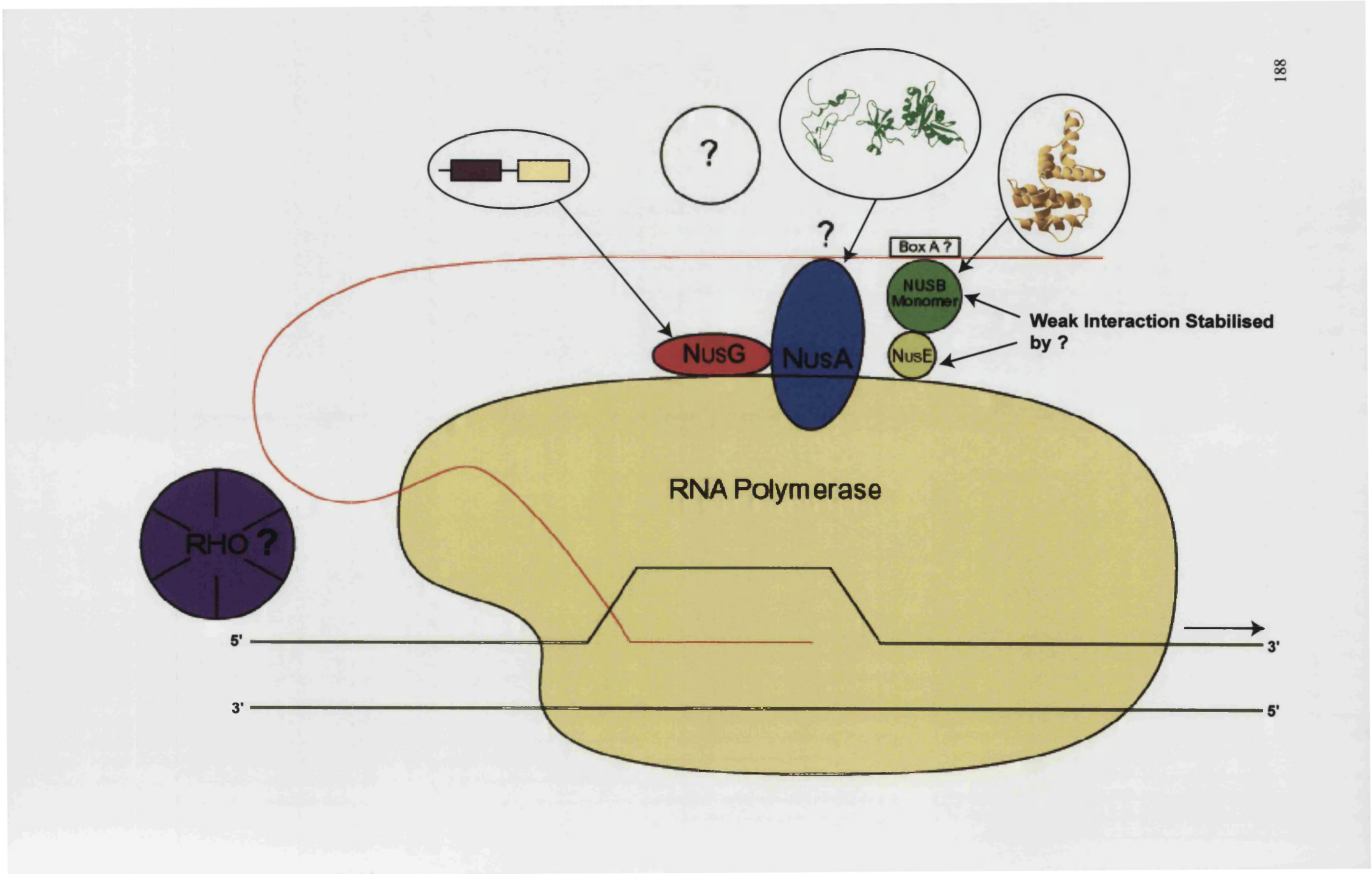
In conjunction with these experiments, *in vitro* transcription assays should be used to determine whether the identified components are capable of assembling and transcribing through Rho-dependent terminators. An approach used by Torres *et al.* (2000) may be useful for the determination of novel proteins required for *M. tuberculosis* antitermination. A modified *EcoR*1 enzyme (capable of specific DNA sequence binding but unable to cleave the DNA) was used to stall the transcribing RNA polymerase (in an *in vitro* transcription assay containing *E. coli* cell free extract) and subsequently allowed for the identification of the factors bound to the RNAP after having transcribed through an *rrn nut* site. This would be a very useful approach for confirming the requirements of

the Nus factors and identifying unknown antitermination components in *M. tuberculosis*. It is important that the mutant, monomeric NusB is used in all of the above experiments so as to fully establish whether NusB takes part in *M. tuberculosis* antitermination as a monomer.

The proposed model, involving the looping of the RNA allowing the Nus factors to maintain contact with both the RNAP and the *nut* site, could possibly be determined using an RNase protection assay on a stalled transcription complex. Should the *nut* sites remain protected from RNase digestion during such a scenario, this would be indicative of the looping of the RNA.

Figure 36 : Proposed Antitermination model in *M. tuberculosis*.

A possible model for antitermination shows the involvement of the four Nus proteins. An *in vitro* assay proving the involvement of Nus proteins in antitermination has not been carried out, however the homology of the Nus proteins to those found in *E. coli* is highly indicative of their involvement. The model also indicates a monomeric NusB interacting with NusE as well as the predicted weak interaction between NusB and NusE. Unknown factors that may stabilise the NusB – NusE interaction are indicated by a question mark. The elongated shape of NusG is shown. A question mark has been placed against Rho as it was not possible to show interactions with NusA and NusG. More thorough studies need to be carried out on the involvement of Rho in *M. tuberculosis* antitermination. Current studies involve the investigation of binding of the NusB – NusE complex with the *rRNA* boxA region and the *rRNA* binding site of NusA (both indicated by question marks). Future studies require the identification of additional factors, possibly involved in antitermination, as has and is occurring in *E. coli*; indicated by the question mark in the white circle (Squires *et al.*, 1993; Torres *et al.*, 2001). The structures of NusA and NusB (solved in this laboratory) and the domain organisation of NusG are indicated.


?
?
Box A ?
NUSB
Monomer
Weak Interaction Stabilised
by ?
NusG
NusA
NusE
RNA Polymerase
RHO ?
5'
3'
3'
5'

References

Web Pages

Format : Author Name, *Article Title*, Publication Date, Associated Institute, Access Date, **<Web Site Address>**

Fact Sheet 104 - Tuberculosis, 2002, World Health Organisation, 23 April 2003, **<www.who.int>**

Through the Microscope : An Internet Gallery of Health Science Images – Disease Carriers – Bacteria : Mycobacterium tuberculosis, New York State Department of Health 23 April 2003, **< www.wadsworth.org/databank/mycotubr.htm>**

COLE, S., *Tuberculist*, 2003, Institut Pasteur, 10 June2003, **< http://genolist.pasteur.fr/TubercuList/>**

ASHCROFT A.E., *An Introduction to Mass Spectroscopy*, 2002, Astbury Centre – University of Leeds, 15 July 2003, **< www.astbury.leeds.ac.uk/Facil/MStut/mstutorial.htm>**

SARREL M., *A History of Tuberculosis*, 2002, New Jersey State, USA, 21 April 2003, **< www.state.nj.us/health/cd/tbhistry.htm >**

STEYN L.H., *The Mycobacteria : Ducks of the Microbial World*, 1999, University of Cape Town, 24 April 2003, **<www.uct.ac.za/depts/mmi>**

Journal Articles and Books

(1998). *DynaPro MS and MSTC : Theory and Data Interpretation*, Protein Solutions Inc, Charlottesville.

Aksoy, S., Squires, C. L. & Squires, C. (1984). Evidence for antitermination in Escherichia coli RRNA transcription. *J Bacteriol* **159**, 260-4.

Allison, T. J., Wood, T. C., Briercheck, D. M., Rastinejad, F., Richardson, J. P. & Rule, G. S. (1998). Crystal structure of the RNA-binding domain from transcription termination factor rho. *Nat Struct Biol* **5**, 352-6.

Altieri, A. S., Mazzulla, M. J., Horita, D. A., Heath Coats, R., Wingfield, P. T., Das, A., Court, D. L. & Andrew Byrd, R. (2000). The structure of the transcriptional antiterminator NusB from Escherichia coli. *Nat Struct Biol* **7**, 470-4.

Artsimovitch, I. & Landick, R. (1998). Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. *Genes Dev* **12**, 3110-22.

Artsimovitch, I. & Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97**, 7090-5.

Bachmann, B. J. (1990). Linkage map of Escherichia coli K-12, edition 8. *Microbiol Rev* **54**, 130-97.

Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905-20.

Barik, S. & Das, A. (1990). An analysis of the role of host factors in transcription antitermination in vitro by the Q protein of coliphage lambda. *Mol Gen Genet* **222**, 152-6.

Berg, K. L., Squires, C. & Squires, C. L. (1989). Ribosomal RNA operon anti-termination. Function of leader and spacer region box B-box A sequences and their conservation in diverse micro-organisms. *J Mol Biol* **209**, 345-58.

Brennan, P. J. & Nikaido, H. (1995). The envelope of mycobacteria. *Annu Rev Biochem* **64**, 29-63.

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

Burns, C. M., Richardson, L. V. & Richardson, J. P. (1998). Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. *J Mol Biol* **278**, 307-16.

Burova, E., Hung, S. C., Sagitov, V., Stitt, B. L. & Gottesman, M. E. (1995). Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. *J Bacteriol* **177**, 1388-92.

Busby, S. & Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *J Mol Biol* **293**, 199-213.

Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry, Part II*, W.H. Freeman, San Francisco.

Carlomagno, M. S. & Nappo, A. (2001). The antiterminator NusB enhances termination at a sub-optimal Rho site. *J Mol Biol* **309**, 19-28.

Carlomagno, M. S. & Nappo, A. (2003). NusA modulates intragenic termination by different pathways. *Gene* **308**, 115-28.

Chan, E. D. & Iseman, M. D. (2002). Current medical treatment for tuberculosis. *Bmj* **325**, 1282-6.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Barrell, B. G. & et al. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537-44.

Coligan, E., Dunn, B. M., Ploegh, H. L., Speicher, D. W. & Wingfield, P. T. (2003). Current Protocols in Protein Science. In *Current Protocols*. John Wiley and Sons, Hoboken.

Condon, C., Squires, C. & Squires, C. L. (1995). Control of rRNA transcription in Escherichia coli. *Microbiological Reviews* **59**, 623-45.

Court, D. L., Patterson, T. A., Baker, T., Costantino, N., Mao, X. & Friedman, D. I. (1995). Structural and functional analyses of the transcription-translation proteins NusB and NusE. *J Bacteriol* **177**, 2589-91.

Dalboge, H., Bayne, S. & Pedersen, J. (1990). In vivo processing of N-terminal methionine in E. coli. *FEBS Lett* **266**, 1-3.

Darlix, J. L., Sentenac, A. & Fromageot, P. (1971). Binding of termination factor RHO to RNA polymerase and DNA. *FEBS Lett* **13**, 165-168.

Das, A. (1993). Control of transcription termination by RNA-binding proteins. *Annu Rev Biochem* **62**, 893-930.

DeVito, J. & Das, A. (1994). Control of transcription processivity in phage lambda: Nus factors strengthen the termination-resistant state of RNA polymerase induced by N antiterminator. *Proc Natl Acad Sci U S A* **91**, 8660-4.

Dolan, J. W., Marshall, N. F. & Richardson, J. P. (1990). Transcription termination factor rho has three distinct structural domains. *J Biol Chem* **265**, 5747-54.

Dombroski, A. J., Brennan, C. A., Spear, P. & Platt, T. (1988a). Site-directed alterations in the ATP-binding domain of rho protein affect its activities as a termination factor. *J Biol Chem* **263**, 18802-9.

Dombroski, A. J., LaDine, J. R., Cross, R. L. & Platt, T. (1988b). The ATP binding site on rho protein. Affinity labeling of Lys181 by pyridoxal 5'-diphospho-5'-adenosine. *J Biol Chem* **263**, 18810-5.

Dormandy, T. (1999). *The White Death : A History of Tuberculosis*, Hambledon Press, London.

Ebright, R. H. (1993). Transcription activation at Class I CAP-dependent promoters. *Mol Microbiol* **8**, 797-802.

Fatkenheuer, G., Taelman, H., Lepage, P., Schwenk, A. & Wenzel, R. (1999). The return of tuberculosis. *Diagn Microbiol Infect Dis* **34**, 139-46.

Friedman, A. M., Fischmann, T. O. & Steitz, T. A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science* **268**, 1721-7.

Friedman, D. I., Baumann, M. & Baron, L. S. (1976). Cooperative effects of bacterial mutations affecting lambda N gene expression. I. Isolation and characterization of a nusB mutant. *Virology* **73**, 119-27.

Friedman, D. I. & Court, D. L. (1995). Transcription antitermination: the lambda paradigm updated. *Molecular Microbiology* **18**, 191-200.

Friedman, D. I. & Gottesman, M. E. (1983). Lytic mode of lambda development. In *Lamda II* (Hendrix , R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A., eds.), pp. 21-51. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.

Friedman, D. I., Jolly, C. T. & Mural, R. J. (1973). Interference with the expression of the N gene function of phage lambda in a mutant of Escherichia coli. *Virology* **51**, 216-26.

Friedman, D. I., Schauer, A. T., Baumann, M. R., Baron, L. S. & Adhya, S. L. (1981). Evidence that ribosomal protein S10 participates in control of transcription termination. *Proc Natl Acad Sci U S A* **78**, 1115-8.

Fu, L. M. & Fu-Liu, C. S. (2002). Is Mycobacterium tuberculosis a closer relative to Gram-positive or Gram-negative bacterial pathogens? *Tuberculosis (Edinb)* **82**, 85-90.

Gentry, D., Xiao, H., Burgess, R. & Cashel, M. (1991). The omega subunit of Escherichia coli K-12 RNA polymerase is not required for stringent RNA control in vivo. *J Bacteriol* **173**, 3901-3.

Geoghegan, K. F., Dixon, H. B., Rosner, P. J., Hoth, L. R., Lanzetti, A. J., Borzilleri, K. A., Marr, E. S., Pezzullo, L. H., Martin, L. B., LeMotte, P. K., McColl, A. S., Kamath, A. V. & Stroh, J. G. (1999). Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal Biochem* **267**, 169-84.

Gill, S. C., Weitzel, S. E. & von Hippel, P. H. (1991). Escherichia coli sigma 70 and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. *J Mol Biol* **220**, 307-24.

Gollnick, P. & Babitzke, P. (2002). Transcription attenuation. *Biochim Biophys Acta* **1577**, 240-50.

Gonzalez-y-Merchand, J. A., Colston, M. J. & Cox, R. A. (1996). The rRNA operons of Mycobacterium smegmatis and Mycobacterium tuberculosis: comparison of promoter elements and of neighbouring upstream genes. *Microbiology* **142 (Pt 3)**, 667-74.

Gopal, B., Haire, L. F., Cox, R. A., Jo Colston, M., Major, S., Brannigan, J. A., Smerdon, S. J. & Dodson, G. (2000). The crystal structure of NusB from Mycobacterium tuberculosis. *Nat Struct Biol* **7**, 475-8.

Gopal, B., Haire, L. F., Gamblin, S. J., Dodson, E. J., Lane, A. N., Papavinasasundaram, K. G., Colston, M. J. & Dodson, G. (2001a). Crystal structure of the transcription elongation/anti-termination factor NusA from Mycobacterium tuberculosis at 1.7 A resolution. *J Mol Biol* **314**, 1087-95.

Gopal, B., Papavinasasundaram, K. G., Dodson, G., Colston, M. J., Major, S. A. & Lane, A. N. (2001b). Spectroscopic and thermodynamic characterization of the

transcription antitermination factor NusE and its interaction with NusB from Mycobacterium tuberculosis. *Biochemistry* **40**, 920-8.

Greenblatt, J. & Li, J. (1981a). Interaction of the sigma factor and the nusA gene protein of E. coli with RNA polymerase in the initiation-termination cycle of transcription. *Cell* **24**, 421-8.

Greenblatt, J. & Li, J. (1981b). The nusA gene protein of Escherichia coli. Its identification and a demonstration that it interacts with the gene N transcription anti-termination protein of bacteriophage lambda. *J Mol Biol* **147**, 11-23.

Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-23.

Gunsalus, R. P. & Yanofsky, C. (1980). Nucleotide sequence and expression of Escherichia coli trpR, the structural gene for the trp aporepressor. *Proc Natl Acad Sci U S A* **77**, 7117-21.

Haldenwang, W. G. (1995). The sigma factors of Bacillus subtilis. *Microbiol Rev* **59**, 1-30.

Henkin, T. M. (1996). Control of transcription termination in prokaryotes. *Annu Rev Genet* **30**, 35-57.

Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G. & Blanquet, S. (1989). Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A* **86**, 8247-51.

Holben, W. E. & Morgan, E. A. (1984). Antitermination of transcription from an Escherichia coli ribosomal RNA promoter. *Proc Natl Acad Sci U S A* **81**, 6789-93.

Horwitz, R. J., Li, J. & Greenblatt, J. (1987). An elongation control particle containing the N gene transcriptional antitermination protein of bacteriophage lambda. *Cell* **51**, 631-41.

Ito, K. & Nakamura, Y. (1996). Localization of nusA-suppressing amino acid substitutions in the conserved regions of the beta' subunit of Escherichia coli RNA polymerase. *Mol Gen Genet* **251**, 699-706.

Jin, D. J., Burgess, R. R., Richardson, J. P. & Gross, C. A. (1992). Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc Natl Acad Sci U S A* **89**, 1453-7.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202.

Kanai, K. (1991). *Introduction to Tuberculosis and Mycobacteria*, Tokyo Southeast Asian Medical Information Cente/International Medical Foundation of Japan, Tokyo.

Keener, J. & Nomura, M. (1996). Regulation of Ribosome Synthesis. In *Esherichia coli and Salmonella : Cellular and Molecular Biology* (Neidhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaetcher, M. & Umbarger, H. E., eds.), pp. 1417-1431. American Society for Molecular Biology, Washington, DC.

Kempsell, K. E., Ji, Y. E., Estrada, I. C., Colston, M. J. & Cox, R. A. (1992). The nucleotide sequence of the promoter, 16S rRNA and spacer region of the ribosomal RNA operon of Mycobacterium tuberculosis and comparison with Mycobacterium leprae precursor rRNA. *J Gen Microbiol* **138 (Pt 8)**, 1717-27.

Komine, Y., Adachi, T., Inokuchi, H. & Ozeki, H. (1990). Genomic organization and physical mapping of the transfer RNA genes in Escherichia coli K12. *J Mol Biol* **212**, 579-98.

Kustu, S., North, A. K. & Weiss, D. S. (1991). Prokaryotic transcriptional enhancers and enhancer-binding proteins. *Trends Biochem Sci* **16**, 397-402.

Kyrpides, N. C., Woese, C. R. & Ouzounis, C. A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem Sci* **21**, 425-6.

Landick, R. & Yanofsky, C. (1987). Isolation and structural analysis of the Escherichia coli trp leader paused transcription complex. *J Mol Biol* **196**, 363-77.

Lau, L. F., Roberts, J. W. & Wu, R. (1982). Transcription terminates at lambda tR1 in three clusters. *Proc Natl Acad Sci U S A* **79**, 6171-5.

Laue, T. M. & Stafford, W. F., 3rd. (1999). Modern applications of analytical ultracentrifugation. *Annu Rev Biophys Biomol Struct* **28**, 75-100.

Lebowitz, J., Lewis, M. S. & Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci* **11**, 2067-79.

Letunic, I., Goodstadt, L., Dickens, N. J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R. R., Ponting, C. P. & Bork, P. (2002). Recent

improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* **30**, 242-4.

Li, J., Horwitz, R., McCracken, S. & Greenblatt, J. (1992). NusG, a new Escherichia coli elongation factor involved in transcriptional antitermination by the N protein of phage lambda. *J Biol Chem* **267**, 6012-9.

Li, J., Mason, S. W. & Greenblatt, J. (1993). Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. *Genes Dev* **7**, 161-72.

Li, S. C., Squires, C. L. & Squires, C. (1984). Antitermination of E. coli rRNA transcription is caused by a control region segment containing lambda nut-like sequences. *Cell* **38**, 851-60.

Linn, T. & Greenblatt, J. (1992). The NusA and NusG proteins of Escherichia coli increase the in vitro readthrough frequency of a transcriptional attenuator preceding the gene for the beta subunit of RNA polymerase. *J Biol Chem* **267**, 1449-54.

Liu, K. & Hanna, M. M. (1995). NusA contacts nascent RNA in Escherichia coli transcription complexes. *J Mol Biol* **247**, 547-58.

Liu, K., Zhang, Y., Severinov, K., Das, A. & Hanna, M. M. (1996). Role of Escherichia coli RNA polymerase alpha subunit in modulation of pausing, termination and anti-termination by the transcription elongation factor NusA. *Embo J* **15**, 150-61.

Loewen, P. C. & Hengge-Aronis, R. (1994). The role of the sigma factor sigma S (KatF) in bacterial global regulation. *Annu Rev Microbiol* **48**, 53-80.

Lund, E. & Dahlberg, J. E. (1977). Spacer transfer RNAs in ribosomal RNA transcripts of E. coli: processing of 30S ribosomal RNA in vitro. *Cell* **11**, 247-62.

Luttgen, H., Robelek, R., Muhlberger, R., Diercks, T., Schuster, S. C., Kohler, P., Kessler, H., Bacher, A. & Richter, G. (2002). Transcriptional regulation by antitermination. Interaction of RNA with NusB protein and NusB/NusE protein complex of Escherichia coli. *J Mol Biol* **316**, 875-85.

Mah, T. F., Li, J., Davidson, A. R. & Greenblatt, J. (1999). Functional importance of regions in Escherichia coli elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. *Mol Microbiol* **34**, 523-37.

Martin, S. R. & Bayley, P. M. (2002). Absorption and circular dichroism spectroscopy. *Methods Mol Biol* **173**, 43-55.

Mason, S. W. & Greenblatt, J. (1991). Assembly of transcription elongation complexes containing the N protein of phage lambda and the Escherichia coli elongation factors NusA, NusB, NusG, and S10. *Genes Dev* **5**, 1504-12.

Mason, S. W., Li, J. & Greenblatt, J. (1992a). Direct interaction between two Escherichia coli transcription antitermination factors, NusB and ribosomal protein S10. *J Mol Biol* **223**, 55-66.

Mason, S. W., Li, J. & Greenblatt, J. (1992b). Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage lambda. *J Biol Chem* **267**, 19418-26.

Mims, C. A., Playfair, J. H., Roitt, I. M., Wakelin, D., Williams, R. & Anderson, R. M. (1993). *Medical Microbiology*, Mosby, St Louis.

Modrak, D. & Richardson, J. P. (1994). The RNA-binding domain of transcription termination factor rho: isolation, characterization, and determination of sequence limits. *Biochemistry* **33**, 8292-9.

Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E. & Greenblatt, J. (1998). Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Mol Cell* **1**, 265-75.

Mogridge, J., Mah, T. F. & Greenblatt, J. (1995). A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein. *Genes & Development* **9**, 2831-45.

Morgan, W. D., Bear, D. G. & von Hippel, P. H. (1983). Rho-dependent termination of transcription. II. Kinetics of mRNA elongation during transcription from the bacteriophage lambda PR promoter. *J Biol Chem* **258**, 9565-74.

Mukherjee, K. & Chatterji, D. (1997). Studies on the omega subunit of Escherichia coli RNA polymerase--its role in the recovery of denatured enzyme activity. *Eur J Biochem* **247**, 884-9.

Nick, H. & Gilbert, W. (1985). Detection in vivo of protein-DNA interactions within the lac operon of Escherichia coli. *Nature* **313**, 795-8.

Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-30.

Nodwell, J. R. & Greenblatt, J. (1991). The nut site of bacteriophage lambda is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. *Genes Dev* **5**, 2141-51.

Nodwell, J. R. & Greenblatt, J. (1993). Recognition of boxA antiterminator RNA by the E. coli antitermination factors NusB and ribosomal protein S10. *Cell* **72**, 261-8.

Noller, H. F. (1991). Ribosomal RNA and translation. *Annu Rev Biochem* **60**, 191-227.

Nooren, I. M. & Thornton, J. M. (2003). Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol* **325**, 991-1018.

Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-17.

Nudler, E. (1999). Transcription elongation: structural basis and mechanisms. *J Mol Biol* **288**, 1-12.

Nudler, E., Avetissova, E., Markovtsov, V. & Goldfarb, A. (1996). Transcription processivity: protein-DNA interactions holding together the elongation complex. *Science* **273**, 211-7.

Olson, E. R., Flamm, E. L. & Friedman, D. I. (1982). Analysis of nutR: a region of phage lambda required for antitermination of transcription. *Cell* **31**, 61-70.

Olson, E. R., Tomich, C. S. & Friedman, D. I. (1984). The nusA recognition site. Alteration in its sequence or position relative to upstream translation interferes with the action of the N antitermination function of phage lambda. *J Mol Biol* **180**, 1053-63.

Opalka, N., Mooney, R. A., Richter, C., Severinov, K., Landick, R. & Darst, S. A. (2000). Direct localization of a beta-subunit domain on the three-dimensional structure of Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A* **97**, 617-22.

Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335**, 321-9.

Pasman, Z. & von Hippel, P. H. (2000). Regulation of rho-dependent transcription termination by NusG is specific to the Escherichia coli elongation complex. *Biochemistry* **39**, 5573-85.

Perry, R. P. (1976). Processing of RNA. *Annu Rev Biochem* **45**, 605-29.

Ponting, C. P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res* **30**, 3643-52.

Ponting, C. P., Schultz, J., Milpetz, F. & Bork, P. (1999). SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res* **27**, 229-32.

Ralston, G. (1993). *Introduction to Analytical Ultracentrifugation*, Beckman Instruments Inc, Fullerton.

Record, M. T., Reznikoff, W. S., Craig, M. L., McQuade, K. L. & Schlax, P. J. (1996). Escherichia coli RNA polymerase, Promoters, and the Kinetics of the Steps of Transcription Initiation. In *Escherichia coli and Salmonella : Cellular and Molecular Biology* (Neidhardt, F. C., Curtiss, R., Ingham, C. J., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaetcher, M. &

Umbarger, H. E., eds.), pp. 792-821. American Society for Microbiology, Washington, DC.

Richardson, J. P. (2002). Rho-dependent termination and ATPases in transcript termination. *Biochim Biophys Acta* **1577**, 251-260.

Richardson, J. P. (2003). Loading Rho to terminate transcription. *Cell* **114**, 157-9.

Richardson, J. P. & Greenblatt, J. (1996). Control of RNA Chain Elongation and Termination. In *Esherichia coli and Salmonella : Cellular and Molecular Biology* (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaetcher, M., Umbarger, H. E., Lin, E. C. C., Reznikoff, W. S. & Riley, M., eds.), pp. 822-848. American Society for Microbiology, Washington, DC.

Riggs, A. D., Newby, R. F. & Bourgeois, S. (1970). lac repressor--operator interaction. II. Effect of galactosides and other ligands. *J Mol Biol* **51**, 303-14.

Roberts, J. W. (1969). Termination factor for RNA synthesis. *Nature* **224**, 1168-1174.

Roland, K. L., Powell, F. E. & Turnbough, C. L., Jr. (1985). Role of translation and attenuation in the control of pyrBI operon expression in Escherichia coli K-12. *J Bacteriol* **163**, 991-9.

Rosenberg, M., Court, D., Shimatake, H., Brady, C. & Wulff, D. L. (1978). The relationship between function and DNA sequence in an intercistronic regulatory region in phage lambda. *Nature* **272**, 414-23.

Salstrom, J. S. & Szybalski, W. (1978). Coliphage lambdanutL-: a unique class of mutants defective in the site of gene N product utilization for antitermination of leftward transcription. *J Mol Biol* **124**, 195-221.

Schmidt, M. C. & Chamberlin, M. J. (1984). Binding of rho factor to Escherichia coli RNA polymerase mediated by nusA protein. *J Biol Chem* **259**, 15000-2.

Schmidt, M. C. & Chamberlin, M. J. (1987). nusA protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. *J Mol Biol* **195**, 809-18.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* **28**, 231-4.

Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**, 5857-64.

Schuster, T. M. & Toedt, J. M. (1996). New revolutions in the evolution of analytical ultracentrifugation. *Curr Opin Struct Biol* **6**, 650-8.

Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* **31**, 3381-5.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U. & Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol* **8**, 27-31.

Sharp, J. A., Galloway, J. L. & Platt, T. (1983). A kinetic mechanism for the poly(C)-dependent ATPase of the Escherichia coli transcription termination protein, rho. *J Biol Chem* **258**, 3482-6.

Sharrock, R. A., Gourse, R. L. & Nomura, M. (1985). Defective antitermination of rRNA transcription and derepression of rRNA and tRNA synthesis in the nusB5 mutant of Escherichia coli. *Proc Natl Acad Sci U S A* **82**, 5275-9.

Shiba, K., Ito, K. & Yura, T. (1986). Suppressors of the secY24 mutation: identification and characterization of additional ssy genes in Escherichia coli. *J Bacteriol* **166**, 849-56.

Skordalakes, E. & Berger, J. M. (2003). Structure of the rho transcription terminator. Mechanism of mRNA recognition and helicase loading. *Cell* **114**, 135-46.

Sorensen, M. A., Jensen, K. F. & Pedersen, S. (1994). High concentrations of ppGpp decrease the RNA chain growth rate. Implications for protein synthesis and translational fidelity during amino acid starvation in Escherichia coli. *J Mol Biol* **236**, 441-54.

Squires, C. L., Greenblatt, J., Li, J. & Condon, C. (1993). Ribosomal RNA antitermination in vitro: requirement for Nus factors and one or more unidentified cellular components. *Proc Natl Acad Sci U S A* **90**, 970-4.

Squires, C. L., Lee, F. D. & Yanofsky, C. (1975). Interaction of the trp repressor and RNA polymerase with the trp operon. *J Mol Biol* **92**, 93-111.

Squires, C. L. & Zaporojets, D. (2000). Proteins shared by the transcription and translation machines. *Annu Rev Microbiol* **54**, 775-98.

Steiner, T., Kaiser, J. T., Marinkovic, S., Huber, R. & Wahl, M. C. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *Embo J* **21**, 4641-53.

Taura, T., Ueguchi, C., Shiba, K. & Ito, K. (1992). Insertional disruption of the nusB (ssyB) gene leads to cold-sensitive growth of Escherichia coli and suppression of the secY24 mutation. *Mol Gen Genet* **234**, 429-32.

Torres, M., Condon, C., Balada, J. M., Squires, C. & Squires, C. L. (2001). Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. *Embo J* **20**, 3811-20.

Van Gilst, M. R. & von Hippel, P. H. (1997). Assembly of the N-dependent antitermination complex of phage lambda: NusA and RNA bind independently to different unfolded domains of the N protein. *J Mol Biol* **274**, 160-73.

van Holde, K. E., Curtis Johnson, W. & Shing Ho, P. (1998). *Principals of Physical Biochemistry*, Prentice-Hall, Upper Saddle River, New Jersey.

Vogel, U. & Jensen, K. F. (1994a). Effects of guanosine 3',5'-bisdiphosphate (ppGpp) on rate of transcription elongation in isoleucine-starved Escherichia coli. *J Biol Chem* **269**, 16236-41.

Vogel, U. & Jensen, K. F. (1994b). The RNA chain elongation rate in Escherichia coli depends on the growth rate. *J Bacteriol* **176**, 2807-13.

Vogel, U. & Jensen, K. F. (1995). Effects of the antiterminator BoxA on transcription elongation kinetics and ppGpp inhibition of transcription elongation in Escherichia coli. *J Biol Chem* **270**, 18335-40.

Vogel, U. & Jensen, K. F. (1997). NusA is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. *Journal of Biological Chemistry* **272**, 12265-71.

Warren, F. & Das, A. (1984). Formation of termination-resistant transcription complex at phage lambda nut locus: effects of altered translation and a ribosomal mutation. *Proc Natl Acad Sci U S A* **81**, 3612-6.

Wilkinson, D. (2000). A Cut Above. Limited proteolysis provides a wealth of protein structural information. *The Scientist* **14**, 1-4.

Wilm, M. (2000). Mass spectrometric analysis of proteins. *Adv Protein Chem* **54**, 1-30.

Wilson, K. & Walker, J. (1994). *Principles and Techniques of Practical Biochemistry.* 4th edit, Cambridge University Press, Cambridge.

Worbs, M., Bourenkov, G. P., Bartunik, H. D., Huber, R. & Wahl, M. C. (2001). An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Molecular Cell* **7**, 1177-89.

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751-8.

Yarnell, W. S. & Roberts, J. W. (1999). Mechanism of intrinsic transcription termination and antitermination. *Science* **284**, 611-5.

Yu, X., Horiguchi, T., Shigesada, K. & Egelman, E. H. (2000). Three-dimensional reconstruction of transcription termination factor rho: orientation of the N-terminal domain and visualization of an RNA-binding site. *J Mol Biol* **299**, 1279-87.

Yura, T., Nagai, H. & Mori, H. (1993). Regulation of the heat-shock response in bacteria. *Annu Rev Microbiol* **47**, 321-50.

Zellars, M. & Squires, C. L. (1999). Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. *Mol Microbiol* **32**, 1296-304.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* **98**, 811-24.

Zhou, Y., Mah, T. F., Greenblatt, J. & Friedman, D. I. (2002). Evidence that the KH RNA-binding domains influence the action of the E. coli NusA protein. *J Mol Biol* **318**, 1175-88.

Zubay, G. (1993). *Biochemistry*. 3rd edit, Wm. C. Brown Publishers, Dubuque, IA.

Appendices

Appendix 1.

Site-Directed-Mutagenesis Primers:

FE22.23AA mutations of Phe and Glu to Ala are highlighted.

Upper Primer **A:**

G GCC CTG CTG **GCC GCT** GCC GAG GTC CGC GGC

Lower Primer **B:**

GCC GCG GAC CTC GGC **AGC GGC** CAG CAG GGC C

PCR Primers:

NusB-GST Forward Primer **C:**

GATCGGATCCATGTCGGACAGAAAGCCGGTTCGCG

NusB-GST Reverse Primer **D:**

GATCGAATTCGACGCGCAGCTCGAGCCGGGTCATCA

NusE + **RBS** Forward Primer **E:**

GATCGAATTCA**TTTCACACAGGAAACAGTATTC**ATGGCGGGACAGAAGAT CCGC

NusE + RBS Reverse Primer **F:**

GATCGAATTCCCCTTTCGTGCCATTGCTCTGTCCAATCTCCTA

NusB – RBS – NusE – His_6 Forward Primer **G** :

GATCGGATCCATGTCGGACAGAAAGCCGGTTCGC

NusB – RBS – NusE – **His_6** Reverse Primer **H** :

GATCGGATCCTCA**GTGGTGGTGGTGGTGGTG**CTCCTGGATGTTGACGTCG AC

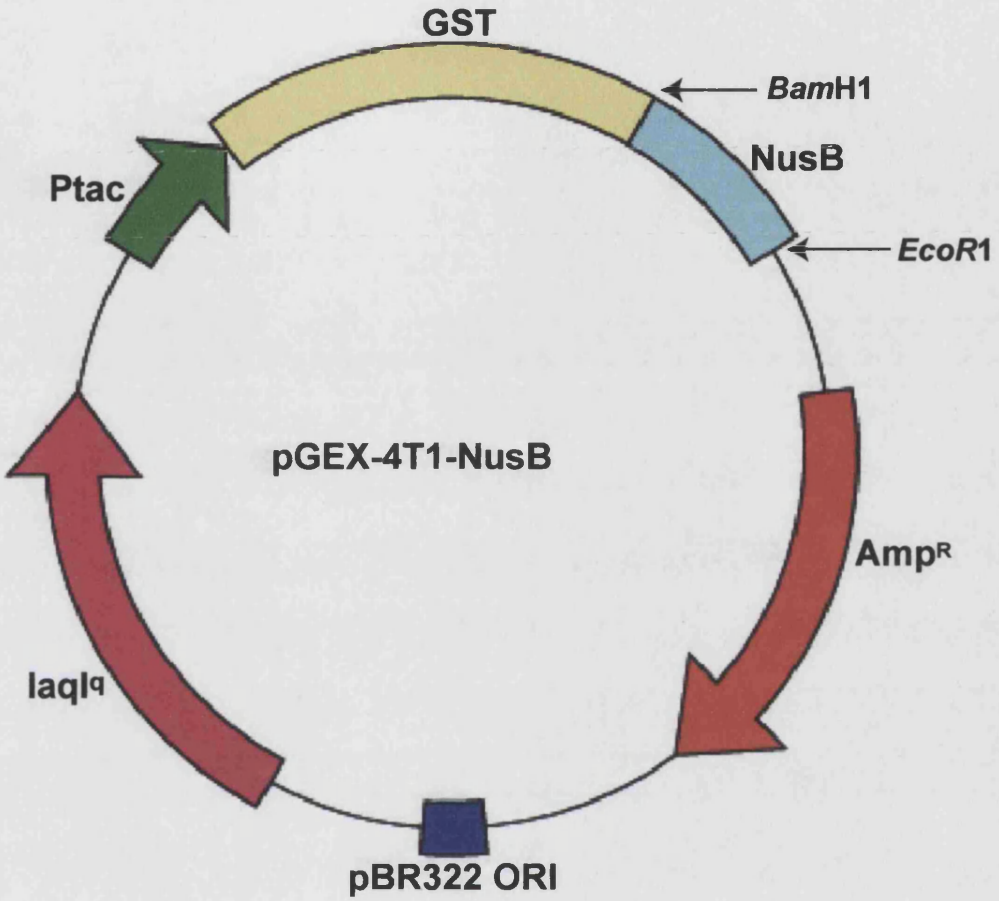
Rho Forward Primer **I:**

CCGGAATTCCATATGACCGATACGGACCTCATTACGGCT

Reverse Primer **J:**

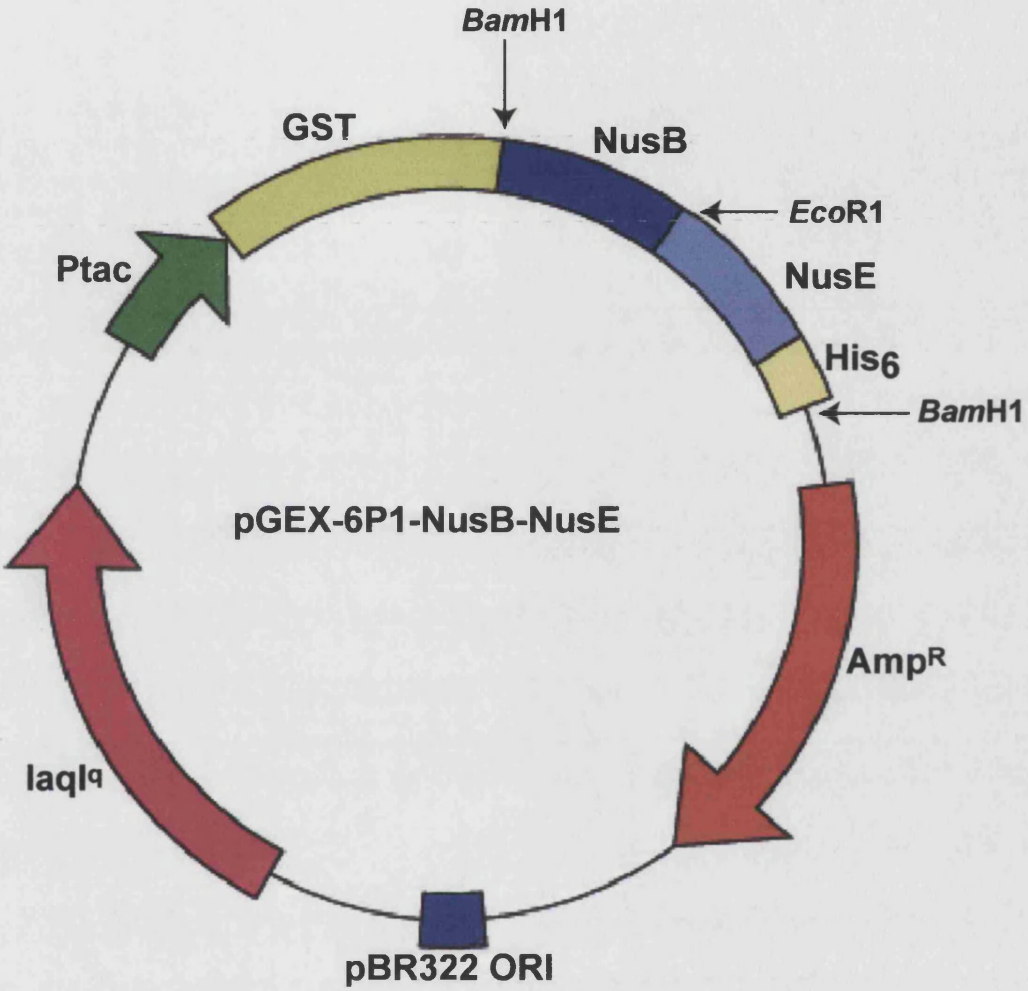
CGCCAAGCTTTCACTCGAGGTCGCTGTCCATGGACCCTGGCGT

Appendix 2a.



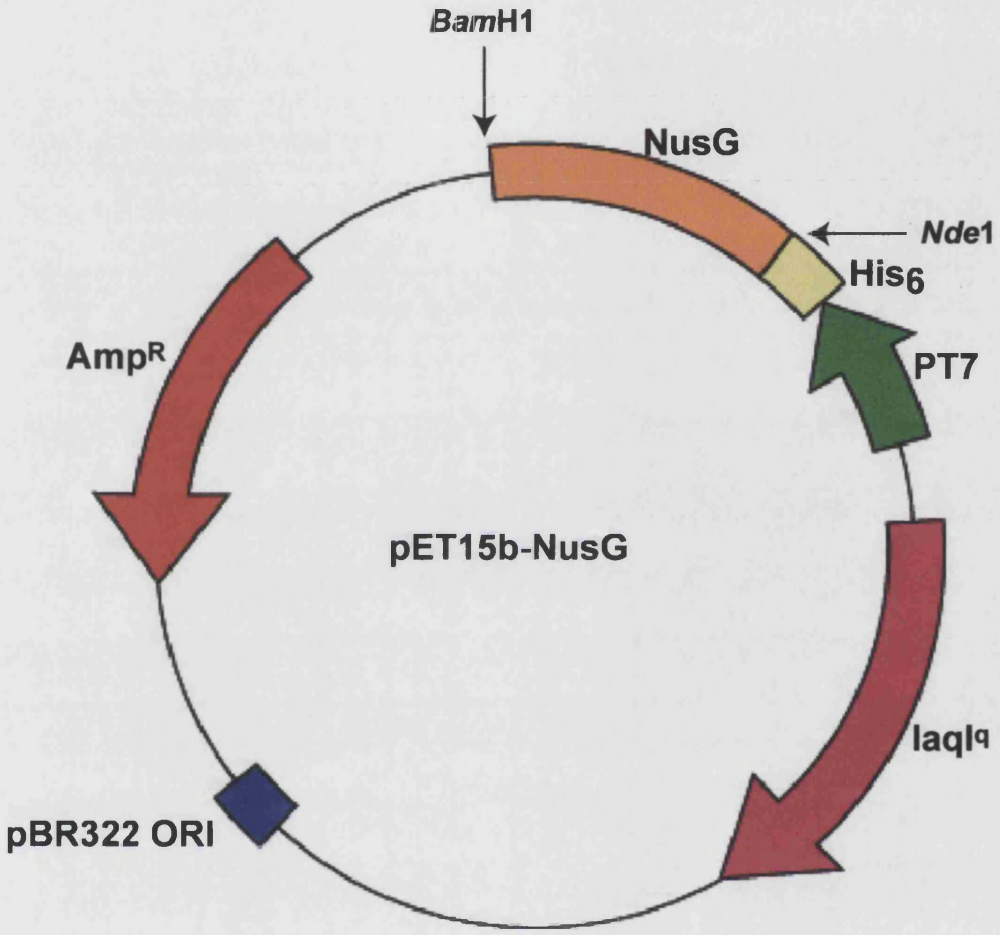
pGEX-4T1 was used for the cloning and overexpression of an N-terminally-GST tagged *M. tuberculosis* NusB. The resulting protein is 43.0 kDa, with a theoretical pI of 6.33 and an extinction coeffeciant of 53580 M^{-1} cm^{-1}. The N-terminally fused GST can be removed by thrombin cleavage.

Appendix 2b.



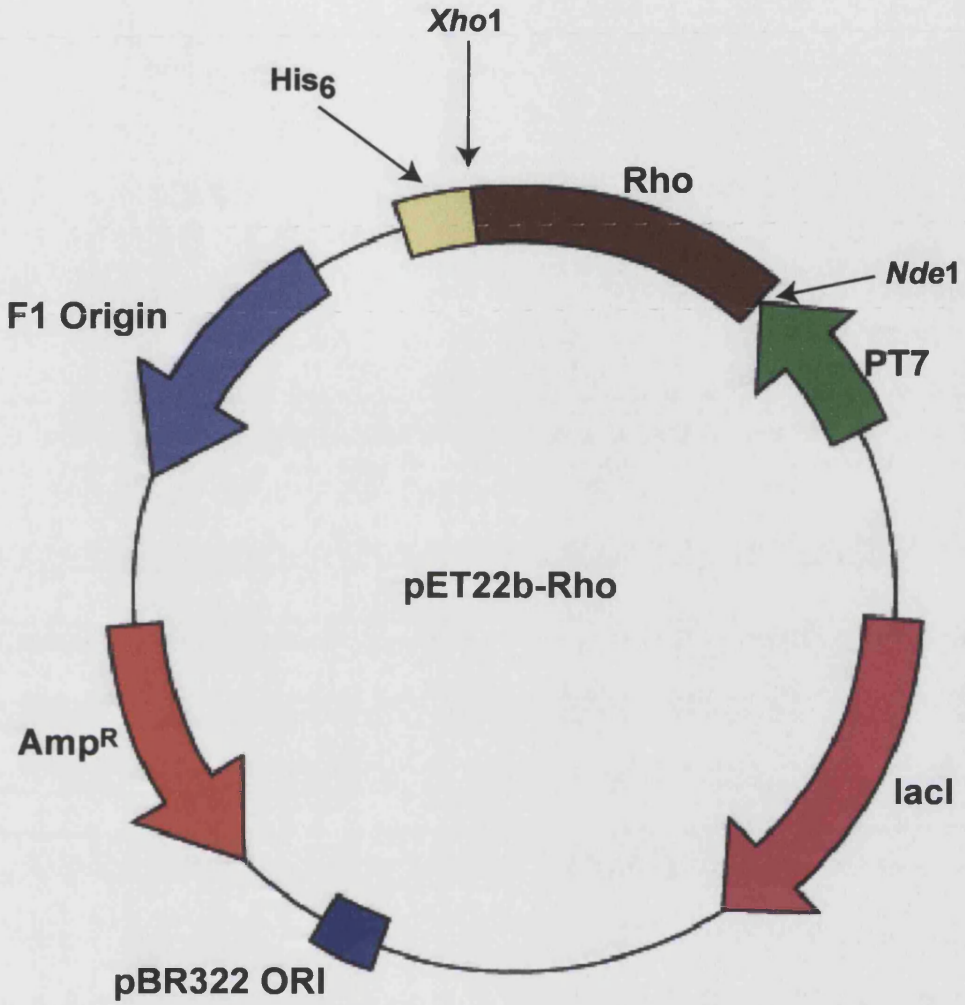
pGEX-6P1 was used for the cloning and overexpression of an N-terminally-GST tagged NusB and a C-terminally-His_6 tagged NusE. The resulting GST-NusB was a 43.0 kDa protein, with a theoretical pI of 6.33 and an extinction coeffecient of 53580 M^{-1} cm^{-1}. The GST can be removed using PreCission Protease (Amersham Biotech).The NusE-His_6 is a 12.4 kDa protein, with a pI of 9.17 and an extinction coeffecient of 3840 M^{-1} cm^{-1}.

Appendix 2c



pET15b was used for the cloning and overexpression of an N-terminally-His_6 tagged NusG. The resulting protein was a 27.6 kDa protein, with a theoretical pI of 5.29 and an extinction coeffecient of 19060 M^{-1} cm^{-1}.

Appendix 2d.



pET22b was used for the cloning and overexpression of a C-terminally-His$_6$ tagged *M. tuberculosis* Rho-termination factor. The resulting protein is a 66.1 kDa protein, with a theoretical pI of 5.8 and an extinction coeffecient of 8960 M^{-1} cm^{-1}.