Name : Thuso Mapotsane

Student Number : 3116368

Supervisor : Professor Wolf-Dieter Schubert

Title : Investigating the Functional Interaction of Transcription Regulator CarD of *Mycobacterium tuberculosis* with Ribonucleic Acid Polymerase

A thesis submitted in partial fulfilment of the requirements for the degree of Masters in Biotechnology in the department of Biotechnology,
University of the Western Cape.

Abstract

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* (*Mtb*). TB mainly affects lungs of patients but other parts of the body can also be affected. It kills approximately 2 million people annually. HIV/AIDS and drug resistance make TB difficult to control. *Mtb* CarD protein forms a physiological complex with Ribonucleic Acid Polymerase (RNAP). This complex causes *Mtb* to undergo dormancy rendering it difficult to control using current antibiotics. CarD and a size-reduced subunit β1 (denoted $\beta1^{m}$ for "minimized") of *Thermus thermophilus* RNAP, in which the central domain has been replaced by a Gly-Gly linker, were produced and purified using affinity nickel nitrilotriaceticacid and glutathione-S-transferase (GST) affinity chromatography techniques respectively. CarD N terminal domain ($CarD^{N}$) was generated from CarD by inserting a stop codon by site directed mutagenesis. CarD was stabilised by adding 5 % (v/v) glycerol to PBS pH 7.4 ensuring protein stability of up to 67 days rather than 2 days without glycerol. $CarD^{N}$ was stable in PBS pH 7.4 without addition of glycerol. This suggests that the CarD C terminal domain may be responsible for CarD instability. To further purify the proteins both anion exchange and gel permeation chromatography techniques were used. CarD and $CarD^{N}$ degrade immediately after anion exchange potentially because of the high ion concentration which partially unfolds the protein making it prone to proteolytic cleavage. GST-pull down assays were used to demonstrate complex formation between RNAP $\beta1^{m}$ and both CarD and $CarD^{N}$ confirming that complex formation is dependent on the N-terminal domain of CarD.

Key Words

Tuberculosis

CarD

CarD/RNAP complex

Drug resistance

Dormancy

Protein-protein interactions

Conventional antibiotics

Mycobacterium tuberculosis

Declaration

I declare that "Investigating the Functional Interaction of Transcription Regulator CarD of *Mycobacterium tuberculosis* with Ribonucleic Acid Polymerase" is my own work that has not been submitted for any degree or examination in any other University, and that all the sources I have used or quoted have been indicated or acknowledged by complete references.

Full name: Thuso Mapotsane **Date:** 18 February 2014

Sign…  **.....**

Acknowledgements

I would like to thank my supervisor Prof. W-D Schubert for allowing me to work in his lab, for his wise and continued guidance and support. I would like to express a word of gratitude and appreciation to Lilia Polle for her early work in this project. This would not be enough if I do not thank BSc Honours student who worked hand in hand with me in this project while learning techniques, Clifford Ntui. I would also like to extend my heartfelt gratitude to Dr. Ntevhe Thovhogi and Mr. Habeeb Bankole who edited my thesis. My colleagues in the structural biology of infectious diseases group, Donne Simpson, Clive Mketsu, Tarriq Firfirey, Jeremy Boonzaier, Mujaahidah Mohamed, Lungelo Mandyoli and Valentine Anye have all given a hand in ensuring I reach my goal. I would like to thank the Botswana government for financial support.

Lastly, I would like to thank my family, best friend Thato Attention Seroke and Pertunia Kganedi Sehata for encouraging me to never give up even when the going gets tough. I give a word of appreciation to my father, mother, siblings, nephews and nieces for support and understanding.

Table of Contents

List of Abbreviations

AIDS	Acquired immunodeficiency syndrome
Amp	Ampicillin
APS	Ammonium persulfate
bp	Base pair
carD	Gene encoding CarD of *M. tuberculosis*
CarD	Dormancy inducing transcription regulator of *M. tuberculosis*
$CarD^N$	CarD N-terminal domain
DksA	RNA polymerase-binding transcription factor of *E. coli*
dNTPs	Deoxynucleotide triphosphates
DSB	Double stranded breaks
E. coli	*Escherichia coli*
EDTA	Ethylene diaminetetra-acetic acid
GST	Glutathione-S-transferase
GS	Glutathione sepharose
HIV	Human immunodeficiency virus
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
iNTP	Initiating nucleoside triphosphate
Kan	Kanamycin
kDa	Kilodalton
LB	Lysogeny broth
LptA	CarD homolog in *Borrelia burgdorferi*

Min	Minute
MES	2-(*N*-morpholino)ethanesulfonic acid
MW	Molecular weight
MWCO	Molecular weight cut off
Mtb	*Mycobacterium tuberculosis*
NADPH	Nicotinamide adenine dinucleotide phosphate
Ni-NTA	Nickel-nitrilotriacetic acid
NTA	Nitrilotriacetic acid
OD_{xxx}	Optical density at wavelength xxx nm
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pGEX-6P-1 or 2	Expression plasmids producing N-terminal GST fusion proteins
RID	RNA interacting domain
PMSF	Phenylmethanesulfonylfluoride
pppGpp	Guanosine pentaphosphate
ppGpp	Guanosine tetraphosphate
RNAP	Ribonucleic acid polymerase
rnaP β1	Gene encoding the β1 subunit of RNAP
RNAP β1	β1 subunit of RNAP
RNAP $\beta1^{m}$	β1 subunit of RNAP "minimized" where the central β1b domain has been replaced by a Gly-Gly linker
rcf	Relative centrifugal force

RelA	*E. coli* enzyme catalysing the synthesis and degradation of ppGpp and pppGpp
RF	Restriction free
rpm	Revolutions per minute
rRNA	Ribosomal RNA
SEC	Size exclusion chromatography
SDS	Sodium dodecyl sulfate
SDM	Site directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TAE	Tris-acetate-EDTA
Taq	*Thermus aquaticus*
TB	Tuberculosis
TEMED	N,N,N,N-Tetramethylethylene-diamine
T_M	Melting temperature
TRCF	Transcription repair coupling factor
Tris-HCl	Tris(hydroxymethyl)aminomethanehydrochloride

List of Figures

Figure 20: Analysis of $CarD^{N}$/ RNAP Complex 54

List of Tables

1.0 Introduction

1.1 Brief History and Transmission of Tuberculosis

1.1.1 Definition

Tuberculosis (TB) is a contagious human disease. In 1882 Robert Koch proved *Mycobacterium tuberculosis (Mtb)* to be the etiological agent of TB (Mcdonough et al. 1993). The bacterium is Gram positive, rod shaped and ~2-4 µm long. *Mtb* is classified as "acid fast" because its cell wall resists staining by acids in the laboratory. The underlying reason for this phenomenon is the mycolic acid constituents of their cell walls which only poorly absorb acid fast stains like carbol fuschin. TB primarily affects the lungs of patients (pulmonary TB) though other parts of the body may also be affected (extra-pulmonary TB) (Bordbar et al. 2010; Deb et al. 2009).

1.1.2 Epidemiology

TB is one of the most lethal infectious diseases globally especially when occurring in combination with Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS) and malaria killing an estimated 2 million people annually (Ordway et al. 2007; Ordway et al. 2010). A third of the world's population harbours *Mtb* bacilli though mostly in a latent or dormant form. About one in twenty of these will develop active disease during their lifetime (Grosset 2003; Clark-Curtiss & Haydel 2003). Overall men are more susceptible to TB than women (Anon 2012), with the exception of women in reproductive age who develop TB more often than their male counterparts (Ordway et al. 2010).

TB, as it is known today, has afflicted humans throughout the ages. Previously it was more commonly referred to as "consumption", due to the patient's body being "consumed" by the

disease, the “white plague”, due to the patients’ pale skin, or “phthisis pulmonaries”, or “Pott’s disease”. *Mtb* has been demonstrated in human remains from ancient Egypt, India and China 15000 to 20000 years ago (Daniel 2006). In Egypt, archaeologists have detected spinal TB in mummies (Barnes 2000).

Rapid urbanization in industrializing countries accompanied by suboptimal living conditions allowed TB to spread rapidly in the 18th century (Daniel 2006) increasing lethality (Mcdonough et al. 1993). Growing knowledge of the disease (Barnes 2000; Daniel 2006) and improving sanitation caused mortality rates to drop significantly in the late 19th century. The discovery of highly effective anti-tuberculosis drugs in the mid-20th century nearly eradicated *Mtb* (Clark-Curtiss & Haydel 2003; Bordbar et al. 2010), only for TB to re-emerge decades later in combination with HIV/AIDS. TB in particular appears to benefit from the HIV-induced suppression of the human immune system, making TB the leading cause of death in HIV/AIDS patients (Daniel 2006). Growing antibiotic resistance of *Mtb* together with HIV increased TB incidences in developed countries in the late 1900’s (Barnes 2000). Developing countries, and especially sub-Saharan Africa, however, bore the brunt of this growing epidemic due to factors such as poor sanitation, malnutrition, lack of drugs and the advent of HIV/AIDS (Mcdonough et al. 1993; Schluger & Rom 1998; Barnes 2000; Anon 2012).

1.1.3 Transmission of *Mtb*

Mtb is transmitted by aerosols from active TB patients to other humans through sneezing, speaking and coughing (Mcdonough et al. 1993; Betts et al. 2002; Fontán et al. 2008). The half-life of *Mtb* within drying droplets is estimated at 6 h (Grosset 2003) while *Mtb*-containing particles may remain suspended in air for several hours aiding their dispersal (Ordway et al.

2007; Grosset 2003). Presumably a single such particle can suffice to elicit a new infection (Betts et al. 2002; Grosset 2003).

Droplets larger than 1-2 µm in diameter are normally eliminated by physical barriers such as the nasopharynx and upper respiratory tract (Schluger & Rom 1998). On the other hand, secretion of antimicrobial peptides with a broad killing spectrum makes the respiratory bronchial epithelium resistant to *Mtb* (Schluger & Rom 1998; Davis & Ramakrishnan 2009).

After arriving in lung alveoli *Mtb* faces three potential fates: Either bacteria are eliminated entirely by the host response, bacteria replicate and initiate clinical disease (pulmonary TB), or bacteria become dormant (latent TB) (Primm et al. 2000). This latent infection can develop into fully blown active TB when dormant bacteria undergo reactivation (Kapoor et al. 2013).

1.2 Immune Response to *Mtb*

The innate immune system is the first line of defence against TB, clearing all invading *Mtb* in ~50% of exposed individuals (Flynn & Ernst 2000; Schluger & Rom 1998) without involvement of the adaptive branch of the immune system (Welin 2011; Flynn & Ernst 2000; Serbina et al. 2001). Clearly the associated margin of error is rather large due to ethical implications of any potential experiments (Ordway et al. 2007; Anon 2012; Serbina et al. 2001). Boosting the innate immune system through immunomodulatory drugs is thus an attractive strategy to limit *Mtb* infection.

Macrophages are the first cell-type of the innate immune system to combat *Mtb* (Davis & Ramakrishnan 2009) – although *Mtb* can under suitable circumstances replicate in resting macrophages. Protective immunity is mediated by T lymphocytes that activate anti-mycobacterial activities of macrophages killing the pathogen. However, despite an active

immune response, *Mtb* is able to persist in a dormant stage. Latent infection is thus the outcome of a continual immune response (Wolf et al. 2008).

When the immune system is suppressed, *Mtb* is no longer suppressed and active disease develops (Kumar & Sanyal 1998; Marino & Kirschner 2004). In fact, *Mtb* makes use of the innate immune system for its pathogenesis *inter alia* by using innate immune receptors to adhere to macrophages such as mannose, Toll-like receptors, pathogen recognition receptors, complement and scavenger receptors (Schluger & Rom 1998; Davis & Ramakrishnan 2009).

1.2.1 Phagocytosis

Mtb passively access lung alveoli by inhalation. Their lipopolisacharides and peptidoglycans activate naïve alveolar macrophages through surface receptors (Davis & Ramakrishnan 2009; Fontán et al. 2008; Ordway et al. 2010) increasing their rate of ingesting and killing microbes. Macrophages increase the levels of major histocompatibility complex II (Davis & Ramakrishnan 2009; Fontán et al. 2008) presenting antigen to T-helper cells and initiating the collaboration between innate and adaptive immunity (Wolf et al. 2008). Other immune cells are recruited to the site of infection around the macrophages forming a caseous material called granuloma (Kapoor et al. 2013).

Upon phagocytosis, bacteria-containing phagosomes fuse with lysosomes containing hydrolytic enzymes such as defensins and lysozyme (Fontán et al. 2008) as well as reactive oxygen generating systems such as nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (Davis & Ramakrishnan 2009) to digest and kill phagocytized bacteria (Wolf et al. 2008). The membrane-integral NADPH oxidase produces superoxide by transferring electrons from intracellular NADPH and coupling these to molecular oxygen to produce the reactive

radical superoxide anion inside the phagosome (Flynn & Ernst 2000; Marino & Kirschner 2004; Keertan et al. 2010; Schluger & Rom 1998).

Evasion mechanisms of *Mtb* have not been fully characterized. *Mycobacteria* do produce ammonia to inhibit phagosome-lysozyme fusion and alkalinizing the lysosomal contents thereby reducing the potency of this fusion complex (Schluger & Rom 1998). While the host immune response kills most bacteria, some do survive but become dormant for extended periods potentially reactivating much later to cause active TB (Fontán et al. 2008; Schluger & Rom 1998).

1.3 Treatment

Common antibiotics used to treat *Mtb* are isoniazid and rifampicin, which together with pyrazinamide form the first line of treatment regimen (Wehrli 2013; Ordway et al. 2010). These antibiotics target actively growing bacteria (Stallings & Glickman 2011; Srivastava et al. 2013). Pyrazinamide, which reduces treatment from nine to six months, targets semi-dormant and old non growing bacilli, exerting its function at acidic pH (Zhang et al. 2003). Pyrazinamide action is slow and it kills less than 76% of the bacterial population (Zhang et al. 2003; Ordway et al. 2010) meaning that it cannot overcome latency entirely.

Initially, first line of protection was 90% effective (Jain et al. 2008). However this relief was short lived as *Mtb* developed resistance against them. This prompted the introduction of second line of protection (**Table 1**). Second line of protection drugs proved to have more side effects than their predecessors in recipients (Kapoor et al. 2013). It is therefore imperative that new therapeutic intervention which is expeditious and effective be identified.

Table 1: First and Second-line Anti-TB Drugs (Jain et al. 2008)

First-line drugs	Second-line drugs
Para-aminosalicylic acid	Viomycin
Isoniazid	Fluoroquinolones
Rifampicin	Ofloxacin
Rifapenline	Cipromycin
Pyrazinamide	Aminoglycosides
Streptomycin	Kanamycin
Ethambutol	Amikacin
	Ethiolnamide
	D-Cycloserine

1.4 Stringent Response

The stringent response is a bacterial survival mechanism (Fontán et al. 2008; Magnusson et al. 2005; Potrykus & Cashel 2008) geared towards amino acid starvation, DNA damage, nutrient starvation and phosphate deprivation (Stallings & Glickman 2011; Stallings et al. 2003). Bacteria synthesize the alarmones guanosine tetraphosphate (ppGpp) (Barker et al. 2001) and guanosine pentaphosphate (pppGpp) (Potrykus & Cashel 2008). The enzyme RelA or Rel*Mtb* (Kapoor et al. 2013) controls both synthesis and degradation of these signalling molecules (Primm et al. 2000) by triggering the stalled ribosome with an uncharged tRNA in the acceptor site to produce ppGpp (Raskin et al. 2007; Brockmann-gretza 2006; Betts et al. 2002). With the aid of the transcription factor DksA, ppGpp directly binds to the RNA polymerase secondary channel near the catalytic centre reducing rRNA synthesis (Ferullo & Lovett 2008). Another enzyme, SpoT, hydrolyses ppGpp.

1.4.1 Small Molecule Effectors and Initiating Nucleoside Triphosphate

Promoters regulate initiation of transcription to produce rRNA (Nickels & Hochschild 2004) by initiating nucleoside triphosphate (iNTP) and ppGpp on the open complex (Paul et al. 2004) in a concentration dependent manner. Increasing iNTP concentrations stimulate rRNA transcription by transiently stabilizing the open complex (Paul et al. 2004). ppGpp, by contrast, inhibits transcription by binding to the secondary channel near the RNAP active site (Rutherford et al. 2007) and potentially base pairing with exposed cytosines in the non-template strand to decrease the lifetime of the open complex (Paul et al. 2004; Brockmann-gretza 2006). Independent of its orientation, ppGpp coordinates Mg^{2+}at either its 3' or 5' end (Vrentas et al. 2008) physically linking it to the transcription regulator DksA (Nickels & Hochschild 2004; Rutherford et al. 2007; Barker et al. 2001).

1.4.2 DksA

The 151 amino acid transcription factor DksA critically regulates rRNA production (Nickels & Hochschild 2004) by binding to the secondary channel of RNAP (Gaal 1997). A crystal structure reveals that the DksA coil-coiled domain extends deep into the secondary channel similar to other Gre transcription factors (Nickels & Hochschild 2004; Paul et al. 2004). Two aspartate residues at the tip of this domain coordinate Mg^{2+}, which is in turn coordinated by ppGpp binding it to the transcription complex (Nickels & Hochschild 2004; Paul et al. 2004; Potrykus & Cashel 2008). Regulation of rRNA production by either ppGpp or iNTP requires a short-lived initiation complex. DksA shortens the lifetime of the open complex of transcription initiation sensitizing it to small molecule effectors (Magnusson et al. 2005; Rutherford et al. 2007; Brockmann-gretza 2006).

1.5 CarD

Protein-protein interactions of transcription regulatory proteins are important in regulating transcription (Gaal 1997). One such transcription regulator, CarD, is found in many bacteria such as *Myxococcus xanthus*, *Borrelia burgdorferi* and *Thermus thermophilus* (Galbis-martı et al. 2004; García-heras et al. 2013) but not in eukaryotes and archaea (Garcı et al. 2010). *Borrelia burgdorferi* codes for LtpA, a CarD homolog, whereas *M. xanthus* produces a CarD ortholog involved in fruiting body production (Yang et al. 2008). All sequenced mycobacterial genomes code for CarD underscoring its physiological relevance (Weiss et al. 2012) with 98.1 % and 95.7 % sequence identity between *M. smegmatis* CarD and those of *Mtb* and *M. leprae* respectively. Functionally, CarD resembles DksA of *E. coli* (Garcı et al. 2010).

1.5.1 Molecular characterization of *Mtb* CarD

CarD of *Mtb* is 162 amino acids in length and consists of an N-terminal domain ($CarD^N$) of 64 amino acids and a C-terminal domain of 98 residues (Stallings et al. 2009). The N-terminal domain shares 60 % sequence identity with RNAP interacting domain (RID) of transcription repair coupling factor (TRCF) of *E. coli* (Westblade et al. 2010) and interacts with the β1-subunit of RNAP (Stallings & Glickman 2011) analogous to TRCF-RID (Westblade et al. 2010; Srivastava et al. 2013). The C-terminal domain, thought to bind DNA (Garcı et al. 2010; García-heras et al. 2013), is essential for CarD stability (Stallings & Glickman 2011; Westblade et al. 2010).

1.5.2 Complex of CarD with RNAP β1-subunit

Transcription regulatory factors control RNAP transcription of DNA to mRNA (Gaal 1997). Two-hybrid assay and immunoprecipitation experiments indicate that *Mtb* CarD binds to the N-terminal domain of the RNAP β1-subunit (RNAP β1) by its RID domain (Stallings et al.

2009). Although TRCF and CarD bind similar sites in RNAP, their functions are distinct such that CarD- and TRCF-RID cannot functionally replace each other (Stallings & Glickman 2011).

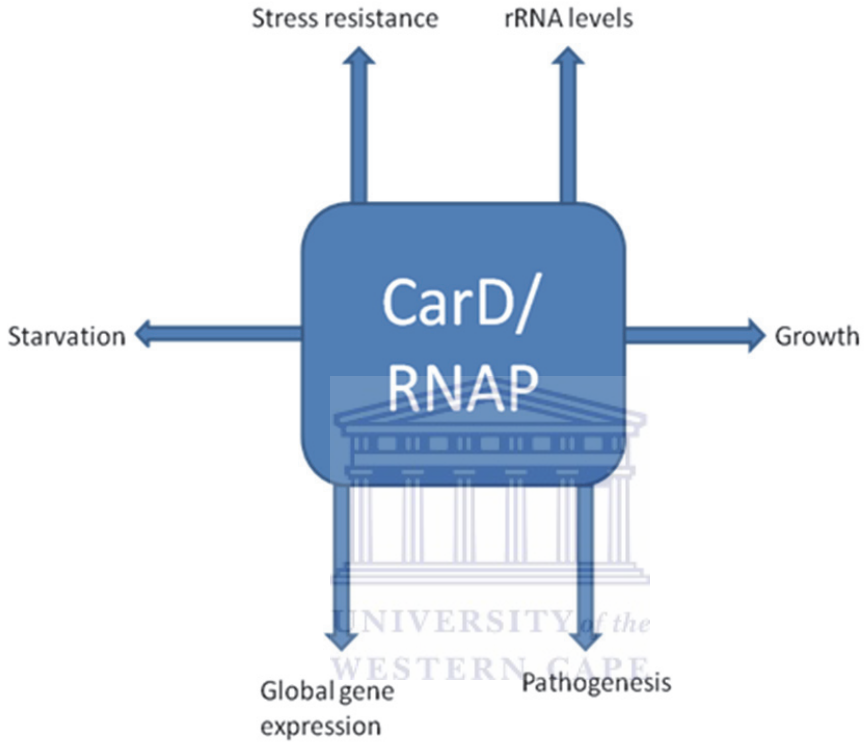

Figure 1: Biophysical Effects of *Mtb* CarD Binding RNAP (Weiss et al. 2012)

CarD binding to RNAP β1 subunit elicits a number of downstream effects (**Figure 1**) that limit transcription to proteins essential for *Mtb* persistence. rRNA production is down regulated to match the declining need of proteins to allow the cell to survive stressful conditions such as starvation (Stallings & Glickman 2011). *Mtb* correspondingly enters a state of dormancy. Altering the CarD/RNAP β interaction could terminate dormancy, sensitizing *Mtb* to physiological stresses (Priya & Megha 2012). CarD also appears to be involved in limiting conflict between transcription and replication (Srivastava et al. 2013).

1.5.3 CarD in *Mtb* Persistence

Mtb persistence mechanisms include the repair of double stranded DNA breaks (DSB) without DNA template (Stallings and Glickman, 2011). CarD was implicated in DNA DSB repair, though its precise role remains to be elucidated. Other functions of CarD in *Mtb* persistence include the stringent response control mechanism, responses to environmental stresses and viability of *Mtb* (Srivastava et al. 2013). CarD is critical for growth as its removal results in retarded growth (Stallings & Glickman 2011; Stallings & Michael 2010).

1.5.4 CarD is Essential for *Mtb* Viability

Deletion of *carD* in both *Mtb* and *M. smegmatis* were unsuccessful (Stallings et al. 2009). Its conditional depletion similarly confirm that CarD is essential for *Mtb* viability as well as for homeostasis and growth (Priya & Megha 2012). Defects in CarD correlate with impaired growth of *Mtb* and sensitivity to genotoxins and hydrogen peroxide (Stallings & Michael 2010).

1.5.5 CarD is Essential for Survival of Nutrient Limitation and Oxidative Stress

Mtb survival in granuloma despite nutrient deprivation implies specialized cellular processes for their persistence (Kapoor et al. 2013; Davis & Ramakrishnan 2009; Ordway et al. 2010). This includes CarD up-regulation during nutrient deprivation and under oxidative stress (Stallings et al. 2009). *Mtb* strains unable to produce CarD are killed by oxidative agents such as ciprofloxacin, bleomycin and hydrogen peroxide, while strains producing CarD are largely unaffected (Srivastava et al. 2013).

1.5.6 CarD is Required for the Stringent Response

Induction of CarD results in reduced RNA levels in Mycobacteria (Stallings et al. 2009). In most bacteria including *E. coli,* stable RNA synthesis is accomplished by the stringent response (Magnusson et al. 2005; Vrentas et al. 2008; Barker et al. 2001) during which ppGpp down regulates synthesis of RNAs related to environmental stress (Primm et al. 2000; Magnusson et al. 2005; Vrentas et al. 2008). *E. coli* cells lacking proteins such as ppGpp-producing RelA are unable to grow on amino acid deficient media (Weiss et al. 2012).

The stringent response in *Mtb* works differently (Stallings et al. 2009) and involves CarD (Stallings & Glickman 2011, Stallings & Michael 2010). Thus CarD alongside ppGpp-producing RelA is essential for stringent response.

1.6 Comparing CarD to TRCF and DksA

DksA of *E. coli* shares some features of CarD (Weiss et al. 2012; Stallings et al. 2010) so much so that CarD can partly substitute for DksA in *E. coli* (Stallings & Glickman 2011). DksA and CarD binding sites on RNAP do, however, not overlap (Priya & Megha 2012) implying that CarD and DksA achieve similar functions through different protein-protein interactions. DksA, furthermore, is non-essential in *E. coli* (Paul et al. 2004) and its effect is evident only under high nutrient conditions – the opposite of CarD (Stallings et al. 2009).

Both TRCF-RID and CarD bind to the β-subunit of RNAP (Srivastava et al. 2013; Westblade et al. 2010) and they share a sequence identity of 60% (Westblade et al. 2010). Nevertheless, CarD cannot substitute for DksA in *E. coli* and TRCF-RID does not have the same effect as CarD in *Mtb*.

1.7 Problem Identification

Mtb has developed resistance to many standard antibiotics making TB one of the deadliest human diseases. To survive stress conditions, *Mtb* enters a dormant state limiting the efficacy of antibiotics. The physiological complex of CarD with RNAP, which initiates the dormant state by down regulating transcription remains to be structurally characterized. Such a structure could help to unravel the disparate roles of CarD in *Mtb* persistence and dormancy and aid the development of drugs to prevent this critical interaction as a first step in overcoming persistence and dormancy. Ideally this would involve a structural and biophysical characterization of CarD on its own as well as in complex with RNAP.

1.8 Research Aim and Objectives

1.8.1 Aim

The overall aim of this project is to elucidate the physiological CarD/RNAP complex of *Mtb*.

1.8.2 Objectives

The project focused on the following objectives.

1) Produce and purify full-length CarD.
2) Stabilize CarD.
3) Generate a clone for $CarD^{N}$. Produce and purify $CarD^{N}$.
4) Produce and purify *Mtb* RNAP $\beta1^{m}$. Clone, produce and purify *Taq* RNAP $\beta1^{m}$.
5) Test for binding between CarD and both *Taq* and *Mtb* RNAP $\beta1^{m}$.
6) Purify CarD/*Mtb* RNAP $\beta1^{m}$ and $CarD^{N}$/*Taq* RNAP $\beta1^{m}$.

2.0 Materials and Methods

2.1 Chemicals and Reagents

General chemicals and reagents were purchased from Sigma or Merck except where otherwise specified**.**

2.2 Primers

Polymerase chain reaction (PCR) and site directed mutagenesis (SDM) primers were designed manually. PCR primers conformed to the following rules:

1) Length: 15 to 30 nucleotides.

2) Melting temperature, $T_m = 4\ (G+C) + 2\ (A+T)$, where G, C, A, T are the number of corresponding nucleotides in the primer.

3) T_m of forward and reverse primers should differ by ≤ 5 °C.

SDM primers, by contrast, were 25 to 45 nucleotides long to achieve $T_m \geq 78$ °C. Altered nucleotides were positioned near the middle of the primers.

Forward and reverse primers for the cloning of *Mtb* RNAP β1 were designed using the restriction-free software, http://www.rf-cloning.org/. DNA sequences for the *rnaP* β1 gene and pGEX-6P-2 vector were provided. The software proposed primers of 51 nucleotides combining sequences respectively complementary to the *rnaP* β1 and to pGEX-6P-2. Primers were purchased from Inqaba Biotech (**Table 2**). The mutations are underlined and shown in bold.

Table 2: List of Primers

Primer	Length (bases)	Sequence
Mtb-RNAP β1 forward	28	5'-ATGGTCGGATCCGGACTCCTTGACGTC -3'
Mtb-RNAP β1 reverse	34	5'-ATCTGACGCGGCCGCCTAGAATTGGCTCAGCTGG-3'
Mtb-RNAP β1-RF forward	40	5'-GTTCTGTTCCAGGGGCCCCTGGGGACTCCTTGACGTCCAG-3'
Mtb-RNAP β1-RF reverse	42	5'-GTCAGTCAGTCACGATGCGGCCCTAGAATTGGCTCAGCTGGC-3'
Mtb-RNAP $\beta1^{m}$-SDM forward	34	5'-TGGTGGTCAGCCAG**GGAGGT**CGTACGGTCGGCGA-3'
Mtb-RNAP $\beta1^{m}$-SDM reverse	34	5'-TCGCCGACCGTACG**ACCTCC**CTGGCTGACCACCA-3'
Mtb-$CarD^{N}$ forward	51	5'-GGTGTTCGTGATGTTGTTGGTCAA**TAA**GAAGGTCTGGATAAAGTTTTTCAG-3'
Mtb-$CarD^{N}$ terminus reverse	51	5'-CTGAAAAACTTTATCCAGACCTTC**TTA**TTGACCAACAACATCACGAACACC-3'

2.3 Plasmids

Names, features and suppliers of plasmids used in the study are listed in **Table 3**.

Table 3: Utilized Plasmids

Plasmid	Size (Kb)	Selection	Tag	Position	Cleavage site	Source
pGEX-6P-1	4.9	Amp^R	GST	N-terminal	PreScission protease*	GE Healthcare
pGEX-6P-2	4.9	Amp^R	GST	N-terminal	PreScission Protease*	GE Healthcare
pET-28a	5.4	Kan^R	His_6	N-terminal	Thrombin	Novagen

PreScission protease is a human rhinovirus C3 protease.

2.3 Recombinant Plasmids

Recombinant plasmids used in this work are summarized in **Table 4** with restriction sites on either side of the gene of interest.

Table 4: List of Recombinant Plasmids Used in the Study

Plasmid	Restriction sites	Source
pGEX-6P-1-CarD	*BamH*I, *Not*I	Previous work by Lilia Polle
pGEX-6P-1-$CarD^N$	*BamH*I, *Not*I	Current work
pGEX-6P-2_*Mtb*-RNAP	*BamH*I, *Not*I	Current work
pET28a_*Taq*-RNAP $\beta 1^m$	*Nde*I, *Hind*III	Rockefeller University, New York, USA

2.4 Expression Cells

A number of competent expression cells were used (**Table 5**).

Table5: Competent Expression Cells Used in the Study

***E. coli* strains**	**Source**
Rosetta 2	Novagen
BL21 (DE3)	Stratagene
BL 21+	Stratagene
DH5α	Stratagene
Gene Hogs	Stratagene

2.5 Preparation of Solutions, Media and Buffers

All buffers, media, and sterile materials were autoclaved for 20 min at 121 °C. Heat-sensitive components were filter-sterilized with 0.2 µM filters and degassed.

2.5.1 Lysogeny Broth and Agar Plates

Lysogeny broth (LB) medium was prepared by dissolving 10 g tryptone, 5 g yeast extract powder and 5 g NaCl in 500 mL deionized water before making up to 1 L and autoclaving. For LB agar additionally 15 g of bacteriological agar was added. The medium was allowed to cool to around 40 °C before adding appropriate antibiotics. Petri dishes were filled with ~25 mL of liquid agar in the laminar flow cabinet, spread by gentle swirling and allowed to set for 1 h. They were stored at 4 °C for later use.

2.5.2 Buffers and Solutions

Table 6: Buffers and Solutions

Buffers/Solutions	**Composition**
Lysis buffer	1 x PBS pH 7.0, Complete EDTA-free Protease Inhibitor Cocktail Tablet and 10 µg/mL DNaseI
PreScission protease cleavage buffer	10 mM Tris/HCl pH 7.0, 150 mM NaCl
GST elution buffer	1 x PBS pH 7.4
Ni-NTA equilibration buffer	1 x PBS pH 7.4
Ni-NTA wash buffer	1 x PBS pH 7.4, 10 mM imidazole
Ni-NTA elution buffer	1 x PBS pH 7.4, 250 mM imidazole
TAE buffer	40 mM Tris-HCl pH 7.5, 20 mM sodium acetate, 1 mM EDTA, adjusted to pH 8.2 with acetic acid
Anion exchange buffer A	20 mM Tris-HCl pH 7.0, 25 mM NaCl
Anion exchange buffer B	20 mM Tris-HCl pH 7.0, 1 M NaCl
Gel filtration buffer	5 % (v/v) glycerol, 1 x PBS pH 7.4
10 x DNA loading buffer	70 % (w/v) sucrose, 0.25 % (w/v) bromophenol blue, 0.1 M EDTA
SDS-PAGE running buffer	25 mM Tris-HCl, 192 mM glycine, 0.1 % (w/v) SDS
SDS-PAGE sample buffer (8 x)	16 mL 10 % SDS, 4 mL glycerol, 2.2 mL 1.5 mM Tris/HCl pH 6.8, 800µL β-mercaptoethanol, 1 spatula tip bromophenol blue

SDS-PAGE separating gel (15 %)	15 mL acrylamide/bisacrylamide 30 %/0.8 % (w/v), 7.6 mL 1.5 M Tris-HCl pH 6.8, 300 µL 10 % (w/v) SDS, 10 mL ddH_2O, 40 µL TEMED, 100 µL 25 % (w/v) APS
SDS-PAGE stacking gel (5 %)	1.5 mL acrylamide/bisacrylamide 30 % (w/v)/0.8 % (w/v), 2.5 mL 0.5 M Tris-HCl, pH 6.8, 5.9 mL ddH_2O, 15 µL TEMED, 25 µL 25 % (w/v) APS
SDS-PAGE Staining solution	0.25 % (w/v) Coomassie Brilliant Blue R-250, 30 % (v/v) ethanol, 10 % (v/v) acetic acid
SDS-PAGE destaining solution	40 % (v/v) ethanol, 10 % (v/v) acetic acid

2.5.3 Liquid Culture

Bacteria were grown in liquid medium using baffled Erlenmeyer flasks for aeration. Flasks with inoculated liquid media and appropriate antibiotics (0.1 µg/mL) were incubated at 37 °C and a rotation speed of 160 rpm overnight. Liquid bacterial cultures were prepared for pre-cultures, protein production, DNA preparation and glycerol stocks.

2.6 Molecular Biology

2.6.1 Polymerase Chain Reaction

The gene encoding the β1-subunit of *Mtb* RNAP (entry Rv0667, gene *rpo B*) was obtained from the Tuberculist database (http://tuberculist.epfl.ch/). To ensure the optimal choice of start and end codons for the construct, the *Mtb* RNAP β1 protein was aligned with *Taq* RNAP β1 in ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) as the latter has successfully been cloned, produced and purified and used in structural analyses (Srivastava et al. 2013). The required gene construct was delimited and the percentage of rare codons calculated using

RACC (http://nihserver.mbi.ucla.edu/RACC/). Forward and reverse primers for *Mtb*-RNAP β1 (**Table 2)** were designed. The amplification reaction is described in **Table 7.**

Table 7: Reaction Mixture for PCR Preparation

Reaction components	**Volume (µL)**	**Concentration**
10 x Phusion GC buffer	5	1 x
Template DNA	2	100 ng/µL
Forward primer	1	10 pmol/µL
Reverse primer	1	10 pmol/µL
dNTP mix	1	10 µM
Phusion DNA polymerase	1	1 U
Nuclease free water	34	
Total volume	50	

Table 8: PCR Reaction Programme

PCR Step	Purpose	Temperature	Duration
Step1	Initial denaturation	98 °C	1 min
Step2*	Denaturation	98 °C	30 s
Step3*	Annealing	50 °C	30 s
Step4*	Elongation	72 °C	1.5 min
Step5	Final elongation	72 °C	10 min

*Steps 2-4 were repeated 35 times

The control experiment without primers was run alongside.

2.4.2 Agarose Gel Electrophoresis

Agarose gel electrophoresis is used to analyse DNA samples. In this study, 1 % (w/v) agarose gels were prepared by dissolving 0.5 g agarose in 50 mL Tris-acetate-EDTA (TAE) buffer (**Table 6**) brought to boil in a microwave for 1 min. The agarose solution was allowed to cool to ~60 °C and 3 x SYBR Green was added as a fluorescent dye to visualize the DNA samples. The samples were prepared by mixing 4 µL of 10 % loading buffer (10 x) (**Table 6**) and 10 µL of DNA (For all DNA samples in this study). This 14 µL mixture was loaded into a sample well, the gel was electrophoresed at 80 V and 2 mA for 60 min before being visualized and photographed under a UV illuminator.

2.6.3 Restriction Free Cloning

Restriction free (RF) cloning is a method whereby a gene of interest may be inserted at any desired location into a circular plasmid, independent of restriction sites, ligation or any modifications in both vector and gene (Ent & Löwe 2006; Ulrich et al. 2012).

The PCR reaction was planned in two phases. First, *Mtb rnaP β1* was to be amplified from *Mtb* genomic DNA (kindly provided by Mr. Zhuo Fang, PhD student, Stellenbosch University). The reaction mixture and PCR programme are identical to conventional PCR as outlined in **Tables 7** and **8**. Second, the amplified product was to be inserted in the vector pGEX-6P-2. The reaction mixture for this second phase of RF cloning is summarized in **Table 9**. The PCR programme is the same as that in **Table 8** apart from the annealing temperature being changed to 62 °C as suggested by the RF server.

Table 9: Reaction Mixture for RF Second Phase

Reaction components	Volume (µL)	Concentration/Amount
5 x Phusion GC buffer	5	1 x
First phase product DNA	10	100 ng/µL
dNTP mix	1	10 µM
pGEX-6P-2	3	50 ng/µL
Phusion DNA polymerase	1	1 U
Nuclease free water	33	
Total volume	50	

To optimise the PCR conditions for the RF and the conventional PCR, the annealing temperature was varied using gradient PCR. The PCR reaction mixture and running programme remained the same as **Tables 7** and **8** respectively except that the annealing temperature ranged from 45 to 65 °C. The primers designed for the conventional PCR (**Table 2**) were used to amplify from the PCR product of the RF first phase providing a higher concentration of template as compared to the genomic DNA. While the reaction mixture remained the same, a two-step PCR was used for amplification in the thermocycler (**Table 10**).

Table 10: Two Step PCR Reaction Programme

PCR Step	Purpose	Temperature	Duration
Step 1	Initial denaturation	98 °C	30 s
Step 2*	Denaturation	98 °C	10 s
Step 3*	Elongation	72 °C	35 s
Step 4	Final elongation	72 °C	10 min

*Steps 2-3 were repeated 35 times.

All amplified products were analysed by agarose electrophoresis.

2.6.4 Site Directed Mutagenesis

Site directed mutagenesis (SDM) is used to replace, delete or insert nucleotides in a target DNA by combining PCR with appropriately designed primers. After amplification, the methylated parental DNA is digested by *DpnI* for 1 h at 37 °C to ensure that only the product DNA is present for downstream applications.

To study the source of CarD instability, the N-terminal domain ($CarD^N$) was to be produced and analysed independently of the C-terminal domain. To generate a corresponding gene construct, a stop codon (TAA) had to be introduced between the coding regions for the N- and the C-terminal domains. Secondary structure prediction was used to identify a suitable position and to ensure that the stop codon would not interrupt an α-helix or β-strand. The primers for $CarD^N$ (**Table 2**) were designed to insert a stop codon (TAA) between codons 64 and 65 of *carD*. The mutation was undertaken by SDM using the Quick Change Kit (QIAGEN). The PCR reaction components were as follows:

Table 11: Reaction Mixture for SDM PCR

Reaction components	**Volume (µL)**	**Concentration**
10 x PCR buffer	5	1 x
CarD plasmid DNA	1	85 ng/ µL
Forward primer	1	0.2 µM
Reverse primer	1	0.2 µM
$MgSO_4$	1	1.0 mM
dNTP's mix	1	0.2 mM
KAPA HIFI DNA polymerase	1	1 U
Nuclease free water	39	
Total volume	50	

The PCR conditions in **Table 12** were used for the CarD^{N} SDM.

Table 12: PCR Reaction Programme

Step 1	Initial denaturation	95 °C	60 s
Step 2*	Denaturation	95 °C	40 s
Step 3*	Annealing	72 °C	40 s
Step 4*	Elongation	74 °C	6 min
Step 5	Final elongation	74 °C	8 min
*Steps 2-4 were repeated 20 times			

The β1-subunit of RNAPs consists of two distinct domains β1a and β1b, where β1b was evolutionarily inserted into β1a. The orientation of β1b relative to β1a requires other subunits of the RNAP complex to stabilize (Westblade et al. 2010; Srivastava et al. 2013). To produce a minimal, conformationally stable *Mtb* RNAP β1, the had to be replaced by a glycine-glycine (GG) linker in analogy to *Taq* RNAP β1 (Westblade et al. 2010; Srivastava et al. 2013). Here SDM was chosen to potentially achieve the substitution in a single step. The conditions used are identical to **Tables 11** and **12**. This minimized RNAP β1 is denoted RNAP β1^{m}.

2.6.5 Cloning

The amplified *Mtb rnaP β1* product as well as pGEX-6P-2 was double digested with *NotI* and *BamH*I to create compatible sticky ends. The resulting fragments were analysed by agarose electrophoresis (**Section 2.3.2**).

Table 13: Double Digest of the Amplified *rnaP β1* Gene

Reagent	**Volume (µl)**	**Final concentrations/amount**
dsH_2O	13	-
10 x Fast Digest enzyme Buffer	2	1 x
Amplified *rnaP β1* DNA product	4	1 µg
NotI	0.5	5 U
BamHI	0.5	5 U

The same experiment was set up for the pGEX-6P-2 plasmid DNA digest except that 3 µg of DNA were used rather than 1 µg.

The reaction as outlined in **Table 14** was used to ligate the insert to pGEX-6P-2.

Table 14: The Ligation Reaction Between *rnaP* β1 and pGEX-6P-2 Plasmid DNA

Reagent	**Volume (µl)**	**Final concentration**
5 x Ligase Reaction Buffer	4	1 x
Insert	9	17 ng/µl
Plasmid	3	6 ng/µl
T4 DNA ligase	0.5	0.1 U
dsH_2O	3.5	-

The ligation mixture was incubated for 3 h at room temperature before transformation.

2.6.6 Transformation of Competent Bacteria

Competent *E. coli* DH5α and *E. coli* BL21+, and/or Rosetta 2 cells were transformed with recombinant plasmid DNA for plasmid amplification and recombinant gene expression respectively (see below). Positive clones were selected for plating the transformation mixture on LB agar plates containing antibiotics corresponding to plasmid resistance genes.

2.6.6.1 Heat Shock Method

Plasmid DNA (1 µL) obtained from ligation reactions was added to 50 µL of competent cells thawed on ice. The cells were incubated on ice for 30 min and heat shocked for 45 s at 42 °C, followed by fast cooling on ice for 2 min. 800 µL of LB medium was added to the cells and the suspension shaken at 37 °C for 1 h on a thermo-mixer. The cells were gently resuspended using a 1 mL pipette. The transformation mixture was evenly spread on LB agar plates containing 0.1 µg/mL ampicillin using sterile glass beads. The plates were incubated at 37 °C overnight.

2.6.6.2 Plasmid DNA Extraction

An isolated colony from transformed *E. coli* cells were transferred to 5 mL of LB medium with 0.1 µg/mL ampicillin and allowed to grow at 37 °C with constant shaking overnight. Plasmid DNA was isolated using GeneJET mini prep kit as per manufacturer's instructions. The concentration of eluted DNA was measured by reading absorption at 260 nm (A_{260}) using a Nanodrop ND-1000 spectrophotometer.

2.6.6.3 DNA Sequencing

All the newly cloned plasmids in this work were sequenced by Inqaba Biotech using the Sanger chain sequencing method. 10 µL of the extracted plasmid DNA was pipetted into an

Eppendorf tube, labelled and sent for sequencing. The results were annotated using the BioEdit software package.

2.6.7 Preparation of Glycerol Stocks

Glycerol stocks were prepared for long-term storage of the cells (BL21+ and Rosetta 2) containing plasmids of interest from the transformation described in section 2.3.6.1. 5 mL LB medium was inoculated with the colony of transformed bacteria. After ampicillin (final concentration = 0.1 µg/mL) was added, the culture was grown to mid log phase (OD_{600} = 0.4 - 0.8) by incubating overnight at 37 °C at 160 rpm. Thereafter, 800 µL of the culture was mixed with 200 µL of sterile 80 % (v/v) glycerol (cryoprotectant) in a 1 mL cryo-vial. The vial was capped tightly and inverted several times for good mixing before being stored at -80 °C.

2.7 Protein Production and Purification

2.7.1 Pre-culture Preparation

50 mL of autoclaved LB was pipetted into a sterile 100 mL Erlenmeyer flask. This was supplemented with 0.1 µL/mL of an appropriate antibiotic for selecting cells containing the transformed plasmid. An autoclaved pipette tip was used to scratch the frozen glycerol stock of transformants from -80 °C freezer and dropped inside the Erlenmeyer flask with 50 mL LB. The covered flask was incubated at 37 °C shaking at 180 rpm overnight.

2.7.2 Main-culture Preparation

The 100 µL pre-culture was mixed with 900 µL of sterile LB in a cuvette. The absorbance was measured to calculate the volume of pre-culture to add to the main-culture to obtain a starting absorbance OD_{600} = 0.1. The formula $C_1V_1 = C_2V_2$ was used, where C and V stand for concentration and volume, and subscripts 1 and 2 denote initial and final states, respectively.

Ampicillin was added to a final concentration of 0.1 µM/mL. The main-culture was incubated at 37 °C until it reached an OD_{600} of between 0.4 and 0.8 (log growth phase). Gene expression was induced by adding 100 µL/1 L (0.1 mM) of isopropyl β-D-1-thiogalactopyranoside (IPTG).

Cultures were incubated overnight at 16 °C and 180 rpm for pGEX-6P-1-CarD and pGEX-6P-2-CarD^{N}. For pGEX-6P-2_*Mtb*-RNAP β1^{m} and pET28a_*Taq*-RNAP β1^{m}, the temperature was adjusted to 25 °C. After overnight incubation, cells were harvested by centrifugation at 7600 relative centrifugal force (rcf) and 4 °C for 15 min. The supernatant was discarded and the cell pellets resuspended in 20 mL lysis buffer per 1 L culture at 4 °C.

2.7.3 Sonication of Cells

To inhibit host serine proteases, phenylmethanesulfonylfluoride (PMSF) was added to the resuspended cells for a final concentration of 0.5 mM prior. Cells were lysed by sonication using 5 cycles of 30 s sonication separated by 30 s breaks on ice to prevent denaturing of protein. To degrade genomic DNA, 0.32 µg/mL of DNAse I (20 µL of a 0.4 mg/L stock) was added to the disrupted cells and incubated on ice for 20 min. The disrupted cells were then transferred to pre-chilled SS34 tubes and soluble protein separated from cell debris and other insoluble proteins by centrifugation at 29000 rcf for 60 min at 4 °C using an SA-600 rotor in a RC6 Sorvall centrifuge.

2.7.4 Protein Purification

Proteins were purified using different methods (see below). To assess the protein purity in this study, 15 % sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE) was used.

2.7.4.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Protein samples were mixed with 8 x sample buffer (10 µL protein + 3 µL sample buffer). For protein denaturation, samples were heated to 95 °C for 5 min. The protein samples were pipetted into a loading pocket of the gel and electrophoresed at 40 mA for 40 to 45 min. The gel was stained for 15 to 20 min in a Coomassie Brilliant Blue R-250 solution and excess stain removed by incubating in a destaining solution overnight at room temperature.

2.7.4.2 Affinity Chromatography

Affinity chromatography utilizes strong and specific binding of affinity tags fused to the N- or C-terminus of the target protein to an immobilized matrix. In this work, CarD, CarD^{N} and *Mtb* RNAP β1^{m} were produced as fusion proteins with an N-terminal glutathione-S-transferase (GST) tag, while *Taq* RNAP β1^{m} was produced with an N-terminal hexahistidine tag (His_6). CarD production and purification buffers were augmented with 5 % (v/v) glycerol to prevent degradation.

2.7.4.2.1 Glutathione-S-Transferase Affinity Chromatography

As CarD was produced as a GST fusion protein it could be purified using GST-affinity chromatography. The glutathione sepharose (GS) beads were equilibrated with 10 x bed volume of 1 x PBS pH 7.4, 5 % (v/v) glycerol in a 12 mL column. The centrifugation supernatant containing soluble CarD was added to the equilibrated GS beads in a 50 mL Falcon tube and incubated at 4 °C on a roller mixer for 4 h to allow coupling of the fusion protein to the matrix. Unbound proteins and cell debris were eluted in a fraction labelled "flow-through". The beads were washed by adding 10 mL PBS with 5 % (v/v) glycerol to the column and allowing the buffer to elute. This was repeated 5 times. CarD-bearing beads were equilibrated with 10 mL cleavage buffer. The beads were resuspended in 5 mL of this buffer

by gentle inversion and 200 µL of PreScission cleavage protease was added to the beads and the mixture incubated overnight at 4 °C on a roller mixer. CarD was eluted from the GS beads with CarD purification buffer and fractions analysed by 15 % SDS-PAGE.

The same protocol was used to purify $CarD^{N}$ and *Mtb*-RNAP $\beta1^{m}$ except that buffers were not augmented with 5 % (v/v) glycerol.

2.7.4.2.2 Nickel-NTA Affinity Chromatography

Nickel-nitrilotriacetic acid (Ni^{2+}-NTA) agarose beads are used to purify recombinant proteins fused to His_6-tags. The histidine residues bind to the immobilized Ni^{2+} with high specificity and affinity. His_6-tagged *Taq* RNAP $\beta1^{m}$ was recovered from the bacterial host cells by cell disruption and centrifugation as described for CarD. 1.5 mL resuspended Ni^{2+}-NTA beads were equilibrated with 100 mL 1 x PBS pH 7.4 in a drip column. The soluble protein fraction was added to the beads in a 50 mL Falcon tube and incubated at 4 °C on a roller mixer for 4 h. The suspension was poured back into the column and the flow-through eluted and collected to check for the presence of unbound fusion protein. The beads were washed five times with 10 mL of 1 x PBS pH 7.4 with 10 mM imidazole. *Taq* RNAP $\beta1^{m}$ was then eluted with 1 x PBS pH 7.4, 250 mM imidazole in five 10 mL fractions and the purity analysed by SDS-PAGE. *Taq* RNAP $\beta1^{m}$ was further purified with size exclusion chromatography (SEC).

2.7.2.3 Ion Exchange Chromatography

Ion exchange chromatography separates protein samples based on their electrostatic properties. In cation exchange chromatography, positively charged protein molecules bind to a negatively charged resin and *vice versa* for anion exchange. After the protein of interest has bound to the beads, all unbound proteins are removed by washing with low ionic buffer. After binding and washing, the bound proteins are eluted by linearly increasing NaCl concentration.

CarD and CarD^{N} were both purified by anion exchange chromatography using a MonoQ HR 10/100 GL column (GE Healthcare) on an ÄKTA purifier (GE Healthcare). Protein samples were dialyzed against buffer A (**Table 6**) overnight using 35 kDa MWCO snake skin pleated dialysis tubing (Thermo Scientific). The column was prepared by spiking it with 20 mL of 2 M NaCl to displace all other ions and then equilibrating with buffer A (**Table 6**). Samples were loaded onto the MonoQ column and differentially eluted using a linear gradient of increasing high salt buffer B (**Table 6**) at a constant flow rate of 0.5 mL/min. Fractions corresponding to absorption peaks were analysed by SDS-PAGE.

2.7.2.4 Size Exclusion Chromatography

Size exclusion chromatography (SEC) – also known as gel permeation chromatography or gel filtration - separates proteins based on their molecular size. In this study, *Taq* RNAP β1^{m} was purified by SEC. The protein was concentrated until 1 mL using a 10 kDa MWCO vivaspin concentrator. A Superdex 75 16/60 column (bed volume 24 mL) was equilibrated with 2 column volumes of degassed running buffer (1 x PBS pH 7.4) on an ÄKTA purifier at 10 °C. 800 µL of protein sample was injected onto the column at a flow rate of 0.5 mL/min and the elution profile was monitored at 280 nm. Peak fractions were analysed by SDS-PAGE.

2.8 Pull-Down Experiments

This assay involves the immobilization of one protein (bait) of a suspected protein-protein interaction on a matrix through an affinity tag and the selective retention of a second protein (prey) through their direct interaction.

2.8.1 Glutathione-S-Transferase Pull Down

Here GST-CarD and GST-CarD^{N} were used as bait while *Taq* RNAP β1^{m} was used as prey.

2.8.2 Ni^{2+}-NTA Pull Down Assay

His_6 tagged *Taq* RNAP $\beta1^m$ was immobilized on Ni^{2+}-NTA beads while CarD and $CarD^N$ were used as prey.

Optimal complex formation for CarD/*Taq* RNAP $\beta1^m$ and $CarD^N$/ *Taq* RNAP $\beta1^m$ were tested by using buffers and salt conditions outlined in **Table 15**.

Table 15: Conditions Explored for Optimal CarD/RNAP Complex Formation

Buffer	Acetate	MES/NaOH	PBS	Tri-HCl	Glycine-NAOH	Glycine-NAOH
Concentration	50	50	1 x	150	100	100
pH	5	6	7.4	8	9	10
NaCl	200	50	40	150	100	30
Na_2HPO_4	0	100	50	150	10	20
Li_2S0_4	10	5	50	20	40	0
Glycerol	5	5	5	5	5	5

Concentrations in mM, except PBS (1 x) and glycerol (% v/v).

2.8.3 Small Scale Pull Downs

All proteins were purified by affinity chromatography using appropriate matrices for each tag – GS beads for GST and Ni-NTA beads for His_6. Bead samples for each protein were split into two equal parts A (bait) and B (prey). Part A was split into six equal fractions in 1.5 mL Eppendorf tubes and each incubated for 5 min with 500 µL of one buffer from Table 15. Tubes were centrifuged at 10500 rcf and the supernatant discarded. B fraction proteins were cleaved/eluted and purified as outlined above and concentrated to 10 mg/mL using a 10 kDa MWCO Vivaspin concentrator. The purified protein fractions were centrifuged at 10500 rcf

for 10 min to remove precipitates, divided into six equal divisions and added to A fraction subdivisions of its corresponding bait protein. 500 µL of corresponding buffers from table 15 were added to respective bait/prey mixture tubes. The bait/prey mixtures were incubated for 4 h at 4 °C on a roller mixer, centrifuged at 10500 rcf and the bead supernatants carefully removed by pipette and transferred to a separate tube. Beads were washed with respective designated buffers by centrifugation at 10500 rcf. For Ni^{2+}-NTA bait protein, potential protein complexes were eluted with 500 µL of 1 x PBS, 250 mM imidazole. GST-tag bait protein was cleaved by adding 50 µL PreScission protease, incubating overnight at 4 °C on a roller mixer and eluting by centrifuging at 10500 rcf at 4 °C for 3 min. This was repeated twice. Samples were analysed in 15 % SDS-PAGE.

2.8.4 Large Scale Pull Downs

For large-scale interaction studies of CarD and $CarD^N$ with *Taq* RNAP $\beta1^m$, production was scaled up to 1 L. To allow optimal protein interaction, 1 x PBS pH 7.4 with 5 % (v/v) glycerol was used as this buffer most closely resembles physiological cytosolic conditions. In contrast to the preceding section, GS beads from 1 L cultures were not subdivided. Washing and elution steps were repeated 5 times with 10 mL of buffer each time. Instead of centrifugation, the buffer was allowed to elute from the drip column and collected for analysis by 15 % SDS-PAGE. Before adding *Taq* RNAP $\beta1^m$ to GST-CarD or GST-$CarD^N$, it was purified by SEC.

3.0 Results

3.1 RNAP β1 Delimitation

At the start of this project, clones for the production of *Mtb* CarD and *Taq* RNAP $\beta1^m$ were available. However, as the two proteins derive from different organisms, data obtained for their interaction may be of limited value in assessing the interaction of the corresponding physiological complex. To overcome this limitation, a clone for the production of *Mtb* RNAP $\beta1^m$ had to be prepared by amplifying the corresponding gene fragment from genomic DNA (Westblade et al. 2010).

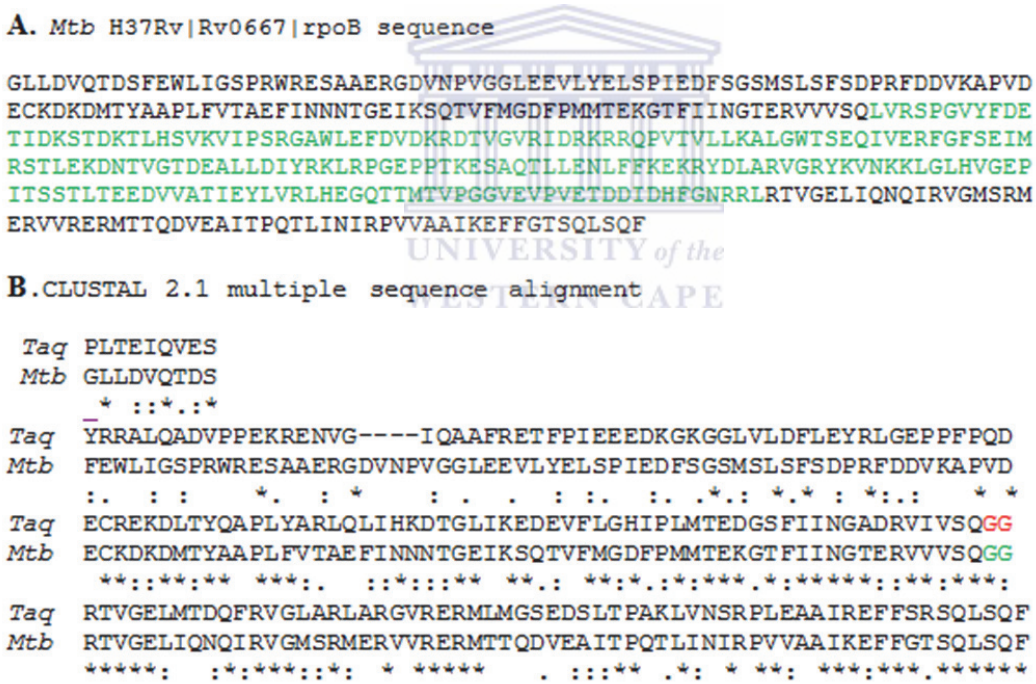

Figure 2: The *Mtb* Gene for RNAP

A: Amino acid sequence for *Mtb* RNAP β1 from Tuberculist. The central green highlighted region corresponds to the β1b domain that was to be replaced by a GG-linker to obtain the minimal, conformationally stable *Mtb* RNAP $\beta1^m$ as previously done for *Taq* RNAP β1.
B: An amino acid sequence alignment of *Taq* RNAP $\beta1^m$ and *Mtb* RNAP $\beta1^m$ (ClustalW). The inserted GG dipeptides (red and green) connect the N- and C-terminal parts of the β1a domains of *Taq* RNAP β1 and *Mtb* RNAP β1 respectively, after removal of the β1b domain. Symbols `*`, `:` and `.` indicate residues, which are conserved, semi conserved or only similar for the two proteins. Spaces indicate non-conserved residues.

The amino acid sequence of *Mtb* RNAP β1 domain is shown in **Figure 2A**. As outlined in section 2.6.4, the domain β1b (green in **Figure 2A**), that has evolutionarily been inserted into the β1a domain had to be replaced by a GG linker (**Figure 2B)** to generate a conformationally stable *Mtb* RNAP $\beta1^{m}$ protein**.** The sequence alignment indicates numerous sequence differences for *Mtb* RNAP $\beta1^{m}$ and *Taq* RNAP $\beta1^{m}$ (**Figure 2B**) which may dramatically affect the interaction with *Mtb* CarD.

Mtb rnaP gene

gga ctc ctt gac gtc cag acc gat tcg ttc gag tgg ctg atc ggt tcg ccg
cgc tgg cgc gaa tcc gcc gcc gag cgg ggt gat gtc aac cca gtg ggt ggc
ctg gaa gag gtg ctc tac gag ctg tct ccg atc gag gac ttc tcc ggg tcg
atg tcg ttg tcg ttc tct gac cct cgt ttc gac gat gtc aag gca ccc gtc
gac gag tgc aaa gac aag gac atg acg tac gcg gct cca ctg ttc gtc acc
gcc gag ttc atc aac aac aac acc ggt gag atc aag agt cag acg gtg ttc
atg ggt gac ttc ccg atg atg acc gag aag ggc acg ttc atc atc aac ggg
acc gag cgt gtg gtg gtc agc cag ctg gtg cgg tcg ccc ggg gtg tac ttc
gac gag acc att gac aag tcc acc gac aag acg ctg cac agc gtc aag gtg
atc ccg agc cgc ggc gcg tgg ctc gag ttt gac gtc gac aag cgc gac acc
gtc ggc gtg cgc atc gac cgc aaa cgc cgg caa ccg gtc acc gtg ctg ctc
aag gcg ctg ggc tgg acc agc gag cag att gtc gag cgg ttc ggg ttc tcc
gag atc atg cga tcg acg ctg gag aag gac aac acc gtc ggc acc gac gag
gcg ctg ttg gac atc tac cgc aag ctg cgt ccg ggc gag ccc ccg acc aaa
gag tca gcg cag acg ctg ttg gaa aac ttg ttc ttc aag gag aag cgc tac
gac ctg gcc cgc gtc ggt cgc tat aag gtc aac aag aag ctc ggg ctg cat
gtc ggc gag ccc atc acg tcg tcg acg ctg acc gaa gaa gac gtc gtg gcc
acc atc gaa tat ctg gtc cgc ttg cac gag ggt cag acc acg atg acc gtt
ccg ggc ggc gtc gag gtg ccg gtg gaa acc gac gac atc gac cac ttc ggc
aac cgc cgc ctg cgt acg gtc ggc gag ctg atc caa aac cag atc cgg gtc

Figure 3: Rare Codon Usage

The DNA sequence for *Mtb* RNAP β1 domain highlighting rare codons with respect to translation in *E. coli*. Rare codons for isoleucine, arginine and proline are respectively colored blue, red and orange

The gene construct encoding *Mtb* RNAP β1 contains six codons considered rare in *E. coli* including four proline codons (**Figure 3**). Generally, mRNA with less than ten rare codons should allow for unhindered translation in BL21 and BL21+ cells (Berg & Silva 1997).

3.1 Amplification of *Mtb rnaP β1* Gene

Two distinct cloning strategies were followed to amplify the *Mtb rnaP β1* gene – restriction free (RF) and conventional (normal) PCR (**Figure 4)**. The reason for trying two alternative

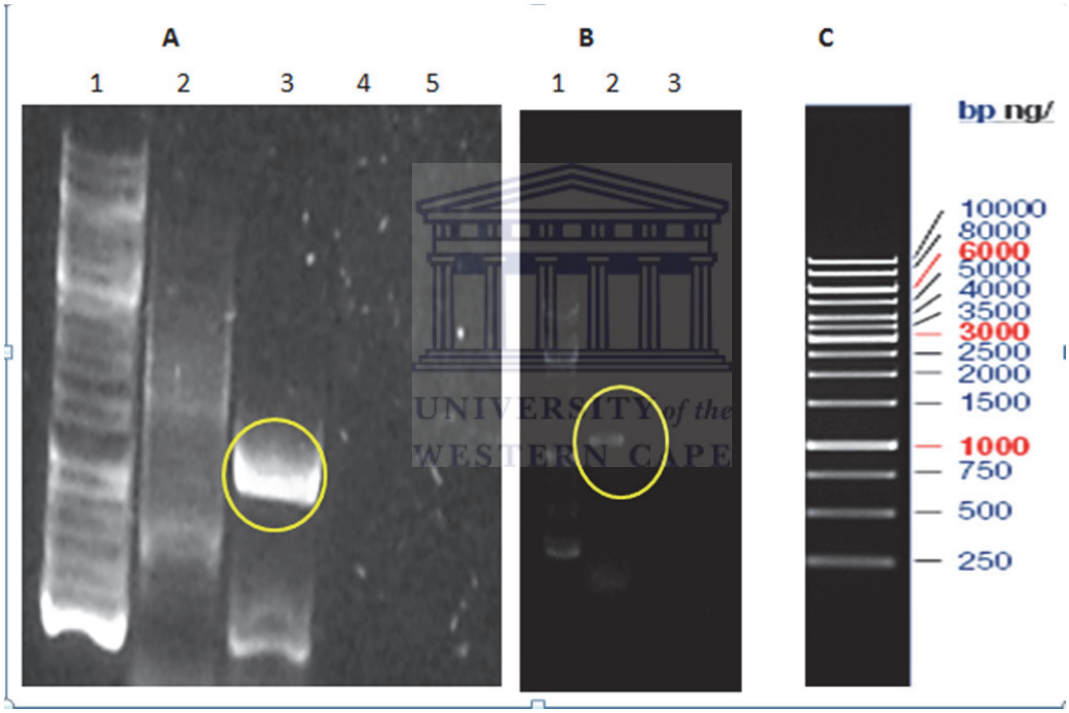

Figure 4: Amplification of *Mtb* r*naP β1* Gene from Genomic DNA

A Lane 1: 1 kb gene ladder. Lane 2: Product from conventional PCR. The expected PCR product of a 1000 bp is not observed. Lane 3: Successful first phase PCR product from genomic DNA for restriction-free (RF) cloning marked by a yellow circle. Lanes 4 and 5: Control experiments for conventional and RF PCR, run without primers. No amplification is observed, as expected.
B, Lane 1: 1kb gene ladder; Lane 2: Repeat of the RF first phase amplification. The successful product is marked by a yellow circle. Lane 3: Control without primers. **C:** 1kb gene ladder

methodologies is that amplification of genes from GC-rich *Mtb* genomic DNA frequently

proves difficult in practice. Following a double track strategy increases the likelihood of achieving successful amplification.

Attempts to amplify the *Mtb rnaP β1* gene from genomic DNA by conventional PCR were not successful for the annealing temperature of 50 °C (**Figure 4A,** lane 2). The gene of interest is 1167 nucleotides long and should hence be visible just above the 1000 bp band of the DNA standard. By contrast, the RF first phase PCR successfully (**Figure 4A** Lane 3: band within yellow circle) and reproducibly (**Figure 4B,** band within yellow circle) amplified the gene from the same genomic DNA at the same annealing temperature.

3.1.1 Gradient PCR

To improve the product yields for the RF first phase PCR, the experimental conditions were optimized. **Figure 5** documents the yields for different annealing temperatures.

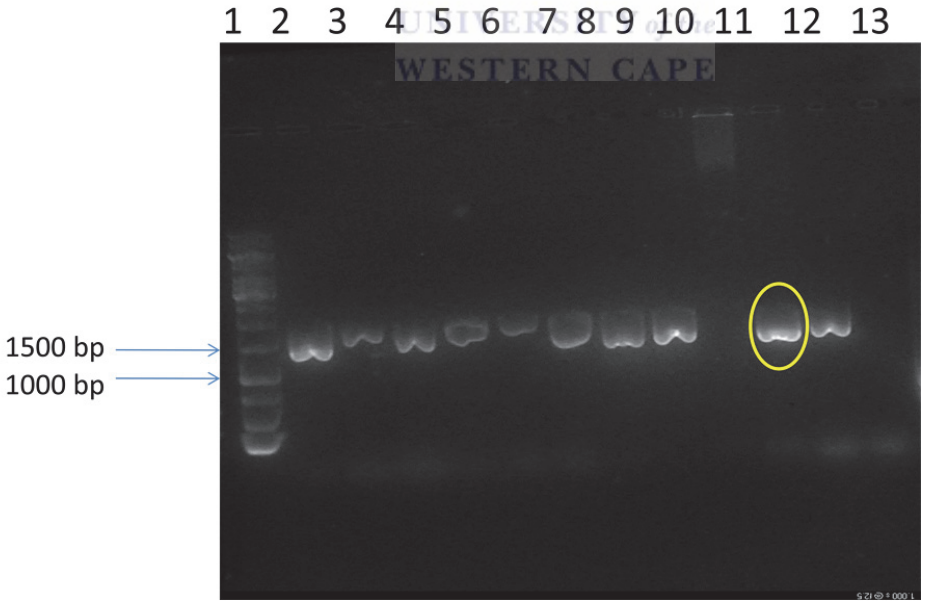

Figure 5: Annealing Temperature Gradient for RF PCR First Phase

Lane 1: 1 kb Marker. Lanes 2-13: PCR products for temperatures 45.0, 46.3, 47.2, 49.9, 52.8, 55.7, 57.0, 59.6, 62.9, 63.8, 64.4 and 65.0 °C. The product lies between has a size between 1500 and 1000 bp. The amplification was successful for all temperatures except for 62.9 °C and 65.0 °C. However 63.8 °C as shown by yellow circle, shows the best result.

The RF first phase PCR yielded the expected products for all annealing temperatures tested but with the highest yield recorded for 63.8 °C (**Figure 5**, band marked by yellow circle, size 1167 bp). Conventional PCR by comparison did not yield an amplified DNA product for all temperatures between 45.0 and 65.0 °C. To proceed with conventional PCR, the RF first phase PCR product (yellow circle, **Figure 5**) was used as template DNA for conventional PCR. This worked particularly well resulting in higher yields of PCR product (**Figure 6**, lane

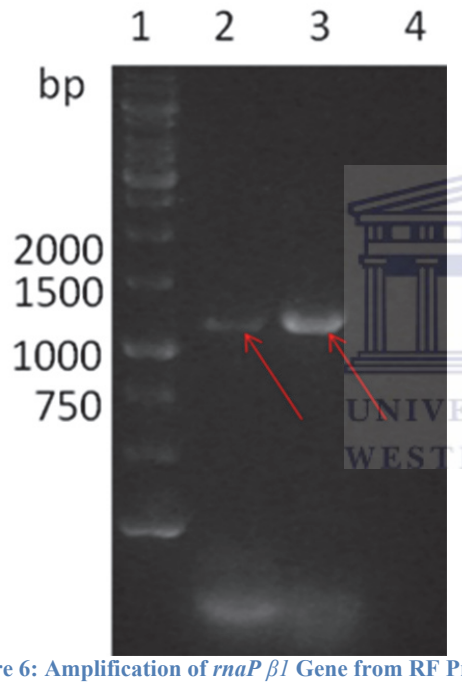

Figure 6: Amplification of *rnaP β1* Gene from RF Product

Lane 1: 1kb gene ladder; Lane 2: RF phase one PCR product amplified from genomic DNA (left red arrow); Lane 3: conventional PCR product amplified from RF product using a two-step PCR (right red arrow); Lane 4: Control without DNA polymerase

3) than RF PCR (**Figure 6**, lane 2): 200 ng/µL for conventional PCR of compared to 124 ng/µL for RF.

The product of conventional PCR (**Figure 6,** lane 3) was successfully digested by restriction enzymes, ligated to expression plasmid, and transformed into DH5α cells. The successful incorporation was confirmed by Sanger sequencing: **Figure 7** illustrates the the beginning of

the corresponding chromatogram. By contrast, the second phase of RF PCR, incorporation of target gene into the production plasmid, was not successful. However, as the target of the experiment had been achieved in the successful cloning of *Mtb* RNAP β1, the RF cloning strategy was not pursued any further.

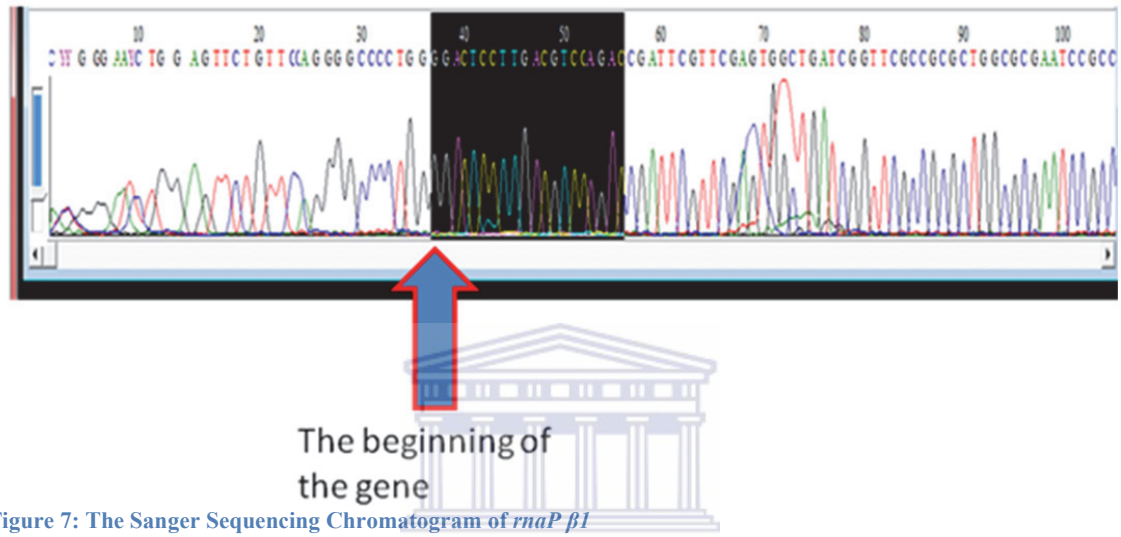

Figure 7: The Sanger Sequencing Chromatogram of *rnaP β1*

The first 20 nucleotides of the gene *rnaP* β1 are highlighted by black shading with the 3' end being pointed out by a large arrow marks. Each colour codes for one of the four nucleotides. The sharp peaks are sharp unambiguously providing the anticipated nucleotide sequence.

Sanger sequencing from the 3' end of the target gene confirmed the first 800 nucleotides of the gene. Diminishing signals thereafter yielded increasingly ambiguous results, as is typical for this technique. A second sequencing reaction from the 5' end complemented the missing data confirming the correct amplification and incorporation of the entire *Mtb rnaP β1* gene into the expression vector.

3.1.2 Generation of the CarD^{N}

To yield a GST-tagged CarD fusion protein, the *carD* gene had previously been cloned into the multicloning site of pGEX-6P-1 (Lilia Polle, personal communication) using the restriction enzymes *BamHI* and *NotI* and incorporating two stop codons before the *NotI*

restriction site. A PreScission protease cleavage site encoding sequence in the pGEX-6P-1 vector precedes the *BamHI* restriction site. **Figure 8** provides a schematic representation of the *carD* construct in pGEX-6P-1 for production in *E. coli.*

Figure 8: Schematic Representation of the c*arD* Gene Construct

Green, orange and white rectangles respectively represent GST, PreScission Protease Cleavage (PPC) site and CarD encoding regions. Red and yellow bars indicate *BamHI* and *NotI* restriction sites, respectively. The two stop codons prior to the *NotI* restriction site are marked in blue.

To generate a construct that will produce the GST-tagged N-terminal domain of CarD ($CarD^N$), a stop codon was inserted between codons 64 and 65 of *carD* by SDM. The PCR product for this reaction was not visible in an agarose gel. The ligated product was, however, transformed into Gene Hog cells, which upon cultivation, cell lysis, DNA isolation and Sanger sequencing confirmed the incorporation of the TAA stop codon at the expected position.

3.2 CarD purification

A one-step purification of CarD by GST affinity chromatography produced fairly pure CarD as evaluated by 15 % SDS-PAGE gel (**Figure 9).** This protein was observed to degrade within two days presumably due to the presence of host-strain specific proteases. This degradation limited progress of further experiments. After testing a range of inhibitors and other additives, 5 % (v/v) glycerol was found to extend the lifetime of CarD to at least 67 days when added to both purification and storage buffers.

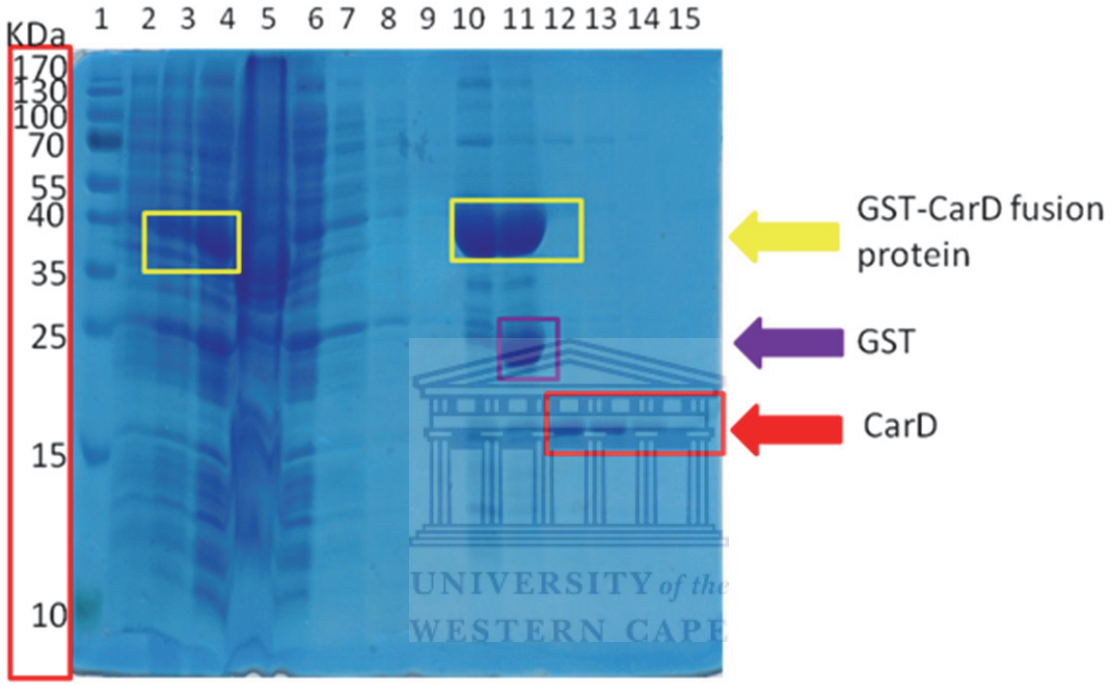

Figure 9: SDS-PAGE Analysis of CarD Purification

Lane 1: Prestained protein marker; Lane 2: *E. coli* full-cell sample prior to IPTG induction; Lane 3: 16 h after induction; Lane 4 Soluble fraction sample after centrifugation at 29000 RCF at 4 °C; Lane 5: Urea-solubilized insoluble fraction after centrifugation; Lane 6: Elution of non-coupled proteins after coupling to GS beads for 4 h; Lanes 7 and 8: Wash fractions after washing with 10 mL of 1 x PBS pH 7.4; Lane 9: Clean GS beads before adding the soluble cell fractiond; Lane 10: GS beads after elution of unbound protein; Lane 11: Partly cleaved fusion protein with PreScission protease. Most fusion protein is still uncleaved; Lanes 12 to 15: CarD elution fractions. Yellow, purple and red boxes and arrows indicate GST-CarD fusion protein, GST and CarD, respectively.

CarD production and purification is demonstrated in **Figure 9.** The GST-CarD fusion protein is highlighted by yellow boxes. The fusion protein is evident in the soluble fraction (lane 4) but not in the insoluble fraction (lane 5) implying that it is highly soluble. Following removal of unbound or non-specifically bound proteins from the column (lane 10), and proteolytic

separation of GST and CarD by PreScission protease (lane 11), GST (purple box) remains bound to the glutathione sepharose. Clearly, cleavage is fairly slow as most fusion protein is still intact (fusion protein band in lane 11 is essentially indistinguishable from that in lane 10 18 h after adding the protease). The small amount of eluted CarD (**Figure 9,** red boxes in lanes 12 to 15 was, nevertheless, fairly pure with small contaminants of fusion protein and another protein around 70 KDa.

Pooling all fractions of eluted CarD (lanes 12-15, **Figure 9**) yielded a total of 6 mg CarD – clearly only a small fraction of the entire protein produced. Anion exchange purification was then attempted using a MonoQ column to remove the remaining impurities**.** The protein was injected onto the column under low salt conditions, resulting in most CarD binding as evidenced by the lack of a peak during the loading phase (**Figure 10A**). CarD was then eluted with a linear salt gradient ranging from 0 to 20 % buffer B (0 to 0.2 M NaCl). Although a strong peak is clearly visible in the elution curve (**Figure 10A**), fractions corresponding to this peak do not reveal a protein band for CarD in the corresponding 15 % SDS-PAGE (**Figure 10B)**, implying that CarD is destabilized or unfolded by the anion exchange column resulting in its rapid degradation.

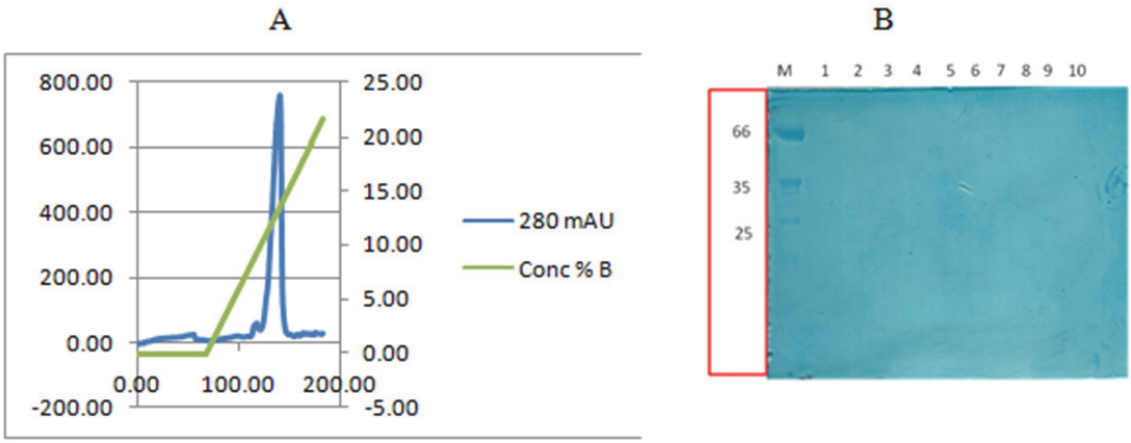

Figure 10: Anion Exchange Chromatography of CarD

A) Anion exchange chromatogram for CarD (blue) showing a peak presumably corresponding to eluting CarD at a buffer B concentration of ~12%. Absorbance is given in milli absorbance units (mAU). The green line plots the theoretical linear gradient of buffer B in %.
B) A 15 % SDS-PAGE analysis of fractions corresponding to the peak in A. Unexpectedly no protein bands are visible implying that protein is rapidly degraded after elution from the column

.

3.3 Purification of the CarD^{N}

As the construct for the CarD^{N} retains the original N-terminal GST-tag, the purification strategy for this domain was identical to that of the full-length protein. The outcome is shown in **Figure 11** below.

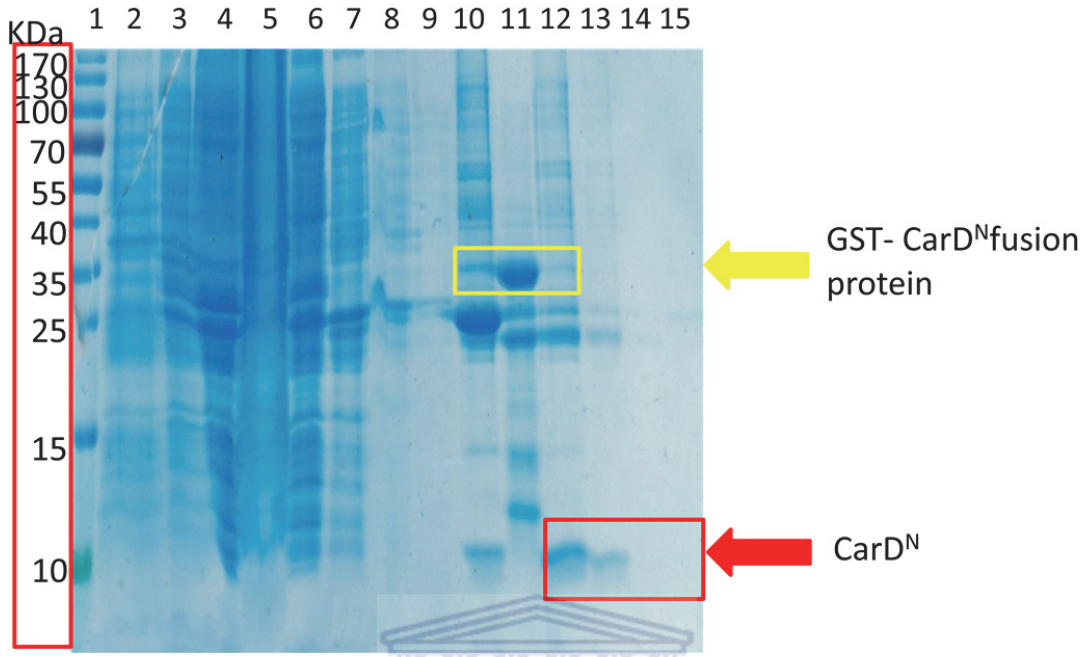

Figure 11: SDS-PAGE Analysis of the Affinity Purification of the CarD N-terminal Domain

Lane 1: Prestained protein marker; Lane 2: Full-cell sample prior to induction with IPTG; Lane 3: Full-cell sample 16 h after induction; Lane 4: Soluble fraction; Lane 5: Urea-solubilized insoluble fraction; Lane 6: GS beads after protein coupling; Lanes 7 and 8: Wash fractions after washing GS beads with 20 mL PBS pH 7.4. Lane 9 and 11: GS beads before and after coupling CarD^N; Lane 10: GST remaining attached to beads after cleavage and elution of CarD^N; Lanes 12-15: Eluted CarD^N. Yellow and red boxes and arrows mark GST-CarD^N and CarD^N.

GST-CarD^N, like GST-CarD is a highly soluble fusion protein of ~35 kDa (**Figure 11**, Lane 5 but not visible in lane 6, insoluble fraction). Proteolytic cleavage of GST-CarD^N fusion protein is significantly faster than that of CarD (**Figure 11,** lane 10). CarD^N was stable for more than 20 days in PBS pH 7.4. Eluted CarD^N was, however, not as pure as CarD (**Figure 9** lanes 12 and 13). The total yield of CarD (**Figure 11**, lanes 12 – 15) was 3.2 mg.

As described for full-length CarD above, purification of CarD^{N} by anion exchange chromatography (**Figure 12A**) allowed for the binding and specific elution of the protein from the column. However, again the protein was rapidly degraded such that no protein bands were visible in the SDS-PAGE analysis as indicated in **Figure 12B** below.

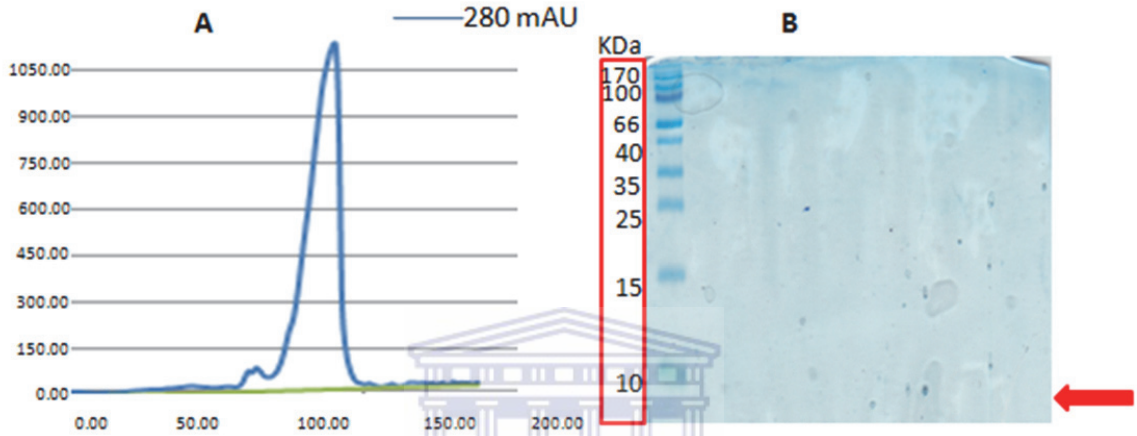

Figure 12: Anion Exchange Chromatography of CarD^{N}

A) Anion exchange chromatogram for CarD (blue). An absorption peak is observed that presumably corresponds to eluting CarD^{N} at a buffer B concentration of ~13%. Absorbance (y-axis) is in milli absorbance units (mAU). The x-axis corresponds to the NaCl concentration (in mM).
B) A 15 % SDS-PAGE analysis of fractions corresponding to the peak in **A**. Surprisingly, no protein bands at about size 7 kDa (red arrow) are visible implying that protein is rapidly degraded after elution.

3.4 Production and Purification of RNAP β1

Despite the successful cloning of *Mtb rnaP β1* gene and the transformation of the plasmids into BL21+ and Rosetta 2 cells, induction of recombinant protein production by IPTG did not result in observable production of the *Mtb* RNAP β1. The results for expression trials in BL21+ cells are documented in **Figure 13**.

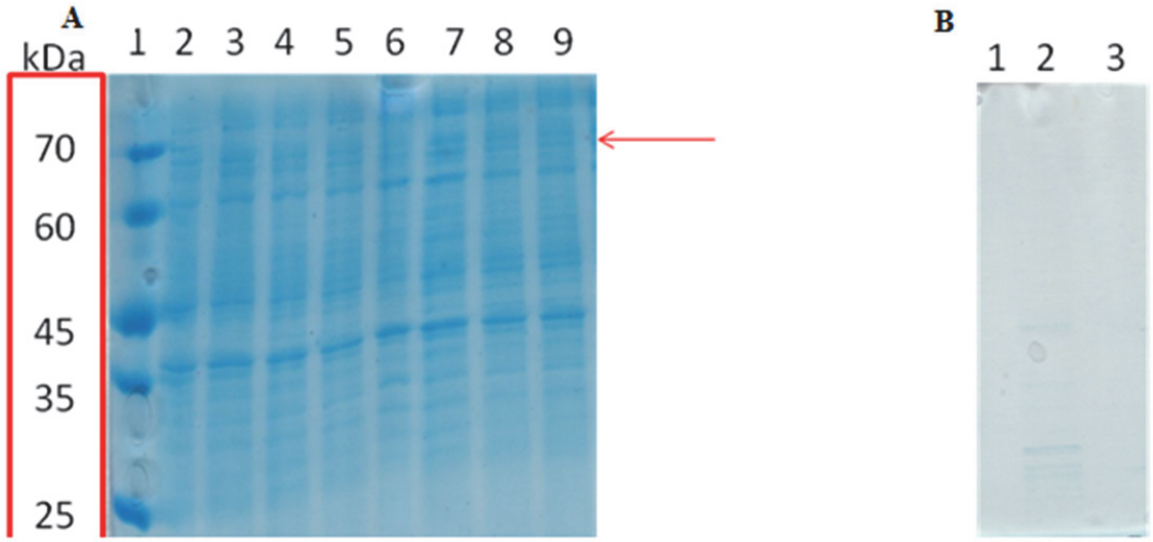

Figure 13: Expression Tests for *Mtb rnaP* gene in BL21+ Cells

A) Lane 1: Unstained protein marker; Lane 2-5: Soluble fractions respectively 0, 3, 10 and 16 h after induction with 0.1 mM IPTG; Lane 6: Insoluble fraction prior to induction; Lanes 7-9: Insoluble fractions respectively 3, 10 and 16 h after induction. The red arrow marks the expected size of the GST-RNAP β1 fusion protein at 70 kDa. **B)** Lanes 1 and 2: GS beads before and after addition of the soluble fraction; 3: Clean GS beads.

Unfortunately, *Mtb* RNAP β1 was not produced after inducing with 0.1 mM IPTG as samples from soluble and insoluble protein fractions 3, 10 and 16 h after induction indicate no increase in a band at the expected size (red arrow in **Figure 13 A**). Occasionally production levels of proteins can be too low to be visible on Coomassie stained SDS-PAGE. In such cases, produced proteins can be enriched using affinity chromatography to reveal residual protein production. The content of lysed cells 18 h after induction was therefore incubated with GS beads, eluted and the beads washed. However, samples of the beads do not indicate any fusion protein bound to the column (**Figure 13 B**), implying that**.**

Nevertheless, *Taq* RNAP $\beta1^{m}$ produced and expressed solubly in BL 21 cells. **Figure 14** shows the purification results.

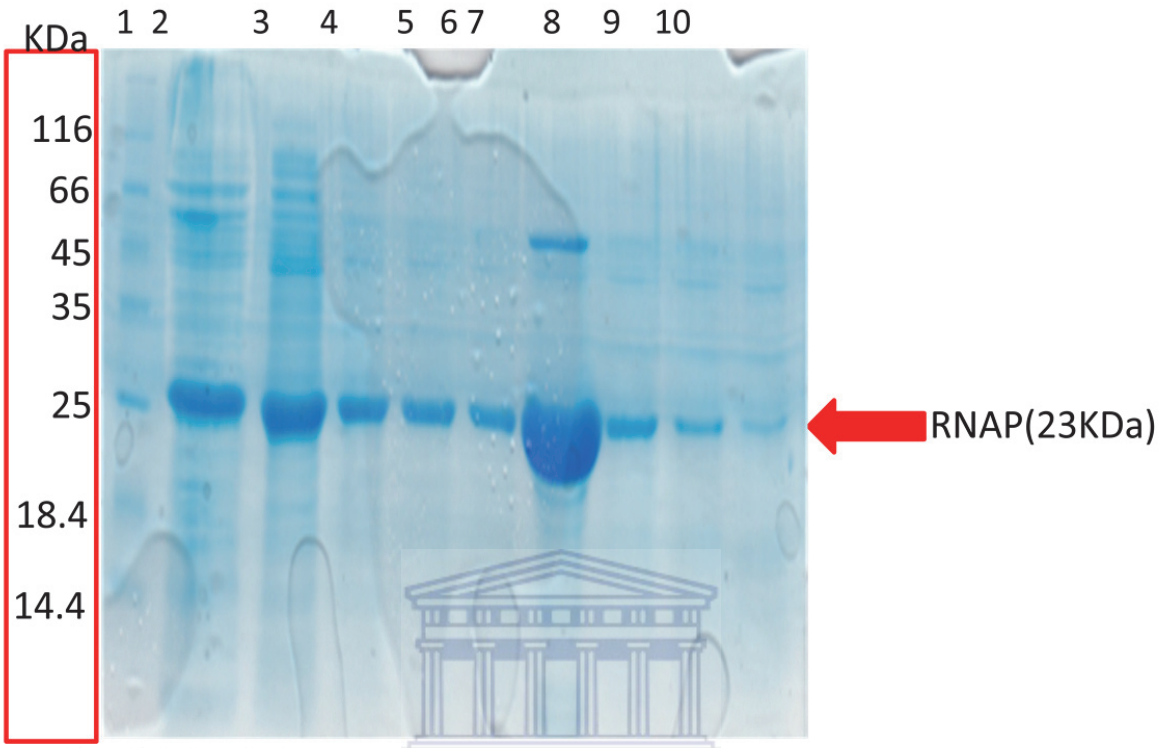

Figure 14: 15 % SDS-PAGE for *Taq* RNAP

Lane 1: Unstainable protein marker, lane 2: proteins that did not bind to the Ni^{2+}-NTA beads after incubation with soluble protein fraction. Lanes 3-6: wash fractions collected after Ni^{2+}-NTA beads were washed with 40 mL of Ni^{2+}-NTA wash buffer (**Table 10**). Lanes 7-10 are samples collected when *Taq* RNAP $\beta1^{m}$ was eluted with Ni^{2+}-NTA elution buffer. The red arrow shows the *Taq* RNAP $\beta1^{m}$ at 23 KDa

The *Taq* RNAP $\beta1^{m}$ comes out as early as wash step with buffer containing 10 mM imidazole. Almost all of the remaining *Taq* RNAP $\beta1^{m}$ was eluted by 10 mL of 250 mM imidazole containing buffer with just above 45 kDa contaminating protein (**Figure 14** lane 7). The *Taq* RNAP $\beta1^{m}$ was further purified using SEC and was pure as analysed by 15 % SDS-PAGE gel (**Figure 15**)

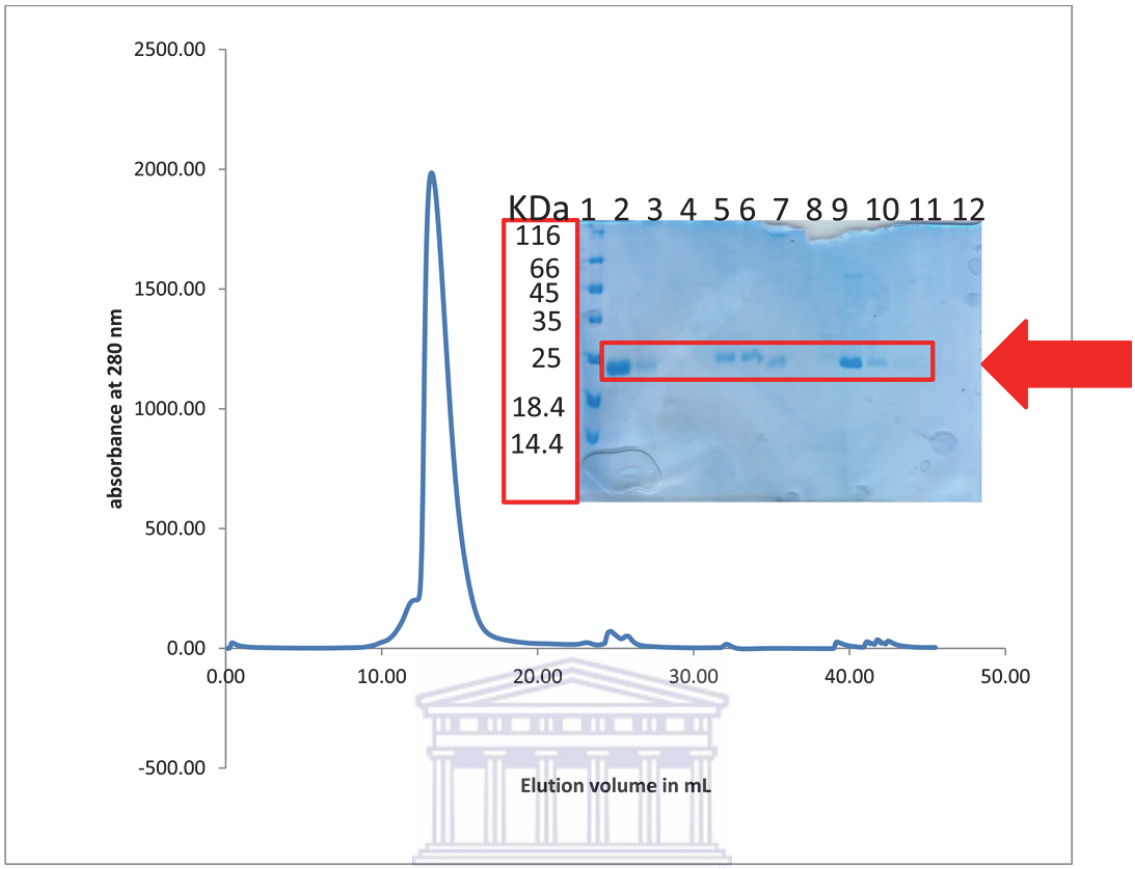

Figure 15: Size exclusion chromatography purification of *Taq* RNAP β1^{m}

SEC of *Taq* RNAP primarily using PBS pH 7.4 reveals a single absorption peak at an elution volume of 12 mL, presumably corresponding to purified RNAP. Insert – SDS-PAGE analysis of the size exclusion purification. Lane 1: Molecular weight marker; Lanes 2 to12: samples from fractions within the absorption peak. Fractions with pure RNAP are marked by a red rectangle and a red arrow.

Taq RNAP β1^{m} eluted from the GPC column in a broad peak covering 12 mL (**Figure 15**). The size corresponds to 23 kDa (red box and red arrow). The purity is improved compared to Ni^{2+}-NTA affinity chromatography. The pooled samples contained a total of 30 mg of *Taq* RNAP β1^{m}.

3.5 CarD/RNAP β1^{m} Complex

As described in the Introduction (**Section 1.5.2**), CarD is assumed to bind to RNAP through its N terminal domain. To investigate and confirm this interaction, two pull-down assays were developed and applied (**Figure 17**).. In the first, *Taq* RNAP β1 (bait) is allowed to bind to

Ni^{2+}-NTA whereupon CarD is added, incubated and analysed for binding (**Section 2.5**). In the second, GST-CarD (bait) is coupled to GS-beads and *Taq* RNAP β1 (prey) is added, incubated and tested for binding. Unexpectedly the former assay, Ni^{2+}-NTA pull down, did not work. The pull-down assay using GST-CarD as bait confirms the expected stoichiometric (1:1) complex between CarD and *Taq* RNAP β1.

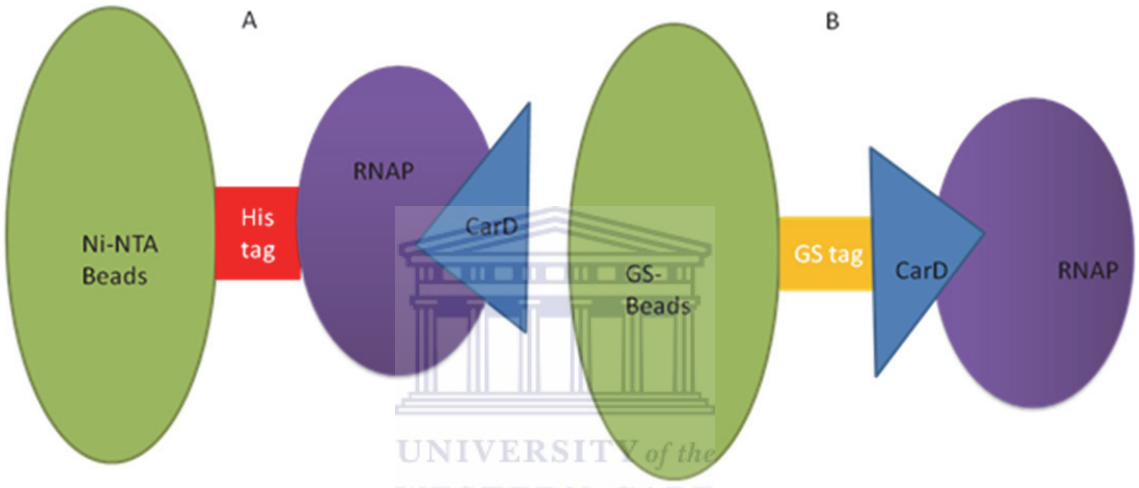

Figure 16: The Schematic Representation of Two Ways of Forming a CarD/RNAP Complex

A) First, RNAP is allowed to bind to the Ni^{2+}-NTA beads (bait) via its His_6-tag. CarD (prey) is then added to the immobilized RNAP and incubated to allow complex formation – presumably via the N-terminal domain of CarD B) GST-CarD fusion protein (bait) is allowed to bind to GS beads via the GST domain. *Taq* RNAP $\beta1^m$ is then added to the immobilized CarD and incubated to allow complex formation between CarD and *Taq* RNAP $\beta1^m$.

Figure 17 depicts the results for the pull-down assays in the form of 15 % SDS-PAGE for complex formation according to the schematic diagrams in **Figure 16 A.** Conditions were varied to optimize CarD complex formation (**Section 2.5.2**). Preparation of both bait and prey proteins were described in the preceding sections (**Figures 9, 11, 14 and 15**).

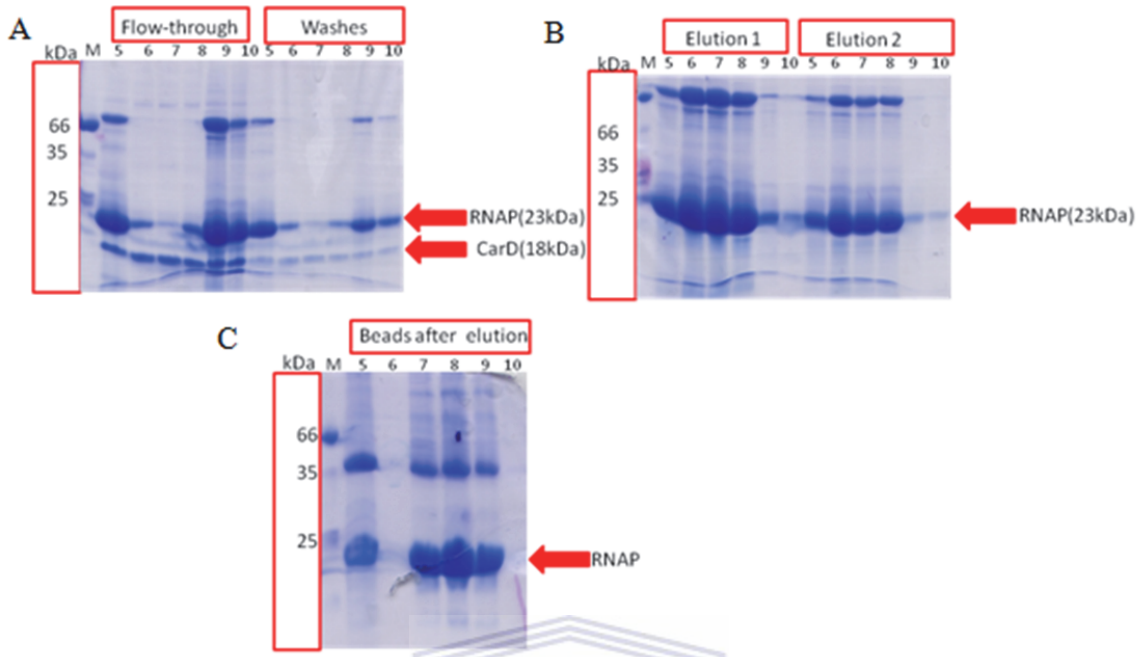

Figure 17: 15 % SDS-PAGE results for Ni-NTA pull down experiments

The number above each lane indicates the pH of the buffer used to form, purify and elute the complex from the Ni^{2+}-NTA beads.
A) Unbound proteins eluted after incubation of CarD with Ni-NTA/*Taq* RNAP $\beta1^{m}$ beads for 3 h using the indicated buffers. Red arrows mark *Taq* RNAP $\beta1^{m}$ (shown as RNAP in the picture) and CarD.
B) Elution of potential *Taq* RNAP $\beta1^{m}$/CarD complexes using the designated buffer enriched with 250 mM imidazole. The step was repeated to ensure complete elution of the coupled proteins. The red arrow marks eluted RNAP.
C) Beads from **B** after elution**.** The red arrow marks RNAP remaining on the beads.

CarD eluted along with other unbound proteins after incubation with *Taq* RNAP $\beta1^{m}$-loaded Ni^{2+}-NTA beads for all conditions tested. CarD was also eluted during the washing steps – as was some of the *Taq* RNAP $\beta1^{m}$ **(Figure 16A).** When beads potentially bearing *Taq* RNAP $\beta1^{m}$/CarD complexes were incubated with buffer containing 250 mM imidazole, only *Taq* RNAP $\beta1^{m}$ eluted **(Figure 16B)** and only RNAP remained on the beads **(Figure 16C).** This observation implied, that *Taq* RNAP $\beta1^{m}$ and CarD do not form a complex under the conditions tested here. To test whether this was always the case, the inverted experiment was

undertaken in which GST-CarD was coupled to the matrix and *Taq* RNAP $\beta 1^m$ added to check for complex formation (**Figure 17**).

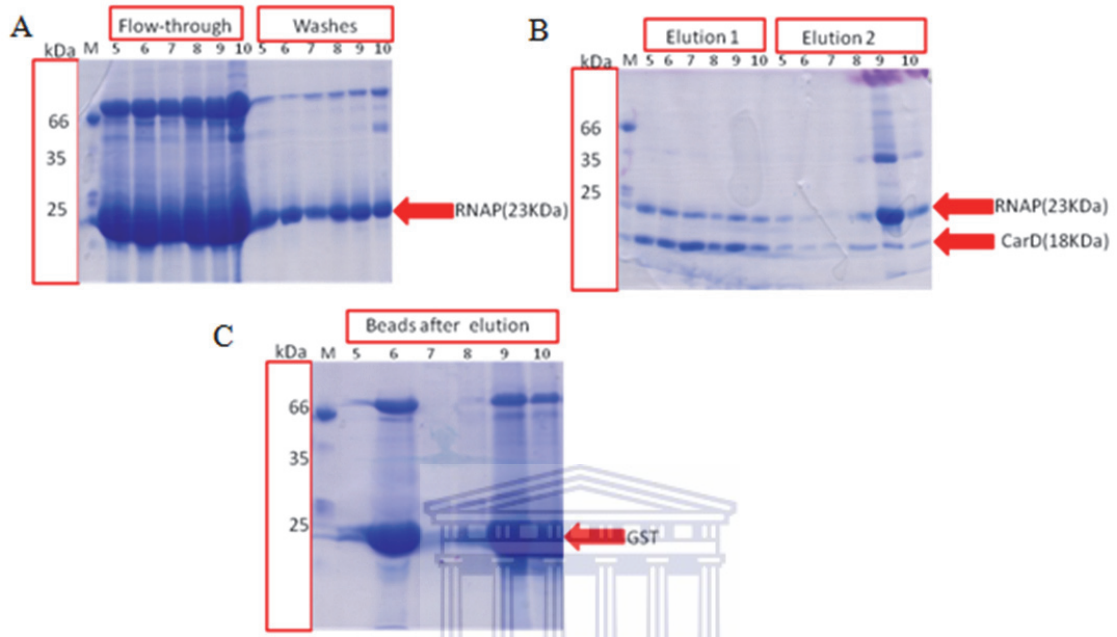

Figure 18: GST-CarD/ *Taq* RNAP $\beta 1^m$ Pull Down Assay

The pH of the buffer used during incubation, purification and elution of potential GST-CarD/*Taq* RNAP $\beta 1^m$ complex from GS beads is shown above each lane.
A) Left half: Unbound protein eluted after incubating *Taq* RNAP $\beta 1^m$ (labeled as just RNAP) with GST-CarD for 3 h. Right half: Elution after beads washing to remove all unspecifically bound proteins. Excess *Taq* RNAP $\beta 1^m$ elutes.
B) Elution of CarD/*Taq* RNAP $\beta 1^m$ complex after proteolytic release from GST bound to GS-beads with PreScission protease overnight. The beads were washed to ensure complete removal of released complex. Red arrows mark *Taq* RNAP $\beta 1^m$ and CarD at 23 kDa and 18 kDa respectively.
C) GS-beads after removal of CarD/*Taq* RNAP $\beta 1^m$. The red arrow marks GST remaining on the beads.

Excess *Taq* RNAP $\beta 1^m$ is seen to elute both in the flow through after incubating *Taq* RNAP $\beta 1^m$ with GST-CarD complex coupled to GS-sepharose as well as in the following wash stages (**Figure 17A**). As expected, GST-CarD remains firmly attached to the GS beads, due to the high affinity of GST for GS. Following cleavage of the GST-CarD fusion protein by PreScission protease, *Taq* RNAP $\beta 1^m$ and CarD are seen to elute in approximately stoicheometric amounts **(Figure 17B)**, indicating that the proteins do form a complex. Note that this complex is observed for all tested buffer conditions. Optimal complex formation is

seen for PBS pH 7.4 buffer with 5 % (v/v) glycerol. The proteolytic release of CarD is seen to have worked quantitatively, with no fusion protein but only GST remaining on the beads (**Figure 17C**).

Larger scale complex formation was attempted using PBS pH 7.4, 5 % (v/v) glycerol (**Figure 18**). Only the elution of Complex was shown here.

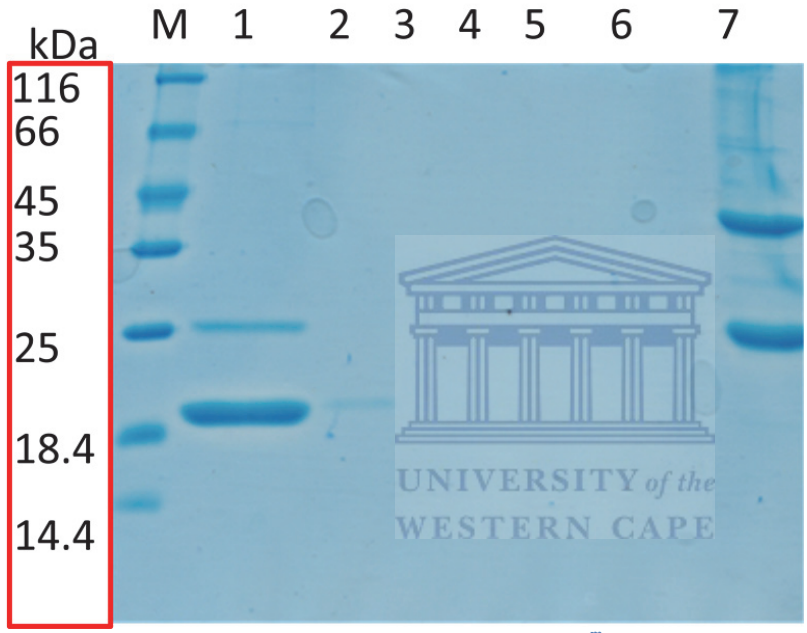

Figure 19: Large-scale Complexation of CarD/ *Taq* RNAP β1^{m}

M: Protein marker; Lanes 1-6: CarD/*Taq* RNAP β1^{m} complex eluted from the GS beads following incubation in the presence of PBS pH 7.4, 5 % (v/v) glycerol. This complex was cleaved from the beads by PreScission protease overnight at 4 °C; Lane 7: GS beads after complex elution

In contrast to the run in Figure 18, this round did not yield a 1:1 CarD/*Taq* RNAP β1^{m} complex as evidenced by a weaker *Taq* RNAP β1^{m} band (at ~26 kDa) compared to CarD (~20 kDa). A stoichiometric complex was recorded only for the first 10 mL (**Figure 19**, lane 2). Lane 7 demonstrates that both GST-CarD fusion protein (45 kDa) and GST (25 kDa) remain bound to the column after elution sample.

The complex CarD^{N}/*Taq* RNAP β1^{m} was produced similarly, using a GST pull down approach (**Figure 20**).

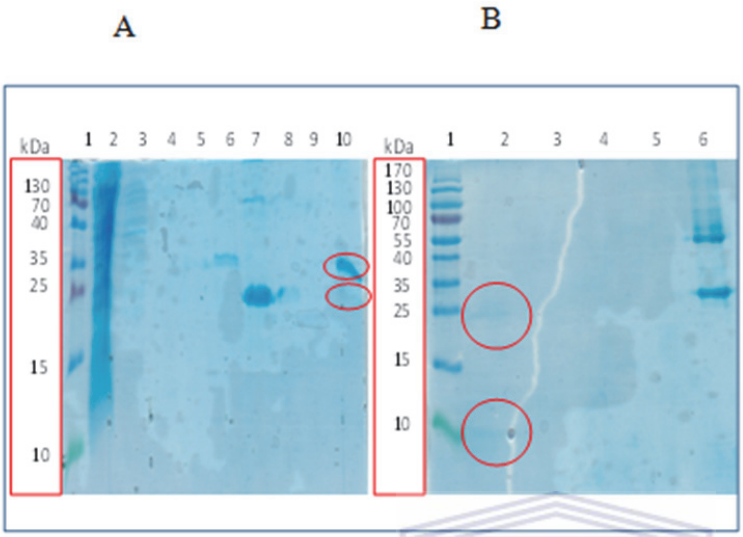

Figure 20: Analysis of CarD^{N}/ RNAP Complex

A) Lane 1: Protein marker; Lane 2: Soluble fraction after coupling of CarD^{N} to the GS beads; Lanes 3 to 5: Wash fractions to remove unbound proteins; Lane 6: GS beads after wash steps; Lanes 7 to 9: Removal of unbound proteins after incubating GS-bound CarD and *Taq* RNAP β1^{m} for 3 h; Lane 10: GS beads after elution of CarD^{N}/*Taq* RNAP β1^{m}. Red circles mark *Taq* RNAP β1^{m} and GST-CarD^{N} fusion protein at 23 and 34 kDa respectfully. **B)** Lane 1: Protein marker; Lanes 2 to 5: CarD^{N}/*Taq* RNAP β1^{m} complex was eluted from column after overnight cleavage with PreScission protease. Red circles mark CarD^{N} and *Taq* RNAP β1^{m} at 7 and 23 kDa respectively in lane 2. Lane 6: Beads after elution of the complex

CarD^{N} was produced as a GST-fusion protein (34 kDa, **Figure 20A** lane 6). In lane 7, *Taq* RNAP β1^{m} is visible as a dark band at 23 kDa. The complex was noted as 2 bands before cleavage from the GS beads (**Figure 20** lane 10) as shown by the circles even though the band associated with *Taq* RNAP β1^{m} is not as bold as that of the CarD^{N} fusion protein. However, the two proteins appeared in almost equal amounts after proteolytic release of the complex from the GS beads (**Figure 20 B** lane 1) as marked in red circles. No band was noted thereafter except for CarD^{N}-GST fusion protein and GST still stuck on the GS beads after elution (**Figure 20B** lane 6).

.

4.0 Discussion

4.1 Codon Usage and Gene Expression

Of a total of 64 possible three-base codons, 61 code for amino acids, two for stop codons and one for N-terminal formyl-methionine. Depending on the organism, however, some codons for the same amino acid are used much more frequently (major codons) than others (rare codons) (Andersson & Sharp 1996). The usage frequency of individual codons correlates with the availability of the corresponding tRNA molecules. Heterologous expression of a gene may thus suffer from reduced efficiency due to previously major codons being rare in the expression host. Alternative outcomes include inhibition of protein synthesis, reduced cell growth, decreased mRNA stability, frame shifts and deletions (Berg & Silva 1997). In *E. coli,* normally no more than 10 rare codons per gene are tolerated (Andersson & Sharp 1996).

The *Mtb rnaP β1* gene contains only six codons considered rare in *E. coli* (**Figure 3**) implying that it should be produced. Nevertheless, despite successful cloning, *Mtb* RNAP β1 was not produced in *E. coli* BL 21^{+} after induction with IPTG. Reasons may include the very high GC-content of *Mtb rnaP β1*. mRNA from GC-rich genes frequently form secondary structures (Berg & Silva 1997) resulting in low expression levels (Andersson & Sharp 1996; Berg & Silva 1997). This may partly explain the known difficulty in expressing genes of *Mtb* in *E. coli*. To overcome this limitation, genes may be codon harmonized by replacing guanines and cytosines near the 5'-end with adenines and thymines (Andersson & Sharp 1996).

4.2 Cloning of *rnaP β1* gene

4.2.1 Restriction Free and conventional Cloning

Restriction free (RF) cloning is a method allowing the insertion of a gene of interest at any desired location into a circular plasmid, independent of restriction sites, ligation or any modifications in both vector and gene (Ent & Löwe 2006; Ulrich et al. 2012). RF cloning is achieved in two independent PCR steps: in the first step the gene of interest is amplified using a set of plasmids, resulting in a linear product containing the target gene (fragment) flanked by 5' and 3' regions complementary to the target plasmid. In the second PCR, the product of the first round anneals to the target plasmid by its flanking regions, resulting in the incorporation of the gene of interest into the expression plasmid after amplification. The advantage of RF cloning lies in facile primer design and a technical simplification compared to conventional cloning, which may suffer from incompatible restriction sites between expression plasmid and insert, as well as inefficiency of unbalanced restriction digestion.

In this study, conventional and restriction free cloning methodologies were successfully combined after neither technique yielded the expected result on its own. The first phase of RF cloning allowed the amplification of *Mtb rnaP β1* from genomic *Mtb* DNA. Conventional cloning was not successful in this step, possibly due to the high GC content of the *Mtb* genome, the large size of the genomic DNA, the shortness of the primers (**Table 2**) and the associated low efficiency of the annealing step. RF cloning primers by contrast are significantly longer (**Table 2**) allowing for higher annealing probabilities.

In a second step, the linear DNA product of the first RF cloning step was successfully combined with primers for conventional cloning to generate a pre-restriction digest product for conventional ligation into an expression plasmid (**Figure 8**). Clearly, the gene of interest

was present in a significantly higher concentration as compared to genomic DNA, allowing for significantly more efficient annealing between primers and template DNA.

Unexpectedly, the second phase of RF cloning was not successful. This may be due to the method's inherent low efficiency for primers annealing to the vector for inserts of above 1000 nucleotides (Ent & Löwe 2006).

The overall successful cloning of *rnaP β1* gene of *Mtb* was confirmed by Sanger sequencing. This serves to underscore that unconventional combinations of techniques may yield results where the methods individually did not.

4.3 Site Directed Mutagenesis

The success of site directed mutagenesis (SDM) depends on the length of region to mutate or delete (Ent & Löwe 2006). SDM was used successfully to generate $CarD^{N}$ from CarD by the insertion of a stop codon. The generation of $CarD^{N}$ by inserting a stop codon after the codon for the 64^{th} amino acid was a relatively simple exercise as only 3 nucleotides needed to be inserted.

However, SDM failed to replace the β1b domain of the successfully cloned *Mtb rnaP β1* gene by a Gly-Gly linker to produce the conformationally stable *Mtb* RNAP $\beta1^{m}$ (see **section 4.2.1**). This experiment entails replacing 600 nucleotides by a mere 6 in a gene of 1167 nucleotides. This strategy has frequently been successful in the past *inter alia* for the equivalent experiment for *Taq* RNAP $\beta1^{m}$ (Westblade et al. 2010). The crystal structure of the complete *Taq* RNAP revealed the distance between the N- and C-termini of the inserted RNAP β1b to be 5.4 Å (Srivastava et al. 2013). Hence a GG-linker would be able to connect the two halves of the β1a subunit upon excision of β1b (Westblade et al. 2010).

The assumed CarD/RNAP protein-protein interaction area is located in the β1a subunit (Stallings et al. 2010). The successful cloning and production of *Mtb* RNAP $\beta1^{m}$ could have allowed the physiological complex with CarD to be studied in more detail as both proteins would derive from the same organism. Comparison of *Taq* and *Mtb* RNAP $\beta1^{m}$ (**Figure 2**) confirms that numerous exchanges of surface residues may result in *Taq* RNAP $\beta1^{m}$ being a poor substitute for *Mtb* RNAP $\beta1^{m}$.

An alternative strategy would have been to amplify the entire plasmid except the region to be deleted by employing primers that introduce a Gly and identical restriction sites to either end of the retained sequence. Using a single restriction enzyme and ligating the resulting circularized plasmid, the β1b domain could be removed – though depending on the restriction site, additional residues could potentially be introduced.

4.4 Protein Production and Purification

4.4.1 CarD and CarD N-terminal Domain

Both CarD and $CarD^{N}$ were soluble when produced and expressed in BL21 *E. coli* cells as indicated in **Figures 9 and 11.** This correlates with the computational prediction of CarD solubility in PBS pH 7.4 buffer (Priya & Megha 2012). The same analysis predicted CarD to be stable in this buffer. Experimentally, this proved not to be the case as CarD significantly degraded within 2 days. Instead, 5 % (v/v) glycerol was found to be required to stabilize CarD for more than 67 days. Why glycerol should stabilize CarD is not entirely clear but an explanation could be that glycerol shields exposed hydrophobic surface patches in CarD (Bondos &Bicknell 2003) increasing its stability (Gekkot & Timasheff 1981) and compactness (Vagenende et al. 2009). At the same time, glycerol is a chaotrope and its

stabilizing effect is highly concentration dependent. Too much glycerol could (partly) unfold CarD while too little would not help to stabilize CarD (Gekkot & Timasheff 1981).

A second purification step of CarD by anion exchange chromatography surprisingly led to its rapid degradation in less than 1 h (**Figure10**). The sudden destabilization of CarD led to the speculation that the selective removal of a cofactor such as an (oligo)nucleotide – shown to bind to the C-terminal domain of CarD (Garcı et al. 2010; Srivastava et al. 2013) – during anion exchange chromatography could be responsible. However, this interpretation had to be revisited when $CarD^N$ was similarly degraded after anion exchange chromatography, despite not being implicated in DNA or RNA binding. An alternative explanation could involve the effect of NaCl. Proteins purified by anion exchange chromatography are eluted at higher NaCl concentrations than the original buffer (**section 2.4.2.3**). For most proteins this would not be a problem, as "salting out" mostly only happens under very high salt conditions. CarD though may be more sensitive to a high ion concentration. Yet a further explanation may involve the high ion concentration in the matrix of the ion exchange column. This could partially unfold inherently unstable CarD increasing its sensitivity to proteolytic cleavage (Gekkot & Timasheff 1981).

$CarD^N$ did not require glycerol for stability, implying that the source of CarD instability involves the C–terminal domain. What is known about the C-terminal domain is that it binds DNA. However, the conclusion that nucleotides are critical to the stability of CarD may be premature.

4.4.2 Production of *Mtb* RNAP β1 versus that of *Taq* RNAP β1^{m}

While *Taq rnaP β1*^{m} in plasmid pET28a produced soluble protein in *E. coli* BL21+ (**Figure14),** *Mtb rnaP β1* in pGEX-6P-2 did not. A band observed in SDS-PAGE analysis (red arrow, **Figure 13**) roughly corresponded to the expected 70 kDa of the GST-*Mtb* RNAP β1 fusion protein. The protein, however, did not bind to GS beads implying that it is not protein of interest. Of course *Mtb* proteins are notoriously difficult to produce in *E. coli* due to their high GC-content (**section 4.1**). Using the related *Mycobacterium smegmatis*, which has a similarly GC-rich genome, as expression host could provide a solution. Except that the replication time of *M. smegmatis* is 3 h (Noens et al. 2011), significantly extending the time required for protein production.

4.4.3 CarD/RNAP Complex

Complex formation between *Taq* RNAP β1^{m} and CarD was investigated by two related yet inverted pull-down assays involving both GST and Ni^{2+}-NTA affinity chromatography to immobilize the bait. This technique is used extensively to confirm suspected interactions and to map interaction sites (Wysocka 2006). After coupling GST-CarD or GST-CarD^{N} fusion protein to the GS matrix, *Taq* RNAP β1^{m} was added and incubated. Unbound protein was eluted, the column rinsed, and the suspected complex eluted after proteolytic release of CarD. In the Ni^{2+}-NTA pull-down, His_6-tagged *Taq* RNAP β1^{m} was coupled to Ni^{2+}-NTA beads, CarD/CarD^{N} added and incubated, unbound proteins removed and the suspected complex eluted with 250 mM imidazole.

The results in **Figures 17, 18 and 19** indicate that the GST pull-down proved superior to its Ni^{2+}-NTA equivalent. It confirms the predicted complex between CarD and *Taq* RNAP β1^{m} (Stallings et al. 2010; Srivastava et al. 2013; Weiss et al. 2012; Priya & Megha 2012;

Westblade et al. 2010). It also confirms that CarD interacts with *Taq* RNAP $\beta1^{m}$ through its N-terminal domain (**Figure 20**) as both CarD and $CarD^{N}$ successfully recruit *Taq* RNAP $\beta1^{m}$. Despite experimental variations and imperfections, a stoichiometry of 1:1 best explains the observations.

Reasons for the failure of the Ni^{2+}-NTA pull-down assay, include the N-terminal His_6-tag of *Taq* RNAP $\beta1^{m}$ which could sterically interfere with complex formation (Srivastava et al. 2013). However, this is an unlikely explanation, as the His_6-tag is similarly present in the successful GST pull-down experiments. Instead it seems more likely that the high imidazole concentration used to wash and elute the suspected complex could actually destabilize the inherently weak complex leading to the premature elution of CarD.

While glycerol does stabilize CarD, it also has the disadvantage of increasing the buffer viscosity. This correspondingly slowed the proteolytic activity of PreScission protease dramatically reducing the cleavage of the GST-CarD fusion protein compared to GST-$CarD^{N}$ which did not require the addition of glycerol for stability (**Figures 18** and **19**).

$CarD^{N}$ appears to be related to a family of RNAP-binding domains including the RNA interacting domain (RID) of transcription repair coupling factor (TRCF). The crystal structure of TRCF/*Taq* RNAP $\beta1^{m}$ (Priya & Megha 2012; Weiss et al. 2012) therefore provides a model of the interaction of these complexes. Isoleucine 108, glutamate 110 and lysine 109 of TRCF-RID β1 are central to the interaction (Westblade et al. 2010). As these residues are conserved in CarD, they may be similarly critical to CarD/RNAP complex formation. A crystal structure of CarD/RNAP β1 could serve to confirm this analogy and could answer questions such as: 1) What is the role of the C-terminal domain of CarD? 2) What forces stabilize the complex? 3) Which residues are important in the interaction?

5.0 Conclusion/Outlook

This work experimentally confirms the previously proposed interaction of CarD from *Mtb* through its N-terminal domain with *Taq* RNAP $\beta 1^{m}$. This complex is critical for *Mtb* to attain a state of dormancy that *inter alia* limits the efficacy of antibiotics in fighting this disease. The crystal structure of this complex has not yet been elucidated, which leaves a void in understanding the molecular basis of how CarD interacts and regulates RNAP. The C-terminal domain also needs to be investigated in more depth to identify its function in dormancy or in the interaction with RNAP.

The CarD/RNAP complex provides an exciting avenue for TB therapeutic intervention. Small molecules could potentially be used to disrupt this complex forcing *Mtb* out of dormancy to restore its susceptibility to conventional antibiotics. The efficacy of this strategy is underscored by the mechanism of rifampicin, which inhibits transcription from the DNA dependent RNAP. This shows that the targeting of RNAP can be utilized in combating *Mtb*, highlighting the crucial importance of this complex.

6.0 References

Andersson, S.G.E. & Sharp, P.M., 1996. Codon usage in the *Mycobacterium tuberculosis* complex. *Microbiology*, 142(14), pp.915–925.

Anon, 2012. Global tuberculosis report. WHO report.

Barker, M.M., Gaal, T. & Gourse, R.L., 2001. Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. *Journal of Molecular Biology*, 305(4), pp.689–702.

Barnes, D.S., 2000. Historical perspectives on the etiology of tuberculosis. *Microbes and Infection*, (2), pp.431–440.

Berg, O.G. & Silva, P.J.N., 1997. Codon bias in *Escherichia coli*: the influence of codon context on mutation and selection. *Nucleic Acid Research*, 25(7), pp.1397–1404.

Betts, J.C. et al., 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Molecular Microbiology*, 43(3), pp.717–31.

Bondos, S.E. & Bicknell, A., 2003. Detection and prevention of protein aggregation before , during , and after purification. *Analytical Biochemistry*, 316(1), pp.223–231.

Bordbar, A. et al., 2010. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Molecular Systems Biology*, 422(6), pp.2010–21.

Brockmann-gretza, O., 2006. Global gene expression during stringent response in *Corynebacterium glutamicum* in presence and absence of the rel gene encoding (p) ppGpp synthase. *BMC Genomics*, 230(7), pp.1–15.

Clark-Curtiss, J.E. & Haydel, S.E., 2003. Molecular genetics of *Mycobacterium tuberculosis* pathogenesis. *Annual Review of Microbiology*, 57(1), pp.517–49.

Daniel, T.M., 2006. The history of tuberculosis. *Respiratory Medicine*, 100(1), pp.1862–1870.

Davis, J.M. & Ramakrishnan, L., 2009. The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell*, 136(1), pp.37–49.

Deb, C. et al., 2009. A Novel in vitro multiple-stress dormancy model for *Mycobacterium tuberculosis* generates a lipid-loaded, drug-tolerant, dormant pathogen. *Plos one*, 4(6), pp.6077–80.

Ent, F. Van Den & Löwe, J., 2006. Short note RF cloning: a restriction-free method for inserting target genes into plasmids. *Journal of Biochemical and Biophysical Methods*, 67(1), pp.67–74.

Ferullo, D.J. & Lovett, S.T., 2008. The stringent response and cell cycle arrest in *Escherichia coli.Plos Genetics*, 4(12), pp.21–23.

Flynn, J.L. & Ernst, J.D., 2000. Immune responses in tuberculosis. *Current Opinion in Immunology*, 12(1), pp.432–436.

Fontán, P. et al., 2008. Global transcriptional profile of *Mycobacterium tuberculosis* during THP-1 human macrophage infection. *Infection and immunity*, 76(2), pp.717–25.

Gaal, T., 1997. Transcription regulation by initiating NTP Concentration: rRNA Synthesis in acteria. *Science*, 278(7), pp.2092–2097.

Galbis-martı, M., Fontes, M. & Murillo, F.J., 2004. The high-mobility group A-type protein CarD of the bacterium *Myxococcus xanthus* as a transcription factor for several distinct vegetative genes. *Genetics*, 167(1), pp.1585–1595.

Garcı, F. et al., 2010. CdnL , a member of the large CarD-like family of bacterial proteins, is vital for *Myxococcus xanthus* and differs functionally from the global transcriptional regulator CarD. *Nucleic Acid Research*, 38(14), pp.4586–4598.

García-heras, F. et al., 2013. High-mobility-group A-like CarD binds to a DNA site optimized for affinity and position and to RNA polymerase to regulate a light- inducible promoter in *Myxococcus xanthus*. *Journal of Bacteriology*, 195(2), pp.378–388.

Gekkot, K. & Timasheff, S.N., 1981. Mechanism of protein stabilization by glycerol : preferential hydration in glycerol-water mixtures?pp.4667–4676.

Grosset, J., 2003. *Mycobacterium tuberculosis* in the extracellular compartment : an underestimated adversary. *Antimicrobial Agents and Chemotherapy*, 47(3), pp.833–836.

Jain, S.K. et al., 2008. Antibiotic treatment of tuberculosis:old problems, new solutions. *Microbe*, 3(6), pp.285–292.

Kapoor, N. et al., 2013. Human granuloma in vitro model, for TB dormancy and resuscitation. *PloS one*, 8(1), p.e53657.

Keertan, D. et al., 2010. The immunology of tuberculosis: from bench to bedside. *Respirology*, 15(1), pp.433–450.

Kumar, R. & Sanyal, S., 1998. Dormancy, persistence and survival in the light of protein synthesis. *Journal of Bacteriology*, 24(12), pp.1234–45.

Magnusson, L.U., Farewell, A. & Nyström, T., 2005. ppGpp: a global regulator in *Escherichia coli. Trends in Microbiology*, 13(5), pp.236–42.

Marino, S. & Kirschner, D.E., 2004. The human immune response to *Mycobacterium tuberculosis* in lung and lymph node. *Journal of Theoretical biology*, 227(1), pp.463–486.

Mcdonough, K.A., Kress, Y. & Bloom, B.R., 1993. Pathogenesis of tuberculosis : interaction of *Mycobacterium tuberculosis* with macrophages. *Microbes and Infection*, 61(7), pp.2763–2773.

Nickels, B.E. & Hochschild, A., 2004. Regulation of RNA polymerase through the secondary channel. *Cell*, 118(3), pp.281–4.

Noens, E.E. et al., 2011. Improved mycobacterial protein production using a *Mycobacterium smegmatis* groEL1 Δ C expression strain. *BMC Biotechnology*, 11(1), p.27.

Ordway, D. et al., 2007. The cellular immune response to *Mycobacterium tuberculosis* infection in guinea pig. *Journal of Immunology*, 179(1), pp.2532–2541.

Ordway, D.J. et al., 2010. Evaluation of standard chemotherapy in the guinea pig model of tuberculosis. *Antimicrobial Agents and Chemotherapy*, 54(5), pp.1820–33.

Paul, B.J. et al., 2004. DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell*, 118(3), pp.311–22.

Potrykus, K. & Cashel, M., 2008. (p)ppGpp: still magical? *Annual Review of Microbiology*, 62(23), pp.35–51.

Primm, T.P. et al., 2000. The stringent response of *Mycobacterium tuberculosis* is required for long-term survival. *Journal of Bacteriology*, 182(17), pp.4889–4898.

Priya, V.G.S. & Megha, U.M.M., 2012. Computational analysis of *M . tuberculosis* - CarD protein. *Plos one*, 6(1), pp.228–311.

Raskin, D.M., Judson, N. & Mekalanos, J.J., 2007. Regulation of the stringent response is the essential function of the conserved bacterial G protein CgtA in *Vibrio cholerae*. *PNAS*, 104(11), pp.4636–4641.

Rutherford, S.T. et al., 2007. Effects of DksA, GreA, and GreB on transcription initiation: insights into the mechanisms of factors that bind in the secondary channel of RNA polymerase. *Journal of Molecular Biology*, 366(4), pp.1243–57.

Schluger, N.W. & Rom, W.N., 1998. State of the art the host immune response to tuberculosis. *American Journal of Respiratory and Critical Care Medicine*, 157(19), pp.679–691.

Serbina, N. V, Flynn, J.L. & Flynn, J.O.A.L., 2001. CD8+ T cells participate in the memory mmmune response to *Mycobacterium tuberculosis*. *Infection and Immunity*, 69(7), pp.4320–4327.

Srivastava, D.B. et al., 2013. Structure and function of CarD, an essential mycobacterial transcription factor. *PNAS*, 245(1), pp.1–6.

Stallings, C.L. et al., 2009. CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. *Cell*, 138(1), pp.146–159.

Stallings, C.L. & Glickman, M.S., 2011. A new RNA polymerase modulator in mycobacteria. *Transcription*, 2(1), pp.15–18.

Stallings, C.L. & Michael, G.S., 2010. Is *Mycobacterium* stressed out?A critical assessment of the genetic evidence. *Cell*, 12(138), pp.1091–1101.

Ulrich, A., Andersen, K.R. & Schwartz, T.U., 2012. Exponential megapriming PCR (EMP) cloning — Seamless DNA insertion into any target plasmid without sequence constraints. *Plos one*, 7(12), pp.1–9.

Vagenende, V., Yap, M.G.S. & Trout, B.L., 2009. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry*, 48(46), pp.11084–11096.

Vrentas, C.E. et al., 2008. Corrigendum to "Still Looking for the magic spot: the crystallographically defined binding site for ppGpp on RNA polymerase is unlikely to be responsible for rRNA transcription regulation. *Journal of Molecular Biology*, 379(5), p.1130.

Wehrli, W., 2013. Rifampin: mechanisms of action and resistance. *Reviews of Infectious Diseases*, 5(3), pp.407–11.

Weiss, L.A. et al., 2012. Interaction of CarD with RNA polymerase mediates *Mycobacterium tuberculosis* viability , rifampin resistance , and pathogenesis. *Journal of Bacteriology*, 194(20), pp.5621–5631.

Welin, A., 2011. Survival strategies of *Mycobacterium tuberculosis* inside the human macrophage.

Westblade, L.F. et al., 2010. Structural basis for the bacterial transcription-repair coupling factor / RNA polymerase interaction. *Nucleic Acid Research*, 1093(10), pp.1–13.

Wolf, A.J. et al., 2008. Initiation of the adaptive immune response to *Mycobacterium tuberculosis* depends on antigen production in the local lymph node, not the lungs. *Journal of Immunology*, 205(1), pp.105–115.

Wysocka, J., 2006. Identifying novel proteins recognizing histone modi W cations using peptide pull-down assay. , 40, pp.339–343.

Yang, X.F. et al., 2008. Differential expression of a putative CarD-Like transcriptional regulator, LtpA, in *Borrelia burgdorferi*. *Infection and Immunity*, 76(10), pp.4439–4444.

Zhang, Y. et al., 2003. Mode of action of pyrazinamide: disruption of *Mycobacterium tuberculosis* membrane transport and energetics by pyrazinoic acid. *The Journal of Antimicrobial Chemotherapy*, 52(5), pp.790–5.