STRUCTURAL STUDIES ON TRANSCRIPTIONAL REGULATORY PROTEIN(S) FROM *Mycobacterium tuberculosis* H37Rv

Thesis
Submitted to
Jawaharlal Nehru University, New Delhi

In partial fulfillment for the award of degree of
DOCTOR OF PHILOSOPHY

By
Shuja Shafi Malik

Molecular & Structural Biology Division
Central Drug Research Institute
Lucknow
(INDIA)

2008

Dedicated to my Teachers

केन्द्रीय औषधि अनुसंधान संस्थान
छत्तर मंजिल, पोस्ट बाक्स 173, लखनऊ-226 001 (भारत)
Central Drug Research Institute
Chattar Manzil, P.O. Box 173, Lucknow- 226 001 (India)
Phone : (0522) 2612411-18(PABX)
Fax : 91- (522) 2623405/2623938/2629504
Gram : CENDRUG
web : http://www.cdriindia.org

CERTIFICATE

This is to certify that the research work embodied in this thesis entitled **"Structural studies on transcriptional regulatory protein(s) from *Mycobacterium tuberculosis* H37Rv"** submitted to Jawaharlal Nehru University, New Delhi by Mr. Shuja Shafi Malik is in partial fulfillment of the requirements for the award of the degree of the Doctor of Philosophy. This work, carried out at Central Drug Research Institute, Lucknow is original and has not been submitted so far, in part or full, for any degree or diploma of any other university.

R. Ravishankar
18/08/08

Dr. R. Ravishankar
(Supervisor)

18/8/8

Shuja Shafi Malik

Acknowledgement

I express my sincere thanks to my mentor Dr. Ravishankar Ramachandran. His deep involvement in my work, healthy discussions and encouragement to always think and work independently has brought me to a stage where I have matured as a researcher. These traits will be always of help in attaining my career goals as well as in practical life.

I want to thank all my teachers since my school days because it is due to their contributions that I have been able to reach this stage.

I express my gratitude to all the scientists of MSB division and CDRI. I also especially thank the faculty of the JNU Pre-Ph. D course work. I thank Dr. Amoghanant Sahasrabudhe (CDRI) and Dr. Shantunu Chaudhary (IGIB, New Delhi) with whom I collaboratively carried out a couple of the experiments. Dr. Vinod Bhakuni is thanked for facilitating the use of the departmental facilities.

The help received from peers in the lab and the institute is invaluable and I am grateful to them. In this context I must acknowledge Sandeepji's role as a senior and a friend. He helped me settle down at the time of joining and also introduced me to basic lab routines. Amit is thanked for his help all through and especially in the computational work. I have shared pleasant moments with Sarvind during my stay here. The other members Ruchir, Gaya, Vandana, Abhishek, Ramesh, Vijay and Saurabh have been a party to some memorable moments. Trainees Sunishtha, Dinesh and Varun are thanked for the contributions they have made during their association.

A few people in CDRI deserve special thanks for their support and being around for me. Gauri deserves a mention for being a great friend. Tripti joined along with me and since that day she has been part of the Ph. D journey. Swapnil, Parul, Rohit, Maya, Amar, Divya, Saqib, Tanvir and Geetika share some cherishable memories.

I want to thank all my friends from my school days till my M. Sc especially Tahseen, Abid, Zahid, Masood, Khalid, Tufail, Tariq, Asif, Nousheen, Nazia, Javaid, SriRam, Chandra, Arshad and Gowhar for their support in helping me realize that I can achieve my dreams.

This thesis is a tribute to the immense contributions of my family. My mother has been a real source of inspiration and her resilience in adversities has taught me to face all difficult circumstances in life. Her soul will be at peace to see me achieve this feat. My father has tirelessly encouraged and supported me through the thick and thin of life. I feel that the completion of my Ph. D is a fruit of all that support. Shahid is an apt definition for a brother because it is due to his contribution that I have been able to reach this stage. I can't enlist all people back home who have been supporting me but the emotional support that I received from my Mamaji, Mami and Danish deserves a mention. Dr. Mamaji and Mami are remembered for their keen interest and support. It will be ungrateful if I don't remember my paternal uncles and aunts for their love and affection. I will always remember Naseem bhaiya, Basharat bhaiya, Irfan Didi and their families for their support. Papa and Mummy are remembered for their blessings. Urwashi, Shefauli and Romysa are remembered for their everlasting enquiries about my completion. I have been blessed with Aashiq in my life a person for whom I don't have words to define. Saima has been always a corner stone to never let me feel having lost in life.

I am also grateful to Dr. C. M. Gupta and Dr. R. Tuli, the ex- and present Directors of CDRI for making available various institutional facilities. I also acknowledge the staff of CDRI for administrative and technical support.

The financial assistance from CSIR is duly acknowledged.

Table of Contents

Chapter 3

Cloning, purification and characterization of MtUsfX (Rv3287c)

Chapter 4

Protein-protein interaction analysis of MtUsfX with MtSigF and MtRsfA

ABBREVIATIONS

MEASUREMENTS

Å	Angstrom
°C	Degree centigrade
h	hour
KDa	kilo Dalton
mA	milli Amperes
µg	Microgram
µM	Micromolar
µm	Micrometer
ng	Nanogram
nM	Nanomolar
nm	Nanometer
O.D.	Optical Density

CHEMICALS AND BUFFERS

NTP	Nucleotide triphosphate
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
GTP	Guanosine triphosphate
TTP	Thymidine triphosphate
CTP	Cytidine triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra acetic acid
IPTG	Isopropyl-β-D-Thiogalactopyranoside
PEG	Polyethylene glycol
SDS	Sodium dodecyl sulphate
Tris	[Tris (hydroxymethyl) aminomethane]

PROTEINS

Rsb	Regulator of sigma B
SpoII	Sporulation factor
FlgM	Flagellation factor
SigF	Sigma factor F
SigB	Sigma factor B
UsfX	Upstream of SigF X
RsfA	Regulator of SigF A
RsfB	Regulator of SigF B

PREFACE

M. tuberculosis is a hugely successful pathogen mainly because of its ability to evade host immune response for long periods of time. The role of transcriptional regulatory networks is one of the important mechanisms that help it to tide over these adverse environments. There are about 190 transcriptional regulators and amongst these proteins about 13 sigma factors also play an important role. When this work was initiated, *Rv3286c* had been annotated in the Tuberculist genomic database as an alternate-σ factor SigF and later reports had suggested it to be regulated by its cognate anti- σ factor UsfX. Initially we started our work with the aims of exploring the structural aspects of interaction of UsfX with its interacting partners SigF and anti-anti-σ factors. We also wanted to identify and characterize the ligand /nucleotide binding properties of UsfX. The global aim of the work is to understand the exquisite regulatory cascade involving SigF and its cognate anti-sigma and anti-anti-sigma factors.

Chapter 1 details the factors governing the maintenance of mycobacterial persistence with a focus on sigma factors and the associated proteins. Details are also provided about the sigma/anti sigma regulatory networks found in other systems.

Chapter 2 covers the various techniques and experimental approaches used to clone, purify and characterize the proteins in the present work.The biochemical and *in silico* methodologies employed to study the protein-protein and protein-ligand interactions are also detailed.

Chapter 3 The results obtained while cloning, purifying and characterization of UsfX are detailed here. Subsequently, the characterization of the nucleotide binding and nucleotide hydrolysis properties have been reported.The results of crystallization experiments involving UsfX are also detailed.

Chapter 4 deals with the analysis of the interactions of UsfX with SigF and anti-anti- σ factor RsfA. The cloning and purification of SigF and RsfA have been reported and also the structural aspects of their interaction with UsfX have been discussed.

A part of the results reported here has been published/communicated:

1. *M. tuberculosis* UsfX (Rv3287c) exhibits novel nucleotide binding and hydrolysis properties.
Shuja Shafi Malik, Amit Luthra, Sandeep Kumar Srivastava and Ravishankar Ramachandran
Biochemical and Biophysical Research Communications (Manuscript in press, 2008)

2. Interactions of *M. tuberculosis* UsfX (Rv3287c) with its cognate sigma factor SigF (Rv3286c) and anti-anti-sigma factor RsfA (Rv1365c).
Shuja Shafi Malik, Amit Luthra and Ravishankar Ramachandran
(*Manuscript communicated*)

CHAPTER 1

INTRODUCTION

Tuberculosis (TB), one of the oldest recorded human afflictions, is a contagious disease which is still a major cause of death and illness worldwide. TB primarily is a disease of the respiratory system but can also be present in various forms, including one that attacks bone and causes skeletal deformities. Hard tissues like bone can be preserved for thousands of years, allowing the certain identification of individuals with bone TB who died even 4000 years ago. The discovery of bones with apparent tubercular deformities in ancient Egypt, in various Neolithic sites in Italy, Denmark and countries in the Middle East indicates that TB was found throughout the world up to 4000 years ago (Issar 2003).In the present times, more than two billion people through out the world are infected with *Mycobacterium tuberculosis* bacilli, the microbe that causes TB. But the people infected with TB do not necessarily become sick with the disease. The immune system "walls off" the TB bacilli which, protected by a thick waxy coat can lie dormant for years. When the host becomes immuno compromised, the chances of having an active infection increase. The synergism between HIV and TB is a lethal combination, each speeding the other's progress. A person who is HIV positive and infected with TB bacilli is more likely become infected with TB than an HIV negative; 1.7 million people died from TB in 2006 including 231,000 people with HIV (www.who.int/tb).

The other aspect of concern regarding *Mycobacterium tuberculosis* is the emergence of drug resistant strains. Last half a century has seen the development of drugs against TB and with that the emergence of resistant strains that resist a single drug and multidrug-resistant strains (MDR-TB) which have been defined as strains which resist at least isoniazid and rifampicin, the two most powerful anti-TB drugs. Of late the emergence of extensively drug-resistant (XDR) strains which resist multiple combinations in settings which are infected with HIV has increased the threat of TB (www.who.int/tb). One of the primary reasons for the emergence of drug-resistant strains is the poor compliance with the long duration therapy (about 6 months) that is currently available against tuberculosis. The reason for lengthy therapy is thought to be due to the presence of persistent and dominant bacilli that cannot be killed by the current TB drugs (Dye et al., 1999). Coupled with the poor compliance this results in incomplete eradication and thus supporting the survival of 'persisters' and dominant bacilli.

1.1 Mycobacterial dormancy and persistence

As explained above the current tuberculosis (TB) chemotherapy while effective in killing growing bacilli is ineffective in killing persistent or dormant bacilli. Consequently a considerable interest has aroused to study mechanisms of persistence and dormancy in mycobacteria and try to develop new drugs that can target persistent and dormant bacilli (Ying, 2004). Before going further into the details, dormancy persistence and latency are briefly defined.

Dormancy: Dormancy has been defined as a reversible state of metabolic shutdown. The term "dormant bacteria" refers to a state where the bacteria with the low metabolic activity remain viable but do not form colonies directly or immediately but can be resuscitated to form colonies under appropriate conditions. Important features of dormant bacteria include inability to grow directly on plates and resistance to various and antimicrobial agents (Kell *et al.*, 1998 and Barer *et al.*, 1999).

Persistence: Persistence has been defined as the phenomenon where otherwise drug-susceptible microorganisms have the capacity to survive indefinitely within mammalian tissues despite continued exposure to the appropriate drug or drugs (McCune *et al.*, 1957).

Latency: In contrast to dormancy which is described as a bacterial property, the term latency refers to the *in vivo* situation where bacteria and host have established a balanced state without causing apparent symptoms in the host (Orme, 2001). Latent infection doesn't say anything about the metabolic state of the organism in the host, it simply states that the host is infected but has not developed symptoms.

Normal TB infection can become latent without any apparent symptoms that can last from a few weeks to lifetime. Because of this characteristic of *M. tuberculosis*, one-third of the world population is estimated to be latently infected with this pathogen. Both bacterial and host factors come into play during latent infection, where the tubercle bacilli residing in the lesions in inactive state could reactivate when the host immune system is compromised as in HIV infection (Yukari *et al.*, 2000). It is worth mentioning that latency and dormancy phenomena are not unique to tubercle bacillus, but occur in other chronic and persistent infections like streptococcal infections, syphilis and *H. pylori* (McDermott, 1958, Rhen *et al.*, 2003). Due to this ability of *M. tuberculosis* of circumventing the host immune system, the interest in the mycobacterial dormancy and persistence has increased. It can be likely

that understanding the mechanism(s) and control of mycobacterial persistence and dormancy may have implications for designing new antituberculous drugs, which can further shorten the lengthy 6 months TB therapy through affecting the persistent or dormant tubercle bacilli (Ying, 2004).

Models of persistence:

Many *in vitro* and *in vivo* models have been developed to explore and understand the various properties associated with the mycobacterial dormancy and persistence. A brief description of a few is discussed:

The mouse model of mycobacterial persistence: The model developed called the Cornell model developed by McDermott, McCune and colleagues at Cornell University provided the first convincing evidence of persistent *in vivo M. tuberculosis* (McCune *et al.*,1956, 1966), although circumstantial evidences had long before suggested existence of persistent or dormant tubercle *in vivo* in humans for long periods of time (Dale, 2003).This model and its variants (Scanga *et al.*,1999, Grosset *et al.*,1992) could firmly establish the disappearance and reappearance of tubercle bacilli in infected mice in response to changes in the treatment patterns. From the results of these studies, the reasons for a long chemotherapy against TB could be established.

The Wayne model of TB "dormancy": Wayne's "*in vitro* model of dormancy" mimicked the oxygen depleted *in vivo* environment faced by tubercle bacilli inside macrophages, granulomas and caseous lesions (Wayne *et al.*, 1976, 1982, 1996). While increase in nitrate reduction and shutdown of protein synthesis were found to be associated with hypoxic shiftdown but responses to heat shock and unchanged bacterial viability were observed (Wayne *et al.*,1998, Hu *et al.*,1998, Hu *et al.*, 2000).While as mouse model helped in establishment of the phenomenon of persistence, Wayne model in addition to helping understand hypoxic stress helped in identification of certain factors that are related to persistence in vivo e.g, isocitrate lyase (ICL) and glycine dehydrogenase (McKinney *et al.*, 2000, Hong *et al.*, 2000).

The rifampin persister model: This model developed by Hu *et al.*, (Hu *et al.*,2000) was the first to study relation between persistence and antibiotic response. The existence of transcriptional activity in persistent bacilli even in presence of antibiotics was identified by the help of this model.

The resuscitation model: This *in vitro* dormancy model was developed to observe the resuscitation properties of the stationary phase cultures whereby it could be

established that early stationary phase culture supernatant could resuscitate dormant bacilli in old batch cultures (Sun *et al.*, 1999, Zhang *et al.*, 2001).

Nutrient starvation model: This model had been developed with the purpose of testing antimycobacterial drugs against persistent bacteria. By using proteome analysis approach a number of proteins that are induced under nutrient starvation conditions were identified (Betts *et al.*, 2002).

In addition to be these models various other conditional expression studies like heat shock (Graham *et al.*, 2002) and specific non-replicating stages (Niranjala *et al.*, 2004) have been studied to identify the factors or processes involved in the development and maintenance of the persistent *M. tuberculosis*. Also focus has been on the identification of factors involved in the macrophage existence and survival of tubercle bacilli (Rengarajan *et al.*, 2005). As a result of all these efforts the main altered environments/stresses that *M. tuberculosis* faces during the course of infection can be summarized as follows. The first stress is exposure to oxidizing agents represented by the reactive oxygen intermediates produced by activated macrophages. The second exposure is to low pH produced as a result of phagosome acidification (Deretic,1999).The third is damage to surface structures due to anti-bacterial activity and detergent like properties of alveolar surfactant (Dieli *et al.*, 2001). The fourth stress is hypoxia encountered both within granulomas and phagosomes and this environmental condition is actually the best candidate for induction of persistence (Wayne *et al.*, 2001).The fifth is nutrient and essential-element starvation. There may be reduced nutrients and essential elements inside phagosomes and granulomas, also during transmission (between expulsion from an infected patient and inhalation by a new host), *M. tuberculosis* must face other environmental stresses such as nutrient starvation, exposure to UV light, dehydration and low temperature (Riccardo *et al.*, 2004).

M. tuberculosis employs a battery of machinery to overcome the unfavorable conditions for its *in vitro* or *in vivo* survival. Various cellular processes must be getting altered in these conditions resulting in the production of specific factors that must help the bacteria in these altered conditions. Most importantly, tubercle bacilli can actively manipulate the host immune system to its own advantage to allow for long term persistence through secretion of bacterial factors or by changing the surface antigenic properties to avoid immune attack (Ying, 2004). The cellular processes that can respond to different environments include energy metabolism, lipid biosynthesis,

RNA synthesis and modification, translation apparatus to name a few. One of the important way of responding to the altered environments is by changes in the patterns of gene expression of those genes whose products are required to combat those changes (Hecker *et al.*, 2001). By studying the expression profiles (Upregulation or Downregulation) of various genes of *Mycobacterium tuberculosis*, various factors have been found to be important during different types of environments which *M. tuberculosis* encounters during the course of its life.

1.2 Sigma factors

As discussed above *M. tuberculosis* is capable of adapting to diverse environments and is even able to circumvent host immune system. The sophisticated infection and adaptation mechanisms can be expected to require complex genetic programmes Transcription initiation is the major step in the regulation of gene expression in bacteria (Browning *et al.*, 2004). Bacteria use their genetic material with great effectiveness to make right products at right time and in correct amounts. This is accomplished by the regulation at every step between gene and function and for the sake of economy the key step is to regulate the initiation of RNA-transcript formation (Browning *et al.*, 2004).

σ factors: σ factors are essential for transcription initiation by virtue of their role in promoter recognition. Each of several sigma factors in cell is required for the transcription of a specific subset of gene operons within their 'regulons' (Mooney *et al.*, 2005). The central part of the bacterial transcription machinery is the multi-subunit DNA-dependent RNA polymerase, which is responsible for all transcription (Ebright, 2000). The core enzyme with a subunit composition of $\beta\beta'\alpha_2\omega$ is competent for transcription, but not for promoter directed transcription initiation. The structural studies have shown the *E. coli* structure is similar to that of found in yeast DNA-dependent RNA polymerase II (Claudia *et al.*, 1999, Zhang *et al.*, 1999). This points to the high levels of conservation among the basic components of transcriptional machinery. For RNA polymerase to begin transcription at a particular promote, it must first interact with a σ subunit to form the holoenzyme (Gruber *et al.*, 2003). The σ subunit has three main functions: (i) to ensure the recognition of specific promoter sequences, (ii) to position the RNA polymerase holoenzyme at a target promoter and (iii) to facilitate unwinding of the DNA duplex near the transcription start site

(Wosten *et al.*, 1998). Bacterial genomes encode a principal σ factor, devoted to the transcription of house-keeping genes and also contain a variable number of alternative σ factors (ASFs) (Gruber *et al.*, 2003, Browning *et al.*, 2004).

1.2.1 Structure and classification of σ factors

σ factors can be categorized into two phylogenetically distinct families related to either σ^{54} or σ^{70} from *Escherichia coli* (Lonetto *et al.*, 1992, Buck *et al.*, 2000, Gruber *et al.*, 2003).While all eubacteria encode at least one σ factor copy of the σ^{70} class, not all of them encode members belonging to the σ^{54} family. Structurally σ^{70} members are composed of up to four conserved regions (Fig 1.1) which can be further divided into subregions (Lonetto *et al.*, 1992, Gruber *et al.*, 2003). Region 1 is located in the amino-terminus and contains subregions 1.1 and 1.2. Region 1.1 has been found to contain negative charge and thereby inhibit the binding of free σ factors to DNA (Dombroski *et al.*, 1993). Region 2 is composed of four subregions of which region 2.4 is required in the recognition of the -10 promoter box and region 2.3 generates the open complex (Tsujikawa *et al.*, 2002, Tomsic *et al.*, 2001) by formation of the 'bubble' with the non template strand resulting in the movement of free template strand into active site of RNA polymerase to allow the RNA synthesis (Gardella *et al.*, 1989, Siegele *et al.*, 1989, Zuber *et al.*, 1989, Panaghie *et al.*, 2000). Region 3 comprises subregions 3.0, 3.1 and 3.2. Subregion 3.0 is implicated in the recognition of the extended -10 promoter elements (Barne *et al.*, 1997, Sanderson *et al.*, 2003).The last one *i.e.*, Region 4 consists of sub-regions 4.1 and 4.2. This latter subregion is responsible for the recognition of the -35 promoter element and for the interaction with numerous transcription activators (Dove *et al.*, 2003, Gruber *et al.*, 2003, Paget *et al.*, 2003). All regions contribute, albeit to different extents, to the binding of the σ factor to RNA polymerase (Sharp *et al.*, 1999; Gruber *et al.*, 2001).

The first bacterial σ factor was discovered in 1969 as the subunit of *E. coli* RNA polymerase that was essential for promoter selection (Burgess *et al.*, 1969). The presence of multiple σ factors was established with the finding that several σ factors were required in *B. subtilis* sporulation. Later it was found that certain stress responses (for example, heat shock) in *E. coli* are dependent on alternate σ factors and whole-genome sequencing has revealed a wide distribution of multiple σ factors. On the basis of their structure and physiological roles, σ^{70} related σ factors have been categorized into four groups (Gruber *et al.*, 2003). ***Group 1*** is composed of principal σ

factors, which are essential genes containing all four conserved regions. **Group 2** σ factors are only found in a limited number of bacterial species (*Proteobacteria, Cyanobacteria* and high-GC Gram-positive bacteria) and consist of primary like σ factors. Although closely related to Group 1, these σ factors are normally not essential under laboratory growth conditions and lack most of region 1. The best characterized *Group 2* σ factors are involved in the transcription of general stress response and stationary phase survival genes. ***Group 3*** σ factors also contain conserved regions 2, 3 and 4, but are more distantly related to principal σ factors. Group 3 σ factors fall into clusters comprising evolutionarily related proteins with similar functions, such as heat shock, sporulation or flagellar biosynthesis. These are sometimes called as 'alternative σ factors'. ***Group 4*** is the largest and most heterogeneous collection of σ factors (Gruber *et al.*, 2003). Also described as extracytoplasmic function (ECF) σ factors, members of this group contain conserved regions 2 and 4 only (Fig 1.2). Although the function of most Group 4 σ factors remains unknown, they are often involved in the response to stress conditions, such as iron limitation, oxidative stress and surface stress. Moreover, several σ factors from this category are important for virulence (Raivio *et al.*, 2001, Bashyam *et al.*, 2004). A more simple classification based on the functional roles has divided members of σ^{70} family into three groups: (i) primary σ factors which are usually essential sigma factors and allow the transcription of house keeping genes: (ii) non-essential primary factor-like σ factors, and (iii) alternative σ factors (Wosten *et al.*, 1998). All eubacterial genomes encode one primary σ factor which is usually essential for the transcription of house-keeping genes. The second group *i.e.*, primary factor-like σ factors are non-essential under standard physiological growth conditions and are highly similar to primary σ factors (Wosten *et al.*, 1998). They can be involved in functions like stationary-phase survival, carbon and nitrogen utilization, antibiotic biosynthesis *etc* (Riccardo *et al.*, 2004). The group of alternative σ factors is the most heterogenous containing numerous subgroups the important ones being stress response-sporulation σ factors and extracellular function (ECF) σ factors. ECF σ factors are environmentally responsive regulators, and bacteria usually contain several members of the ECF family that control a variety of functions in response to specific extracellular environmental signals (Missiakas *et al.*, 1998).

The complete genome sequence of *M. tuberculosis* revealed a genome of about 4.4Mb, possessing about 3924 open reading frames accounting for approximately 91% of the potential coding capacity. Analysis of the sequences have

resulted in attributing precise functions to 40 % of the predicted proteins, some information is available for 44% and the remaining 16 % might account for specific mycobacterial functions (Cole *et al.*, 1998). A re-annotation suggests that the genome contains 3995 genes coding for proteins (ORFs) and 50 RNA genes (45 tRNA, 3 rRNA and 2 stable rRNA) (Camus *et al.*, 2002). Expectedly given the complexity of environmental and metabolic variants that *M. tuberculosis* faces during its life cycle, an extensive repertoire of regulatory proteins has been identified. The *M. tuberculosis* genome encodes about 190 transcriptional regulators: 13 σ factors, 11 two-component systems, 5 unpaired response regulators, 11 protein kinases, and more than 140 other putative transcriptional regulators (Riccardo *et al.*, 2004). *M. tuberculosis* σ factors have been found to be involved in diverse functions and are discussed in next section.

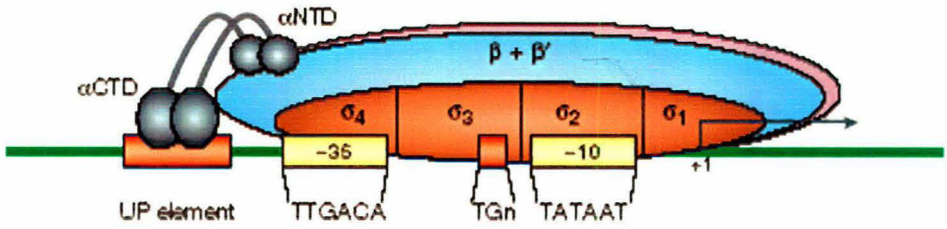


Fig 1.1: Cartoon illustration of the RNA polymerase and its interactions at promoter. The contacts of C-terminal domain of α subunit with the UP element of promoter and that of N-terminal in assembling β and β' subunits is shown. The four regions of σ factor are shown as σ_1, σ_2, σ_3 and σ4 and their role in the initiaition of transcription by binding to DNA is illustrated. The consensus sequences for the -35 (TTGACA), extended –10 (TGn) and -10 (TATAAT) elements are shown. (Adapted from [redacted] Browning *et al.*, 2004)

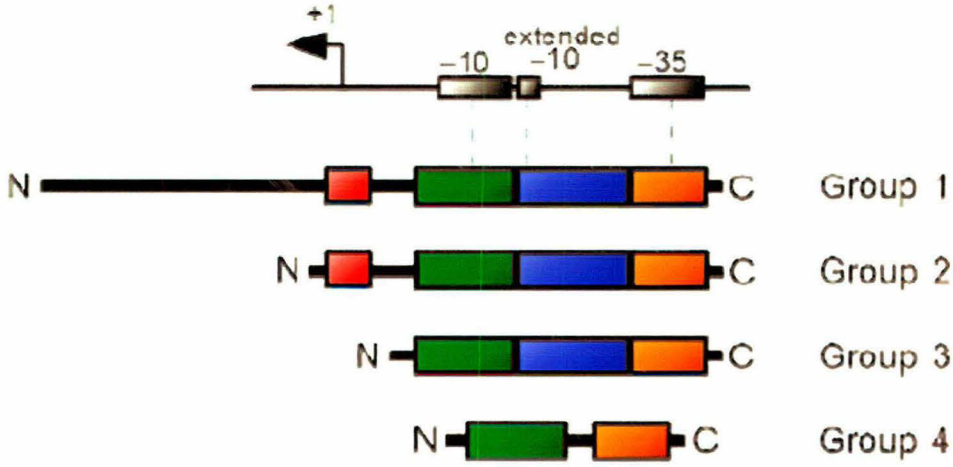


Fig 1.2: Schematic overview of conserved regions in σ factors.
Structural organization of σ factors of Groups 1, 2, 3 and 4. The broken lines link σ factor regions to the promoter element they recognize. The arrow indicates the transcription start site (Adapted from Sebastien *et al.*, 2006).

1.2.2 Mycobacterial σ factors

Mycobacterium tuberculosis encodes 13 σ factors falling into the all groups of σ^{70} families described in previous section. Like other σ factors, mycobacterial sigma factors have their own specificity allowing the transcription of a particular subset of genes. Genes belonging to a defined regulon often participate in related cellular functions. Therefore, temporal variation in active σ factor populations must be a powerful way for *M. tuberculosis* to modulate its gene expression profiles in accordance with physiological requirements, helping to survive in altered environments and thus achieving a successful infection.

The primary σ factor in *M. tuberculosis* is $\boldsymbol{\sigma^A}$ which controls the expression of all the essential genes and has been found to be essential for bacterial viability in *Mycobacterium smegmatis* (Gomez *et al.*, 1998). In addition to its role as essential σ factor, σ^A has been found to be responsible for enhanced intracellular growth and an increased resistance to superoxides, suggesting that this σ factor also modulates the expression of virulence genes (Shiping *et al.*, 2004).

$\boldsymbol{\sigma^B}$ has been predicted to be the sole representative of the Group 2 although its physiological role is still unclear. *M. tuberculosis sigB* knockouts have been found to be more sensitive to surface stress, heat shock, oxidative stress and exposure to vancomycin (Sebastien *et al.*, 2006). σ^B has also been found to be important for adaptation to stationary phase or to growth in nutritionally poor environments (Mukherjee *et al.*, 2005).

The other σ factors σ^C, σ^D, σ^E, σ^F, σ^G, σ^H, σ^I, σ^J, σ^K, σ^L and σ^M belong to Group 4 where σ^F belongs to the subgroup that also contains the stress response-sporulation σ factors in bacilli and streptomycetes (DeMaio *et al.*, 1996) and the rest to the subgroup of the extracellular function (ECF) σ factors. $\boldsymbol{\sigma^C}$ has been found to important for the maintenance of the cell wall properties and and the toxicity levels of *M. tuberculosis* thus contributing to the proliferation of the pathogen in host tissues (Sun *et al.*, 2004). $\boldsymbol{\sigma^D}$ has been suggested to be part of $Rel_{M.\ tuberculosis}$ regulon (Dahl *et al.*, 2003) and is induced during starvation (Betts *et al.*, 2002). $Rel_{M.\ tuberculosis}$ is a gene required for virulence and produces hyperphoshphorylated guanidine (p)ppGpp which is used as a molecule in stringent response and modulates bacterial gene expression for long-term survival under starving conditions (Magnusson *et al.*, 2005).The role of $\boldsymbol{\sigma^E}$ has been found to be in regulating responses to heat shock, oxidative stress and

SDS and vancomycin-mediated cell surface stress. Genes regulated by σ^H include its own structural gene and genes encoding σ^E, σ^B, DNA repair proteins, enzymes involved in thiol metaboloism such as thioredoxin and thioredoxin reductase and enzymes involved in cysteine and molydopterin biosynthesis (Sebastein *et al.*, 2006). σ^J has been shown to be strongly induced in stationary phase cultures and during growth (Hu *et al.*,2001) in human macrophages (Cappelli *et al.*, 2006), however a confirmatory role has been found to be in resistance to oxidative stress (Hu *et al.*, 2004). σ^L has been found to be important for maintaining the lethality of *M. tuberculosis* (Hahn *et al.*, 2005; Dainese *et al.*, 2006). No comprehensible information is available for σ^G, σ^I, σ^K and σ^M.

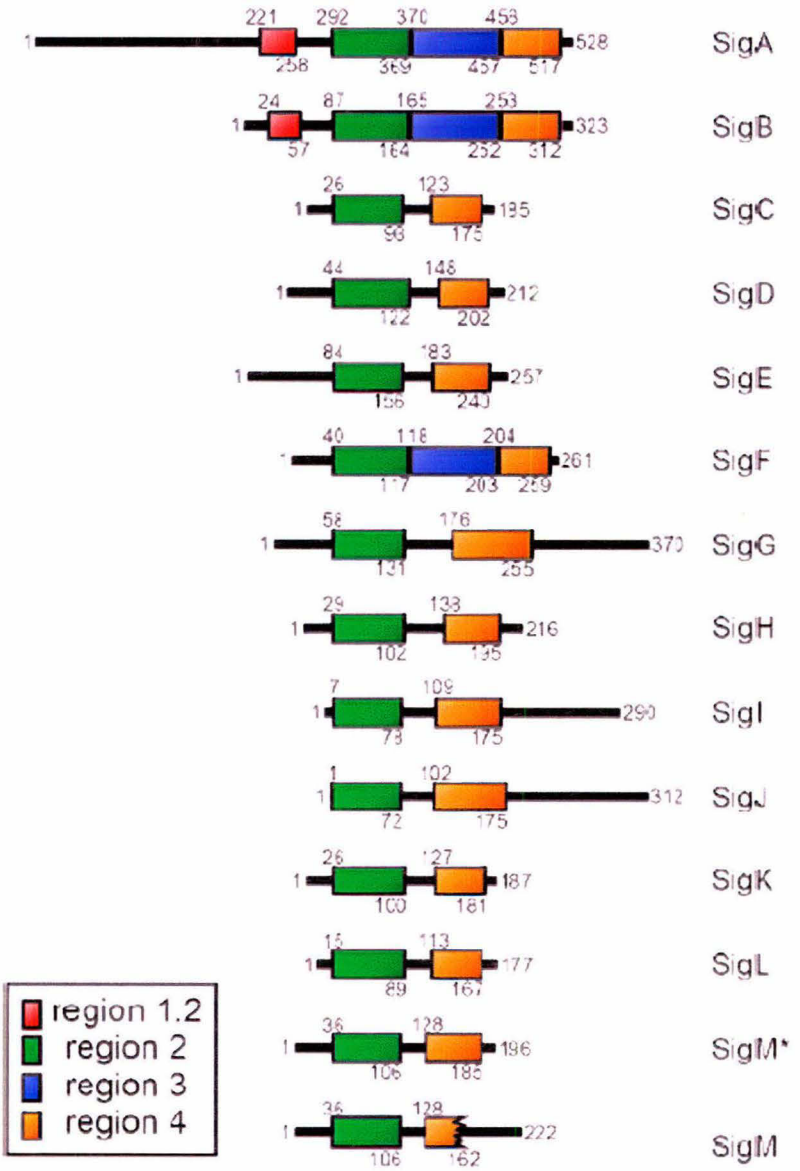


Fig 1.3: Schematic overview of conserved regions in *M. tuberculosis* σ factors. Numbers at the top and bottom of the coloured boxes indicate the residues at the beginning and end of each conserved region. The amino-terminus of each protein is represented by '1' and the number of the last C-terminus residue is shown on the right side of each σ factor. σ^{M*} corresponds to the sequence from *M. tuberculosis* CDC1551(Adapted from Sebastien *et al.*, 2006).

1.2.3 *M. tuberculosis* σ^F

The role of σ^F has been thoroughly debated and studied in *M. tuberculosis*, primarily for the reason that this alternate σ factor has been found to be homologous to *Bacillus subtilis* stress-response sigma factor, SigB, as well as to its developmental σ factor, SigF. *M. tuberculosis* σ^F is preceded by an open reading frame *usfX* encoding a protein that shows statistically significant homology to *Bacillus subtilis* anti-sigma factors RsbW and SpoIIAB (This is detailed in later sections) (DeMaio *et al.*, 1997). The presence of SigF was reported for the first time by DeMaio and co-workers and identified to be present only in slow growers like *M. bovis* BCG and *M. avium* in addition to *M. tuberculosis* but not in rapid growers like *M. smegmatis* and *M. abscessus*. They proposed that *M. tuberculosis* SigF is involved in survival outside the host like in expectorated sputum where it is required for adaptation to oxidative stress, low nutrient levels and low temperature (DeMaio *et al.*, 1996).They also observed that SigF is closely related to *S. coelicolor* and *B. subtilis* sporulation factors. The studies to solve the question of *M. tuberculosis* SigF relatedness to either *B. subtilis* SigB or SigF *i.e*, whether it was a stress response regulator or a developmental factor were initiated by same workers in 1997. They concluded that on the basis of gene organization and antigenic closeness, *M. tuberculosis* SigF is closely related to *B. subtilis* SigB. Further in their experiments they observed that SigF over-expression in *M. bovis* BCG was partially toxic to the cells apparently due to the expression of SigF-dependent genes which may prevent cellular growth due to the involvement of these genes in program of growth arrest in response to starvation or stress (DeMaio *et al.*, 1997).

Theresa and co-workers made efforts at identifying the conditions that induce the *sigF* expression. They found that the exposure of antibiotics like ethambutol, rifampin, streptomycin and cycloserine resulted in the over-expression of *sigF* in their experimental setup, the level of expression being proportional to the amount of antibiotic added. They further found that level of expression was 100-fold in late-stationary-phase compared to that in exponential growth. The expression during exponential phase was found to be induced by a number of other stresses like anaerobic metabolism/oxidative stress, nitrogen depletion, stationary phase growth and cold shock. These workers also suggested that as both stationary phase and antimicrobial exposure lead to *sigF* induction, therefore, SigF-dependent genes may

be conferring a protective effect against some antibiotics. SigF was postulated to contribute to both genotypic and phenotypic tolerance towards antibiotics. Phenotypic tolerance by way of SigF-mediated entry into a stationary growth state resulting in physiological alterations that render antibiotic targets non-essential or genetic tolerance by up-regulating certain genes that overcome target inhibition produced by low levels of the drug (Theresa *et al.*,1999).

A comprehensive analysis of the conditions that contribute to the over-expression of *sigF* came from the mutational analysis where a *sigF* null mutant was challenged by various altered conditions. Temperature shift, oxidative stress, entry into stationary phase and macrophage infection didn't elicit a survival difference between the mutant and the wild type. It was therefore, concluded that although these conditions may be inducing the expression of *sigF*, the increased expression is not necessary for the bacterial survival under these conditions. Instead from the mouse survival data from the same mutant, it was concluded that *sigF* gene has a role in virulence in whole animal. The loss of *sigF* didn't prevent the mutant strain from producing a lethal infection but death was significantly delayed in the mouse infected by the mutant strain (Ping *et al.*, 2000).

The expression of genes controlled by SigF was analyzed by microarray analysis with wild type and *sigF* mutant by Geiman and co-workers (Geiman *et al.*, 2004). In their studies these workers found that in the mutant *sigF*, 187 genes were relatively underexpressed in early stationary phase, 277 in late stationary phase and only 38 in exponential growth phase. Going into the functional role of these underexpressed genes, those underexpressed in exponential phase were found to be mostly hypothetical proteins. The genes that were underexpressed during stationary phase include those involved in energy metabolism (such as electron transport, fermentation, anaerobic metabolism, and polysaccharide synthesis and degradation), nucleotide synthesis, and central intermediary metabolism (including oxidoreductases, arylsulfatases, methyl transferases, acyltransferases, monooxygenase flavin adenine dinucleotide [FAD] binding protein, and a glutamine amidotransferase). This broad range of genes whose expression was influenced in part by SigF supported its role as a stress-response regulator. Also genes involved in biosynthesis and structure of cell envelope were found to be down regulated and also those involved in biosynthesis and degradation of surface polysaccharides and lipopolysaccharides. Most important in their observations these workers found *sigF* itself to be down-regulated in absence

of functional SigF protein indicating a part autoregulation for *sigF* expression. Other transcription related proteins whose expression was found to be related to *sigF* expression include *sigC* gene which has been found to be essential for the lethality in mouse model, repressors/activators from the MarR, GntR and TetR family of DNA binding regulators. On the basis of these findings the authors have suggested the role of SigF in a hierarchial network of *M. tuberculosis* gene regulation. On the basis of a large gene pool found to be SigF dependent, the same workers were able to derive a putative *M. tuberculosis* SigF consensus sequence, $GGTTTCX_{18}GGGTAT$ and was found to be similar to the *B. subtilis* stationary-phase and stress response sigma factor, SigB. It was also observed by these authors that that the conservation of the putative -35 hexamer varied from 43 to 93%, suggesting a role for some residues to be critically important for recognition.

Conclusively, *Mycobacterium tuberculosis* possesses a repertoire of sigma factors; the probable roles of most have been explored in great detail. The alternate sigma factor SigF has been implicated in stress response and important for maintaining the lethality of the infection. It has also been found to be important for expression of approximately 500 genes most of which are expressed in the stationary or stress phase. The co-relation of down-regulation of *sigF* with the absence of functional SigF indicates an auto regulatory mechanism.

1.3 Transcription regulatory mechanisms

After a description about the initiation of transcription through sigma factors and an elaboration of the mycobacterial sigma factors, a look into the mechanisms involving the transcription regulation through sigma factors is described. A crucial point in consideration of microbial gene regulation is to understand that the active RNA polymerase is in short supply inside the cell as much of it is channeled into copying of genes that encode stable RNAs needed for translation. Similarly, some of the RNA polymerase is bound non-productively to the cell's DNA, consequently the amount of free RNA polymerase available for transcription is limited. Additionally, the σ factor supply is limited, so there is intense competition between different promoters for RNA polymerase holoenzyme (Ishihama, 2000, Maeda *et al.*, 2000). So in addition to the regulation of transcription at σ factor level there are few other mechanisms that operate to ensure a prudent distribution of RNA polymerase between competing promoters. These involve promoter DNA sequences, small ligands, folded bacterial chromosome structure, transcription factors and of course the σ factors (Browning *et al.*, 2004). These different mechanisms are utilized in different ways to alter profiles of gene expression in response to environmental change and thereby allowing variation in the level of genes known as 'fine tuning'. However different mechanisms are used in some cases where instead of fine tuning is effective 'on-off' switches are created.

Promoter sequences: Bacterial cells contain about 2000 promoter sequences and the differences in their sequences act as powerful drivers in unequal distribution of RNA polymerase between different transcription units (Salgado *et al.*, 2001). The promoters with near consensus sequences function more efficiently but the presence of non-consensus sequences balances the activity of each promoter against the other promoters. Although differences in promoter sequence elements provide a useful way to control a wide range of promoter activities, these differences only provide a 'static regulation' that can not be modulated according to environmental conditions. So most 'adaptative' responses are modulated by *trans*-acting factors.

Small ligands: Small ligands provide a mechanism by which RNA polymerase can respond quickly and efficiently to the environment. Guanosine 3',5' biphosphate (ppGpp) is synthesized when amino-acid availability is restricted to the extent that translation is also limited (Chatterji *et al.*, 2001). It has been proposed that ppGpp

controls expression of translation machinery in response to sudden starvation, similarly ATP availability has been predicted to control expression in response to growth rate (Gaal *et al.*, 1997, Schneider *et al.*, 2002).

Folded chromosome: Bacterial chromosomes are highly compacted by supercoiling and by their interactions with proteins and RNA. In *E. coli* a dozen important proteins are involved in this compaction. These so-called nucleoid proteins are abundant in the cell and most of them bind nonspecifically to DNA but some bind with weak specificity occupying sites that are distributed throughout the chromosome. The binding of these nucleoid proteins to DNA and the resulting folding of the bacterial chromosome must affect the distribution of RNA polymerase between promoters. As there is not a conclusive idea about the basic unit of bacterial chromatin structure, no general rules have been established. The best-understood case is H-NS protein (histone-like nucleoid-structuring protein) which completely silences gene expression by forming extended nucleoprotein structures e.g. at the *proU* and *bgl* promoters (Schnetz *et al.*, 1995, Jordi *et al.*, 2000, Petersen *et al.*, 2002).

Transcription factors: Transcription factors are predicted to bind to promoters and either up- or downregulate transcription. The *E. coli* genome contains more than 300 such proteins and half of them have been functionally characterized. Some of these proteins control large genes, whereas others control just one or two genes. When a transcription factor binds to a promoter, it can activate or repress transcription initiation. Some transcription factors function solely as activators or repressors, whereas others can function as either according to the target promoter (Perez-Rueda *et al.*, 2000). These factors couple the expression of genes to environmental signals and they are in turn regulated by control of their expression or their activity. The DNA-binding affinity of transcription factors is modulated by small ligands, the concentrations of which fluctuate in response to nutrient availability or stress. An example of this kind of regulation is the reduction of DNA-binding affinity by the Lac repressor by allolactose which signals the presence of lactose in growth medium. Also the activity of some transcription factors (Response Regulators) is modulated by covalent modifications e.g. some like NarL bind to their target DNA only when phosphorylated by its cognate sensor kinase. These sensor kinases are located in the inner bacterial cell membrane and are regulated by extra-cellular signals (Stock *et al.*, 2000).

σ factors: The role of sigma factors in the initiation of transcription has been already discussed, so they become one of the primary tools in the regulation of gene expression. The alteration of gene expression by σ factors occurs through the competition between the principal sigma factor and the alternative sigma factors for the RNA polymerase. As these alternative σ factors accumulate in response to specific signals, they compete with σ^{70} for RNA polymerase and enable it to recognize promoters that control genes that assist the cell in coping with the altered conditions in response to which these alternate σ factors are synthesized. As different σ factors regulate cellular responses to different stresses, consequently their activities are tightly regulated. Regulation of alternative σ factors occurs at transcriptional, translational and post-translational level. Some of the regulation is at the level of σ factor synthesis, but in many cases regulation is effected by anti-sigma factors which modulate the activity of a σ factor independently of its transcription and translation (Hughes *et al.*, 1998). Many anti-sigma factors sequester their cognate σ factor so that it is not free to combine with RNA polymerase. The activity of the anti-sigma factor is then regulated by ligand binding, covalent modification or proteolysis (Helmann 1999) (Discussed in next section). In some cases, the anti-sigma factor sequesters its cognate σ factor to the cell membrane, where it is sensitive to extracellular signals. In other cases, the activity of the anti-sigma factor controls proteolytic degradation of its cognate σ factor (Hengee *et al.*, 2003). In most of the bacteria σ^{54}, an unrelated member of the σ^{70} family renders alternate specificity to RNA polymerase by recognizing different elements. In addition these σ factors have an absolute requirement for an activator. Such activators carry specific conserved modules that drive an ATP-dependent remodeling of the RNA polymerase promoter complex, which results in formation of open complex (Buck *et al.*, 2000).

Overall, the regulation of transcription occurs at multiple steps with each step important in its own way. In the next part we are detailing the role of anti-sigma and anti-anti sigma factors, as my thesis work is mainly involved in the study of these proteins.

1.4 Anti-σ factors : Posttranslational regulation of σ factors

Anti-σ factors provide a means of regulating gene expression at the transcription level acting as an inhibitor protein by holding σ factor in a reversible interaction. An anti-σ factor is defined by its ability to prevent its cognate σ factor to compete for core RNA polymerase. The result of the action of an anti-σ factor is the inhibition of the transcription from a given set of promoters specific for its cognate σ factor. Anti-σ factors are themselves regulated by either secretion from the cell (i.e, FlgM export through the hook-basal body), sequestration by an anti-anti-σ factor (i.e, phosphorylation regulated partner switching modules), or interaction with extracytoplasmic proteins or small molecule effectors (i.e, transmembrane regulators of extracytoplasmic function σ factors). A description of few well studied anti-σ factor systems is given to get an idea about the factors involved in the interaction and the mechanism by which they regulate the transcription.

Bacteriophage T4 AsiA Protein: Anti-$\sigma^{70}$of E. coli: The AsiA protein of T4 bacteriophage has been purified and found to inhibit transcription of T4 DNA by *E. coli* RNA polymerase (Orsini *et al.*, 1993). During growth after infection of *E. coli*, the genes of bacteriophage T4 are differentially transcribed according to when they are needed in phage development. The T4 promoters are assigned to one of three classes depending on when in the infection state they are recognized: early, middle or late. The early promoters are transcribed immediately after infection by the *E. coli* σ^{70}- containing RNA polymerase holoenzyme. AsiA is an early gene product of 10kDa that was found to co-purify with *E. coli* from T4-infected cells (Stevens 1972, 1973, 1976). The purified AsiA protein was found to inhibit transcription of T4 DNA by *E. coli* RNA polymerase (Orisini *et al.*, 1993). Also AsiA has been found to associate with the modified RNA polymerase (Modified by *mod* and *alt* early gene products) of *E. coli* and results in the start of transcription from middle promoters and inhibition of transcription from early promoters (Hinton *et al.*, 1996). This transcription facilitating role is played by AsiA in presence of other protein MotA (Guild *et al.*, 1998). One of the middle genes, gene 55, encodes an alternative σ factor, gp55 that is specifically required to transcribe late promoters (Geiduschek *et al.*, 1991). Because *E. coli* σ^{70}–dependent transcription is needed throughout the T4 infection cycle, it will normally outcompete T4 gp55 for RNA polymerase but an AsiA-dependent mechanism has evolved in which AsiA weakens σ^{70}–core RNA

polymerase interactions to an extent that allows both middle and late transcription promoter activity to occur simultaneously. Thus AsiA presents an example of a temporal regulator of gene expression as it plays dual roles as a positive and a negative regulator of σ^{70}–dependent transcription.

FlgM anti-σ factor: Regulator of flagellar biosynthesis: The regulation of flagellar genes of *Salmonella typhimurium* is similar to that of bacteriophage genes in that the genes are organized in a transcriptional hierarchy. The Class 2 genes synthesized by the action of transcription factor gene products of Class 1 encode the proteins needed for the structure and assembly of the basal body hook structure as well as a flagellar specific alternative σ factor, σ^{28}, necessary for transcription from Class 3 promoters (Liu *et al.*, 1995, Ohnishi *et al.*, 1990). The Class 3 genes encode proteins involved in the final stages of flagellar assembly like flagellar motor force generators, flagellin subunit genes and chemotactic signal transduction pathway. In *E. coli*, the transcription of a flagellin subunit gene *fliC* has been found to be coupled to the formation of an earlier stage of flagellar assembly, the hook-basal body structure (HBB). If any of the necessary genes required for the formation of intermediate assembly is defective or absent the expression of *fliC* gene is blocked. *flgM* was found to be the negative regulatory gene preventing the expression of Class 3 genes defective in HBB formation. Same form of negative regulation has been found true in *S. typhimurium* for all Class 3 genes (Lord *et al.*, 1996, Gillen *et al.*, 1991, 1991). The inhibitory role of the *flgM* gene product was confirmed with the experiments with purified FlgM. Purified FlgM protein was found to be able to inhibit Class 3 transcription of the *fliC* promoter in levels stoichiometric to the σ^{28} holoenzyme (Oshini *et al.*, 1992). The structural characterization of FlgM-σ^{28} interaction has revealed FlgM to be an unstructured protein in solution which attains a partially structured conformation on complex formation (Hughes *et al.*, 1998).With further experiments it was conclusively proved that FlgM acts by binding directly to σ^{28} to prevent both its interaction with core RNA polymerase and interaction of free σ^{28} to Class 3 promoter DNA (Kutsukake *et al.*, 1994). FlgM senses the signal for its functioning from the flagellar export system.When the HBB structure is complete, the transport machinery switches over to the transport of the Class3 proteins which is regarded as the signal by FlgM. This is recognized as the signal for the completion of HBB formation and for being export-competent to be transported by flagellin subunits. In the presence of HBB complex, FlgM has been found in spent growth

medium while as in case strains having defective gene required for HBB formation, FlgM has been localized in cytoplasm (Kutsukake 1994, Hughes *et al.*, 1993). In this way FlgM is removed from the cytoplasm and σ^{28} is relieved free for transcription providing a means of temporal regulation of gene expression.

Recently, the crystal structure of the σ^{28}/FlgM complex from *Aquifex aeolicus* has been reported, where σ^{28} has been found to be tightly packed where the important core binding determinants have been found to be occluded by FlgM (Margareta *et al.*, 2004).

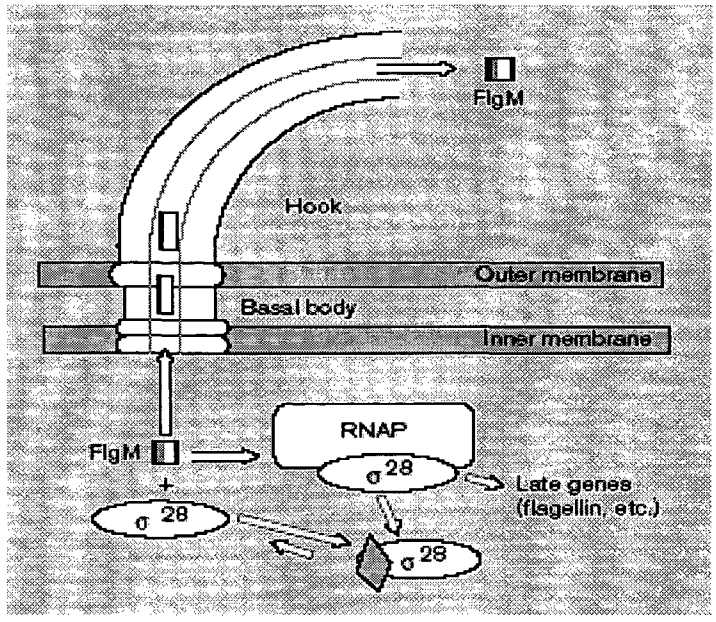


Fig. 1.4: Regulation of late flagellar genes by FlgM anti-σ factor.
FlgM (square) inhibits the activity of σ^{28} either by binding to free σ^{28} or by inducing dissociation of the σ^{28} holoenzyme. When bound to σ^{28}, FlgM adopts a partially folded structure (diamond). Once the hook and basal body structure is complete, FlgM can be exported from the cell by the flagellar export machinery. FlgM passes through the hollow inner core of the basal–body, hook, and flagellar filament structures. Once σ^{28} transcription commences, flagellin becomes a competing substrate for this export channel and FlgM levels in the cell again begin to rise (Adapted from Helmann *et al.*, 1999).

DnaK, An anti- σ factor regulator of heat-shock response: The heat-shock response of *E. coli* involves the coordinated expression of heat-shock proteins that either act as chaperones by preventing protein turnover by maintaining proper folding or as proteases by facilitating protein turnover (Hendrick *et al.*, 1993, Gottesman *et al.*, 1997). In *E. coli* an alternative σ factor, σ^{32}, is induced following heat shock which then helps in the transcription of heat-shock proteins (Grossman *et al.*, 1984, 1987). Regulation of the heat-shock response has been found to be achieved by changes in concentration of σ^{32} in the cell by a posttranscriptional control (Straus *et al.*, 1987, Voelker *et al.*, 1996). The DnaK (Hsp70 homolog), DnaJ (Hsp40 homolog) and GrpE proteins make up the prokaryotic Hsp70 system. They are themselves heat-shock proteins assisting in refolding of proteins following heat-shock and also help in folding of nascent polypeptide chains and protein translocation (Polissi *et al.*, 1995). Biochemical evidences have suggested a mechanism for binding and release of proteins including σ^{32} by DnaK and its regulation by ATP hydrolysis, DnaJ and GrpE (Karzai *et al.*, 1996). Dnak, DnaJ and GrpE can bind σ^{32} independently, but with conclusive evidences it has been established that DnaJ binds σ^{32} and presents it to ATP-bound DnaK. The binding of GrpE into this complex allows entry of ATP to dissociate the complex (Reid *et al.*, 1996, Packschies *et al.*, 1997).

As a result of heat shock, σ^{70} bound to RNA polymerase dissociates, gets inactivated and forms aggregates (Blaszczak *et al.*, 1995). Under these conditions, σ^{32} remains active and associates with free core to direct transcription from σ^{32}-dependent promoters. DnaK, DnaJ, and GrpE are induced by the heat-shock response and act on the σ^{70} aggregates to refold σ^{70} into active protein. As protein aggregate substrates for the DnaK, DnaJ, and GrpE proteins are removed, they are free to act on σ^{32}. Under normal growth conditions, in the presence of ATP, DnaJ induces DnaK to compete with RNA polymerase for binding σ^{32} and to form a stable DnaJ-σ^{32}-DnaK-ADP complex after ATP hydrolysis (Liberek *et al.*, 1995). GrpE can interact with this complex to release active σ^{32}, which can reassociate with RNA polymerase or present unfolded σ^{32} to the HflB (FtsH) protease for degradation (Hughes *et al.*, 1998).

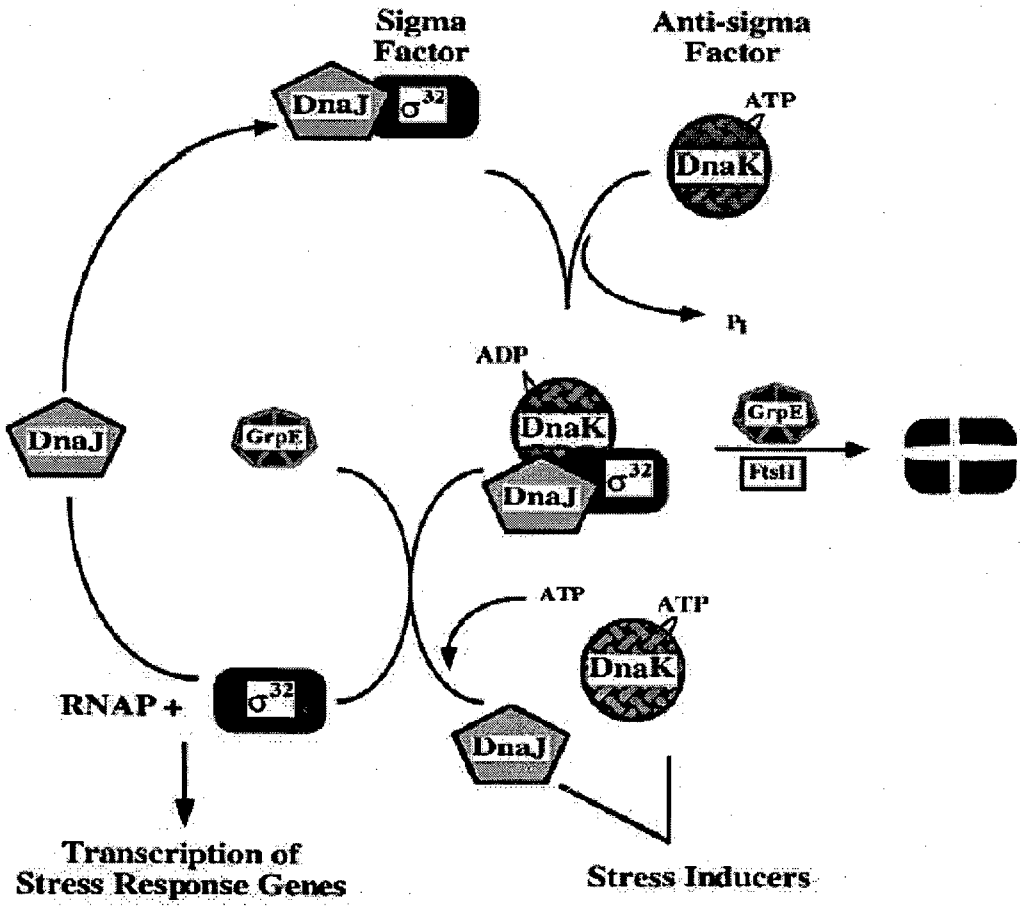


Fig. 1.5: Regulation of heat-shock response of *E. coli*.

Under normal growth conditions, DnaJ in presence of DnaK protein binds the heat-shock–specific alternative σ^{32} factor to form a stable DnaJ- σ^{32}-DnaK-ADP complex after ATP autohydrolysis by DnaK. GrpE directs this complex to FtsH protease for degradation of σ^{32} and complex dissociation. Following heat-shock, σ^{70} bound to RNA polymerase becomes inactivated and separates from the active core RNA polymerase complex and forms aggregates. The DnaK, DnaJ, and GrpE act on the σ^{70} aggregates to refold σ^{70} into active protein. The FtsH protease acts on proteins that have become denatured as a result of heat shock. σ^{32} is unaffected by heat shock and is free to interact with core RNA polymerase and transcribe σ^{32} dependent operons (Adapted from Hughes *et al.*, 1998).

SpoIIAB Anti-σ factor in B. subtilis sporulation regulation: Sporulation a process activated by known stress signals such as nutrient starvation, in *B. subtilis* proceeds through a series of morphological stages that are regulated by the activity of a cascade of different σ factors, each required for the expression of a set of operons necessary for transition from one developmental stage to other (Haldenwang 1995, Errington 1996 and Stragier *et al.*, 1996). σ^E, σ^F, σ^G and σ^K are the four σ factors that play a role in sporulation. Sporulation causes nuclear division followed by an asymmetric septum formation resulting in the formation of two diploid cells; the mother cell and the forespore. This results in the activation of compartment specific σ factors, σ^F in forespore and σ^E in mother cell. Thereafter forespore becomes engulfed within the mother cell and causes the activation of another σ factor, σ^G, within the forespore (Hughes *et al.*, 1998). Although σ^F has been found to be active in forespore only, by immunofluorescence methods its presence has been detected in the predivisional cell and post division in both the mother cell and forespore (Lewis *et al.*,1996, Straiger *et al.*, 1996). The activity of σ^F is controlled by regulatory proteins SpoIIAB, SpoIIAA, SpoIIE and purine nucleotides. SpoIIAB is an anti-σ factor that inhibits σ^F activity in the predivisional cell and mother cell by direct interaction (Duncan *et al.*, 1993, Lewis *et al.*, 1996). In the mother cell, SpoIIAB has kinase activity that inactivates SpoIIAA by phosphorylation thereby inhibiting its anti-anti- σ factor activity (Magnin, 1996). This way SpoIIAB can act both as an anti-σ factor for σ^F and as an anti-anti-anti-σ factor for its antagonistic anti-anti-σ factor. The key to σ^F activity in the forespore is the activity of SpoIIE (Margolis *et al.*,1991). In the forespore SpoIIE is active where it dephosphorylates SpoIIAA-PO_4 leading to accumulation of SpoIIAA which releases σ^F from the SpoIIAB-σ^F complex and gene expression leading to development of spore is initiated by σ^F–bound RNA polymerase.

In the last few years many reports have come up explaining the structural basis of these protein-protein interactions and giving details about their functioning in the regulatory mechanisms. The crystal structure of the *Bacillus stearothermophilus* anti-σ factor with the sporulation σ factor σ^F has been reported (Campbell *et al.*, 2002). In total 17 residues of σ^F have been found to make favorable interactions with SpoIIAB involving mostly hydrophobic interactions in addition to electrostatic and van der Waal's interactions. The alignment analysis of SpoIIAB:σ^F structure with other σ^{70} type σ factors found in *B. subtilis*, has revealed a maximum of 59% homology in the case of interacting residues, thus providing a structural insight into the highly specific

nature of σ factor/anti- σ factor interactions. Also these authors proposed a 'zipper model' explanation for SpoIIAA 'induced release' model (Duncan *et al.*, 1996) where they found that the interaction of SpoIIAB with σ^F partially occludes the nucleotide binding site in one chain. This allows SpoIIAA to gain foothold on the other side distal from σ^F binding site and SpoIIAA engages its binding sites in a sequential manner beginning with the ATP pocket of SpoIIAB thereby gradually displacing σ^F due to increasing steric clashes. Arg-20 of SpoIIAB was found to be the residue involved in the interaction with both the proteins. A 'docking model' for the release of σ^F was proposed on the basis of biochemical evidences (Margaret *et al.*, 2003). These authors also showed the relevance of Arg-20 but argued that SpoIIAA docks onto the SpoIIAB on one side and thereby dislodges σ^F but not in a sequential way. The crystal structure (Shoko *et al.*, 2004) of ATP and ADP bound forms of *B. stearothermophilus* SpoIIAB/SpoIIAA complex confirmed the 'docking model'. In their opinion there are actually two docking events involved occurring two separate surfaces dislodging σ^F and it is the final docking steps that makes SpoIIAA amenable for phosphorylation. The phosphorylation acts to provide an electrostatic flag that alters the affinity of the protein for its binding partner.

Conclusively, the σ factor/anti- σ factor interactions are highly specific in nature with the specificities determined by intrinsic properties of the proteins. The regulation by another set of proteins the anti-anti-σ factors provides a higher order level of regulation where cells can program the expression of their genes in response to changing stimuli.

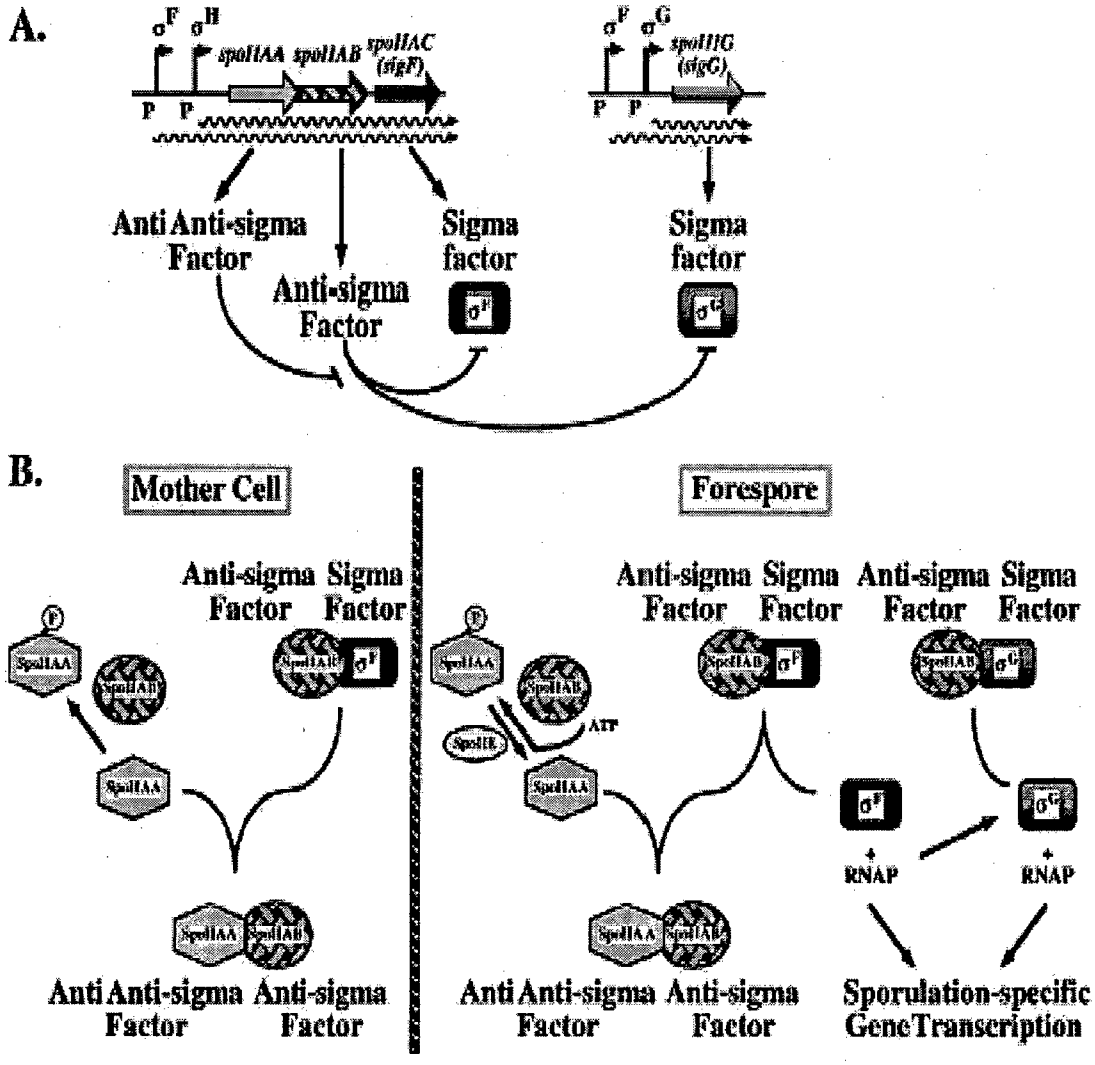


Fig. 1.6: Regulation of sporulation-specific expression in *B. subtilis*.
(A) Induction of sporulation leads to expression of the *spoIIA* operon, *spoIIAA-spoIIAB-spoIIAC*, from a σ^H-dependent promoter. The *spoIIAC* gene encodes σ^F, the *spoIIAB* gene encodes anti-σ^F factor and the *spoIIAC* gene encodes an anti-SpoIIAB factor (anti-anti-σ^F). σ^F is autoregulatory and transcribes the *spoIIAA-spoIIAB-spoIIAC* operon as well as the *spoIIIG* operon. *spoIIIG* gene encodes the sporulation-specific sigma factor, σ^G.
(B) In the mother cell, SpoIIAB inactivates SpoIIAA by phosphorylation. SpoIIAB is free to bind σ^F to prevent σ^F -dependent gene expression.In the forespore, SpoIIE protein is active. SpoIIE is a phosphorylase that dephosphorylates SpoIIAA-PO_4. Unphosphorylated SpoIIAA binds SpoIIAB to prevent its interaction with σ^F leaving it free to transcribe sporulation-specific genes. (Adapted from Hughes *et al.*, 1998)

RsbW Anti-σ factor in B. subtilis stress response regulation: *B. subtilis* responds to various stresses by the expression of stress-induced genes utilizing a stress-response-specific alternative σ factor σ^B. The transcription factors σ^B and σ^F are closely related and the regulatory mechanisms affecting their activity are strikingly similar. σ^B activity is regulated by anti-σ factor RsbW and its antagonist RsbV (Kang *et al.*, 1996, Yang *et al.*, 1996). The σ^B pathway actually is a partner-switching cascade that utilizes phosphorylation to alter the binding partner-specificity of the proteins that are ultimately responsible for activation of σ^B. Interactions between RsbW and both σ^B and RsbV have been demonstrated by gel-filtration chromatography, co-immunoprecipitation, chemical cross-linking and yeast two hybrid system (Hughes *et al.*, 1998). There are two distinct stress response pathways; energy stress and environmental stress, but in each case the key signal is the phosphorylation state of RsbV (Dufour *et al.*, 1994, Alper *et al.*, 1996). RsbW is a kinase, and during normal exponential growth, RsbW inactivates RsbV by phosphorylation. RsbW is then free to bind σ^B and inhibit σ^B-dependent transcription. Under stress response, RsbV-PO_4 is dephosphorylated, and it is this form of RsbV that binds the RsbW-σ^B complex to release σ^B and allow transcription from σ^B-dependent promoters. A drop in the levels of ATP forms one of the important stress signals and the cell responds by inhibiting RsbW activity by the dephosphorylation of RsbV and RsbV catalyzed release of σ^B from the RsbW-σ^B complex. RsbW has been found interact with σ^B independently of ATP, while as ATP has been found to strongly inhibit binding between RsbV and RsbW (Alper *et al.*, 1996). In this system RsbP which acts as an energy-responsive phosphatase activated RsbV (Delumeau *et al.*, 2002), while RsbP is itself activated by RsbQ. The exact mechanism has not been yet defined but structural studies on RsbQ have shown it to be an α/β-hydrolase (Brody *et al.*, 2001) that is capable of binding some small yet to be identified hydrophobic molecule (Kaneko *et al.*, 2005) that may be used to activate RsbP (Jon *et al.*, 2007).

A separate regulatory mechanism allows induction of σ^B-dependent expression by stresses that have no direct effect on cell's ATP levels. In these cases, other regulatory proteins RsbR, RsbS, RsbT, RsbU, RsbV, RsbW and RsbX that are co-transcribed from the σ^B operon have been identified to affect RsbW and RsbV activities (Kang *et al.*, 1996, Voelker *et al.*, 1995, 1996 and 1996). Of these RsbU has been identified as a positive effector of σ^B-dependent transcription (Voelker *et al.*, 1995, 1995, Wise *et al.*, 1995). RsbU is a phosphatase of RsbV whose activity is

opposite to that of kinase activity of RsbW (Yang *et al.*, 1996) and has been implicated in the release of σ^B from RsbW-σ^B complex. RsbU activity in turn is controlled by RsbT which acts as a kinase, which in turn is regulated by phosphorylation state of RsbS. The phosphorylation state of RsbS is dependent on the opposing RsbT kinase and RsbX phosphatase activities. Two partner switching modules have been proposed for for the regulation of σ^B. One module, RsbU-RsbV-RsbW, composed of a phosphatase (U)-antagonist (V)-kinase(W), responds to energy stress through ATP levels, while the other module, RsbX-RsbS-RsbT, also composed of a phosphatase (X)-antagonist (S)-kinase(T) responds to environmental stress. Recently attempts have been made to characterize the role of the RsbR and the presence of a supramolecular protein assembly, the 'stressosome' has been reported. In a series of gel-filteration experiments on the interactions of RsbR, RsbS and RsbT, RsbR and RsbS proteins have been reported to form high molecular weight complex that can also bind RsbT (Chien-Cheng *et al.*, 2003). These complexes have been subsequently isolated from wild type *B. subtilis* (Delumeau *et al.*, 2006). The stressosome represents a signaling hub where environmental stress encompassing signals from small molecules, light, protein and peptide fragments (Hecker *et al.*, 2001) are integrated to effect a single signaling outcome (Jon *et al.*, 2007).

Bacillus subtilis appears to have exploited the anti-σ factor control mechanism to three areas of gene regulation viz., sporulation, energy stress and environmental stress by gene duplication of a key regulatory module: the phosphate-antagonist-kinase module. The fundamental processes involved in these regulatory networks involve various protein-protein interactions and protein ligand especially nucleotide interactions. The role of protein-nucleotide interactions in these regulatory networks apparently acts as a steric or electrostatic plug that alters the affinity of a protein for its binding partners and facilitates partner switching cascades.

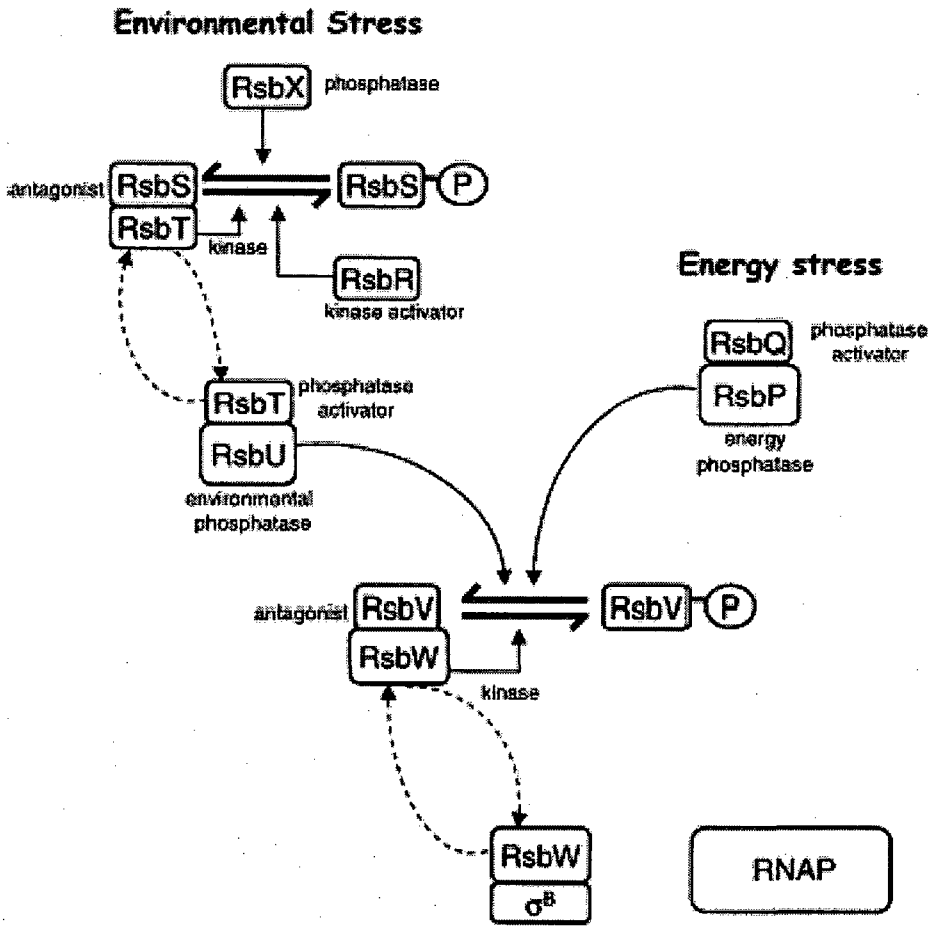


Fig. 1.6: Model for the regulation of σ^B

In the absence of stress, σ^B is held inactive in a complex with its antisigma factor, RsbW, which also inactivates RsbV by phosphorylation through its kinase activity. Depending on the nature of the imposed stress, either RsbP or RsbU (the phosphatases for RsbV-P) is activated. The RsbW: σ^B complex is disrupted by RsbV, with which RsbW forms an alternative complex (dashed arrows). Free σ^B can interact with RNA polymerase and direct the synthesis of the general stress response genes. In the environmental stress pathway RsbU is activated by RsbT, which, in the absence of stress, is believed to be inactivated by forming a complex with RsbS. During stress, the kinase activator RsbR facilitates the phosphorylation of RsbS by RsbT, which is released from the inhibitory complex formed with RsbS (dashed arrows). RsbX, the phosphatase for RsbS-P, has been shown to be responsible for the feedback mechanism by which the level of σ^B Activity returns to the prestress levels.(Adapted from Chien-Cheng *et al.*, 2003)

1.5 Posttranslational regulation of mycobacterial σ factors

The posttranslational regulation of σ factor activity has been reported in *Mycobacterium tuberculosis* also. To date four anti- σ factors have been identified in the genome of *M. tuberculosis* ; RshA, RseA, RslA and UsfX. RshA a σ^H-specific antagonist that senses redox potential through specific cysteine residues and therefore RshA binds σ^H in reduced environments. Under oxidizing conditions the disulphide bonds are formed resulting in the release of that becomes free to interact with RNA polymerase (Song *et al.*, 2003). RslA is a transmembrane protein and has been shown to bind σ^L (Hahn *et al.*, 2005) specifically and is able to inhibit σ^L dependent transcription *in vitro* (Dainese *et al.*, 2006). Because of being a transmembrane protein, RslA has been speculated to respond to the signals from cell envelope or from external environment (Sebastien *et al.*, 2006). RseA has been found to code for σ^E specific anti-σ factor but the signals and potential effectors regulating their interaction are still unknown (Sebastien *et al.*, 2006).

The role of UsfX as the anti-σ factor of σ^F has been debated since the identification of the σ^F operon (DeMaio *et al.*, 1997). The gene for *usfX* (Rv3287c) is located upstream of *sigF* (Rv3286c) and the two are co-transcribed. The physical interaction of UsfX and σ^F has been demonstrated and UsfX has been found to inhibit σ^F -dependent transcription from σ^F promoter. UsfX has been reported by same workers to be regulated by two antagonistic proteins: RsfA and RsfB. Both anti- anti-σ factors have been found to bind to UsfX and disrupt the UsfX-σ^F interaction ensuring the release of σ^F and transcription from σ^F-dependent promoters (Beaucher *et al.*, 2002). While as RsfA has been shown to respond to changes in redox potential RsfB is believed to respond in phosphorylation dependent manner but no evidences are yet available for this prediction. Parida and co-workers have shown the presence of other anti-anti-σ factors regulating σ^F and have identified their interactions with UsfX and σ^F (Parida *et al.*, 2005).

We carried out our work for exploring the structural aspects of UsfX interaction with its sigma factor and anti-anti-factor and identifying the nucleotide binding /nucleotide hydrolysis properties of UsfX.

CHAPTER 2

MATERIALS & METHODS

2.1 Introduction

Working on any scientific problem involves a number of methodologies at different stages of the experiment. In the following pages of this chapter the reader is introduced to various techniques that have been employed in the execution of the planned experiments. These include analysis of gene sequences, cloning of genes, protein purification, techniques used for biochemical and structural characterization of the proteins. Also the computational approaches employed have been detailed. A description of the media, strains, vectors and chemicals used in this study is detailed in Annexure.

2.2 Database searches, sequence alignment studies and phylogenetic analysis

Analysis of the sequences were carried out to try and identify signature motifs as also homologous proteins corresponding to Rv3286c (SigF), Rv3287c (UsfX) and Rv1365c (RsfA) from *M. tuberculosis* H37Rv. The analysis was preformed using the BLAST program (http://www.ncbi.nlm.nih.gov/blast/BLAST.cgi). Searches with BLASTP (Altschul *et al.*, 1997, Altschul *et al.*,2005) and Position specific iterated BLAST (PSI-BLAST) were performed against the available protein databases particularly *Protein Data Bank* (http://www.rcsb.org/) and the Non-redundant protein sequences (nr) database. Specific sequence motifs conserved in the respective proteins with respect to other known sigma factors, anti-sigma factors and anti-anti-sigma factors were identified. The phylogenetic analysis of the aligned sequences from various sources was carried out using the *ClustalX* program (Thompson *et al.*, 1997) using 10000 bootstrap trials for the analysis. The phylogenetic tree was viewed with the help of *TreeView32*.

Identification of conserved residues: To measure the level of sequence variation at different positions in the sequence alignment, the frequency of occurrence of the amino acid residues at each position was determined by the calculation of positional entropy at each position. The degree of amino acid variability was performed using *The Scorecons Server* (Valdar, 2002) (http://www.ebi.ac.uk/thronton-srv/databases/cgi- bin/valdar/scorecons_server.pl).

The entropy is known as the Shannon informational entropy and is calculated by

$$S = -\sum_{a}^{K} P_a \log_2 P_a \qquad (1)$$

where K = number of residue types

$P_a = n_a/N$

N = number of residues in column

n_a = number of residues of type a

The entropy scores calculated from *The Scorecons Server* are normalized for Shannon's entropy so that conserved (low entropy) columns score 1 and diverse (high entropy) columns score 0.

2.3 Primers and vectors

Primers for PCR amplification of the proteins were designed using the sequences of UsfX (Rv3287c), SigF (Rv3286c) and RsfA (Rv1365c) genes from *M. tuberculosis* H37Rv available in the Tuberculist database. The primers were designed using the OLIGO program. The primers used for cloning were as follows:

UsfX

Forward primer	5'-CTGGGTACCGCCATGGCAGACTCGGATTTA-3'
Reverse primer	5'-CGGGATCCCACCTGCTCGATGCCG-3'

SigF

Forward primer	5'-CGGGATCCCATGGCTGCGCGCGCT-3'
Reverse primer	5' CCCAAGCTTCTCCAACTGATCCCGTAGGCGTGC-3'

RsfA

Forward primer	5'-TGGATCCATGGACCCGACTCAGGCAGGTTC-3'
Reverse primer	5'- CCCAAGCTTGTAGCAGACAGCGCGGACTCC-3'

Co-expression Vector

UsfX

Forward primer	5'-CTGGGTACCGCCATGGCAGACTCGGATTTA-3'
Reverse primer	5'-CGGGATCCTCACACCTGCTCGATGCCG-3'

SigF

Forward primer 5'-AGCCCATGGATCCCCGCGCGCGCTGC -3'

Reverse primer 5'CCCAAGCTTTCACTCCAACTGATCCCGTAGGCGT-3'

rbsfusion primer 5'CGGGATCCAAGAAGGAGATATACATATGCGGGC-3'

Mutant UsfX

ΔK9A

Primer1 5'-ATGGCAGACTCTGATTTACCCACC**GCG**G-3'

Primer2 5'-CGGGATCCTCACACCTGCTCGATGCCG-3'

ΔK83A

Primer1 5'-GTGGTCGATCCGCGA**GCA**GACGAAGTT-3'

Primer2 5'-CACCACAACTT**CGT**CTGCTCGCGGATC-3'

Underlined nucleotides denote the sites recognized by corresponding restriction enzymes. UsfX has KpnI and NcoI in forward and BamHI in reverse; SigF contains BamHI and NcoI in forward and HindIII in reverse while as RsfA contains BamHI and NcoI in forward and HindIII in reverse. For co-expression vector the primers for amplification of *usfX* were same as used for the cloning while as SigF was cloned using NheI in forward primer and HindIII in reverse primer. BamHI was introduced in the forward primer of the ribosomal binding site (rbs) fusion primer. UsfX and SigF were cloned into pET21d and pQE31 respectively while for co-expression SigF was first cloned into pRSETB to fuse it with the ribosomal binding site.

Bold and underlined nucleotides denote the sequence of alanine replacing lysine in mutant *usfX*.

2.4 Cloning and over-expression

2.4.1 UsfX

The *usfX* gene was PCR amplified [Denaturation – 97°C (2 min), annealing – 59°C (1 min), extension – 72°C (1 min 30 sec), 35 cycles] from *M. tuberculosis* genomic DNA using the primers described above using a GeneAmp PCR system 2400 (Perkin Elmer). The PCR reaction product was electrophoresed in 1% agarose gel and 0.438 Kb product was excised from the gel and purified using DNA gel band extraction kit. The PCR product was digested with NcoI and and BamHI cloned into T7 promoter based expression plasmid pET21d at the same sites. All the digestion and ligation steps were carried out as per standard protocol (Sambrook *et al.*, 1989). The clones were screened by restriction digestion. The final confirmation was done by DNA sequence analysis. Here onwards the protein has been named as MtUsfX. The clone of *usfX* in pET21d was transformed into *E. coli* host cells compatible for the T7-based expression plasmids. Single colonies were grown at 37°C in 10 ml of LB and YT media containing 100μg/ml ampicillin to an A_{600} of 0.4, 0.6 and greater than 1.0 induced with 0.3mM to 1mM IPTG and grown for further 4-12 hrs. The level of induction was monitored using 12% SDS-PAGE (Laemmli, 1976). Over-expression was achieved in C41(DE3) by inducing with 0.3mM of IPTG at an A_{600} of 0.8 and growing for 4hrs.

2.4.2 SigF

PCR amplification [Denaturation – 97°C (2 min), annealing – 55°C (1 min), extension – 72°C (1 min 30 sec), 35 cycles] of *sigF* gene was performed from *M. tuberculosis* genomic DNA using the respective primers. The PCR product was purified as described above and was then digested with BamHI and HindIII and cloned into T5 promoter based expression plasmid pQE31. Screening of clones was done by restriction digestion and finally confirmed by DNA sequence analysis. Here onwards the protein has been named as MtSigF. The clone of *sigF* in pQE31 was transformed into *E. coli* host cells compatible for the T5 and T7 promoter based expression plasmids. Single colonies were grown at 37°C in 10 ml of LB and YT media containing 100μg/ml ampicillin to an A_{600} of 0.4, 0.6 and greater than 1.0 induced with 0.3mM to 1mM IPTG and grown for further 4-12 hrs. The level of induction was monitored using 12% SDS-PAGE. Over-expression was achieved in JM109 and

C41 (DE3) by inducing with 0.5mM of IPTG at an A_{600} of greater than 1.0 and growing for 8hrs.

2.4.3 RsfA

The *rsfA* gene was PCR amplified [Denaturation – 97°C (2 min), annealing – 59°C (1 min), extension – 72°C (1 min 30 sec), 35 cycles] from *M. tuberculosis* genomic DNA using the primers described above using a GeneAmp PCR system 2400 (Perkin Elmer). The PCR reaction product was electrophoresed in 1% agarose gel and 0.387 Kb product was excised from the gel and purified using DNA gel band extraction kit. The PCR product was digested with NcoI and and BamHI cloned into T7 promoter based expression plasmid pET21d at the same sites. All the digestion and ligation steps were carried out as per standard protocol (Sambrook *et al.*, 1989). The clones were screened by restriction digestion. The final confirmation was done by DNA sequence analysis. Here onwards the protein has been named as MtRsfA. The clone of *rsfA* in pET21d was transformed into *E. coli* host cells compatible for T7 promoter based expression plasmids. Single colonies were grown at 37°C in 10 ml of LB and YT media containing 100µg/ml ampicillin to an A_{600} of 0.4, 0.6 and greater than 1.0 induced with 0.3mM to 1mM IPTG and grown for further 4-12 hrs. The level of induction was monitored using 12% SDS-PAGE. Over-expression was achieved in *E. coli* Origami (DE3) using 2X YT medium was induced in mid log phase (corresponding to an absorbance of 600nm of 0.6-0.7) by IPTG and grown for further after 12hrs. The cultures were also supplemented by 5µg/ml tetracycline and 50µg/ml kanamycin to maintain the Origami strain.

2.4.4 Co-expression construct of UsfX-SigF complex

usfX and *sigF* are in a single operon in *Mycobacterium tuberculosis* and the stop codon of *usfX* lies inside the ORF of *sigF*. A co-expression vector was constructed where the two genes were cloned in a way as they are present in *Mycobacterium tuberculosis* genome. This facilitated the large scale purification of UsfX-SigF complex for biochemical and biophysical studies.

Following steps were involved in the construction of the vector:

Amplification of usfX gene: The *usfX* gene was PCR amplified from *M. tuberculosis* genomic DNA as described in section 2.4.1 except for reverse primer carried a stop codon immediately at end of *usfX* sequence. The amplified product was digested with BamHI and used for ligation.

Cloning of sigF gene into pRSETB vector with n-terminal His tag: The *sigF* gene was PCR amplified from *M. tuberculosis* genomic DNA and cloned into pRSETB by the methodology as described in section 2.4.2 with the difference that a different set of primers was used (Section 2.3).

Fusion of ribosomal binding site of pRSETB with sigF gene: A forward primer was designed for the PCR amplification of *sigF* gene along with the ribosomal binding site (rbs) of the vector. A BamHI restriction site was also introduced in the forward primer. The reverse primer used introduced a stop codon immediately at the end of *sigF* sequence. The PCR amplified fragment was digested with BamHI and used for ligation. The product is named as *SigFrbs*.

Ligation of SigFrbs with usfX gene: The digested products of *SigFrbs* and *usfX* gene were ligated at 16°C using 2% PEG8000 as a crowding agent in one of the reactions. The ligation reaction products were electrophoresed in 1% agarose gel and 1.3kb product was excised from the gel and purified using DNA gel band extraction kit . The product was named as *lUS*.

Cloning and expression: Using *lUS* as template, a PCR amplification reaction [Denaturation – 97°C (2 min), annealing – 55°C (1 min), extension – 72°C (2 min 30 sec), 35 cycles] was performed using forward primer of *usfX* and reverse primer of *sigF*. The PCR amplified product was digested with restriction enzymes NcoI and HindIII and cloned into same sites in pET 21d. The clone was confirmed by restriction digestion and DNA sequence analysis. The clone has been named as *pUS*. The constructed vector was transformed into *E. coli* host cells compatible for the T7-based expression plasmids. Single colonies were grown at 37°C in 10 ml of LB and YT media containing 100μg/ml ampicillin to an A_{600} of 0.4 - 0.6 and with 0.3mM to 1mM isopropyl-1-thio-β-D-galactopyranoside, and grown for further 4-16 hrs. The level of induction was monitored using 12% SDS-PAGE. Over-expression was achieved in C41 (DE3) by inducing with 0.3mM of IPTG and growing for 8hrs.

2.4.5 Mutant UsfX

A double mutant for the two lysine residues was generated by using the primers given in section 2.3. The mutations were introduced by modifying the residue specific nucleotides in the oligonucleotides. A PCR based strategy was used for site-directed mutagenesis in which the products of the first reaction acted as template as well as primer for second reaction. A nucleotide change was introduced in the primer so as to

change lysine to alanine. The conditions for the PCR amplification were as Denaturation – 97°C (2 min), annealing – 59°C (1 min), extension – 72°C (1 min), 35 cycles. For ΔK9A change in the nucleotide sequence was introduced in the forward primer and reverse primer used was as used for cloning. The 438bp PCR product was electrophoresed in 1% agarose gel, excised from the gel and purified using DNA gel band extraction kit. Purified fragment was used as template for the introduction of K83A mutation. For ΔK83A mutation the change was introduced as explained in the scheme above. The two PCR products were run on 1.5% agarose gel and the fragments of 249bp and 156bp were excised from the gel, purified and used in a mixture as template for amplification from the cloning primers.

The further steps for cloning and expression were the same as explained in section 2.4.1.

2.5 Purification

2.5.1 MtUsfX and its mutant

One-liter medium of YT containing 100μg/ml ampicillin were inoculated with 1% inoculum and grown at 37°C until the A_{600} reached 0.8.The cultures were induced by the addition of 0.3mM IPTG and grown for 4hrs. The cells were harvested by centrifuging at 7000 rpm for 10 minutes, resuspended in buffer A (50mM Tris-Cl, pH 8.5, 50mM NaCl,10mM Imidazole), and lysed by sonication at 16 % output power, 50 % pulsar duty cycle with a pulse time of 8 min. The cell lysate was centrifuged at 14000 rpm for 20min at 4°C to remove the cell debris. The clear supernatant was loaded on a Ni^{2+}-IDA column pre-equilibrated with Buffer A (50mM Tris-HCl pH 8.5, 50mM NaCl, 10mM Imidazole). The column was washed with 5 column volumes of buffer A. The protein was eluted using a linear gradient of buffer B containing 500mM Imidazole. The protein eluted at 35-55% Imidazole (175-275 mM). Purity was monitored by electrphoresing the samples on a 12% SDS-PAGE. Fractions with less contamination were pooled and precipitated with 60% ammonium sulphate. Precipitate was dissolved in minimum volume of gel filtration running buffer (50mM Tris-Cl, pH 8.5, 50mM NaCl, 5mM EDTA, 2mM β-mercaptoethanol).Trace contaminants were removed by gel filtration on a Superdex S-75 HR10/300 column on AKTA FPLC system. For protein to be utilized for ATPase assay slight modification was made in the purification procedure. After elution from Ni^{2+}-IDA column the protein was concentrated in a 10kDa cut-off centricon and run on S-75 column in 50mM HEPES,pH 7.5 and 50mM KCl.

2.5.2 MtSigF

One-liter medium of YT containing 100μg/ml ampicillin were inoculated with 1% inoculum and grown at 37°C until the A_{600} reached greater than 1.0. The cultures were induced by the addition of 0.3mM IPTG and grown for 8hrs. The cells were harvested by centrifuging at 7000 rpm for 10 minutes, resuspended in buffer A (50mM Tris-Cl, pH 8.5, 50mM NaCl, 10mM β-mercaptoethanol and 10mM Imidazole) and lysed by sonication at 16 % output power, 50 % pulsar duty cycle with a pulse time of 12 min. The cell lysate was centrifuged at 14000 rpm for 30min at 4°C to remove the cell debris. The supernatant was loaded on a Ni^{2+}-IDA column pre-equilibrated with Buffer A (50mM Tris-HCl pH 8.5, 50mM NaCl, 10mM β-mercaptoethanol, 10mM Imidazole). The column was washed with 5 column volumes of buffer A. The protein

was eluted using a stepwise gradient of buffer A containing 50mM, 100mM, 150mM, 200mM and 300mM Imidazole. The protein eluted at 200mM Imidazole. Purity was monitored by electrphoresing the samples on a 12% SDS-PAGE. Fractions with less or no contamination were pooled and precipitated with 80% ammonium sulphate. Precipitate was dissolved in minimum volume of gel-filtration running buffer (50mM Tris-Cl, pH 8.5, 50mM NaCl, 5mM EDTA, 10mM β-mercaptoethanol). Trace contaminants were removed by gel filtration on a Superdex S-75 HR10/300 column on AKTA FPLC system.

2.5.3 MtRsfA

One-liter medium of YT containing 100μg/ml ampicillin, 5μg/ml tetracycline and 50μg/ml kanamycin were inoculated with a 1%inoculum and grown at 37°C until the A_{600} reached greater than 1.0.The cultures were induced by the addition of 0.5mM isopropyl-1-thio-β-D-galactopyranoside and grown for 12hrs. The cells were harvested by centrifuging at 7000 rpm for 10 minutes, resuspended in buffer A (50mM Tris-Cl, pH 8.5, 50mM NaCl) and lysed by sonication at 16 % output power, 50 % pulsar duty cycle with a pulse time of 12 min. The cell lysate was centrifuged at 14000 rpm for 20min at 4°C to remove the cell debris. The supernatant was loaded on a Ni^{2+}-IDA column pre-equilibrated with Buffer A (50mM Tris-HCl pH 8.5, 50mM NaCl, 10mM Imidazole). The column was washed with 5 column volumes of buffer A. The protein was eluted using a stepwise gradient of buffer A containing 50mM, 100mM, 150mM, 200mM and 300mM Imidazole. The protein eluted at 150mM Imidazole. Purity was monitored by electrphoresing the samples on a 12% SDS-PAGE. The purified fractions were pooled and concentrated in a 30KDa centricon and loaded into the gel filtration column Superdex-75 pre-equilibrated with 50mM Tris-HCl pH 8.5, 50mM NaCl and 5mM EDTA.

2.5.4 UsfX-SigF complex

One-liter medium of YT containing 100μg/ml ampicillin were inoculated with a 1%inoculum and grown at 37°C until the A_{600} reached 0.3-0.4.The cultures were induced by the addition of 0.3mM IPTG and grown for 6hrs. The cells were harvested by centrifuging at 7000 rpm for 10 minutes, resuspended in buffer A (50mM Tris-Cl, pH 8.5, 50mM NaCl, 10mM $MgCl_2$, 1mm ATP, 10mM Imidazole and 12%glycerol), and lysed by sonication at 12 % output power, 50 % pulsar duty cycle with a pulse time of 8 min. The cell lysate was centrifuged at 14000 rpm for 20min at 4°C to

remove the cell debris. The supernatant was loaded on a Ni^{2+}-IDA column pre-equilibrated with Buffer A (50mM Tris-HCl pH 8.5, 50mM NaCl, 10mM $MgCl_2$, 1mm ATP, 10mM Imidazole). The column was washed with 5 column volumes of buffer A. The protein SigF complexed with UsfX was eluted using an increasing gradient of Imidazole in buffer A in a batch wise manner. The proteins (US) eluted in the range of about 200mM Imidazole. The eluted fractions of US complex were pooled and concentrated in a 30KDa centricon and loaded into the gel filtration column Superdex-75 pre-equilibrated with 50mM Tris-HCl pH 8.5, 50mM NaCl, 10mM $MgCl_2$ and 1mM ATP. The purification of the complex could be achieved in the absence of 10mM $MgCl_2$ and 1mM ATP also.

Protein Estimation

Protein concentration was determined by the Bradford method (Bradford, 1976) using BSA as a standard.

2.6 Pull-down binding assay

The pull-down assay is an *in vitro* method used to determine physical interaction between two or more proteins. Pull-down assays are useful for both confirming the existence of a protein:protein interaction predicted by other research techniques (e.g., co-immunoprecipitation, yeast two-hybrid and density gradient centrifugation) and as an initial screening assay for identifying previously unknown protein:protein interactions. The minimal requirement for a pull-down assay is the availability of a purified and tagged protein (the bait) which will be used to capture and 'pull-down' a protein-binding partner (the prey). The co-expression vector acted as a tool for pull-down assay for UsfX-SigF interaction analysis. The presence of N-terminal 6x Histidine tag in MtSigF made it to act as a bait protein. The bacterial culture of *pUS* transformed C41(DE3) was used for protein purification as described in section 2.4.4. Control experiment of non-his tagged MtUsfX purification was utilized to rule out non-specific binding of MtUsfX to Ni^{2+}-IDA column. The presence of MtUsfX in the complex was confirmed by nucleotide binding of the UsfX-SigF complex.

2.7 *in vitro* protein-protein interaction analysis

The interaction analysis between MtUsfX and MtRsfA was performed by setting up *in vitro* reactions between equimolar quantities of MtUsfX and MtRsfA in 50mM Tris-HCl pH 8.5, 50mM NaCl. Two sets of reactions were set one containing 5mM DTT and one without DTT. The mixtures were incubated on ice for 2-3hrs.The formation of complex was monitored by gel-filtration analysis of the reaction mixtures.

2.8 Structural characterization

2.8.1 Size-exclusion chromatography (SEC)

Gel filtration experiments were carried out on a Superdex 75 HR 10/300 column (manufacturer's exclusion limit 70 kDa) with AKTA Fast Performance Liquid Chromatography (GE Biosciences). The running buffers used during purification were the same as indicated in the purification protocols above. The column was equilibrated and run with 50mM Tris-HCl pH 8.5, 50mM NaCl and 5mM EDTA for all the molecular weight analysis experiments. For molecular weight analysis 200 µl (containing at least 0.4 mg protein) of the sample was loaded on the column and run in buffer containing 50mM Tris-HCl, pH 8.5, 50mM NaCl and 5mM EDTA maintained at 25°C at a flow rate of 0.3 ml/min and eluted protein was detected at 280 nm. The column was calibrated with various molecular weight standard markers like Albumin (66kDa), Ovalbumin (43kDa), Chymotrypsinogen (23.5kDa) and Ribonuclease (13.7kDa) (GE Biosciences). The partition coefficient (K_{av}) was plotted as a function of log MW of the standard proteins according to the equation

$$K_{av} = V_e\text{-}V_o/V_t\text{-}V_o \qquad (2)$$

where V_e is the elution volume of the protein, V_o is the void volume of the column and V_t is the total column volume.

Gel filtration chromatography was also used to monitor changes in MtRsfA as a result of reaction with DTT. 0.5mg protein in 200ul buffer containing 50mM Tris-HCl pH8.5 and 50mM NaCl and 10mM DTT were run in gel filtration column Superdex 75 under the conditions above.The control run was done in buffer without DTT. The conformational changes were detected in molecular Stokes radius. Molecular Stokes radii (R_S) were estimated from partition coefficient (Kav) by plotting a calibration curve of known proteins albumin (35.5 Å), ovalbumin(30.5Å), chymotrypsinogenA (20.9 Å) and ribonucleaseA (16.4 Å).To estimate the individual peak position

accurately of native and reduced protein, a Gaussian deconvolution algorithm (Keneth *et al.*, 1995) was used, using software Origin version 5.0 (Origin lab).

2.8.2 MALDI-TOF

The protein solution of MtUsfX was characterised on a mass spectrophometer MICRO-MASS QUATTRO II mass spectrometer (Micro mass, Altricem, United Kingdom) to verify its precise molecular weight. The samples were precipitated overnight with 10% trichloro-ethanol (TCA) on ice, washed twice with ice cold acetone and pelleted by centrifuging at 14000rpm. The pellet was dried at room temperature and dissolved in MilliQ water and subjected to mass analysis.

2.8.3 Densitometric analysis

Densitometric analysis was used as one of the tools to determine the stoichiometry in protein-protein interaction analysis of MtUsfX with MtSigF and MtRsfA. Fractions from the gel filteration column were analyzed by SDS-PAGE and stained with Coomasie blue. The gels were then scanned and the intensity of the bands in the peak fractions were quantified. Since Coomassie blue staining is roughly proportional to the protein mass in the band, we compensated for the different molecular mass of each protein by dividing the observed intensity by the molecular mass. Two intensities were measured for each protein, and the average calculated. Finally, we calculated the stoichiometry of the complex by dividing the average intensity of MtUsfX over MtSigF or MtRsfA. Densitometry analysis was done utilizing Image QuantTL software (GE Biosciences).

2.8.4 Tryptophan fluorescence

Identification of emission maxima of intrinsic tryptophan fluorescence: Fluorescence spectra were recorded with a Perkin Elmer Life Sciences LS 50B spectrofluorimeter in a 5 mm path length quartz cell at 25°C. The spectra were recorded at 25 °C with excitation at 285nm and fluorescence monitored from 300-420nm. Typically the spectra were acquired for 0.5μM protein samples dissolved in 50mM Tris-HCl pH8.5 and 50mM NaCl with a scan speed of 200nm/sec, each spectra representing an average of two accumulations. The reading for every sample was taken in triplicates.

Solute accessibility studies: Solvent phase quenchers are often employed to investigate the accessibility of fluorescent probes and to trace the micro-environments around these probes. We used these studies to trace the micro-environment around

tryptophan of MtUsfX. Further quenching studies were used to trace the changes in solvent accessibilities as a result of nucleotide binding. Acrylamide was used as the quencher for these studies. Fluorescence dynamic quenching of a fluorophore during its excited-state lifetime occurs as a result of the random collision of quenching ligand (Q) and the fluorophore (Bujalowski *et al.*, 1994). The deactivation of the excited state in absence of a static component is described by equation (Stern *et al.*, 1919)

$$F_0/F=1+K_{SV}[Q] \quad (3)$$

where F_0 and F are the fluorescence intensities in the absence and presence of the quenching ligand at concentration [Q], K_{SV} is the Stern-Volmer quenching constant. When plotted in this fashion, the data from a single fluorescent species produce a straight line with a slope that is proportional to accessibility.

The efficiency of solute quenching (γ) is defined as $\gamma = k_q^b/k_q^f$, where k_q^b and k_q^f are the bimolecular quenching rate constants for the free and bound fluorophores respectively, (in our case tryptophan residues in MtUsfX protein and free tryptophan in solution). If the effect of two different quenching ligands (Q1 and Q2) is studied, the ratio of two different quenching efficiencies for Q1 and Q2 in terms of their Stern-Volmer constants can be defined as (Bujalowski *et al.*, 1994, George *et al.*, 1991)

$$\gamma_1/\gamma_2 = (K_1^b/K_2^b) / (K_2^f/K_2^f) \quad (4)$$

The ratio γ_1/γ_2 reflects the relative accessibilities of the tryptophan to Q_1, as compared to Q_2, which can be a consequence of steric hindrances and/or interactions of quencher with fluorophore surroundings (Bujalowski *et al.*, 1994). Acrylamide was used as the reference quencher to which the quenching efficiency of I^- was compared.

For studying the UsfX-SigF interaction studies were performed with 1μM MtUsfX by increasing the molar ratio from a 100uM MtSigF solution from 0-2.5 with an increment of 0.25 μM in each reaction

2.8.5 Circular Dichroism Measurements

CD measurements were made with a Jasco J 810 spectropolarimeter calibrated with ammonium (+)-10-camphorsulfonate. The spectra were calculated from 1 μM protein samples dissolved in a 20mM Tris-Hcl pH8.5 and 30mM NaCl buffer. Typical spectra were recorded from 190 to 250 nm for far UV at a scan speed of 10nm per min., with each spectrum representing an average of 3 accumulations. Each spectrum was an average of two scans. During acquisition the samples were maintained at a regulated

temperature (25-30°C). The results were expressed as molar ellitipicity, [θ], or the residue ellipticity, $[\theta]_{MRW}$, are calculated from the measured θ (in degrees)

$$[\theta] = \theta * 100 * Mr/c * l \quad (5)$$

$$[\theta]_{MRW} = \theta * 100 * Mr/c * l * N_A \quad (6)$$

where θ is the measured ellipticity in degrees, c is the protein concentration in mg/ml, l is the pathlength in cm, and Mr and MRW are the protein molecuar weight and mean residue weight,respectively. N_A is the number of amino acids per protein. $[\theta]_{MRW}$ has units degrees * cm^2 * $dmol^{-1}$ .The values obtained were normalized by subtracting the baseline recorded for the buffer.

Thermal stability analysis: The thermal denturation experiments with the individual proteins and the complexes were performed in the above said spectropolarimeter with the temperature maintained with the help of PFD-425S Peltier temperature controller in the instrument. The samples were scanned from 25-100 °C with the ellipticity values at 222nm automatically recorded at 2 °C intervals. The folded fraction of protein at any temperature was determined as follows:

$$([\theta]^{obs}-[\theta]^{den})/([\theta]^{nat}-[\theta]^{den}) \quad (7)$$

$[\theta]^{obs}$ denotes the ellipticities at any temperature, $[\theta]^{den}$ at highest temperature, $[\theta]^{nat}$ at lowest temperature, respectively.

2.9 Fluorescence based nucleotide binding assay

The quenching of intrinsic tryptophan fluorescence as a consequence of ligand binding has been used as prelude to quantify the extent of ligand binding to protein. The extent to which nucleotides bind MtUsfX was determined by monitoring the fluorescence emission of a fixed concentration of protein and either titrating or pre-incubating with the given nucleotide concentration. We preferred the pre-incubation method to avoid the dilution caculations. Fluorescence spectra were recorded with excitation performed at a wavelength of 290nm and the emission values were taken those at 349nm. Background emission was eliminated by subtracting the signal from either buffer alone or bufer containing the appropriate quantity of substrate (Picard-Jean *et al.*, 2007). The binding can be described by eqn (8),

$$K_d = ([P][N])/[P.N] \qquad (8)$$

Where K_d is the apparent dissociation constant, [P] is the concentration of the protein, [P.N] is the concentration of the complexed protein and [N] is the concentration of unbound nucleotide.

The proportion of nucleotide-bound protein as described by eqn (8) is related to the measured fluorescence emission intensity by eqn (9),

$$\Delta F/ \Delta Fmax = [P.N]/[P]_{tot} \qquad (9)$$

Where ΔF is the magnitude of the difference between the observed fluoresence intensity at a given concentration of nucleotide and the fluorescence intensity in the absence of nucleotide, ΔFmax is the difference at infinite [Nucleotide], [P] is the total protein concentration.

If the total nucleotide concentration, $[N]_{tot,}$, is in large molar excess relative to $[P]_{tot,}$ then it can be assumed that [Nucleotide] is approximately equal$[Nucleotide]_{tot}$. Equations (8) and (9) can be combined to give eqn (10),

$$\Delta F/\Delta Fmax=[Nucleotide]_{tot}/(K_d+[Nucleotide]_{tot}) \qquad (10)$$

The K_d values were determined from non-linear least-squares regression analysis of titration data by eqn (10).

The stoichiometry of binding was established from a linear version of the Hill equation,

$$\log(\Delta F/\Delta Fmax-\Delta F)= n_H \log[Nucleotide]-\log K' \qquad (11)$$

where n_H (Hill Constant) is the order of the binding reaction with respect to ligand concentration and K'is the concentration of nucleotide that yields 50% of ΔFmax.

Correction for inner filter effect: Correction for apparent decrease in fluorescence of MtUsfX due to inner-filter effect of absorption of incident light by nucleotide was corrected by equation

$$F_{corr} = F_{obs}/10^{-m[N]} \quad (12)$$

where N is nucleotide concentration and *m* is a constant determined for ratios of at saturated nucleotide concentrations[at least 10 fold dissociation constant].

2.9.1 Effect of salts on binding affinity

The dependence of nucleotide binding on ionic strength was examined for the case of ATP. All ATP binding experiments were carried out in 50mM Tris-HCl pH8.5 at 25 °C by studying the effect on binding of 1mM ATP with the appropriate amount of KCl added to generate the ionic strength of buffer to 500mM. The results were plotted as increase in %ΔF as a function of ionic strength *I*. The control experiments were done by titrating the protein with KCl alone.

2.9.2 Role of divalent ions in binding

The role of divalent ion $MgCl_2$ in the binding of ATP to MtUsfX was monitored by observing the enhanced quenching of tryptophan fluorescence by a fixed concentration of ATP in presence of increasing concentration of $MgCl_2$. In order to determine the role of $MgCl_2$ in ATP binding, the enzyme in presence of 1mM ATP in 50mM Tris-HCl pH8.5 was treated with an increasing concentration of $MgCl_2$ from 0-30mM.The monovalent ionic strength of the medium was held at 0.050 M by adding the appropriate quantity of KCl. The results were plotted as increase in %ΔF as a function of increasing concentration of $MgCl_2$. $MgCl_2$ alone also produced decrease in fluorescence intensity so standard column graph was plotted in which fluorescence change in presence of ATP alone, $MgCl_2$ alone and in presence of both ATP and $MgCl_2$ was plotted.

2.10 NTPase activity assay

Purified MtUsfX was assayed for nucleotide hydrolysis activity (ATPase/GTPase) by analysing the the release of [^{32}P]Pi from [γ-^{32}P]NTP (Yamanaka *et al.*, 2000). 0.5 ug protein from gel filteration peak fraction was assayed in reaction mixtures (15µl) consisting of 1µCi [γ-^{32}P]NTP, in 50mM HEPES-KCl buffer pH7.4 and 10mM $MgCl_2$ at 37°C for varying time. The reaction was terminated by addition of 0.75 M

NaH_2PO4 (pH 3.65). Portions of the terminated reaction mixture (1µl) were spotted onto a silica gel thin-layer chromatography plate, which was developed in 1M LiCl and 0.35 M NaH_2PO4 (pH 3.65). The plate was autoradiographed to identify hydrolyzed products. In this solvent unhydrolysed ATP appears as a retarded spot towards the base of the plate and released [^{32}P]Pi as fast moving upper spot. The intensity of the spots was calculated was measuring the densities of the spots. Densitometry analysis was done utilizing Image QuantTL software (GE Biosciences). The activity was expressed as % activity (%ATP or GTP hydrolysed). Briefly, the density of both unhydrolysed and hydrolysed spots was estimated and % activity was calculated as % hydrolysed of the total intensity of two spots.

$$\% \text{ Activity} = (I_{Pi} / I_{Total})*100 \quad (13)$$

where I_{Pi} is the intensity of the inorganic phosphate band (fast moving band on TLC) and I_{Total} is the total intensity i.e, unhydrolyzed ATP and and the released inorganic phosphate.

2.10.1 Substrate specificity analysis

To determine the substrate specificity of MtUsfX competition experiments were carried out with unlabelled ATP and GTP at concentrations ranging from 20-100µM. Reaction mixtures (15µl) containing 50 mM HEPES-KCl buffer pH7.4 and 10mM $MgCl_2$ and 1µCi [γ-^{32}P]NTP were incubated for 45 min at 37°C. Concentrations of unlabelled nucleotides ranging from 20-100µM were added into the reaction mixture. Competition experiments were also performed with ATPγS, a slowly hydrolyzing analog of ATP to observe the effects in presence of a competitive analog. The results were plotted as retained % activity in presence of respective competitor/analog.

2.10.2 Effect of salts

Reaction mixtures (15µl) containing 50 mM HEPES buffer pH7.4 and 10 mM $MgCl_2$ and 1µCi [γ-^{32}P]NTP were incubated for 45 min at 37°C at NaCl and KCl concentration ranging from 0 to 400 mM. Extent of hydrolysis was plotted as % activity as a function of salt concentration.

2.10.3 Role of divalent ions

For divalent cations, reaction mixtures (15 µl) containing 50mM HEPES-KCl buffer pH7.4, 1µCi [γ-^{32}P]NTP and 10 mM divalent cations each of $CaCl_2$, $CoCl_2$, $MgCl_2$, $MnCl_2$, $NiCl_2$, $SrCl_2$ and $ZnCl_2$ were incubated for 45 min at 37°C. Extent of hydrolysis was plotted as % activity as a function individual cation.

2.11 Crystallization

2.11.1 Preparation

The presence of extraneous substances during setting up of crystallization may act as initial nuclei for crystal formation. Control of excessive nucleation by filtration and general cleanliness has been shown to be conducive to the formation of better crystals (Chayan *et al.*, 1993, Blow *et al.*, 1994). All buffers, crystallization reagents, solutions were filtered through 0.22μm filters. Protein solutions were centrifuged at 10,000g for 5 min to settle extraneous debris prior to setting up drops. Microscopic cover slips used in vapour diffusion method needed special cleaning treatments. These cover slips were thoroughly washed with cleaning solution, rinsed with MilliQ, dried and then coated with dilute siliconizing solution.

2.11.2 Crystallization trials

Preliminary crystallization trials of MtUsfX, MtSigF, MtRsfA, UsfX-SigF complex and UsfX-RsfA complex were performed using hanging drop vapor diffusion method in a "Sparse matrix" (Jancarik and Kim, 1991), "Crystal screen 2" (Jancarik and Kim., 1991; Cudney *et al.*, 1994) and "Mixed precipitant screens" (Mazeed *et al.*, 2003). The basis for utilizing sparse matrix and "Crystal screen 2" is that each methodology evaluates ~ 50 unique combinations of pH, salts and precipitants which cover a diverse range of chemical entities. "Mixed precipitant screens" contains a combination of sixty-four precipitant mixtures the individual components having mechanistically distinct properties. The "Mixed precipitant screens" was used at a 1:1 dilution initially. For crystallization of protein-protein complexes in addition to the above screening methods those described by Sergei were utilized (Sergei 2002, Sergei *et al.*, 2006). Both these methods involve using an enlarged screen around those crystallizing agents which have been reported for protein-protein complexes in Protein Data Bank. Initial crystallization trials for all the forms were set up by mixing 2 μl of 2-12 mg/ml protein in 50 mM Tris-HCl, pH 8.5, 50 mM NaCl and 2 mM EDTA with an equal volume of reservoir solution at different temperatures ranging from 4°C to 25°C. For crystallization of MtSigF the buffer contained 5mM DTT to prevent protein aggregation. For crystallization of UsfX-SigF complex the buffer was supplemented with 10mM $MgCl_2$ and 1mM ATP. Crystallization attempts using sitting drop method was also employed in later stages to screen for initial conditions.

2.11.3 Exploration of leads for MtUsfX

No leads in crystallization trials of MtSigF, MtRsfA, UsfX-SigF complex and UsfX-RsfA were observed for further exploration. Following crystal leads were obtained in case of MtUsfX in "Mixed precipitant screens":

A. 4M NaCl, 5% Isopropanol – pH 7.5

B. 2M Na Formate, 2.5% PEG 3350 and 15% Isopropanol – pH 8.5

C. 1M Li_2SO_4, 15% MPD and 0.1M $MgSO_4$ – pH 4.5

Crystal leads obtained from condition "A" were very small crystallites too tiny to be used for diffraction so further optimization became essential. Optimizations were performed by screening the constituent concentrations over a wide range, varying the overall pH and temperatures. Nucleotide co-factors and combination of nucleotide and Mg^{2+} were also added into the crystallization solutions.

Various inorganic and organic additives were used in an attempt to improve crystalline conditions. Additives were used in range from 5 to 10mM concentration in a whole range of pH.

2.11.4 Seeding techniques for MtUsfX

Seeding techniques can be classified into two categories based on the size of the seeds:

• Microseeding—transfer of submicroscopic seeds, too small to be distinguished individually.

• Macroseeding—transfer of a single crystal, usually 5–50 μm.

Microseeding (Thaller *et al.*, 1981, 1985) was used to improve the crystallizing conditions of MtUsfX as the crystals obtained in the above conditions were too small for the macro-seeding technique. For the microseeding experiments, a complete drop of MtUsfX in all the above three described conditions were crushed with the help of capillary jets and transferred to a microcentrifuge tube and mixed in 50 μl well solution, centrifuged at 6000g for 5 min to exclude large nuclei, precipitate etc. Serial dilutions of the seeds were prepared upto a factor of 10^{-7} and crystallization trials at all dilutions were performed by pipetting the seed solution into crystallization drop. Streak seeding was also performed with the help of a rabbit whisker. In this method usually a plucker (an animal whisker or a seeding wand) is used, which is touched or stroked over the surface of the parent crystal to dislodge and trap the nuclei. The whisker is then drawn through the new drop, depositing the seeds in a streak line.

2.12 *In silico* structural analysis

2.12.1 Molecular modeling and energy minimization

Models for MtUsfX, MtSigF and MtRsfA were generated using the PHYRE (Protein Homology/analogy Recognition Engine) server at http://www.sbg.bio.ac.uk/~phyre/index.cgi. Subsequently, the DISCOVER program implemented in the *InsightII* (M/s Accelrys Inc.) program was used for the refinement of the raw models. The DISCOVER program performs functions like energy minimization, template forcing, torsion forcing and calculates properties such as integration energies and mean square deviations. Accordingly this can be used for energy refining a model-built structure, evaluate the conformations available to a model and also evaluate configurations or chemical perturbations of a system. *Discover* employs the steepest descent, conjugate gradients and Newton-Raphson algorithms. Energy minimizations were run initially using steepest gradient and then with conjugate gradient.

2.12.2 Protein-protein interaction analysis

The models of the protein-protein complexes were generated by superposing the individual models with the known structural homologs. This was done utilizing the BIOPOLYMER module of InsightII. The model for UsfX-SigF complex was developed by superposing the individual models of MtUsfX and MtSigF by structural superimposition of Sequence Conserved Regions (SCRs) with the structural homolog from *Bacillus stearothermophilus* (PDB ID:1L0O). The model coordinates were merged after the alignment and used as the raw model for refinement and energy minimization as before. Steric clashes were removed by visualization and adjustment using *InsightII* and minimized further. Similarly the model for UsfX-RsfA complex was developed by using the structural homolog from *Bacillus subtilis* (PDB ID:1TH8). The final models were refined by minimizing with both steepest decent and conjugate gradient algorithms.

2.12.3 Nucleotide docking

Active site/binding site identification of MtUsfX was carried out by a superposition of the *B. stearothermophilus* SpoIIAB-SigF complex structure (PDB:1L0O) with the model of MtUsfX. Docking experiments were carried out with ATP as well as other three nucleotides GTP,CTP and TTP for *in silico* analysis of the active site using AutoDock v 3.0.5 (Morris *et al.*, 1998). AutoDock provides an automated procedure for predicting

the interaction of ligands with biomacromolecular targets. Docking procedure can be divided into five steps:

1. Macromolecule/Target Preparation
2. Ligand Preparation
3. Generating Grid Maps
4. Running Autodock
5. Analyzing Results

Macromolecule preparation involves removal of all non-protein atoms and addition of all / polar hydrogens. This is followed by addition of Kollman united atom charges and solvation parameters defined for each residue in Addsol. The model of MtUsfX that had refined by energy minimization was chosen as the target molecule for docking analysis.

Ligand preparation involves addition of all hydrogens followed by Gasteiger charge calculation. Non-polar hydrogens and their corresponding charges are merged with the parent heavy atom. Rotatable bonds in the ligand are identified with Autotors and appropriate records are written to an output file. All the nucleotide ligands were sketched and converted to 3D and optimized in Builder module of *InsightII.* Optimized ligands were converted into respective pdbs. Prior to docking studies, the ligands translation, rotation and internal torsions were assigned.

Grid map consists of a three dimensional lattice of regularly spaced points, surrounding and centered on some region of interest. The analysis was done by setting central grid point around Trp106, a residue identified by fluorescence spectroscopy experiments to be in the active site vicinity. The number of points set in the x, y and z dimensions were as 76*76*76 centered on Trp106 with 0.4 Å spacing making a total of 456533 grid points because AutoDock adds one point in each dimension as the central grid point. These parameters covered an area of about 70 percent of the target molecule.

Running Autodock is finally accomplished with help of docking parameter file which defines the map files to be used, the ligand molecule to move, what its center and number of torsions are, where to start the ligand, which docking algorithm to use and how many runs to do. Four different docking algorithms are currently available in AutoDock: SA, the original *Monte Carlo* simulated annealing; GA, a traditional Darwinian genetic algorithm; LS, local search; and GALS, which is a hybrid genetic algorithm with local search. The GALS is also known as a Larmarckian genetic

algorithm, or LGA, because children are allowed to inherit the local search adaptations of their parents. Each search method has its own set of parameters, and these must be set before running the docking experiment itself. These parameters include what kind of random number generator to use, step sizes, *etc.* The most important parameters affect how long each docking will run. In simulated annealing, the number of temperature cycles, the number of accepted moves and the number of rejected moves determine how long a docking will take. In the GA and GALS, the number of energy evaluations and the number of generations affect how long a docking will run. In our docking run we used Larmarckian genetic algorithm with 100 runs.

Evaluation of the docked structures found at the end of each run can be done from the energies of these docked structures and their similarities to each other. The similarity of docked structures is measured by computing the root-mean-square-deviation, rmsd, between the coordinates of the atoms. UsfX-nucleotide docked complexes were evaluated for their energies using the scoring function implemented in the program.

2.12.4 Molecular Dynamics simulations

MD simulations were used to analyze the UsfX-RsfA interaction for the conformational changes that may have occurred as a result of complex formation.Two types of UsfX-RsfA models were taken as the starting material for performing simulations, one in which the Cys73-Cys109 disulfide bond was intact and the other in which the said bond was broken. MD simulations for both the models were performed using the Discover module in the *InsightII* program. The simulations were performed within a sphere of 10 Å radius with the protein complex solvated with water molecules. The pH was set in the calculations to 8.5 using the appropriate option in the program. This reflects the pH of the buffer in which the experiments were performed. All atoms were considered explicitly and their interactions were computed using the CVFF force field. The time step in MD simulation was 1fs and frames were record after each 100fs. All simulations began with 1000 iterations of SD energy minimization. MD simulations were performed at 300K for 200ps with the first 25ps taken as equilibration phase and trajectories of this time period were not used for final calculations. A computer cluster consisting of SGI ORIGIN350 servers and SGI OCTANES were used for the computational analysis.

CHAPTER 3

CLONING, PURIFICATION & CHARACTERIZATION OF MtUsfX(Rv3287c)

3.1 Introduction

Technological advances in sequencing techniques have enabled the sequencing of a large number of whole genomes while concomitant advances in the field of bioinformatics are helping us in understanding and utilizing this plethora of information. Characterization of a new protein, amongst other things, involves identifying its relation with its homologs from other sources. Biochemical and biophysical characterization of a protein requires the isolation of highly purified protein in large amounts. Recombinant DNA technology has helped in the development of a number of cloning and expression systems for obtaining good quality and quantity of proteins for structural studies. In this chapter the analysis of the gene sequence of *usfX* (Rv3287c) is discussed in the first section. The methodologies employed for cloning the *usfX* gene and for purification of the MtUsfX protein are discussed. Various structural properties and the nucleotide binding/hydrolysis properties of the protein properties have been detailed

3.2 Results and discussion

3.2.1 Sequence alignment studies and phylogenetic analysis

The Tuberculist (http://genolist.pasteur.fr/Tuberculist/) (Cole *et al.*, 1998) database annotates Rv3287c as *rsbW* because of its homology to the anti-sigma factor Regulator of SigB W of *Bacillus subtilis*. In later reports it was renamed as UsfX (Upstream of SigF) because of the firm identification of the Rv3287c gene as being exactly upstream of SigF with the initiation codon of MtSigF starting within MtUsfX (DeMaio *et al.*, 1997). Protein sequence homology searches of the MtUsfX using BLASTP against the available protein databases yielded hits with a value worse than the threshold. Position specific iterated BLAST (PSI-BLAST) against the *Protein Data Bank* (PDB) and other data bases also gave worse than threshold value results. PSI-BLAST against the non-redundant protein sequences (nr) database showed homologies with COG 2172 (Anti-sigma regulatory factor, Serine/Threonine Protein Kinase), cd00075 (Histidine kinase-like ATPases), PRK 03660 (anti-sigma F factor) of Conserved Domain Database and similar clusters (Marchler-Bauer A *et al.*, 2007).

The protein database of Clusters of Orthologous Groups (COGs) phylogenetically classifies proteins on the basis of the orthologous relations between them. Orthologs can be defined as the evolutionary counterparts that are related by vertical descent *i.e.*, they have a common ancestor; consequently they retain the same function in the course of evolution. COG2172 contains the cluster of anti-sigma regulatory factors/Serine-threonine protein kinases. Overall there is a very low global sequence homology among the members from diverse backgrounds (Fig 3.1A). However, within mycobacteria the UsfX sequences are rather well conserved with homology of up to 83 % for the respective proteins (Fig 3.1B). This suggests that MtUsfX is an anti-sigma factor with a common ancestral origin with even the sequence characteristics conserved within the mycobacteria family. Cd00075 of the Conserved Domain Database contains histidine kinase-like ATPases whose main features consist of ATP and Mg^{2+} binding sites. MtUsfX can be expected to possess such properties because anti-sigma factors are known to regulate/alternate their binding with sigma factors and anti-anti sigma factors by phosphorylation of their cognate anti-anti sigma factors.

The phylogenetic analysis provides an insight into the evolution of mycobacterial anti-sigma factors sharing a common ancestor with other anti-sigma factors, but having diverged as a separate group (Fig 3.2). Also clustering of MtUsfX in COG2172 points towards its orthologous behavior. Overall the sequence analysis suggests that the mycobacterial anti-sigma factors (taking MtUsfX as a representative because of high sequence homology) while maintaining their primary functions might have evolved novel modes of regulating the role of their interacting partners in transcription regulation.

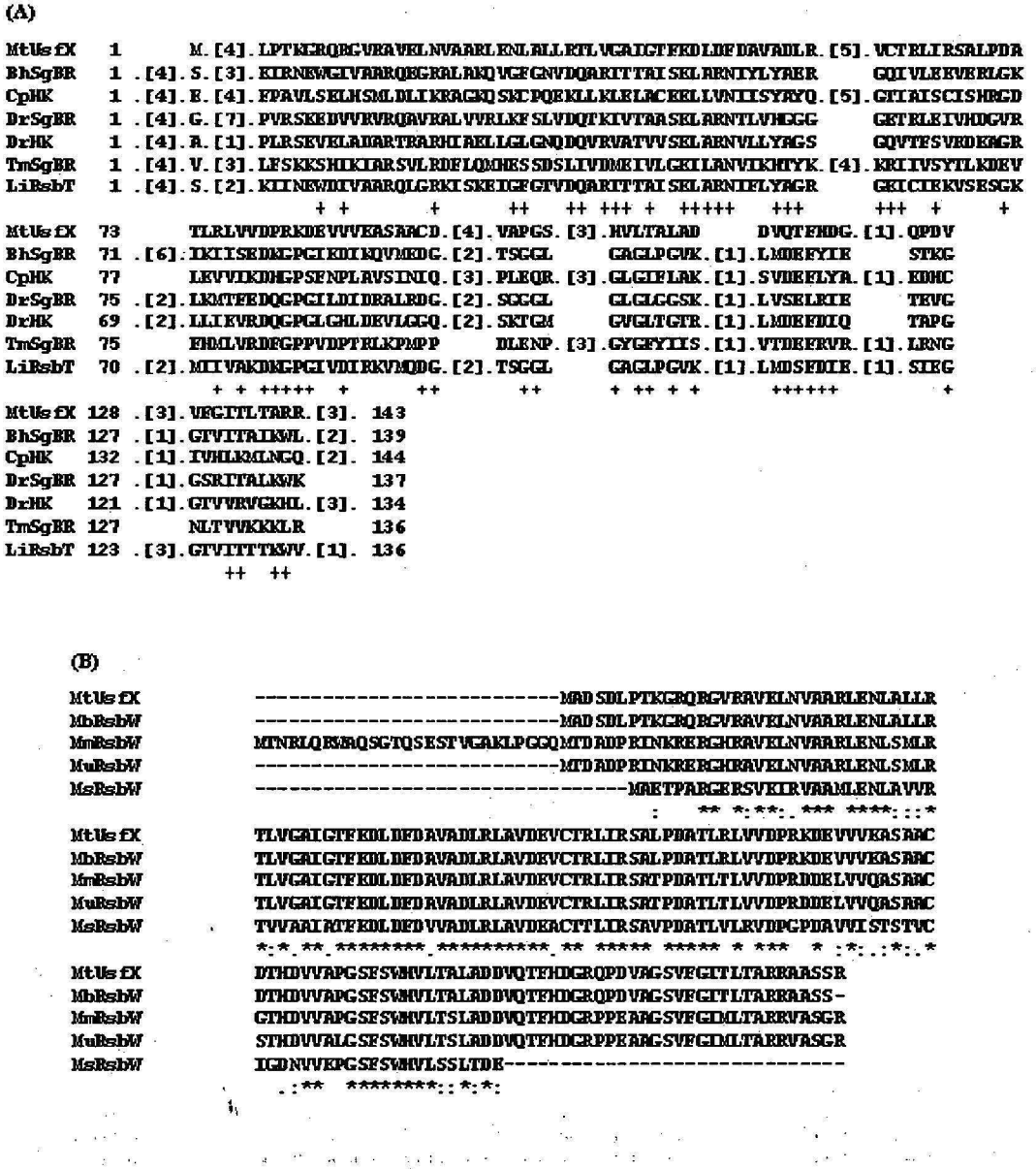


Fig 3.1: Multiple sequence alignment of MtUsfX.
A.With members of COG2172: BhSgBR: *Bacillus halodurans* Regulator of SigB activity, **CpHK:***Chlamydophila pneumoniae* Sigma regulatory factor-histidine kinase, **DrSgBR:***Deinococcus adiodurans* SigmaB regulator RsbT, **DrHK:***Deinococcus radiodurans* Sensor histidine kinase/response regulator, **TmSgBR:** *Thermotoga maritima* SigmaB regulator, **LiRsbT:** *Listeria innocua* SigmaB regulator.**B. With anti-sigma factors from mycobacteria: MbRsbW:** *Mycobacterium bovis,* **MmRsbW:** *Mycobacterium marinum,* **MuRsbW:** *Mycobacterium ulcerans,* **MsRsbW:** *Mycobacterium smegmatis.*

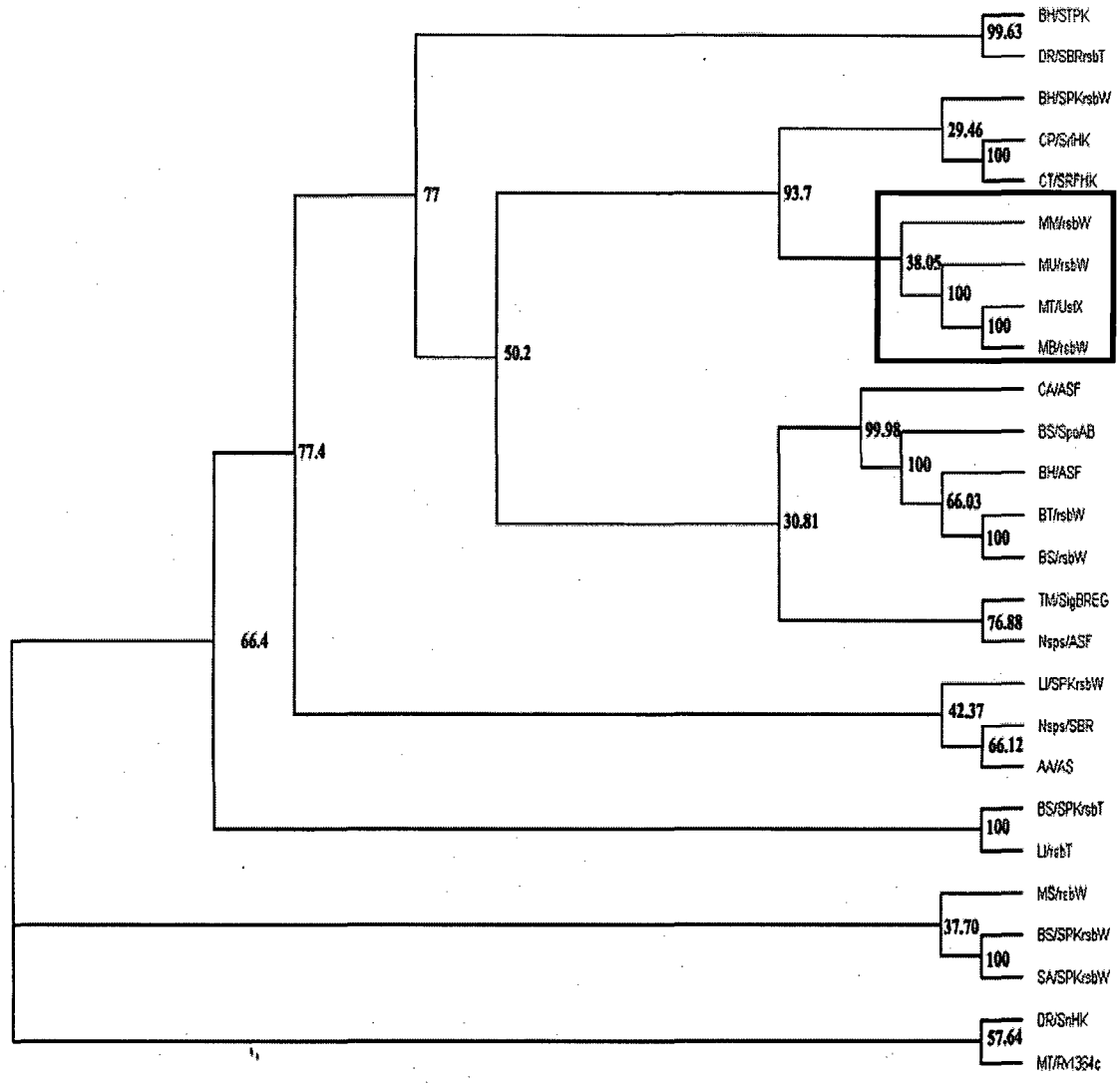


Fig 3.2: Phylogenetic analysis of Sigma-factor regulatory proteins.

The mycobacteria anti-sigma factors cluster as separate group. **BH/STPK** (*Bacillus halodurans*: Switch protein/serine-threonine kinase), **CP/SrfHK** (*Chlamydophila pneumoniae*: Sigma regulatory factor-histidine kinase), **BH/SPKrsbW** (*Bacillus halodurans*: serine-protein kinase RsbW, **BH/ASF** (*Bacillus halodurans*: anti-sigma F factor), **BS/SPKrsbT** (*Bacillus subtilis*: Anti-sigma-B factor rsbT), **BS/SPKrsbW** (*Bacillus subtilis*: Serine-protein kinase rsbW/Anti-sigma-B factor), **BS/SpoAB** (*Bacillus subtilis*: Anti-sigma F factor /Stage II sporulation protein AB), **CA/ASF** (*Clostridium acetobutylicum*: anti-sigma F factor),**CT/SRFHK** (*Chlamydia trachomatis*: sigma regulatory factor-histidine kinase), **DR/SBRrsbT** (*Deinococcus radiodurans*: sigma-B regulator RsbT, **DR/SnHK** (Deinococcus radiodurans: Sensor histidine kinase/response regulator), **MT/Rv1364c** (*Mycobacterium tuberculosis*: Uncharacterized protein Rv1364c/MT1410), **SA/SPKrsbW** (*Staphylococcus aureus*: Serine-protein kinase RsbW),**TM/SigBREG** (Thermotoga maritime: Sigma-B regulator, putative), **Nsps/ASF** (*Nostoc sp*: Anti-sigma B factor), **Nsps/SBR** (*Nostoc sp*: Sigma-B activity negative regulator), **LI/rsbT rsbT** (*Listeria innocua*: Regulator of SigB),**LI/SPKrsbW** (*Listeria innocua*: Serine-protein kinase RsbW), **MT/UsfX** (*Mycobacterium tuberculosis*: UsfX), **BT/rsbW** (*Bacillus stearothermophilus*: Anti-sigma factor F), **BS/rsbW** (*Bacillus subtilis*: Anti-sigma factor), **AA/AS** (*Aquifex aeolicus* : Anti-sigma factor), **MB/rsbW** (*Mycobacterium bovis*: rsbW), **MM/rsbW** (*Mycobacterium marinum*: rsbW), **MU/rsbW** (*Mycobacterium ulcerans*: rsbW), **MS/rsbW** (*Mycobacterium smegmatis*: rsbW)

3.2.2 Cloning, over expression, purification and structural properties

3.2.2.1 Cloning, over expression and purification

The 438bp *usfX* gene was PCR amplified from *M. tuberculosis* genomic DNA and cloned into the T7-based expression plasmid pET21d (Novagen) at NcoI and BamHI to acquire a C-terminal 6x Histidine tag. The clones were confirmed through restriction digestion (Fig 3.3A) and the integrity of clone was confirmed by sequencing.

The expression plasmids harboring the gene for MtUsfX were introduced into *E. coli* host strains like BL21 (DE3) and C41(DE3) that contain T7 RNA polymerase gene under the control of *lac*UV5 promoter. MtUsfX expressed both in BL21 (DE3) and C41(DE3) and the conditions for growth were optimized at 37 °C by adding 0.3mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) to an OD_{600} of ~0.6 and growing for further three hours. OD_{600} of ~0.6 or greater than that was the critical factor for growth as host cells at lower than this OD were intolerant to expression of protein. The protein was soluble in a broad pH as well as salt concentration range but highest percentage about 60-70% soluble protein was observed in 50mM Tris-Cl, pH 8.5, 50mM NaCl. Soluble protein could be obtained only by growing in C41(DE3) and utilizing 2X YT medium. The protein could be identified as a band of ~22kDa on 12%SDS-PAGE. This was approximately 8kDa greater than the calculated mass 15.5kDa±0.8kDa tags. There was no band corresponding either to C41(DE3) alone or C41(DE3) transformed with pET21d without the insert when induced with IPTG.

The C-terminal His-tagged protein was purified from a Ni^{2+}-IDA column (GE Biosciences). The column was equilibrated with buffer containing 50mM Tris-Cl, pH 8.5, 50mM NaCl and 10mM Imidazole (Buffer A). The protein was eluted using a linear gradient of buffer B (Buffer A containing 500mM Imidazole). The protein eluted at 35-55% Imidazole in buffer B. The protein was highly pure with minor trace contaminants as observed on the SDS-PAGE (Fig 3.3B). The final purification step was performed by precipitating the protein with 60% ammonium sulphate. The precipitate was dissolved in minimum volume of gel filtration running buffer (50mM Tris-Cl, pH8.5, 50mM NaCl and 5mM EDTA). Trace contaminants were removed by gel filtration utilizing Superdex-75 size exclusion column mounted in an AKTA-FPLC system (GE Biosciences).

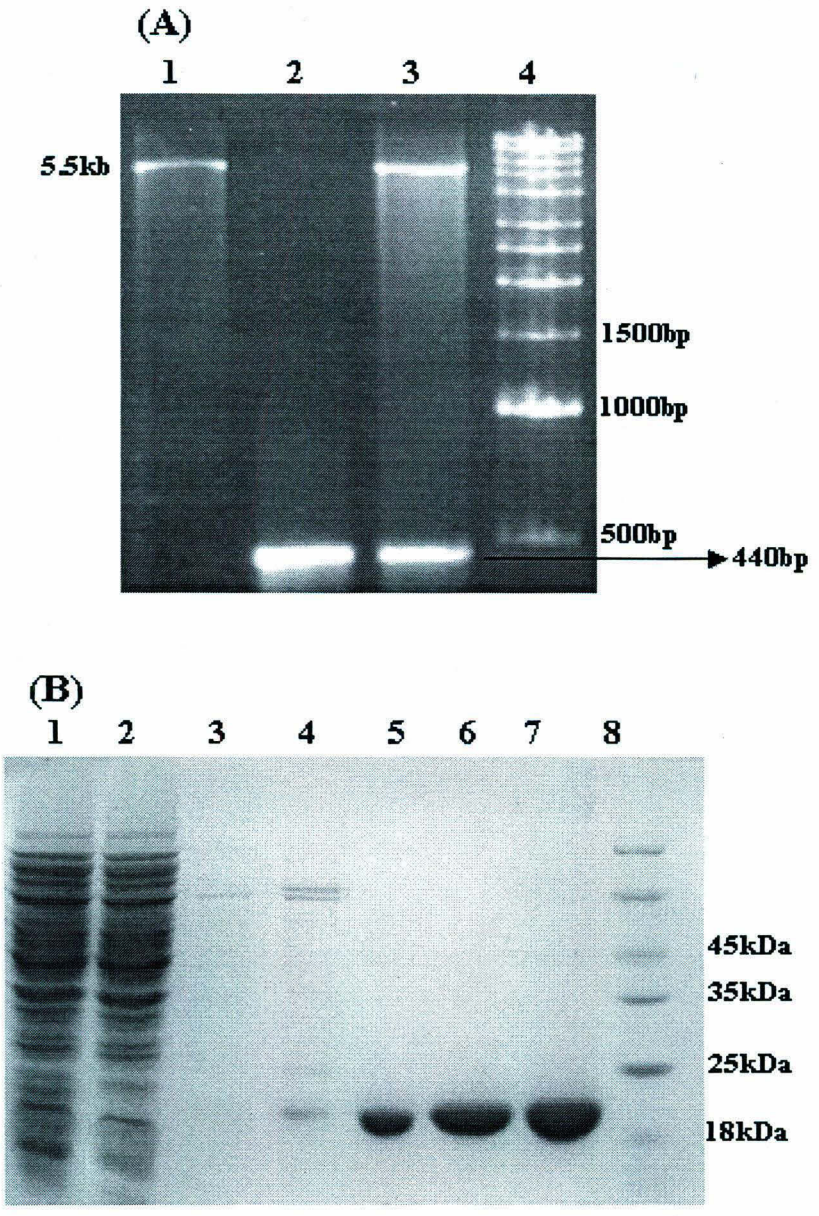


Fig 3.3: Cloning and purification of *M. tuberculosis* UsfX in pET21d.
(A). Cloning: Lane1: NcoI and HindIII digested pET21d, Lane2:440bp *usfX* PCR amplified fragment digested with same enzymes, Lane3: Restriction digestion of *usfX* clone in pET21d, Lane 4:1 Kilobase DNA marker.
(B). Purification: Lane1:Load, Lane2: Flowthrough, Lane3:Wash1, Lane4:Wash2, Lanes5,6 and 7: Purified fractions From Ni^{2+}-IDA column, Lane8:Low Molecular Weight Protein Marker.

3.2.2.2 Molecular weight determination and oligomeric association

The soluble MtUsfX when electrophoresed on SDS-PAGE showed a retardation of about 8 kDa in the mobility giving a molecular mass higher than that of calculated from the amino acid sequence of the protein. The actual mass of the protein was determined by MALDI-TOF analysis and came out to be 15.05±0.5kDa which is in accordance with the expected mass (Fig 3.4A). The retardation of the protein in SDS-PAGE analysis may be because of interactions of MtUsfX with nucleotides as phosphorylated proteins have been found to show retardation in SDS-PAGE run (Sung-Jin *et al.*, 1997).

Size exclusion chromatography was used to probe the oligomeric nature of MtUsfX. Every protein shows a characteristic elution pattern in a gel filtration column depending upon the molecular mass and hydrodynamic radius of the protein. Superdex-75 column was calibrated with proteins of known molecular weight and the estimations were made from K_{av} (partition coefficient) which was plotted as a function of logMW. The protein eluted at a volume of 10.9 ml (Fig 3.4B) in the size exclusion column which corresponds to a molecular weight of about 33 kDa. From the molecular mass of MtUsfX (15.5kDa) we conclude that MtUsfX exists as a dimer in solution. The quaternary association was found to be quite robust in gel-filtration experiments as the dimer was found to be stable even upto 0.7M salt concentration.

Chemical cross-linking was also utilized to study the oligomerization pattern of MtUsfX. Analysis of the chemical cross-linking reactions by bifunctional cross-linking agent glutaraldehyde using SDS-PAGE revealed two bands. The lower band corresponds to the monomer of MtUsfX. The upper band corresponds to a molecular weight of approximately 44kDa which corresponds to dimeric MtUsfX considering that the monomer runs at about 23kDa (Fig 3.4C). The dimeric band is weak in intensity and apparently results from the cross-linking reaction not being complete under conditions of the experiment.

Conclusively, MtUsfX is a dimeric protein in solution with the monomer subunit molecular weight of about 15.5kDa and the subunit interactions stabilized mostly by hydrophobic interactions as deduced from its resistance to high salt levels.

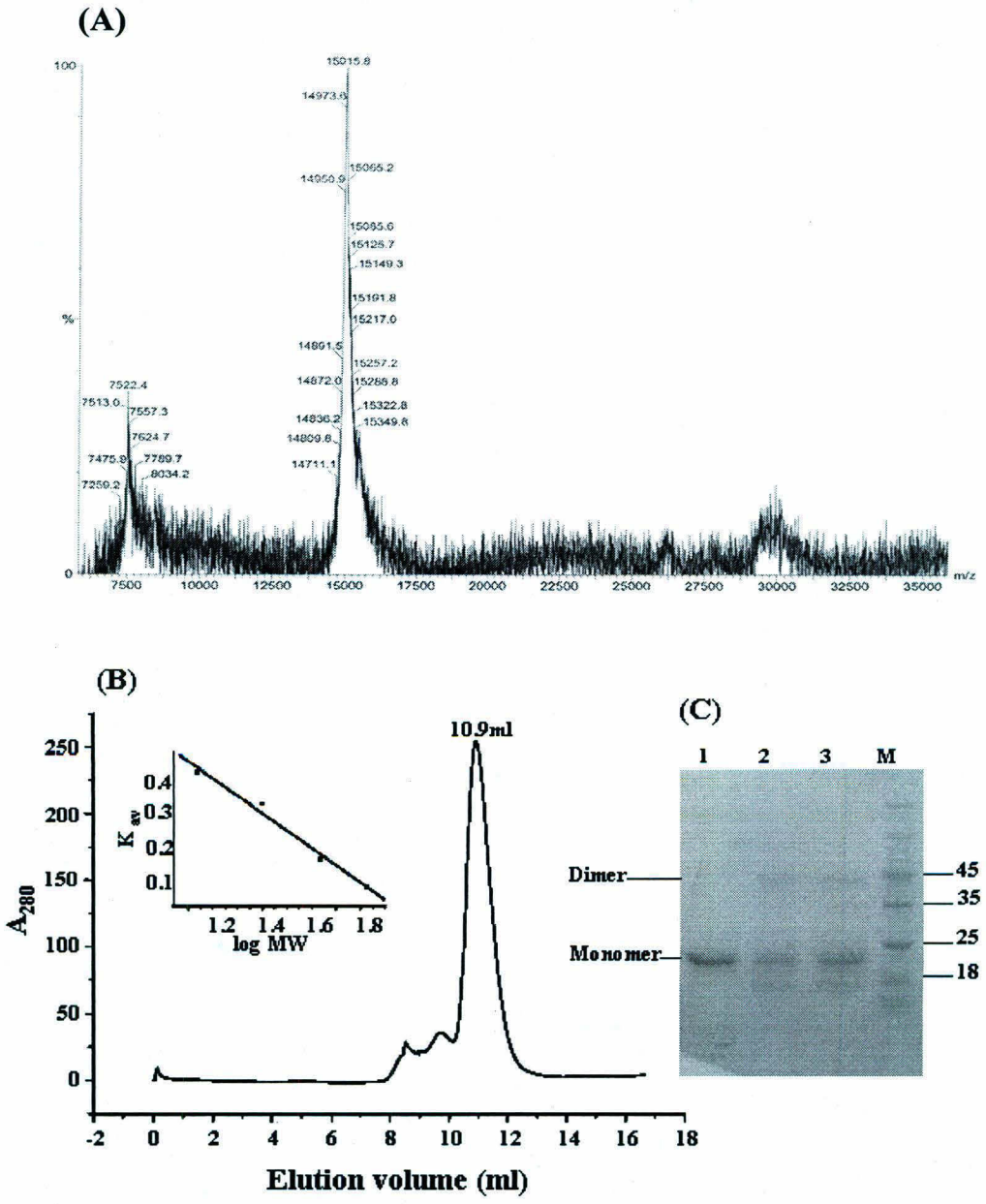


Fig 3.4: Molecular weight determination and oligomeric association of MtUsfX.

(A) MALDI-TOF analysis of purified MtUsfX. The result is near to the expected molecular weight of 15.5kDa.

(B) Size-exclusion chromatogram.The elution volume points towards a molecular weight of ~ 33kDa.

(C) SDS-PAGE analysis of chemical cross-linking reactions. Lane 1: Uncrosslinked protein, Lanes 2 and 3: Cross-linked protein.

3.2.2.3 Intrinsic tryptophan fluorescence

Steady-state tryptophan fluorescence has been extensively used to obtain information on the structural and dynamic properties of proteins as the spectral parameters of fluorescence emission such as position, shape (spectral bandwidth) and intensity are dependent on the electronic and dynamic properties of the chromophore environment. The nature of the microenvironment around tryptophan residues in a protein can be assessed from their fluorescence emission spectra (Bujalowski *et al.*, 1994). The spectrum of MtUsfX is characterized by a fluorescence maximum at ~349nm (Fig.3.5A). The observed λmax at 349nm in the case of MtUsfX suggests that Trp106 (the only Trp residue in the protein) exists in a partly shielded environment away from the solvent. The shielded nature of the residue was further verified by guanidium hydrochloride denaturation experiments involving MtUsfX. At ~2.5M concentration of the denaturant, MtUsfX gives an emission maxima at 353nm which points towards solvent exposure of the tryptophan (Fig.3.5B) on partial denaturation of the protein.

Measurement of the accessibility towards fluorescence quenchers also gives an idea about the location of the tryptophan molecules in the protein. Solvent accessibility was determined by measuring the quenching of fluorescence produced by the solvent phase quencher acrylamide (Section 2.8.4; Eqn. 3). The Stern-Volmer plots (Stern *et al.*, 1919) for fluorescence quenching of MtUsfX and free tryptophan are shown in Fig.3.5C. The obtained quenching constants are 17.3 M^{-1} and 11.3 M^{-1} for tryptophan and MtUsfX, respectively. Although K_{SV} for MtUsfX quenching with acrylamide is lower compared to that observed for free tryptophan, it differs by a factor of only ~ 1.5. This supports that Trp106 in MtUsfX is partially shielded from the environment.

Additional information about the nature of the location and microenvironment of tryptophan residue in a protein can be obtained by studying the quenching of protein fluorescence by ionic quenchers. Ionic quenchers provide information about the polarity of the environment surrounding the tryptophans in proteins. We chose potassium iodide as the ionic quencher to probe the polarity of the environment utilizing the fact that both acrylamide and I^- have the same high quenching efficiency of indol fluorescence and I^- is smaller in size and has a higher diffusion coefficient compared to acrylamide (Lakowicz, 1983). The Stern-Volmer constants together with relative efficiencies are listed in Table 3.1. There is a marked decrease (~3.5 times) in

the accessibility towards KI (K_{SV}=3.2) as compared to acrylamide (K_{SV}=11.3). Also the relative solute quenching efficiency of I^- ($\gamma_{I^-}/\gamma_{acrylamide}$ = 0.45) is 2.2 times less than that of acrylamide. This suggests that there are negatively charged groups of amino acids in close proximity to tryptophan residues in MtUsfX.

Thus the only tryptophan of MtUsfX(Trp106) is in a partially shielded environment relatively less accessible to the solvent as free tryptophan would have been and bulk of the atmosphere around the tryptophan is composed of negatively charged molecules.

Table 3.1 Effective quenching parameters of free tryptophan and MtUsfX protein tryptophan fluorescence by acrylamide and I^-.

	Acrylamide		I^-	
Parameter	**Trp**	**UsfX**	**Trp**	**UsfX**
$K_{SV\ (eff)}$ (M^{-1})	17.3	11.3	10.94	3.2
$(\gamma_{quencher}/\gamma_{acrylamide})_{eff}$	1		0.45	

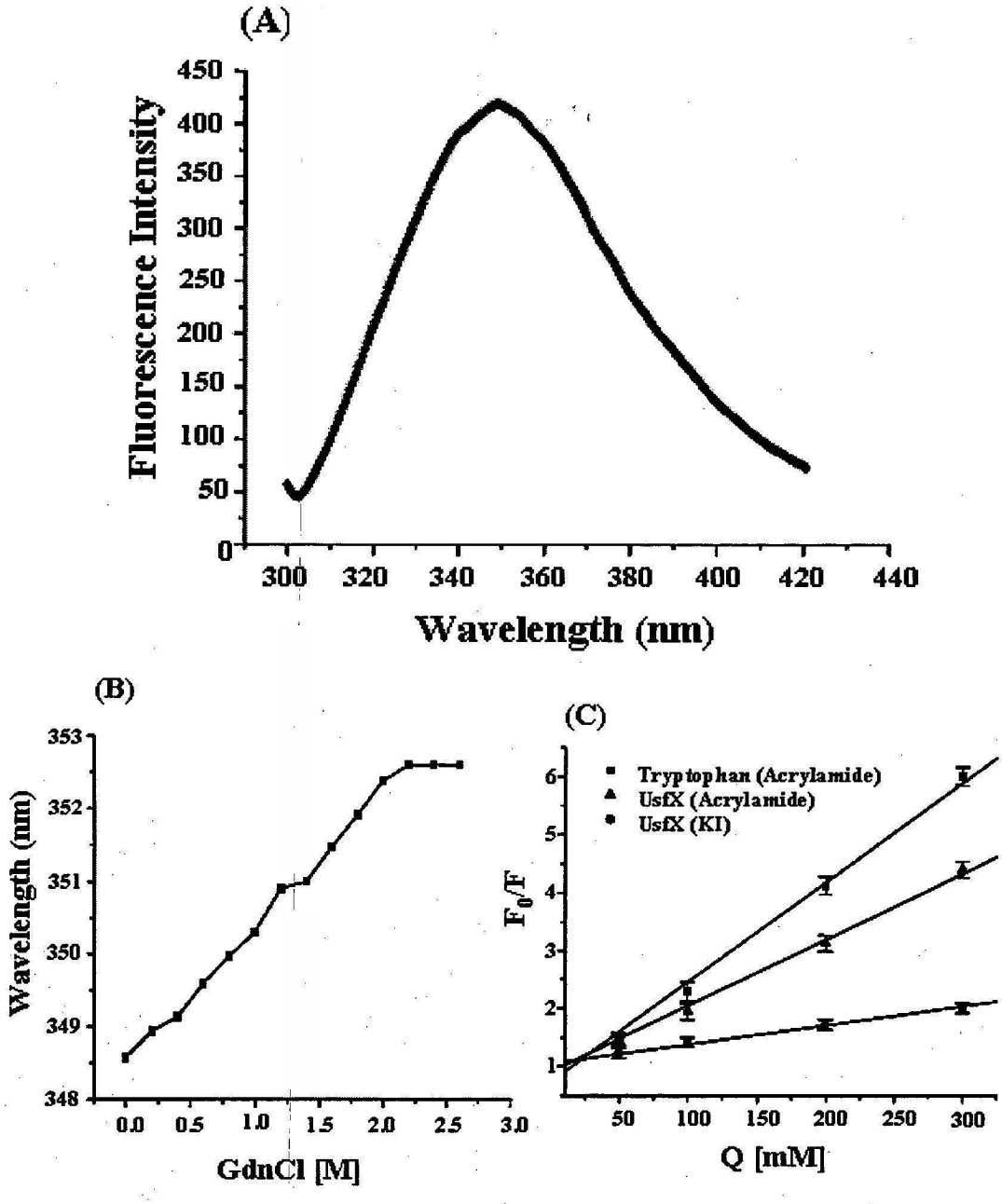


Fig 3.5: Intrinsic tryptophan fluorescence of MtUsfX.

(A) Intrinsic fluorescence emission spectra otained for MtUsfX showing emission maxima at 349nm.

(B) Guanidium Hydrochloride denaturation of MtUsfX showing change in emission maxima with increasing concentration of Guanidium Hydrochloride.

(C) Stern-Volmer plots for quenching of the tryptophan fluorescence of free trytophan with acrylamide, MtUsfX with acrylamide and MtUsfX with KI.

3.2.2.4 Circular Dichroism Measurements/Secondary structure

Circular Dichroism spectroscopy is used to identify three common secondary structure motifs (alpha-helix, beta sheet, and random coil) by virtue of their distinctive CD spectra in the far-ultraviolet region (170-260 nm). The secondary structure content of proteins can be estimated from their CD spectra using computer algorithms. The analysis of the far UV CD spectra (Fig. 3.6A) of MtUsfX for the estimation of secondary structure content was performed by using the K2D (Merelo *et al.*, 1994) software. Secondary structure predictions were also obtained using the JPRED2 package (Cuff *et al.*, 1998) and 3DPSSM Fold prediction server (Kelley *et al.*, 2000). MtUsfX is predicted to consist of three helical regions and five sheet regions linked by short loops with an overall helical content of 33% and sheet content of 23%. The calculated results from CD spectra are more or less in conformation with the predicted results (45% helix and 25% sheet).

Thermal stability of MtUsfX was assessed by monitoring the changes in the spectrum at a fixed wavelength of 222nm with the increasing temperature. The fitting of the data into the sigmoidal model (Solid line Fig.3.6 (B)) returns the midpoint of the transition as 70.5±1 °C. An important parameter of thermal stability is the onset of unfolding which can be defined as the temperature at which 5% of the protein is unfolded. MtUsfX shows onset at about 55°C. From the CD spestroscopy data it can be concluded that MtUsfX is a thermostable protein, a characteristic which may be related to its functioning in varied environments *M. tuberculosis* faces during its life cycle.

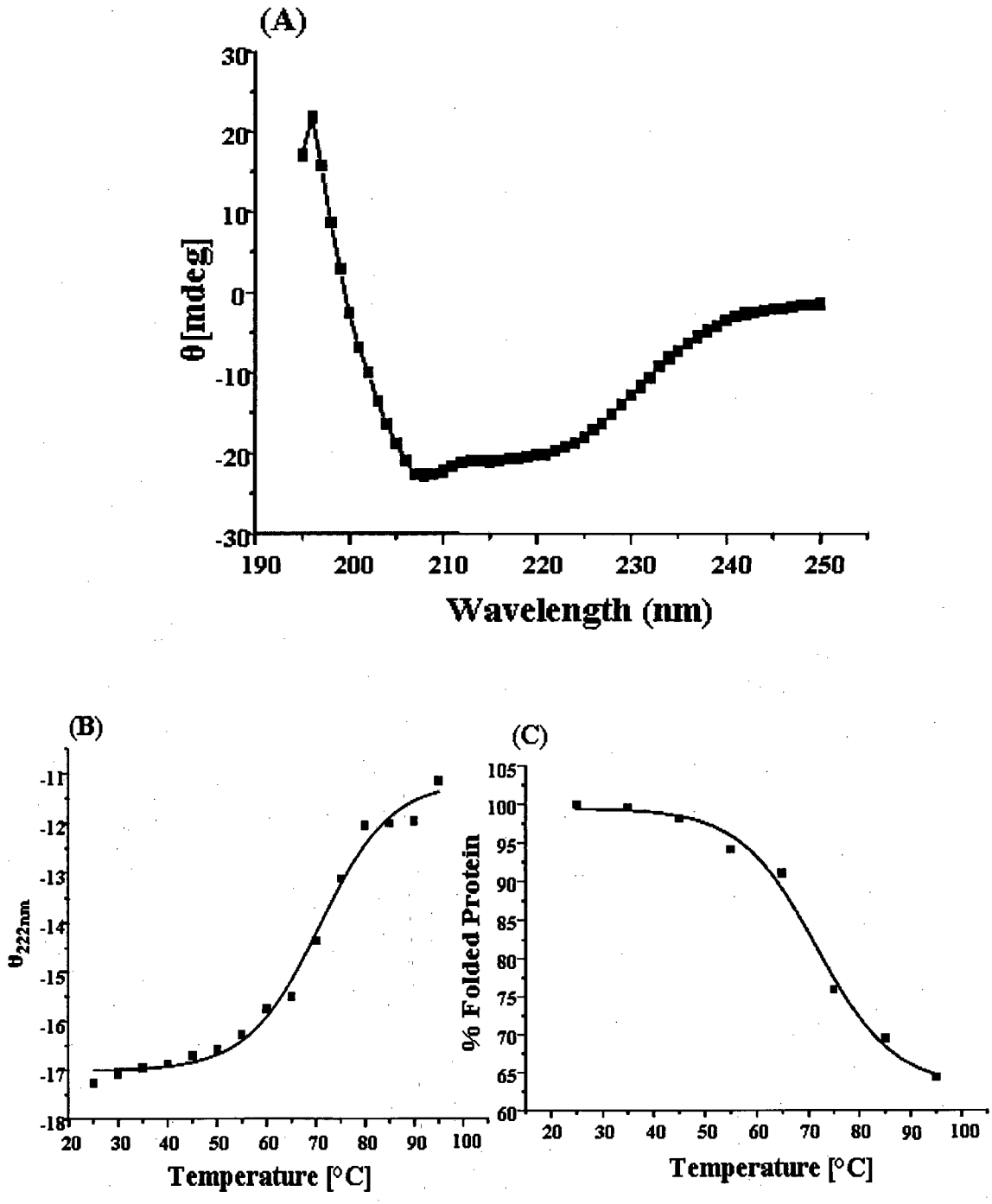


Fig 3.6: Secondary structural properties of MtUsfX.
(A) Far-UV CD spectra of MtUsfX.
(B) Thermal transition of MtUsfX. The signal change of CD at 222nm was measured as a function of increasing temperature.
(C) The percentage folded protein measured as a function of temperature, giving an onset of unfolding in the range of 55-60°C.

3.2.2.5 Crystallization of MtUsfX

Crystallization trials of MtUsfX were performed as described in Section 2.11.3. Much success was not met in the crystallization of MtUsfX with very tiny crystals (Fig 3.7) obtained in 4M NaCl, 5% Isopropanol – pH 7.5, a condition described in the "Mixed precipitant screens" (Mazeed *et al.*, 2003).The further optimization by screening the constituent concentrations over a wide range, varying the overall pH and temperatures also did not bring any change in the quality of the crystals. Utilization of the nucleotide co-factors and combination of nucleotide and Mg^{2+} as additives in the crystallization solutions also was ineffective in bringing about a change in the size of the crystals. Additional inorganic and organic additives were also used in an attempt to improve crystalline conditions but the results were unaltered as above.

The primary step in the crystallization is the nucleation i.e., the formation of centers for the growth of crystals. A paradoxical situation that is often encountered during crystallization is that the optimal conditions for nucleation of the crystals are at times not the ideal ones to support their subsequent growth. This is because spontaneous nucleation is more likely to occur when the levels of super-saturation are high, whereas slow, ordered growth of large crystals is favored by lower levels. A situation where the crystals do not grow beyond a limit needs to be tried for uncoupling the nucleation from growth to satisfy the distinctly different requirements of the two events. Seeding is one such technique that can be used for the separation of nucleation and growth. We utilized the microseeding technique as described in section 2.11.4. Again no improvements in the quality of the crystals could be achieved (Thaller *et al.*, 1981, 1985, Terese *et al.*, 2003). Formation of the crystals is accompanied by decrease of entropy including the conformational entropy of the surface residues which become ordered in the crystal lattices during formation of crystals. It has been hypothesized that surface amino acids like lysines and glutamates with a high conformational entropy are thermodynamically unfavorable and hence impede crystallization (Kenton *et al.*,2001, Agnieszka *et al.*, 2002) There are two lysine residues in MtUsfX: K9 and K83.The mutants for these lysines (ΔK9A and ΔK83A) were generated by replacing them with alanine. Simultaneously a double mutant was also created fot these lysines but that also did not have any effect on the crystallization of MtUsfX.

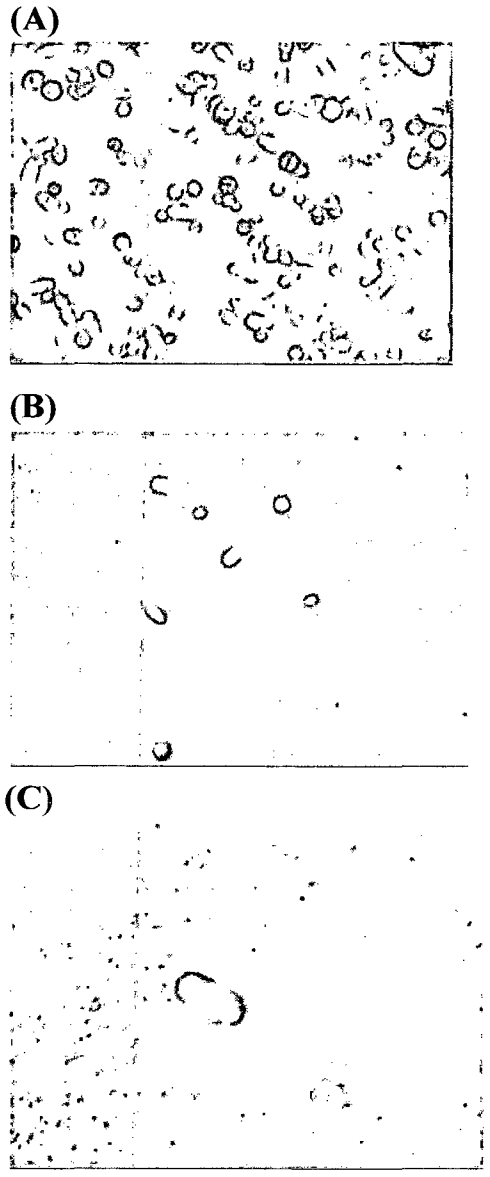


Fig 3.7:Crystals of *M. tuberculosis* UsfX.
The pictures (B) and (C) are higher resolution pictures of (A).

3.2.3 Nucleotide binding properties of MtUsfX

The anti-sigma factors are known to regulate the activity of their cognate sigma factors by sensing and responding to the levels of various signaling molecules. One important stress signal is a drop in the ATP levels in the cell (Energy stress) which occurs in both carbon source limitation and entry into the stress phase. The ATP functions either by influencing the interaction of sigma factor with anti-sigma factor or the anti-sigma and anti-anti sigma factor interaction (Alper *et al.,* 1994, 1996). MtUsfX has been designated as a stress phase anti-sigma factor (De Maio *et al.*, 1997) and exists in a genomic cluster with its sigma factor MtSigF. The functional similarity of MtUsfX with known anti-sigma factors like RsbW of *B. subtilis* and SpoIIAB of *B. stearothermophilus* was used as a premise for identifying the nucleotide binding properties of MtUsfX. The location of the natural chromophores of the protein (tryptophan residues) in close proximity to the nucleotide-binding site provides an opportunity to directly study the presence of nucleotide binding site by monitoring the changes in the intrinsic tryptophan fluorescence of the protein. Both purine and pyrimidine nucleotides have been shown to be effective quenchers of endogenous protein fluorescence (Gordon *et al.*,1970).

3.2.3.1 Fluorescence based nucleotide binding assay

Using fluorescence spectroscopy strong nucleotide binding properties were identified in MtUsfX. The binding of all the purine and pyrimidine nucleotide tri-phosphates (NTPs) was found to result in a significant decrease in the fluorescence emission intensity of MtUsfX. Titrations with nucleotide di-phosphates also produced a significant change in fluorescence intensities. The titration curves for the interaction of MtUsfX with ATP, GTP and ADP are shown in Fig 3.8. The decrease in the emission intensity was a consequence of nucleotide binding and not a result of ATP hydrolysis reaction of MtUsfX, kinetics of fluorescence quenching were studied. The binding buffer was supplemented with 10mM $MgCl_2$ to give the conditions required for hydrolysis reaction (Fig3.8D). After a fixed time interval (100sec) the effect of addition of NTP and NDP was monitored in time drive based experiments. Immediate and approximately equivalent quenching of fluorescence was observed on addition of both tri- and di- phosphate ligands with no further decrease in case of tri-phosphate reactions.

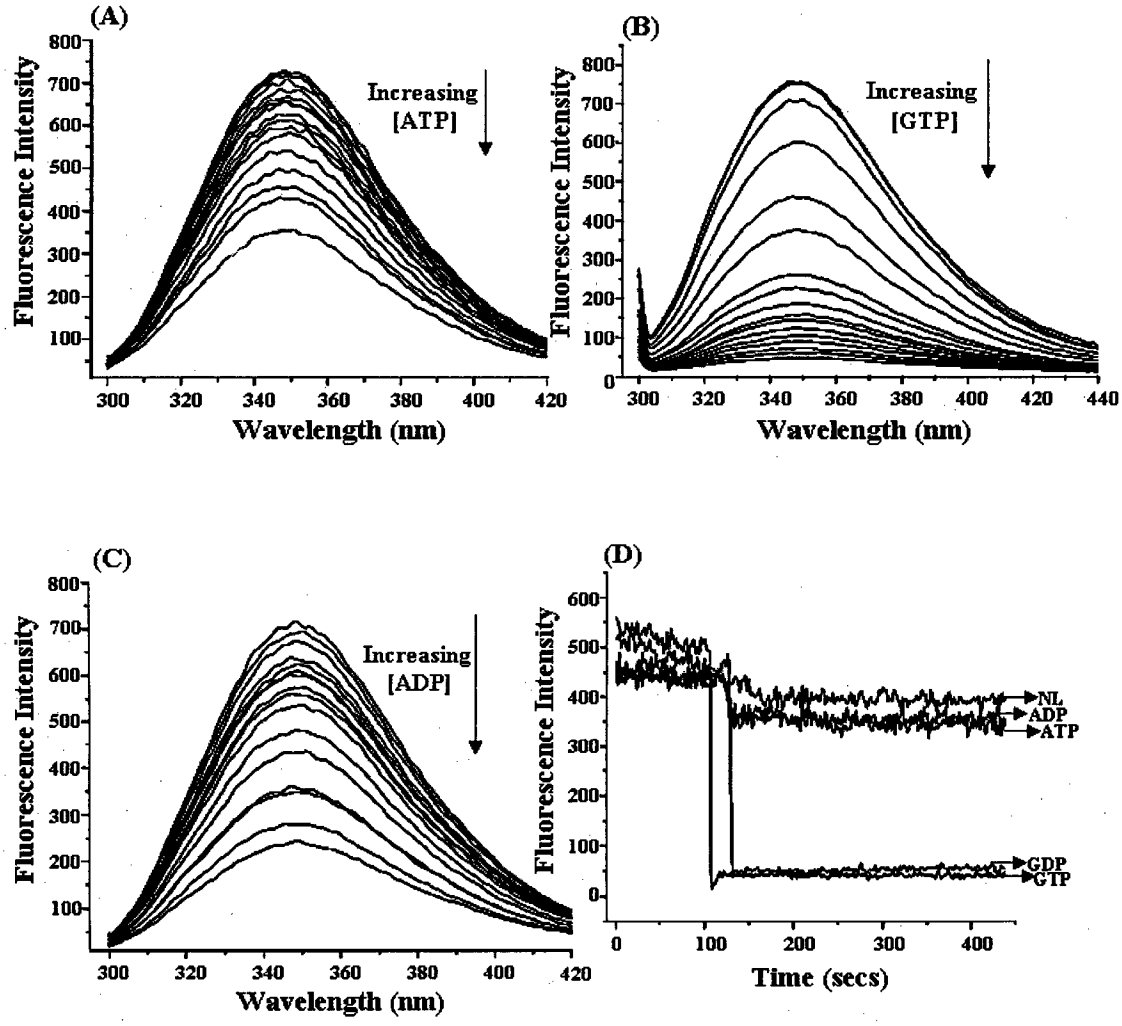


Fig 3.8: Steady state and kinetics of quenching of trytophan fluorescence of MtUsfX due to nucleotide binding .(A), (B) and **(C).** Effect of ATP, ADP and GTP. The spectra were scanned from 300nm to 440nm at an excitation wavelength of 290nm. **(D)** The kinetics based analysis clearly shows that quenching of fluorescence is a consequence of nucleotide binding and not a result of the ATP hydrolysis reaction.

Saturation isotherms or binding isotherms provide a means of quantifying the binding properties like molecular affinity and selectivity of ligands for their interacting partners. A saturation isotherm can be defined as a measure of the concentration dependent recognition behavior of the system. The saturation isotherms for the binding of NTPs and ADP to MtUsfX were generated by plotting percent change in fluorescence intensity versus nucleotide concentration and are shown in the Fig 3.9.The saturation isotherms allow for the comparison of binding properties and the most important parameter that can be calculated is K_d : Dissociation constant. Dissociation constant describes the affinity between the ligand and the protein molecule i.e. how tightly a ligand binds to a particular protein. Dissociation constant corresponds to the ligand concentration at which the binding sites on the protein molecule are half occupied i.e. the concentration of ligand at which the concentration of liganded protein equals concentration of unliganded protein. The dissociation constants (K_d) for all the nucleotides and ADP were generated from a fit of Equation 10 to the saturation isotherm as outlined in Section 2.9. Approximately 98% of intrinsic tryptophan fluorescence is available for quenching as found by interaction with 1mMGTP where a complete saturation of fluorescence quenching is achieved (Bougie *et al.*, 2004). Binding for all the four nucleotides and ADP was identified with a very high K_d value for adenosine nucleotides observed in comparison to the other three naturally occurring nucleotides (Table 3.2). The binding of nucleotides other than ATP is unlike the explained system of *Bacillus subtilis* which acts via ATP/ADP alone.

Another property of ligand binding that can be calculated by treatment to saturation isotherm is the Hill coefficient, n_H (Section 2.9 ; Hill equation : Equation 11). Hill coefficient measures the degree of cooperativity between subunits that bind in multi-subunit proteins. The most common usage of n_H is an estimate of the minimal number of binding sites available for the ligand binding interaction. A binding stoichiometry of two per protein molecule was calculated for the binding of nucleotides to MtUsfX (Fig 3.10). A stoichiometry of two can be explained in terms of the dimeric nature of MtUsfX. The Hill coefficients for the binding of all nucleotides (Table 3.3) are less than two which indicates the absence of cooperativity for the MtUsfX-nucleotide interaction. The lack of cooperativity can be explained in terms of the function of the anti-sigma factors molecule to sense the ATP levels inside the cell. The anti-sigma factor RsbW of *B. subtlilis*, a homolog of MtUsfX is bound by two anti-anti sigma

factor molecules, and once the levels of ATP inside the cell increase, the two anti-anti sigma factor molecules are simultaneously phosphorylated thus releasing itself for binding to SigB and stopping the stress phase transcription. As the phosphorylation reactions have to occur simultaneously, the binding sites for nucleotides must be behaving independently of one another.

Conclusively, MtUsfX like other anti-sigma factors has a nucleotide binding pocket with affinities for all the naturally occurring nucleotides which is unlike the other known systems like that of *B. subtilis*. The absence of cooperativity in ligand binding in the dimeric MtUsfX seems a plausible property for maintaining the functional integrity as an anti-sigma factor. The unusual properties in terms of affinity for other nucleotides in addition to ATP and ADP can be attributed to the different survival conditions in which MtUsfX is supposed to regulate the function of MtSigF.

Table 3.2: Binding constants and stoichiometry values of nucleotide binding.

K_d values, calculated ΔFmax values ($\Delta F_{max(calc)}$) and Hill coefficients (n_H) for the interaction of MtUsfX with various nucleotide ligands. The correlation coefficients (R^2) are for the fits of Equation 10 to the data in Figure 3.9.

Ligand	K_d (μM)	$\Delta F_{max(calc)}$	n_H	R^2
ATP	1300±50	210±5	1.8±0.1	0.99
GTP	200±20	430±5	1.9±0.2	0.97
TTP	200±20	470±5	1.8±0.2	0.97
CTP	180±20	420±5	1.9±0.2	0.98
ADP	1900±50	205±5	1.6±0.1	0.97

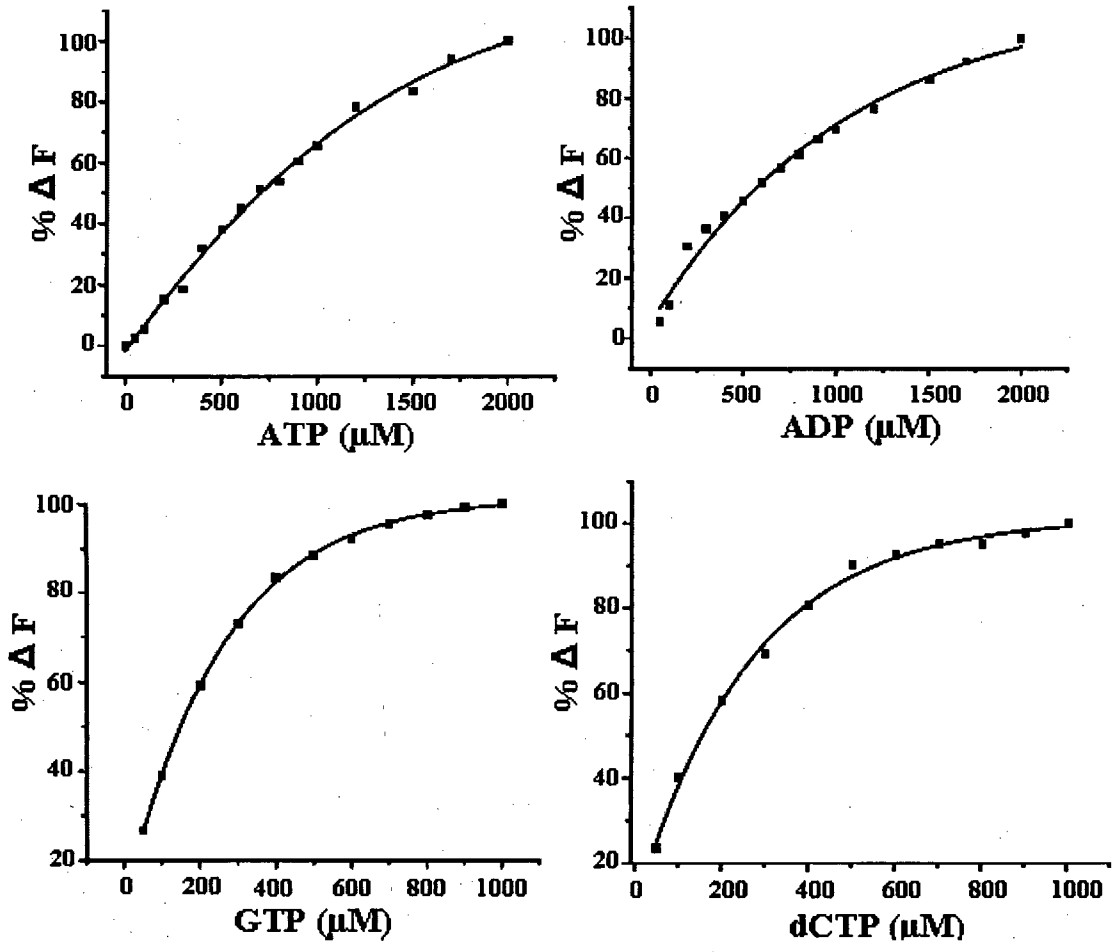


Fig 3.9: Saturation isotherms for nucleotide binding to MtUsfX.

The plots were generated by plotting the percent change in intrinsic tryptophan fluorescence intensity at 349 nm as a function of added nucleotides.

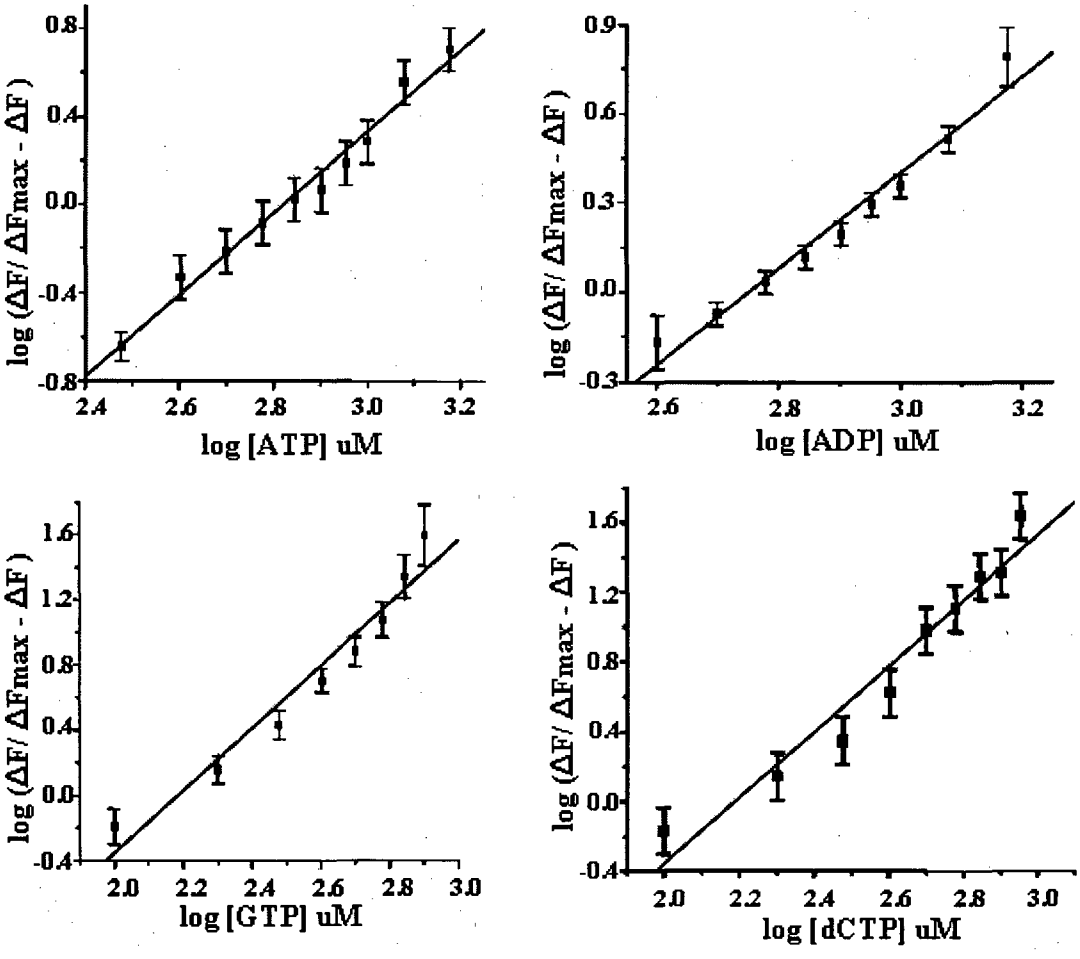


Fig 3.10: Hill plots for ATP, ADP, GTP and dCTP binding to MtUsfX.

3.2.3.2 Differential binding of ATP

Adenosine nucleotides as compared to other nucleotides behaved differentially in the binding site as evident from the ΔF_{max} and K_d values. This could not be attributed to the purine ring of the ATP because like ATP, the GTP molecule also has a purine ring but the binding values of GTP also were like that of other nucleotides. Thus it could be concluded that the binding modes adapted by ATP are different as compared to the other nucleotides. We analyzed the energy transfer rates for the interaction of MtUsfX with ATP and GTP. Complete quenching of fluorescence energy was seen with 1mM GTP while only 45% energy was quenched by 1mM ATP. At 500 μM of nucleotide concentration only 17% fluorescence intensity is quenched by ATP, while as 87% is quenched in case of GTP. The large reduction in the fluorescence in the case of GTP suggests that the nucleotide is in close juxtaposition to the Trp106 in the protein chain. We subsequently calculated the Stern-Volmer constants (K_{SV}) to probe the change in the accessibility of the Trp residue in the presence of the different ligands compared to free MtUsfX. The Stern-Volmer plots for fluorescence quenching of free MtUsfX and in various liganded forms are shown in Fig 3.11. The K_{SV} for ATP-*UsfX* complex at 500μM concentration is 2.3 fold as compared to the GTP-*UsfX* complex at the same concentration (Table 3.4). Also the values for ADP-*UsfX* are almost similar to that for ATP-*UsfX*. This data indicates that the tryptophan in GTP-*UsfX* complex is not readily accessible to the solvent as compared to that in case of ATP-*UsfX* complex at the same ATP concentration. Thus it can be concluded that the nucleotides other than ATP adapt a differential binding mode as concluded from their low K_d values and high energy transfer rates and reduced solvent accessibilities around Trp106. It can be concluded that the binding of nucleotides in the active site is through their ring moieties but the significance of this differential binding of NTPs and differential response that may be created in the protein can have an *in vivo* role which needs to be probed further.

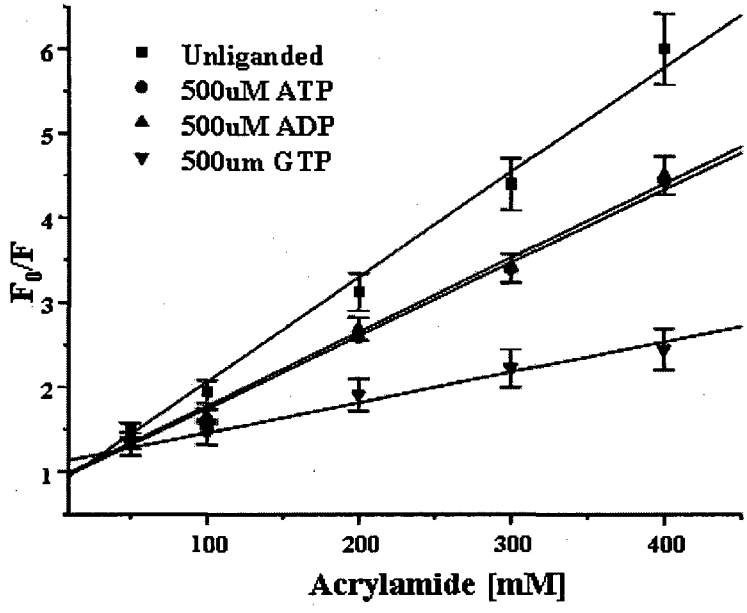


Fig 3.11: Stern-Volmer plots for acrylamide quenching of the tryptophan fluorescence of MtUsfX in liganded and unliganded forms. Data shown are for 0.5 μM unliganded protein and 0.5 μM protein pre-incubated with 500 μM of nucleotide indicated.

Table 3.3: Stern-Volmer constants

K_{SV} for acrylamide quenching of tryptophan fluorescence of MtUsfX in the unliganded and liganded forms. The decrease in fluorescence at 500μM ligand concentration is indicated by $(\Delta F/F_0)_{500\mu M}$ while the maximal decrease in fluorescence is $(\Delta F/F_0)_{max}$.

Ligand	$K_{SV(500\mu M)}$ (M^{-1})	$(\Delta F/F_0)_{500\mu M}$	$(\Delta F/F_0)_{max}$
NIL	9.53	-	-
ATP	8.6	0.17	0.43
ADP	8.9	0.19	0.44
GTP	3.58	0.87	0.98

3.2.3.3 Role of ionic interactions in nucleotide binding

Monovalent and divalent ions contribute to nucleotide binding usually by helping in stabilizing the binding of the phosphate tail of the nucleotide. The dependence of nucleotide binding of MtUsfX on ionic strength was examined for the case of ATP. Fig 3.12A shows the effects of increasing ionic strength, on the percentage change in relative fluorescence (ΔF) for ATP binding to the free enzyme. There is a gradual increase in %ΔF as *I* is increased from 50 mM, the ionic strength of the standard buffer, to 500 mM. There is an increase of about 15% ΔF as a result of increasing salt concentration. No change in fluorescence was observed when the protein was titrated with KCl alone. So KCl can be expected to help in the proper juxtaposition of the nucleotide in the binding site of MtUsfX.

There was a profound effect of $MgCl_2$ on the fluorescence intensity of MtUsfX when the concentration of $MgCl_2$ was increased from 0-100mM. On addition of increasing aliquots of $MgCl_2$ there is quenching of fluorescence intensity indicating a specific binding site for the ion. On increasing the $MgCl_2$ concentration from 0 to 20 mM into a 1mM ATP and protein solution, a synergistic effect of ATP and $MgCl_2$ on the fluorescence intensity is observed (Fig 3.12B). An increase of about 40% ΔF at 10mM $MgCl_2$ is observed which points towards enhanced binding of ATP in presence of $MgCl_2$. $MgCl_2$ at this concentration produced a fluorescence change of ~30%, while ATP alone produced about 50% change. The change in fluorescence can be explained as the enhancement of binding and not the individual effects as the maximum quenching is achieved at 10mM $MgCl_2$ only and further increase after 20mM in $MgCl_2$ concentration does not produce further change in fluorescence intensity. It can be concluded that 10mM $MgCl_2$ is the optimum concentration for proper binding of the nucleotide. The enhanced quenching in the presence of the magnesium ion and nucleotide indicates that the presence of the ion apparently enhances the binding of the nucleotide probably by stabilizing the active site conformation.

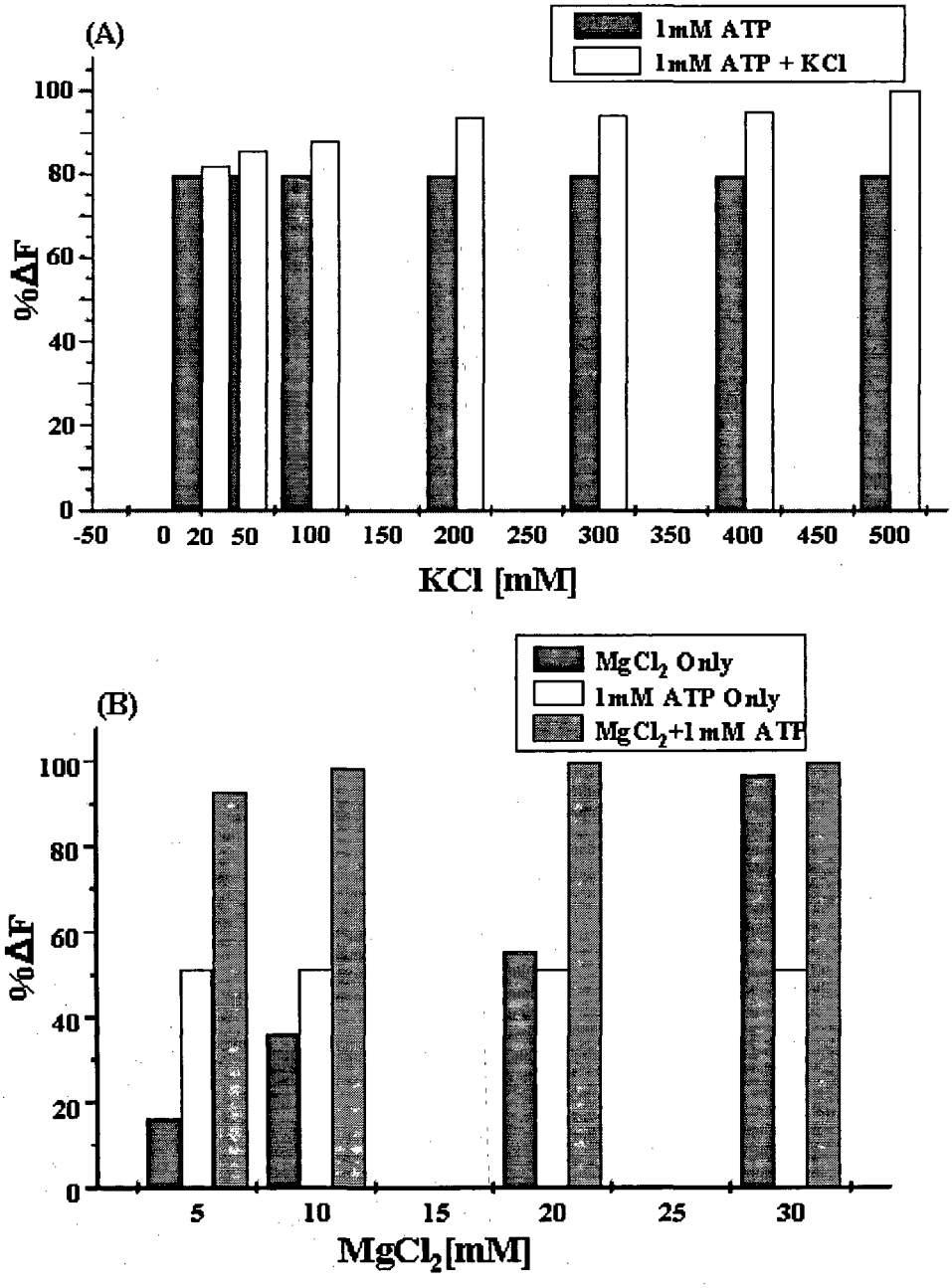


Fig 3.12: Effect of mono- and divalent ions on nucleotide binding.
(A) Graph showing the enhanced quenching of fluorescence a consequence of stronger binding of ATP in presence of increasing KCl concentration. **(B)** The effect of increasing $MgCl_2$ on the binding of ATP to MtUsfX was investigated. Both $MgCl_2$ and ATP contribute to fluorescence quenching individually but produce a synergistic effect on fluorescence quenching pointing towards increased binding of ATP in presence of $MgCl_2$.

3.2.4 *In silico* analysis of nucleotide binding

In the absence of good quality crystals of MtUsfX we used the *in silico* modeling and docking approaches to computationally evaluate the binding of nucleotides. The fold-prediction analysis of the MtUsfX carried out using the 3DPSSM server picked out 4 proteins; the first hit was with the *B. stearothermophilus* SpoIIAB (PDB: 1L0O) which showed about 14% homology. The nucleotide binding sites of Hsp90 (PDB: 1A4H), nucleotide binding domain of human integrin CR3 (PDB: 1ID0) and alpha ketoacid dehydrogenase kinase from murine source (PDB: 1GJV) were the other hits and exhibit around 12% homologies. The model generated for MtUsfX was structurally compared with other ATP binding anti-sigma factors. Fold prediction using the 3D-PSSM server gave the closest match to PDB 1L0O. A close-up of the nucleotide binding site after superposition of the MtUsfX model and PDB 1L0O shows that Trp106 of MtUsfX is located in the NTP binding site. Thus the modeling studies also support the earlier fluorescence experiments which suggested that Trp106 exists in the vicinity of the NTP binding site (Fig 3.13A). The Genetic algorithm implemented in the program was used in the calculations and the docked complexes were ranked against the scoring functions. Choosing Trp106 which is the only tryptophan in MtUsfX as the docking centre because of our results from nucleotide binding assay we could find the nucleotide binding site in the vicinity of tryptophan, indeed tryptophan forming a part of the binding site. All the four nucleotides viz., ATP, GTP, CTP and TTP docked at the same site. The modeling experiments suggest that the nucleotides interact primarily through their ring moieties with the phosphates facing away from the Trp106 (Fig 3.13 B& C). The other residues apparently involved in the interaction include Ser105, Arg63 and Arg66.The calculations predict that GTP binds with a stronger affinity to the protein compared to ATP and also that it's binding mode exhibits distinct differences. An analysis of the accessible surface area of the Trp106 shows that in the *apo* protein it is 144 $Å^2$ while it reduces on the binding of ATP to ~101 $Å^2$. In the case of GTP it is further reduced to 53 $Å^2$. This co-relates rather well to the Stern-Volmer constants obtained for the *apo*, ATP and GTP - bound proteins respectively where the Trp106 was found to be more accessible in the case of the ATP bound complex compared to the GTP bound form. Thus the differential binding of ATP and GTP to the protein apparently arises from altered

binding modes and consequently resulting in altered properties of active site of MtUsfX.

Another interesting result that came out of the modeling and sequence analysis was the identification of an Mg^{2+} binding site close to the nucleotide binding site (Fig 3.13D). As noted in earlier section it was observed that the presence of Mg^{2+} is important for nucleotide binding and activity. The observed quenching of fluorescence on addition of Mg^{2+} and a very low accessibility for the ionic quencher KI suggests that the divalent ion should bind close to the active site in the vicinity of Trp106. A conserved metal ion binding motif in mycobacterial UsfX homologs could be identified on the basis of modeling studies and sequence analysis. In the human integrin CR3 structure (Jie-Oh *et al.*, 1995) a DXSXS motif, where X represents any amino acid, was earlier identified to be involved in metal ion binding. A superposition of the crystal structure with the present MtUsfX model along with the sequence analysis lead to the identification of a conserved XGSFS motif in mycobacterial UsfX homologs where X is mainly a P or L (Fig 3.14). An examination of the superposed Mg^{2+} ion position reveals that this motif is in an advantageous position to interact with the bound metal ion. No such motif has been reported in other characterized anti-sigma factors. In fact the mycobacterial anti-sigma factors form a separate phylogenetic cluster (Fig 3.2) and suggest that these proteins might have evolved novel modes of specific inhibition of their cognate sigma factors.

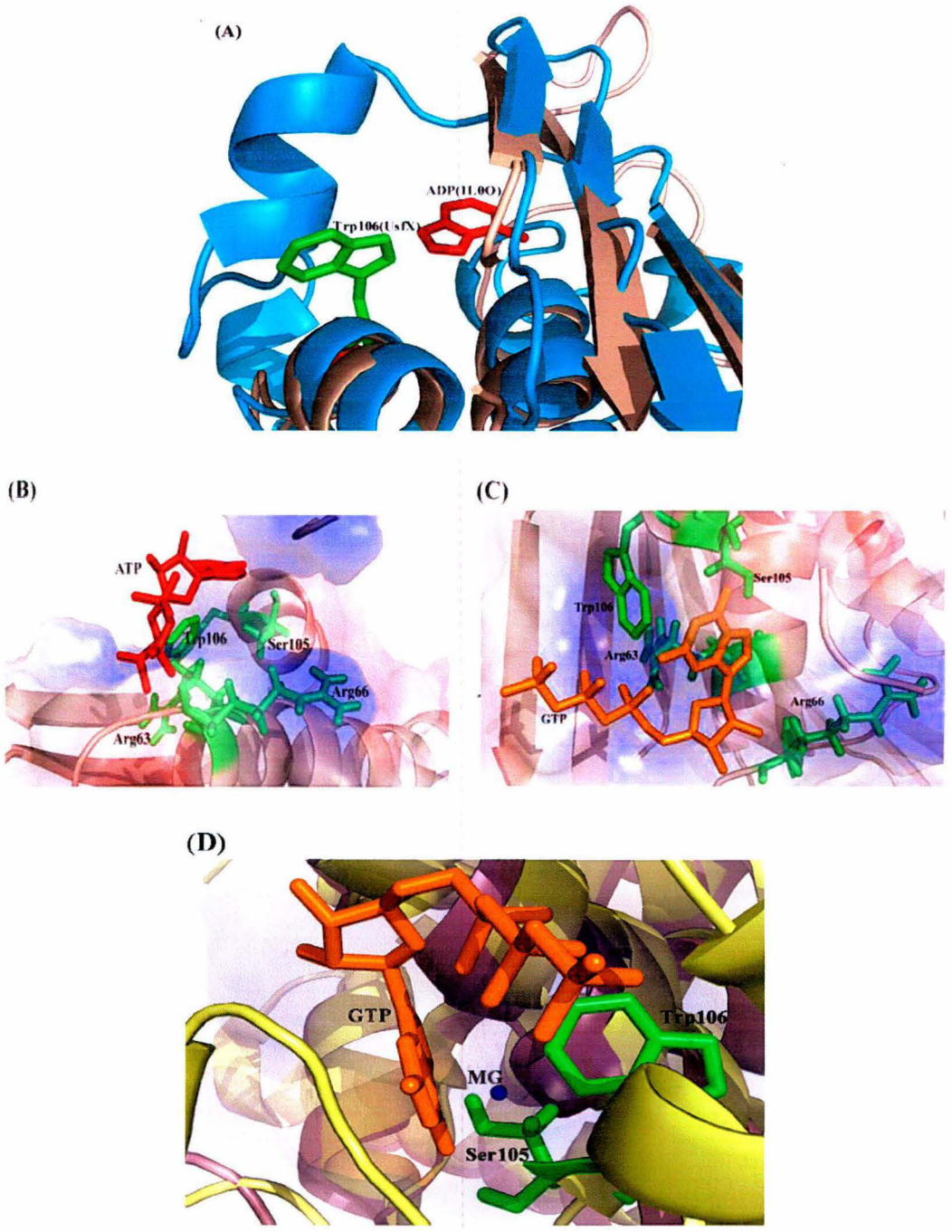


Fig 3.13: Nucleotide binding site of MtUsfX

(A) Structural superimposition of modeled MtUsfX (brown) with 1L0O (cyan). The nucleotide binding site of 1L0O is proximal to the MtUsfX Trp106 in the superposition.

(B) & (C) Docked ATP and GTP respectively in the MtUsfX nucleotide binding site.

(D) Mg^{2+} metal ion interactions in the active site. The metal ion interacts with residues from the conserved motif XGSFS.

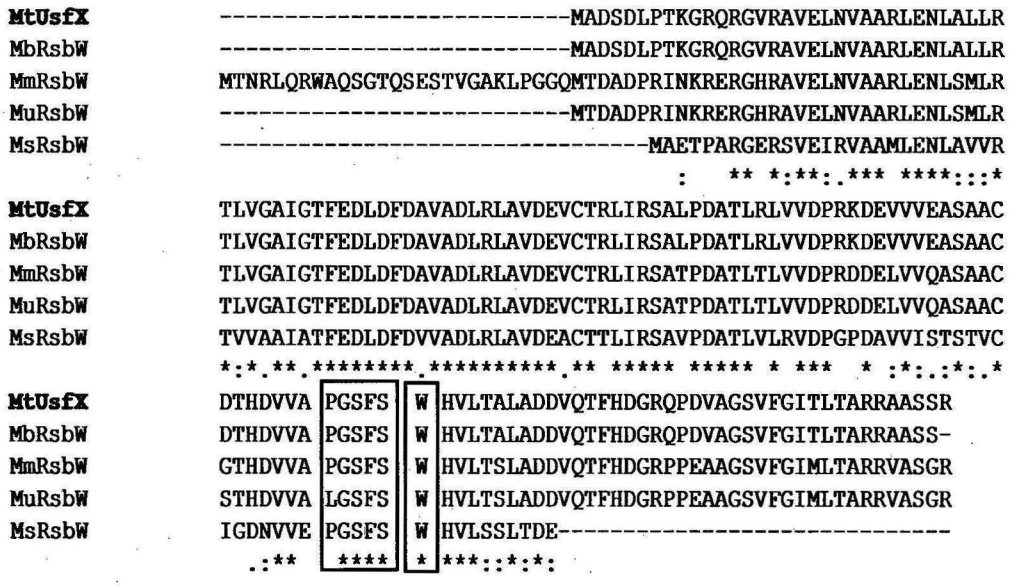


Fig 3.14: Multiple sequence alignment of MtUsfX with anti-sigma factors of mycobacterium family. The highlighted sequences depict the putative metal ion binding site XGSFS and the presence of conserved Trp in the nucleotide binding site. (**Mb** : *Mycobacterium bovis*, **Mm**: *Mycobacterium marinum*, **Mu**: *Mycobaterium ulcerans*, **Ms**: *Mycobaterium smegmatis*)

3.2.5 NTPase assay of MtUsfX

In other known anti-sigma factors, ATP binding has been associated with the phosphorylation of their anti-anti-sigma factors (Dufour *et al.*,1994). The probable role for this phosphorylation is to bring about the release of anti-sigma factor from anti-sigma and anti-anti sigma factor complex in the presence of increasing concentrations of ATP. The kinase activity of MtUsfX has been a subject of speculation in terms of phosphorylation of its anti-anti-sigma factor RsfB. Beaucher *et al.*, 2002 have reported the existence of autophosphorylation activity in MtUsfX but no kinase activity for RsfB as the substrate. More recently Greenstein *et al.*, 2007 have reported kinase activity of MtUsfX for the universal kinase substrate MBP (Myelin Basic Protein). No bands for autophosphorylated MtUsfX bands could be seen. We tried the kinase assay with MBP but could not detect either autophosphorylation or the kinase activity. Then we went on to identify and characterize the nucleotide hydrolysis activities of MtUsfX. The hydrolysis activity was performed by observing for the release of γ-$^{32}P_i$ in a hydrolysis reaction utilizing the [γ-^{32}P] NTP as the substrate and the activity was monitored over time periods of upto 8 hrs. MtUsfX was found to possess both ATPase and GTPase activities with the rate of ATP hydrolysis 4 times higher than that of GTP (Fig 3.15). Although the binding affinity of GTP for MtUsfX was found to be higher, the higher hydrolysis rate in case of ATP can be explained on the basis of the differential binding of the nucleotides. GTP is more embedded in the active site and that apparently positions phosphate residues away from the residues involved in hydrolysis (Section 3.2.2.3) while in case of ATP which is not too buried into the active site, the phosphate moieties are relatively free to get hydrolyzed. We tested the same reactions mixtures for autophosphorylation also by running on SDS-PAGE and no signals could be detected for autophosphorylated bands of MtUsfX. Thus MtUsfX possesses nucleotide hydrolysis property with the ability to use more than one nucleotide as the substrate for hydrolysis but the actual role of hydrolysis in terms of kinase activity or any other activity needs to be probed further.

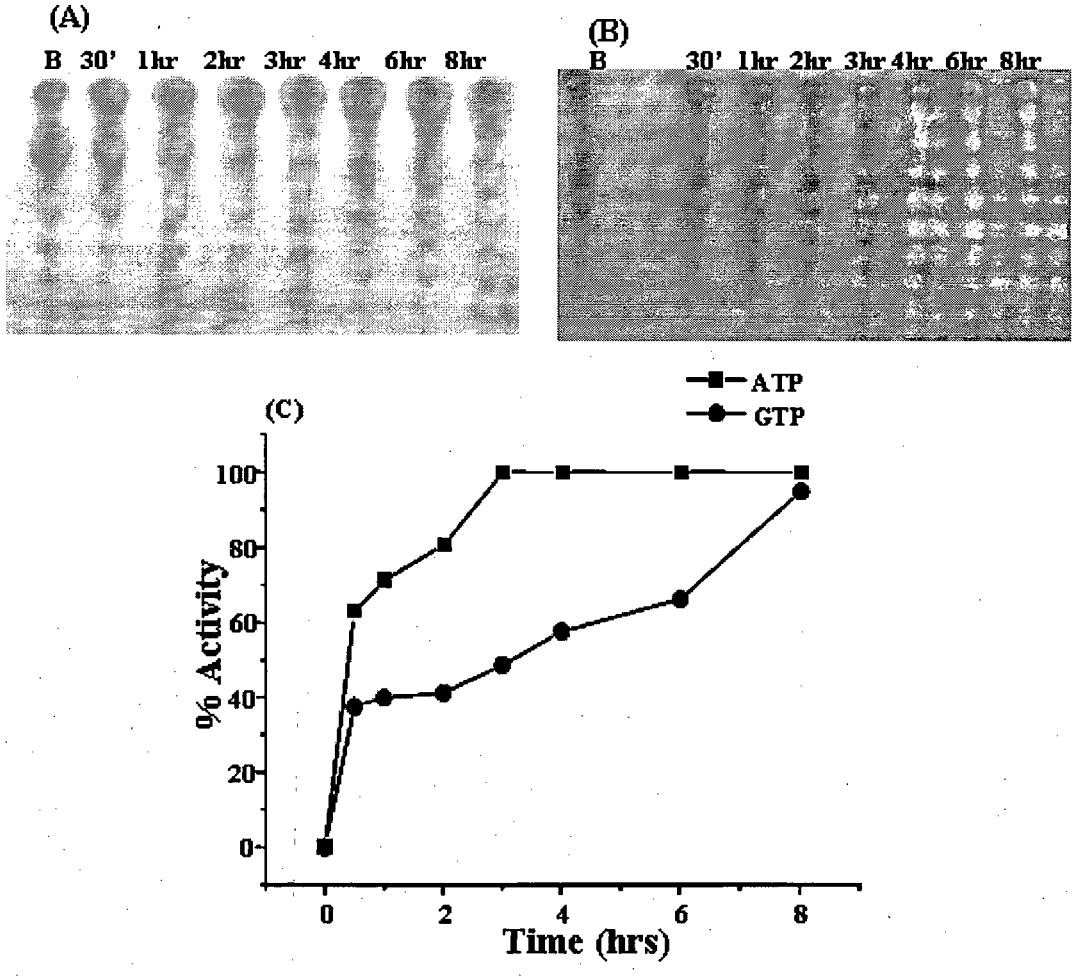


Fig 3.15: ATP and GTP hydrolytic activity of MtUsfX.
(A) Autoradiograph of ATPase activity shows complete hydrolysis attained within 2 hours. **(B)** GTP hydrolysis autoradiograph showing reaction continuing even upto 6 hours. **(C)** % Activity plotted as a function of time.

3.2.5.1 Substrate specificity

The property of hydrolyzing both ATP and GTP prompted us to screen for the substrate specificity of MtUsfX. To determine the substrate specificity, competition experiments were carried out with unlabelled ATP and GTP at concentrations ranging from 20-100μM. As shown in Fig 3.16, at even 20μM of competitor concentration there is a 55% loss of activity in the case where unlabelled ATP is used as a competitor. On the other hand unlabelled GTP used as a competing substrate resulted in only 20% reduction in ATPase activity. The higher ATPase activity observed in the presence of GTP shows higher specificity of MtUsfX for ATP. Thus even though both nucleotides can be used as substrate like other known anti-sigma factors MtUsfX uses ATP as preferable substrate, but utilizing GTP may be related to some unexplored function of this protein. Competition experiments were also performed with ATPγS, a slowly hydrolyzing analog of ATP. ATPγS was found to result in enhanced reduction by about 20% as compared to unlabelled ATP alone. The enhanced reduction can be attributed to blocking of the hydrolysis reaction by binding of ATPγS to the nucleotide binding sites.

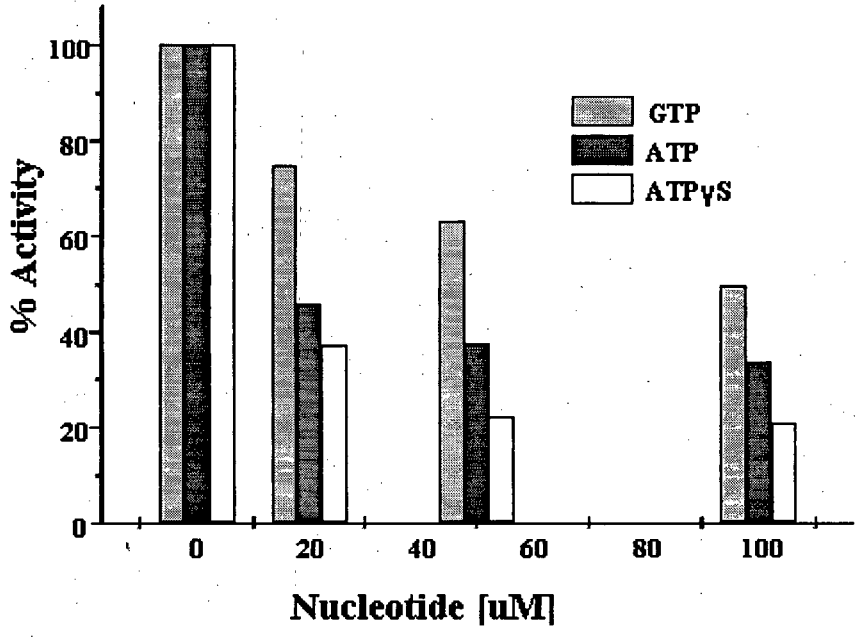


Fig 3.16: ATPase assay of MtUsfX in the presence of unlabelled nucleotide competitors.

3.2.5.2 Role of divalent ions

The role of divalent ions in nucleotide hydrolysis reaction is to properly orient the β and γ phosphate groups of nucleotides. Experiments to probe the requirement of divalent ions for the activity of MtUsfX demonstrated that the activity is maximal in the presence of Mg^{2+} (Fig 3.17A). In the case of Co, Ca and Mn salts the activity is similar while it is abolished in the presence of the Ni, Zn and Sr salts as the protein precipitates. As in nucleotide binding Mg^{2+} was found to enhance the binding of the nucleotide, the requirement for hydrolysis activity further confirms the importance of Mg^{2+} in functioning of MtUsfX.

3.2.5.3 Role of monovalent ions

The increasing concentrations of monovalent ions were found to result in decrease in activity with NaCl having more inhibitory effect than KCl (Fig 3.17B). This probably results from the disruption of the active site and the interactions with the Mg^{2+} ions which are necessary for activity in the presence of high salt concentrations. The complete absence of salt for optimal activity can't be predicted because our reaction mixture contained approximately 20mM of KCl; further reduction in salt concentration leads to precipitation of the protein.

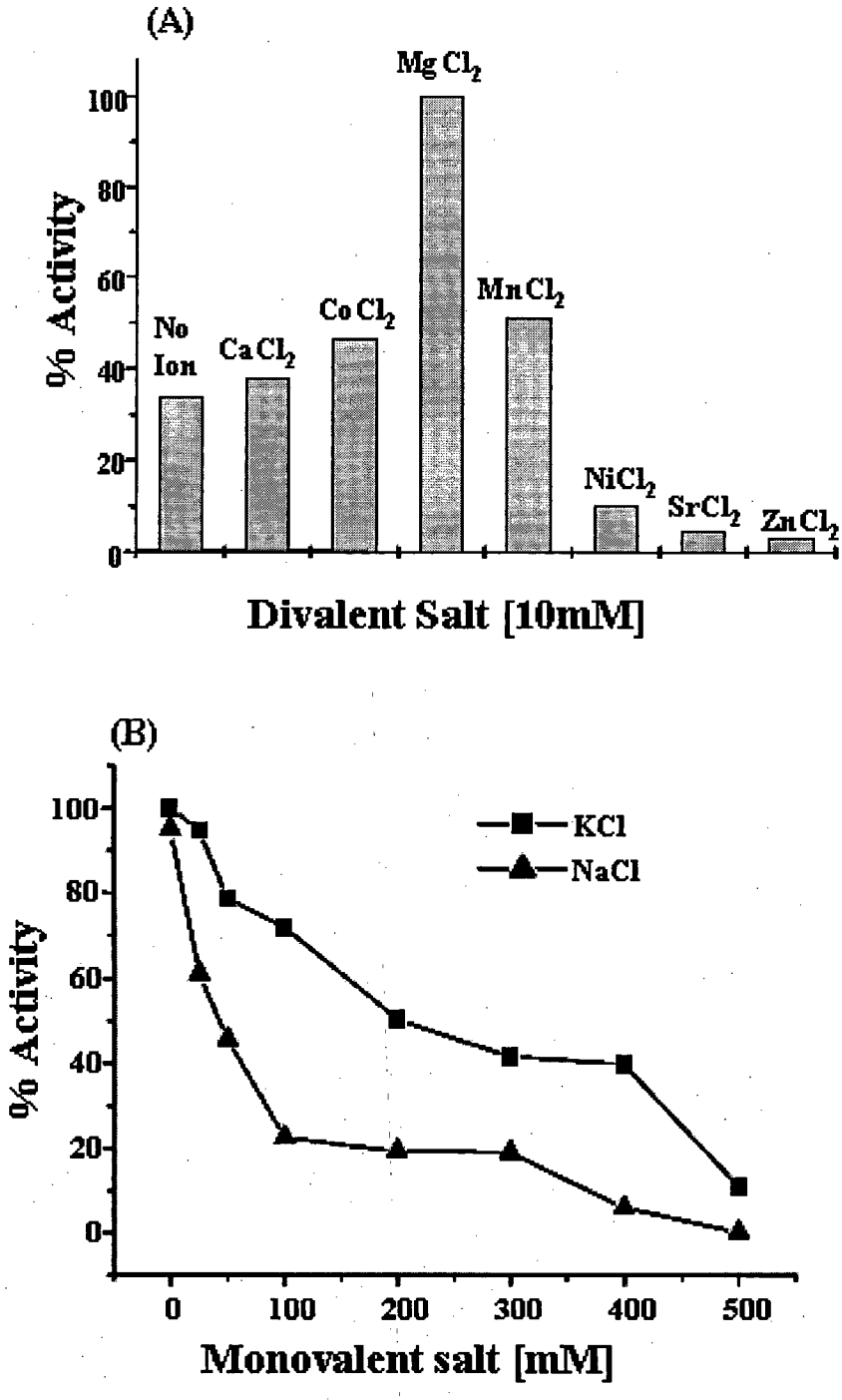


Fig 3.17: (A) Divalent ion specificity for MtUsfX ATPase activity. (B) Effects of salt concentration on the ATPase activity of MtUsfX.

3.3 Summary

The anti-sigma factor MtUsfX (Rv3287c) from its sequence and phylogenetic analysis can be classified to be a member of a highly conserved class of mycobacterial anti-sigma factors. The gene for MtUsfX was cloned into *E. coli* based expression system vector and and the soluble protein was purified and found to exist as dimer in solution. Exploiting the changes in intrinsic tryptophan fluorescence as a consequence of ligand binding, nucleotide binding properties were identified in MtUsfX. MtUsfX doesn't have any of the known conserved motifs of nucleotide binding, therefore utilizing intrinsic tryptophan fluorescence had an advantage in that it helped in 'locating' the binding site of the protein. The protein was found to bind to a variety of nucleotides including GTP, ATP, CTP, TTP and ADP with a stoichiometry of 2 nucleotide molecules per dimer. Two non-cooperative nucleotide binding sites per MtUsfX molecule were identified. Adenosine nucleotides were found to behave differently from other nucleotides. The nucleotide binding properties were rationalized by *in silico* modelling and docking approaches with the binding of the nucleotides identified to be through their ring moieties. A conserved metal ion binding motif XGSFS in mycobacterial UsfX homologs was found in the nucleotide binding site expectedly facilitating the ligand binding. Dual nucleotide hydrolysis activity was identified in MtUsfX with the ATPase activity about 4-fold higher than the activity with GTP.

Conclusively, MtUsfX is a dimeric protein with two independent nucleotide binding sites per subunit and possessing ATP and GTP hydrolysis properties.

CHAPTER 4

PROTEIN-PROTEIN INTERACTION ANALYSIS OF MtUsfX WITH MtSigF AND MtRsfA

4.1 Introduction

Protein-protein interactions are an intrinsic part of cellular processes. Many such interactions are transient but control and regulate a number of functions of the cell. The interactions among proteins can have a number of effects like altering kinetic properties, forming a new or altered binding site; change the specificity of a protein for its substrate or simply inactivating the binding partner. The analysis of the protein-protein interactions in terms of the factors contributing to their interaction, the consequences of the interaction and the extent to which interactions take place inside the cells helps in understanding the significance of these interactions.

The regulation of gene expression in *M. tuberculosis* is performed by an exquisite cascade of transcription factors. An important component of these proteins consists of sigma factors, anti-sigma factors and anti-anti-sigma factors which interact *via* direct physical interactions. In this chapter we detail the cloning and purification of the interacting partners of MtUsfX *viz.*, sigma factor MtSigF (Rv3286c) and anti-anti-sigma factor MtRsfA (Rv1365c). The co-expression of MtUsfX and MtSigF for the *in vivo* formation of UsfX-SigF complex and the *in vitro* conditions for the formation of the UsfX-RsfA complex have been elucidated. The purification of the UsfX-SigF and UsfX-RsfA complexes are discussed as also the factors contributing to their formation and stability.

4.2 Results and Discussions

4.2.1 UsfX-SigF interaction analysis

4.2.1.1 Cloning, over expression and purification of MtSigF

The 786bp *sigF* gene was PCR amplified from *M. tuberculosis* genomic DNA and cloned into the T7-based expression plasmid pET21d (Novagen) at NcoI and HindIII sites to acquire a C-terminal 6x Histidine tag and T5-based expression plasmid pQE31 (Qiagen) at BamHI and HindIII sites to acquire an N-terminal 6x Histidine tag. The clones were confirmed through restriction digestion (Fig 4.1A) and the integrity of clones was confirmed by sequencing.

The expression plasmids were introduced into *E. coli* host strains that contain T7 RNA polymerase gene [like BL21 (DE3) and C41(DE3)] and strains that contain T5 RNA polymerase gene [like JM109 and XL1 Blue]. The expression of MtSigF in *E. coli*

could not be achieved upon induction in the exponential phase in any of the host-vector combinations. The final expression of MtSigF was attained by utilizing the pQE31 construct and C41(DE3) host and was achieved only when induced at the late growth stage *i.e* OD_{600} greater than 1. This could be due to the toxic nature of MtSigF as reported earlier in case of B. *subtilis* (Campbell *et al.*, 2000). The conditions for growth were optimized at 37 °C by adding 0.3mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and growing in 2X YT medium for further six hours. Most of the MtSigF protein was found to be present in the insoluble fraction. The addition of 10mM β-mercaptoethanol into the buffer containing 50mM Tris-Cl, pH 8.5, 50mM NaCl enabled the solubilisation of the protein upto ~30%. β-mercaptoethanol possibly prevents self aggregation that might occur on account of a single cysteine residue present in MtSigF. The protein could be identified as a band of ~30kDa on 12%SDS-PAGE. This was approximately equal to the calculated mass 28.7kDa±0.8kDa tags. There was no band corresponding either to C41 (DE3) alone or C41 (DE3) transformed with pQE31 without the insert when induced with IPTG. The N-terminal His-tagged protein was purified from a Ni^{2+}-IDA column (GE Biosciences). The column was equilibrated with buffer containing 50mM Tris-Cl, pH 8.5, 50mM NaCl, 10mM Imidazole and 10mM β-mercaptoethanol (Buffer A). The protein was eluted using a linear gradient of buffer B (Buffer A containing 500mM Imidazole). The protein eluted at 30-45% Imidazole in buffer B. The protein was highly pure as observed by SDS-PAGE analysis (Fig 4.1B) and the final purification step was performed by precipitating the protein with 60% ammonium sulphate. The precipitate was dissolved in minimum volume of gel filtration running buffer (50mM Tris-Cl, pH8.5, 50mM NaCl, 10mM β-mercaptoethanol and 5mM EDTA). Trace contaminants were removed by gel filtration utilising Superdex-75 column. The protein eluted at a volume of 11.2ml in the above said column and from the calibration of the column, this elution volume corresponds to a molecular weight of ~30kDa and agrees well with the predicted molecular weight of 29.6kDa for the construct. The molecular mass analysis supports the monomeric nature of MtSigF protein in solution which is in accordance with the *in vivo* existence of sigma factors. In the absence of β-mercaptoethanol from the elution buffer most of MtSigF was found to elute as soluble aggregates in the void volume of the gel filtration column (Fig 4.1C).

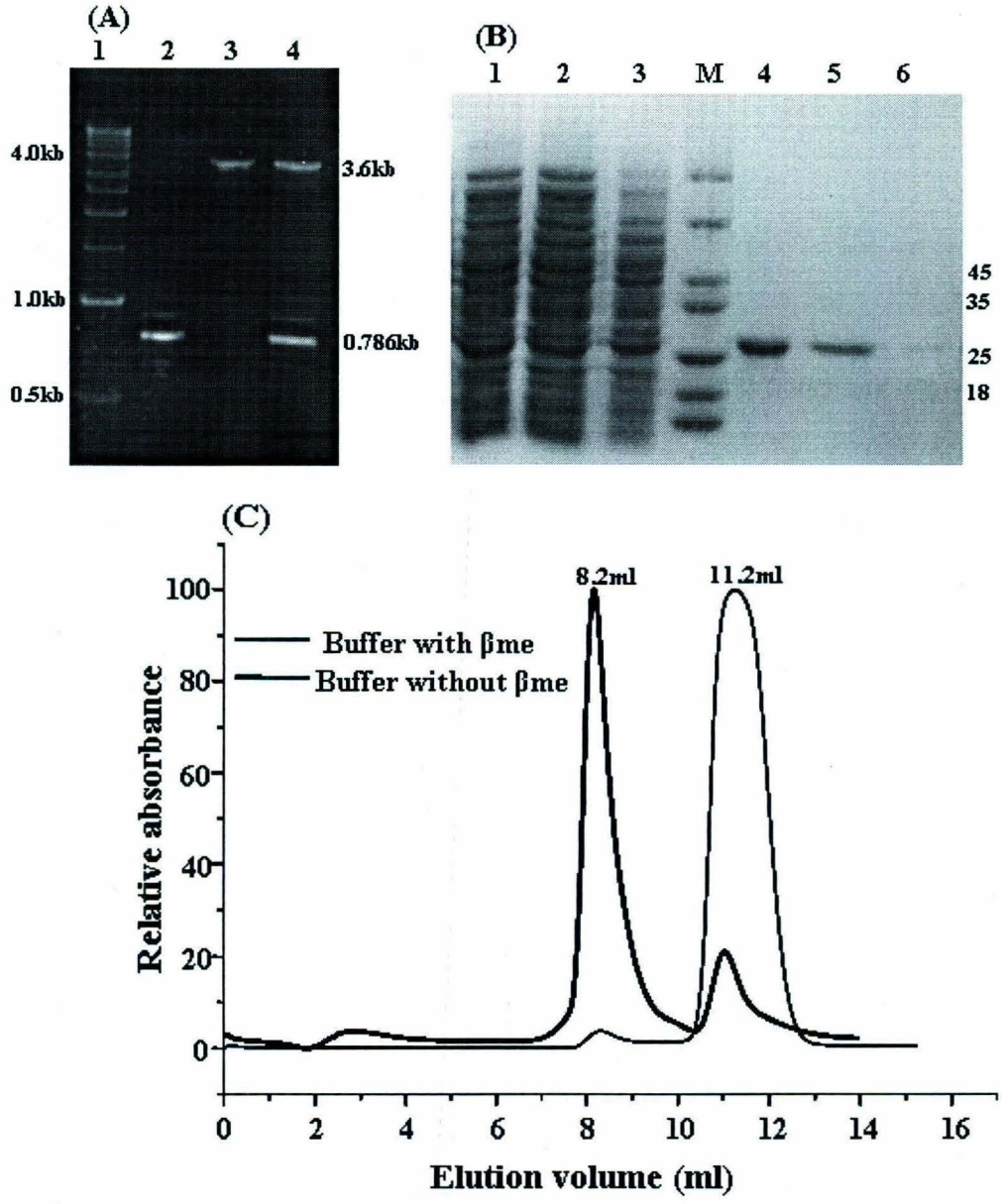


Fig 4.1: (A) Cloning of *M. tuberculosis* SigF in pQE31.
Lane1: 1Kilobase DNA marker, Lane 2: BamHI and HindIII digested pQE31, Lane3:786bp *sigF* PCR amplified fragment digested with same enzymes, Lane3: Restriction digestion of *sigF* clone in pQE31.

(B) Purification of MtSigF.
Lane1:Load, Lane2:Flowthrough, Lane3:Wash1, Lane4:Wash2, Lanes5,6 and 7: Purified fractions From Ni $^{2+}$-IDA column, Lane8:Low Molecular Weight Protein Marker.

(C) Size-exclusion chromatogram. The elution volume of 11.2ml points towards a molecular weight of ~ 30kDa for MtSigF. The elution volume at 8.2ml shows the existance of soluble aggregates of MtSigF in absence of β-mercaptoethanol in buffer.

4.2.1.2 Co-expression and purification of UsfX-SigF complex

The problems encountered with the expression and solubility of MtSigF in the earlier stages of the work was one of the factors for attempting the co-expression of MtUsfX and MtSigF. The co-expression vector was constructed as described in section 2.4.4. The co-expression vector *pUS* placed MtUsfX and MtSigF in a transcriptional operon similar to that found in MtSigF operon *in vivo*. Each gene contains its own ribosomal binding site (rbs), with the termination of transcription achieved in each gene with the help of stop codons in respective gene sequences. This system added an N-terminal His_6-tag on MtSigF (Fig 4.2A). The final construct was cloned in pET21d at NcoI and HindIII in such a way so as to keep the operon under the control of a single inducible T7 RNAP promoter. The integrity of the size of the construct was confirmed by restriction digestion and final confirmation performed by sequence analysis (Fig 4.2B). *E. coli* C41 (DE3) cells transformed with this plasmid and induced transcribed both genes on a single RNA. The expression of UsfX-SigF complex was achieved at 37°C by adding 0.3mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) to an OD_{600} of ~0.4 and growing for further three hours. Since induction occurs in the exponential phase, where ATP is presumably in sufficient concentration, we expected the two proteins to assemble into a complex *in vivo*, but the proteins were translated separately and separated as two separate proteins on SDS-PAGE (Fig 4.2C). The complex was purified in a buffer containing 50mM Tris-Cl pH 8.5, 50mM NaCl, 10mM $MgCl_2$ and 1mM ATP in a batch wise manner from Ni^{2+}-IDA column with equilibration at 10mM Imidazole, washing with 30mM and 70mM Imidazole and finally elution with a 200mM Imidazole. The protein-protein interactions were strong enough to allow for the purification of the complex based on the N-terminal His_6-tag on MtSigF alone in the co-expression vector. The untagged MtUsfX from the solution remained as a complex with MtSigF through the purification steps. The existence of the two proteins as a complex could also be verified from the elution pattern in size-exclusion chromatography. Further the major fraction of MtSigF eluted in the void volume as aggregates while in the presence of its interacting partner MtUsfX no peak for aggregates could be detected (Fig 4.3A).

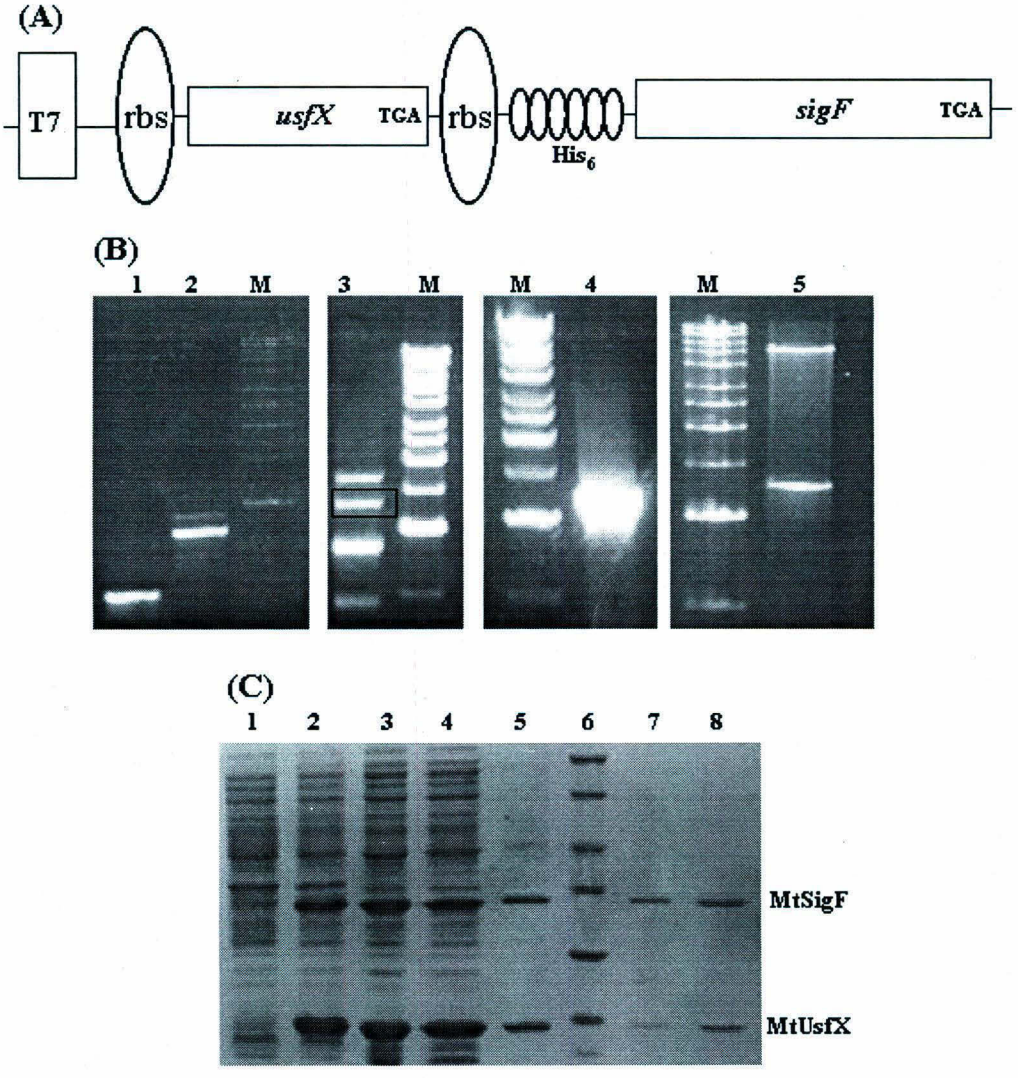


Fig 4.2: Co-expression and purification of MtUsfX and MtSigF; UsfX-SigF Complex formation.

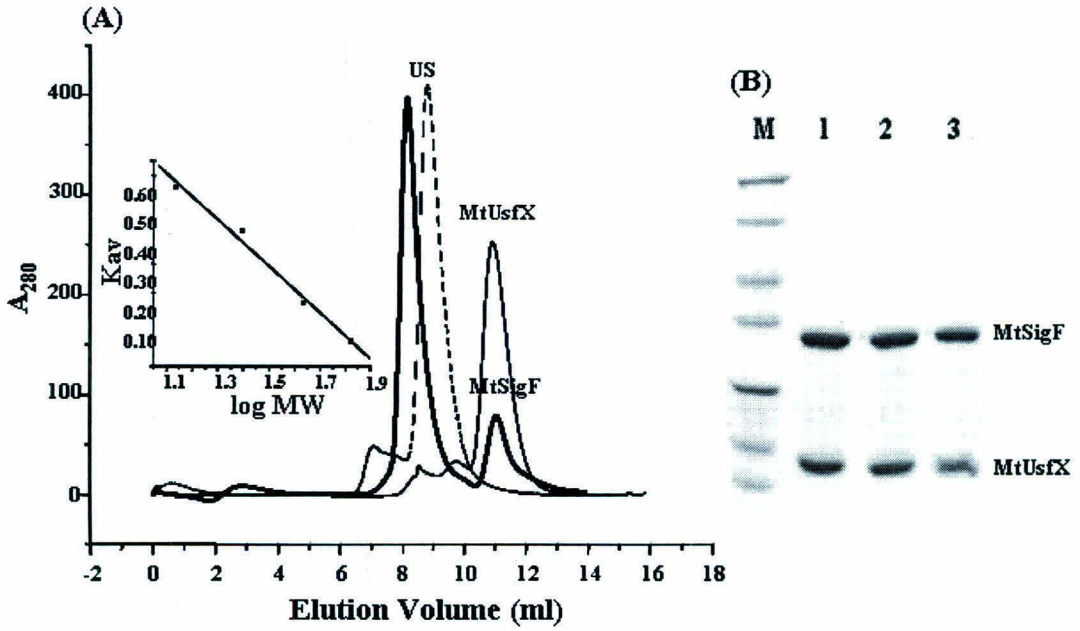
(A) The genes coding for MtUsfX and MtSigF were placed in a transcription operon under T7 promoter. Each gene contains its own ribosomal binding site. (*pUS* construct)

(B) Fusion of *usfX* and *sigF* and cloning into pET21d. *Lane1.* Digested *usfX. Lane2.* Digested *SigFrbs*. *Lane 3.* Ligation of *usfX* and *SigFrbs Lane 4.*Band of 1.25 kb was excised from ligation reaction and PCR amplified with *usfX* (Forward) and *sigF* (Reverse) primers. *Lane 5.* Confirmation of the clone in *pET21d.*

(C) Cells transformed with the fusion vector were induced for two to three hours and proteins purified from the solube fraction of lysed cells. *Lane 1.* Uninduced cells. *Lane 2.* Induced cells. *Lane 3.* Soluble fraction applied to a Ni-NTA column. *Lane 4.* Flowthrough *Lane 5.* 50mM Imidazole eluate. Lanes 7&8. Purified fractions with 300mM Imidazole.

4.2.1.3 Stoichiometry analysis

The analysis of stoichiometric ratios in which the proteins interact can give mechanistic information about the functional importance of the interaction. In order to investigate the stoichiometry of the UsfX-SigF complex we examined the results of gel filtration analysis. The protein complex eluted at a volume of 8.9ml which corresponds to a molecular weight of about 67kDa. Using the molecular mass of MtSigF (28.7kDa + 1.54kDa Tags) and MtUsfX (15.55kDa) along with our early finding that MtUsfX exists as a dimer, it was concluded from the gel filtration analysis that UsfX: SigF complex exists in a stoichiometric ratio of 2:1 with a molecular weight of 60.3kDa *i.e* the MtUsfX dimer interacts with one chain of MtSigF. Densitometry analysis of the SDS-PAGE gels also yielded a stoichiometry of UsfX:SigF (1.75:1) interaction which agrees well with the above result. The molecular mass of MtUsfX is almost twice as that of MtSigF while the intensity of the two bands is almost equal thus showing that MtUsfX is twice as that of MtSigF (Fig 4.3). A similar stoichiometry is observed in the case of the *B. stearothermophilus* SpoIIAB-SigF complex too (Campbell *et al.*, 2002). The ratio is also expected in terms of the individual oligomeric status of the proteins and can be concluded that the anti-sigma factor MtUsfX interacts with its sigma factor MtSigF to block the transcription.



(C)

Protein	Monomeric mass	Sample	Observed Intensity	Observed Intensity/mass	Calculated stoichiometry
MtSigF	28.76kDa	Lane2 Lane3	21 16	1.28	
MtUsfX	15.6kDa	Lane2 Lane3	19 16	2.24	
Complex					**1.75/1**

Fig 4.3 A: Gel filtration and densitometry analysis of UsfX-SigF complex.

(A) A Superdex-75 column was used for molecular weight determination of the UsfX-SigF complex. The protein complex eluted at a volume of 8.9ml which corresponds to a molecular weight of about 67kDa.Standard molecular weight markers were used for calibrating the column.

(B) SDS-PAGE analysis of the peak fractions from gel filtration.

(C) Calculated stochiometry of UsfX-SigF based on PAGE and densometric scanning.

4.2.1.4 Structural studies on UsfX-SigF interaction

The analysis of the structural properties of a protein-protein complex or the changes involved in the interacting partners can give an idea about the extent of interactions and the forces involved in the interaction. Analyzing the strength of the UsfX-SigF complex at high salt concentrations, we found the interaction between the proteins to be stable up to 0.6M salt concentration suggesting that hydrophobic forces are largely involved in stabilizing the protein-protein interactions. The observed strength of the complex at high ionic strength may have arisen due to the functions of these proteins under stress conditions. The complex is stable in the absence of the nucleotide too. Excluding $MgCl_2$ and ATP from the size exclusion buffer didn't have any effect on the stability of the complex as evident from the elution profile. We also tried probing complex formation of individually purified proteins *in vitro,* but were unable to get a stable complex in the presence/absence of nucleotide. It is therefore probable that additional *in vivo* interactions/factors may be responsible for generation of the UsfX-SigF complex.

Circular Dichroism spectroscopy can be used to probe the changes in the secondary structural elements due to molecular interactions. The CD spectra obtained for both MtUsfX and the UsfX-SigF complex are typical of those seen for proteins with a moderately helical content, whereas the strikingly different spectrum observed for MtSigF alone is indicative of high helical content (Fig 4.4A). A helical content of 70% was calculated for MtSigF with K2D server which is near about similar to the predictions with JPRED2 package and 3DPSSM Fold prediction server which predict an overall helical content of 60%. From these spectra it can be concluded that the interaction between these two proteins is accompanied by gross structural changes in MtSigF with the helical regions playing a greater role in the interaction (Table 4.1). Thermal denaturation analysis using the two state model showed a higher thermal stability for the UsfX-SigF complex as compared to MtSigF (Fig 4.4B). This increased resistance to high temperatures can be attributed to the complex formation with MtUsfX which was found to have a high T_m. Overall the interaction between MtUsfX and MtSigF is accompanied by a change in the secondary structure of MtSigF and the complex is stabilized by the more stable MtUsfX.

Table 4.1 : Predicted and calculated secondary structure of MtUsfX and MtSigF.

Structural Element	Method	MtUsfX	MtSigF	UsfX-SigF Complex
α-Helices	3D PSSM	31	59	n.d*
	JPred	27	58	n.d
	Calculated	45	70	53
β-Sheets	3D PSSM	22	5	n.d
	JPred	23	n.r*	n.d
	Calculated	25	n.r	20

*n.d, not done; n.r, no results

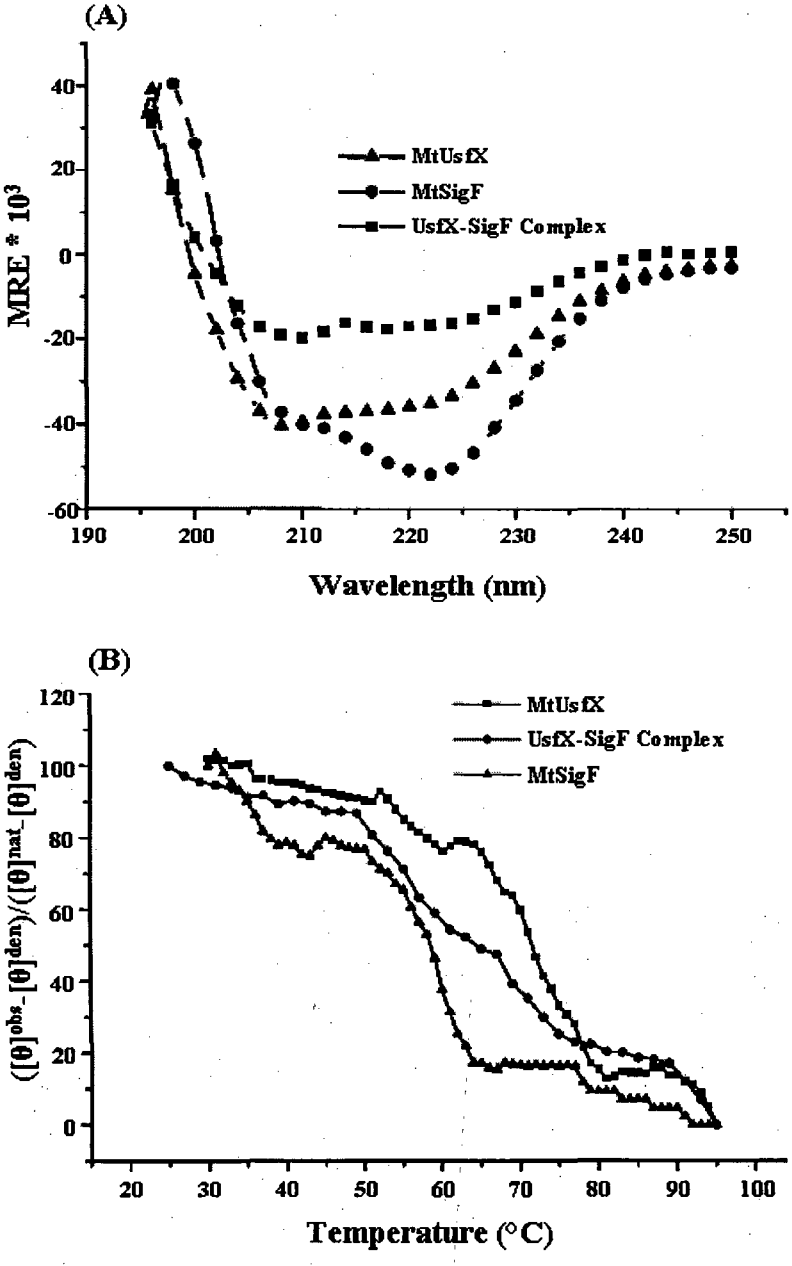


Fig 4.4: CD spectroscopy studies with MtUsfX, MtSigF and UsfX-SigF complex.

(A) Secondary structural changes are observed particularly in MtSigF as a consequence of complex formation.

(B) Thermal denaturation profiles of the two proteins and the protein complex.

Fluorescence spectroscopy studies can also be used to draw information about changes in tertiary structures/conformations of proteins due to their interactions with other proteins by analyzing changes in the micro-environments of their tryptophan residues. Typical intrinsic fluorescence spectra obtained for MtUsfX (Trp106), MtSigF (Trp112), and the UsfX-SigF complex are shown in (Fig 4.5A). MtSigF is characterized by a fluorescence maximum at about 339nm. The emission maximum of MtSigF is blue shifted relative to that of free L-tryptophan, for which the emission maximum under the same conditions is observed at 354nm. It can be concluded that the tryptophan molecule in MtSigF is in a hydrophobic environment not significantly exposed to the solvent as a fluorescence emission λmax at 330-340 nm, with an increased fluorescence yield is characteristic of tryptophan in such environment (Lakowicz, 1983). The fluorescence maximum observed for the UsfX-SigF complex at 341.5nm is blue shifted by approximately 8nm when compared to MtUsfX and slightly red-shifted (by 2.5nm) when compared to that for MtSigF. From these fluorescence spectra it can be concluded that there is a change in the environment of at least one of the two tryptophan species (either of MtSigF or MtUsfX) upon complex formation. From the fluorescence maxima of the individual proteins, it appears that it is the Trp112 of MtSigF which must be slightly solvent exposed as a result of UsfX-SigF interaction. This can be explained from the nucleotide binding properties of MtUsfX also, where strong changes in the intrinsic tryptophan fluorescence as a consequence of nucleotide binding are detected and hence the only tryptophan present in MtUsfX is part of the nucleotide binding site. The nucleotide binding properties of other anti-sigma factors are significant in terms of them phosphorylating their anti-anti-sigma factors and an induced release mechanism has been proposed where the interaction of one binding partner (anti-anti-sigma factor) induces the release of the other (sigma factor), hence the binding sites for two proteins can't be expected to be so close by (Margeret *et al.*, 2003).

Analysis of the stoichiometry of the UsfX-SigF interaction was also performed by observing the effect of increasing the molar ratio of MtSigF *vis-à-vis* the MtUsfX on the wavelength of maximum intrinsic fluorescence was probed following the methodology reported in another protein system (Renshaw *et al.*, 2002). The fluorescence maximum initially shifts from 348 to 342.5 nm with increasing MtSigF concentration; however no further change in the fluorescence maximum was observed at a molar ratio of 1-1.5. Subsequently, at higher concentrations of MtSigF, the

maximum again shifts towards 339 nm which is the fluorescence maximum of MtSigF alone in solution (Fig 4.5B). The data strongly suggests that the two proteins interact to form a 1:1 molar complex *i.e* the stoichiometry ratio will be 2:1 because of dimeric nature of MtUsfX.

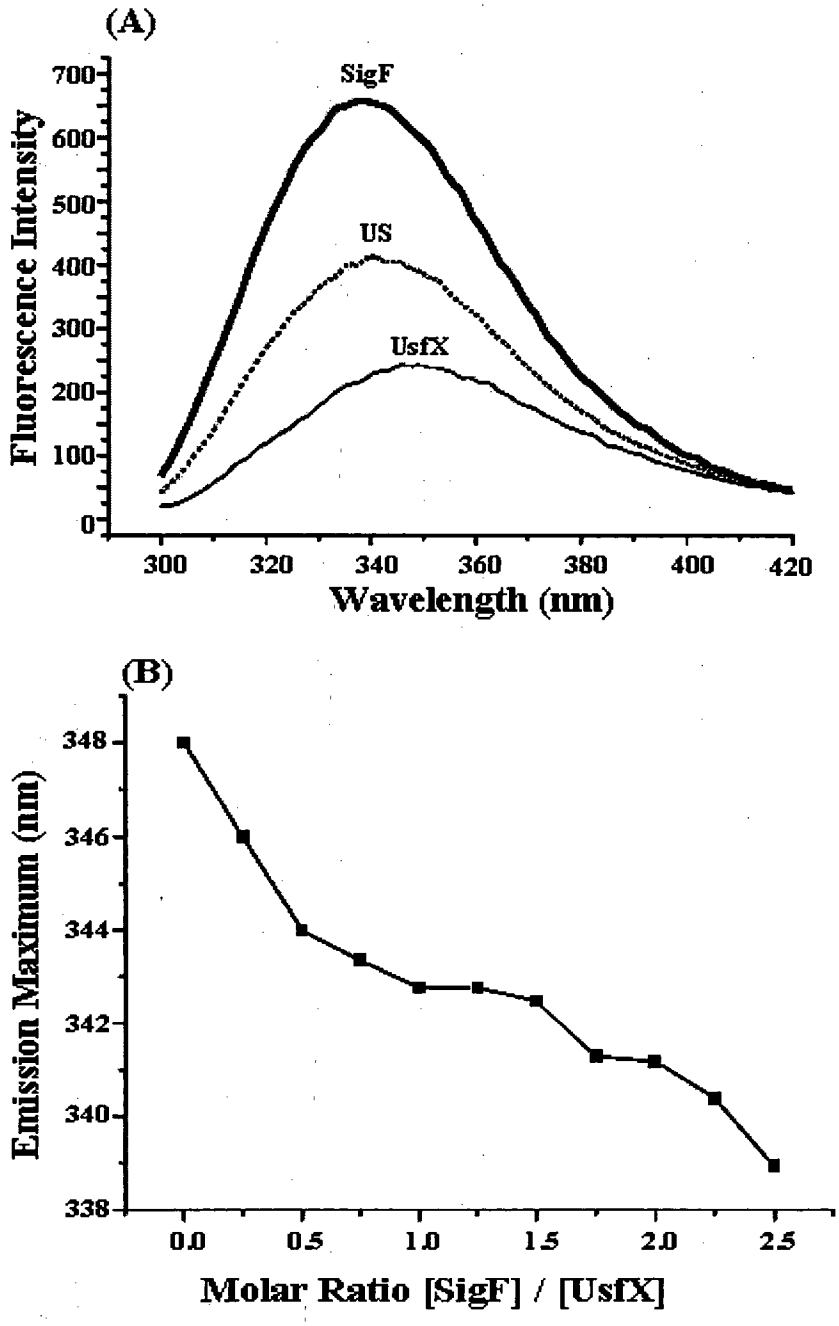


Fig 4.5: Tryptophan fluorescence properties of MtUsfX, MtSigF and UsfX-SigF complex.

(A) Emission spectra obtained for MtUsfX, MtSigF and UsfX-SigF complex. Clearly one of the two tryptophans undergoes a change on complex formation. **(B)** An example of the change in the wavelength of maximum intrinsic fluorescence observed on increasing the molar ratio of MtSigF to MtUsfX, which clearly suggests that the two proteins form a tight complex at a molar ratio of 1:1 (stoichiometry ratio will be 2:1 because of dimeric nature of MtUsfX).

4.2.1.5 *In silico* analysis of UsfX-SigF interaction

Attempts were made to crystallize UsfX-SigF complex as described in section 2.11.2. No promising conditions could be detected in any of the crystal screens utilized. We therefore resorted to the molecular modeling approach to analyze the interaction between MtUsfX and MtSigF and rationalize our experimental data with structural evidences. As described models were generated with the help of PHYRE server. MtUsfX could not be modeled for residues 92-100 while as the model for MtSigF could be made only for residues 110-167 which was found to be the interacting part with MtUsfX (Fig 4.6A). The region encompassing residues 30-42 from MtUsfX and that of 155-167 of MtSigF were found to form the interface of the UsfX-SigF complex. The analysis of the the MtSigF/MtUsfX interface reveals that it is dominated by the presence of hydrophobic residues in both the proteins. The residues of MtUsfX involved in the interaction consist of highly hydrophobic residues leucine (3), isoleucine (1), valine (1), phenylalanine (1), less hydrophobic residues; alanine (2), glycine (2), tyrosine (2) and a lone charged arginine residue. Similarly the residues of MtSigF that are mainly involved in the interaction consist of leucine (2), isoleucine (1), valine (1), alanine (2), serine (1) and glycine (1) and three charged residues ; two glutamines and one arginine. The hydrophobic indices calculated also show higher values of hydrophobicity for these stretches in both the proteins as compared to the overall hydrophobicity of the protein. This is in conjecture with our gel filtration experiments were we had found the UsfX-SigF complex to be stable even at high salt concentrations and thus involvement of hydrophobic forces in the interaction (Fig 4.6 B, C & D). Another aspect that could be rationalized from the model of UsfX-SigF complex was regarding the topography of the MtUsfX in terms of the locations of the MtSigF binding site and nucleotide binding site. We had concluded from our fluorescence data (Section 4.2.4) that the two binding sites have to be at a distance from one another. From the modeled UsfX-SigF complex it is evident while the MtSigF binding site is comprised of helix α1 (Residues 30-42), the nucleotide binding site is formed by α4 (Residues 105-112) having a significant structural distance. Thus MtUsfX possesses the structural features required to perform the functions of an anti-sigma factor.

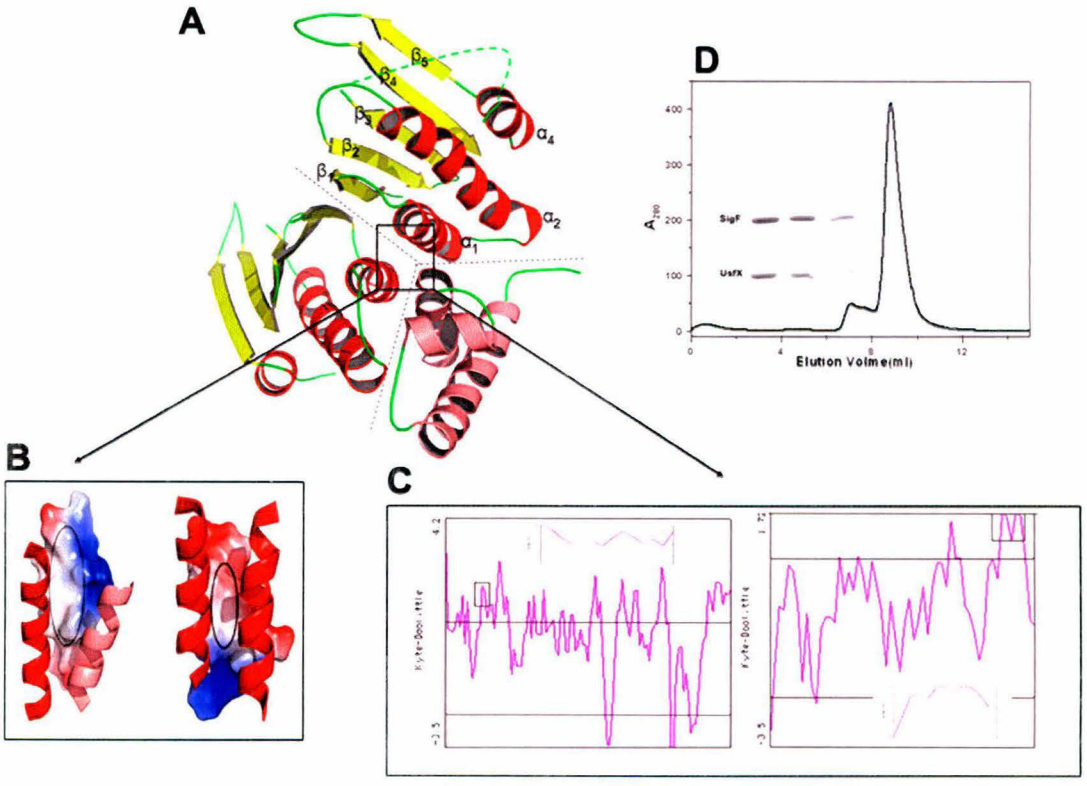


Fig 4.6: Modelled UsfX-SigF complex.
(A) Dimer of MtUsfX in complex with MtSigF.
(B) Surface diagram for MtUsfX dimer interface and UsfX-SigF interface.
(C) Hydrophobicity index for modeled MtUsfX and MtSigF. The interacting parts have a very high hydrophobicity index as shown in the inset Figures.
(D) Gel-filtration profile and SDS-PAGE in presence of 0.6M salt concentration. No change in elution profile and stoichometry is observed at high salt concentration.

4.2.1.6 Nucleotide binding properties of UsfX-SigF complex

We probed the nucleotide binding properties of the UsfX-SigF complex for analyzing the location of the binding sites and observe any changes that may have occurred at the nucleotide binding site of MtUsfX as a consequence of its complex formation with MtSigF. MtSigF alone didn't show any changes in fluorescence intensity when titrated with increasing concentrations of nucleotide. The titration curves and saturation isotherms for the interaction of UsfX-SigF complex with ATP and ADP are shown in Fig 4.7. No significant change was observed in the nucleotide binding properties of MtUsfX which suggests that they bind with nearly the same affinities in the *apo* MtUsfX and UsfX-SigF complex (Table 4.1). From the similar binding affinities for ATP and ADP, it can be concluded that the presence of a nucleotide in the nucleotide binding pocket of MtUsfX is not essential for its binding to cognate sigma factor MtSigF. From the *in vivo* formation of this complex it can be concluded that there may be other factors involved in the complex formation. The K_d and ΔF_{max} for MtUsfX and UsfX-SigF complex are identical (Table 4.2) thus showing the binding of the two ligands to MtUsfX has no effect on one another and it can be concluded that the location of the nucleotide binding site is apparently distal to the protein-protein interaction interface.

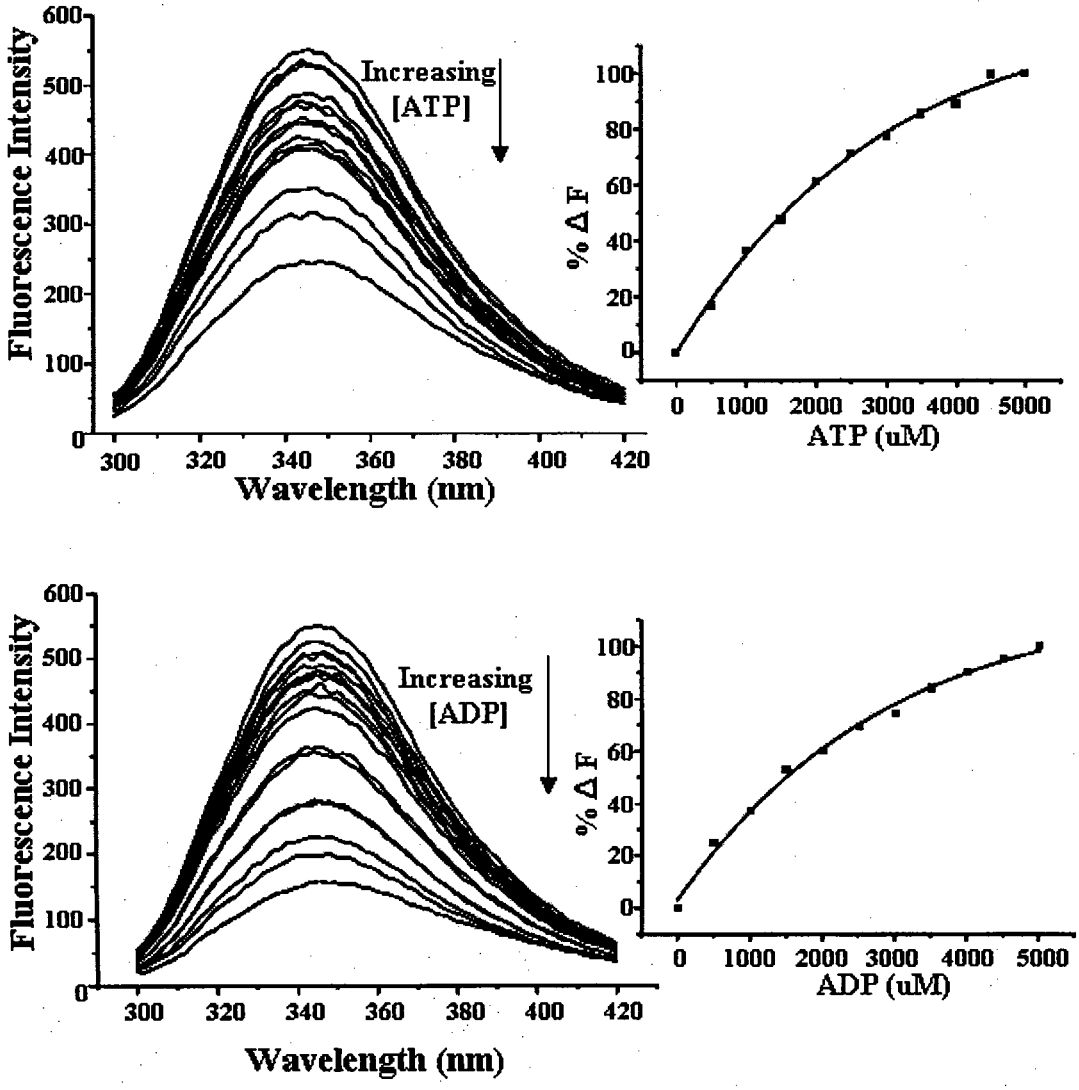


Fig 4.7: Nucleotide (ATP and ADP) binding properties of UsfX-SigF.
Left panel: Titration curves showing decrease in emission intensity with increasing nucleotide concentration.
Right panel: Saturation isotherms: The percent change in fluorescence intensity at 343nm (The emission maxima of UsfX-SigF complex) plotted as a function of nucleotide concentration.

Table 4.2 K_d values and calculated ΔF_{max} values ($\Delta Fmax_{(calc)}$) of MtUsfX and UsfX-SigF complex showing complex formation has little change on nucleotide binding properties of MtUsfX.

Ligand	K_d (μM)		$\Delta F_{max(calc)}$	
	MtUsfX	UsfX-SigF Complex	MtUsfX	UsfX-SigF Complex
ATP	1300± 50	1250±50	210±10	220±10
ADP	1900± 50	1550±50	205±10	215±10

4.2.2 UsfX-RsfA interaction analysis

4.2.2.1 Sequence and phylogenetic analysis of MtRsfA

The sequence and phylogenetic analysis highlighted some interesting aspects of MtRsfA (Rv1365c).This protein along with Rv3687c are the two anti-anti sigma factors that have been reported to bind to MtUsfX; albeit under distinct conditions (Beaucher *et al.*, 2002). Both proteins are called RsfA and RsfB respectively for 'regulator of sigma factor F'. While MtRsfA binds to MtUsfX under reducing conditions, RsfB binds in an ATP/ADP dependent manner. MtRsfA is characterized by the presence of four cysteine residues *viz.*, Cys73, Cys74, Cys86 and Cys109 out of which Cys73 and Cys109 by virtue of a disulphide bridge have been implicated in the anti-sigma antagonist activity. To our knowledge there are no other known anti-anti-sigma factors of this class which regulate the sigma factor activity by monitoring changes in the redox potential although anti-sigma factors have been reported that regulate activity of their cognate sigma factors in a oxidation potential dependent manner (Taeksun *et al.*, 2003). For performing sequence analysis the first step involved was searching various databases using MtRsfA as the query sequence. A combination of search tools BLASTP and PSI-BLAST were utilized to search for the homologs. Position specific iterated BLAST (PSI-BLAST) (Altschul *et al.*, 1997, 2005) search against the non-redundant protein sequences (nr) database resulted in the identification of about 100 sequences after three iterations. Additional iterations did not result in the identification of more sequences. Subsequent attention was focused on sequences with high E values ($>10^{-20}$). MtRsfA is found to be clustered in COG1366 of the Clusters of Orthologous Groups (COG) database of Conserved Domain Database along with six other *M. tuberculosis* genes *viz.*, Rv0516c, Rv1364c, Rv1904, Rv2638, Rv3687c and Rv0914c which have been designated as probable anti-sigma factor antagonists. The BLASTP search against the mycobactrial databases with MtRsfA as query sequences yielded hits with sequence identities upto 50%. A multiple sequence alignment with the CLUSTALW (Higgins *et al.*, 1994, Larkin *et al.*, 2007) program was performed to identify conserved residue patterns. The Cys73 of MtRsfA has replaced the serine in this anti-anti sigma factor family, as also reported earlier. A manual analysis of the protein sequences in the alignment revealed the existence of many proteins with cysteine residues present in place of conserved serines (Fig 4.8). We separated all these proteins with the criteria of having at least

three Cys residues out of Cys73, Cys74, Cys86 and Cys109 in MtRsfA. For the sake of convenience we name this as the CC-(11)-C-(22)-C pattern where 11 and 22 refer to the sequence spacing between the respective Cys residues. Interestingly all the proteins with such residues belonged to the members of mycobacteria where they have been classified as conserved hypotheticals, regulatory proteins or signal proteins (Fig 4.9). While a representative protein with CC-(11)-C-(22)-C pattern was observed in all members a SC-(11)-C-(22)-C pattern (where S is serine) could be observed in *M. tuberculosis, M. bovis* and *M. marinum* (Table 4.2). *M. bovis* BCG1427c shares a 95% homology with MtRsfA. *M. tuberculosis* Rv2638 and *M. bovis* BCG2665 share a unique pattern of CC-(11)-C-(22)-G and like their CC-(11)-C-(22)-C counterparts are highly homologous. The specific patterns of sequence present in these anti-anti sigma factors may reflect the underlying varied roles of the proteins. The SC-(11)-C-(22)-C pattern can be either for functioning by the phosphorylation/dephosphorylation mechanism as well by forming a Cys74-Cys109 bridge in these proteins. Likewise the *M. tuberculosis* and *M. bovis* proteins with a CC-(11)-C-(22)-G pattern may be suited to perform in altered pathways in response to different environmental conditions.

Table 4.3 Cysteine and serine containing anti-anti sigma factors from various mycobaterial sources. Position 73 cysteine is replaced by serine in some members.

Organism and protein name	Cysteine Position wrt *M .tb* RsfA			
	73	74	86	109
***M. tuberculosis* RsfA**	C	C	C	C
***M. tuberculosis* Rv 2638**	C	C	C	G
***M. tuberculosis* Rv 1904**	S	C	C	C
M.bovis (BCG1427c)	C	C	C	C
M.bovis (BCG2665)	C	C	C	G
M.bovis (BCG1943)	S	C	C	C
M.marinum (MMAR_2061)	C	C	C	C
M.marinum (MMAR_2800)	S	C	C	C
M.ulcerans (MUL_3282)	C	C	C	C
M.ulcerans (MUL_2925)	S	C	C	C
M.avium (MA 2742)	C	C	C	C

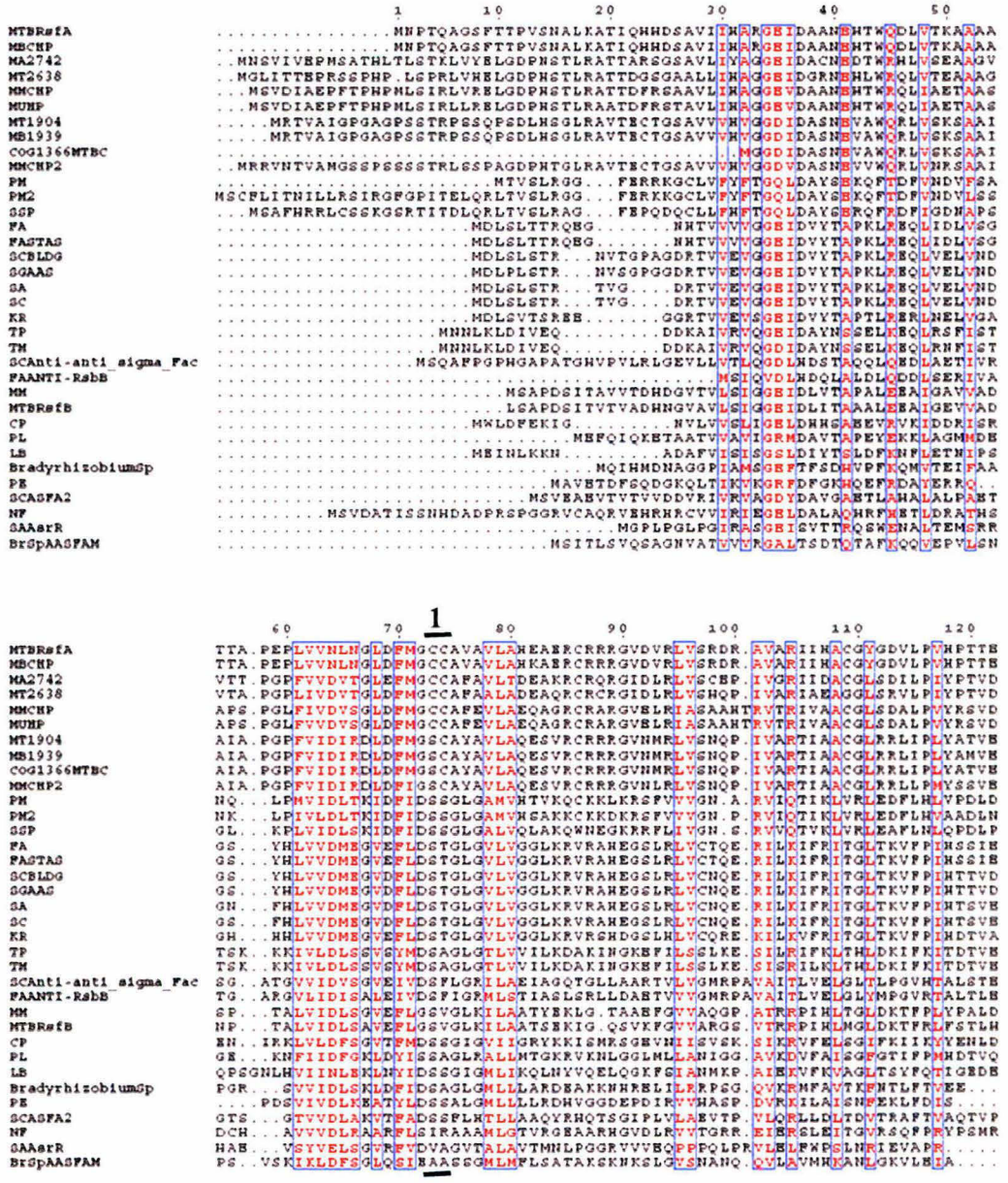


Fig 4.8 Multiple sequence alignment of *Mycobacterium tuberculosis* RsfA with anti-anti-sigma factor members from various microbial sources.

Full form for the codes used in Multiple sequence alignment on previous page.

MTBRsfA (*Mycobacterium tuberculosis* H37Rv: RsfA), **SAAsrR** (*Streptomyces ambofacien* :ArsR), **SCAnti-anti sigma Fac** (*Streptomyces coelicolor*: Anti-sigma factor antagonist), **SCASFA2** (*Streptomyces coelicolor*: anti-sigma factor antagonist), **LB** (*Leptospira biflexa* : Putative anti-sigma factor antagonist), **FAANTI-RsbB** (*Frankia alni*: Antagonist protein rsbS), **PE** (*Pseudomonas entomophila*: putative anti-sigma F factor antagonist), **FA** (*Frankia alni* :Anti-sigma-B factor antagonist), **FASTAS** (*Frankia sp.*: anti-sigma-factor antagonist (STAS) domain protein), **COG1366MTBC** (COG1366: Anti-anti-sigma regulatory factor *Mycobacterium tuberculosis* C), **SCBLDG BldG** (*Streptomyces clavuligerus*: BldG), **SGAAS** (*Streptomyces griseus*: putative anti-sigma factor antagonist), **Bradyrhizobium Sp** (*Bradyrhizobium sp*:anti-sigma-factor antagonist), **BrSpAASFAM** (*Bradyrhizobium sp*:putative anti-sigma-factor antagonist family), **MBCHP** (*Mycobacterium bovis* :Conserved hypothetical protein), **MT1904** (*Mycobacterium tuberculosis* H37Rv:hypothetical protein Rv1904), **MB1939** (*Mycobacterium bovis* : hypothetical protein Mb1939), **MA2742**(*Mycobacterium avium* subsp. Paratuberculosis: hypothetical protein), **MT2638** (*Mycobacterium tuberculosis* H37Rv:hypothetical protein Rv2638), **MMCHP** (*Mycobacterium marinum* : conserved hypothetical protein), **MUHP** (*Mycobacterium ulcerans* : hypothetical protein), **SA** (*Streptomyces avermitilis* : anti-sigma factor antagonist), **SC** (*Streptomyces coelicolor* :anti-sigma factor antagonist), **MMCHP2** (*Mycobacterium marinum* : conserved hypothetical protein), **PM** (*Prochlorococcus marinus* : anti-anti-sigma regulatory factor), **TP** (*Thermotoga petrophila* : anti-sigma-factor antagonist), **SSP** (*Synechococcus sp*: anti-anti-sigma regulatory factor),**NF** (*Nocardia farcinica* :putative anti-anti-sigma factor), **MM** (*Mycobacterium marinum* : anti-anti-sigma regulatory factor), **TM** (*Thermotoga maritime* : anti-sigma factor antagonist, putative), **KR** (*Kineococcus radiotolerans* : anti-sigma-factor antagonist), **PL** (*Pelodictyon luteolum* : anti-anti-sigma regulatory factor, SpoIIAA) **PM2** (*Prochlorococcus marinus*:anti-anti-sigma regulatory factor), **CP** (*Clostridium perfringens* :anti-sigma F factor antagonist), **MTBRsfB** (*Mycobacterium tuberculosis* H37Rv: RsfB).

```
MT RsfA          --------------MNPTQAGSFTTPVSNALKATIQHHDSAVIIHARGEI
MB BCG1427       --------------MNPTQAGSFTTPVSNALKATIQHHDSAVIIHARGEI
MT Rv2638        -MGLITTEPRSSPHPLSPRLVHELGDPHSTLRATTDGSGAALLIHAGGEI
MB BCG2665       -MGLITTEPRSSPHPLSPRLVHELGDPHSTLRATTDGSGAALLIHAGGEI
MU MUL3282       -MSVDIAEPFTPHPMLSIRLLRELGDPHSTLRAATDFRSTAVLIHAGGEV
MM MMAR2061      -MSVDIAEPFTPHPMLSIRLVRELGDPHSTLRATTDFRSAAVLIHAGGEV
MU MUL2925       ---MNTVAMGSSPSSSSTRLSSPAGDPHTGLRAVTECTGSAVVVHVGGDV
MM MAR2800       MRRVNTVAMGSSPSSSSTRLSSPAGDPHTGLRAVTECTGSAVVVHVGGDV
MB BCG1943       ---MRTVAIGPGAGPSSTRPSSQPSDLHSGLRAVTECTGSAVVVHVGGDI
MT Rv1904        ---MRTVAIGPGAGPSSTRPSSQPSDLHSGLRAVTECTGSAVVVHVGGDI
MB BCG3746       --------------------------------------------------
MT RsfB          -------------------------LSAPDSITVTVADHNGVAVLSIGGEI
MM MMAR5181      -------------------------MSAPDSITAVVTDHDGVTVLSIGGEI
MU MUL4427       ----------------------MNATAKSPNALAIDTRSEDSLVVLSVEGAI

                                                    1          2
MT RsfA          DAANEHTWQDLVTKAAAATTAPEPLVVNLNGLDFMGCCAVAVLAHEAERC
MB BCG1427c      DAANEHTWQDLVTKAAAATTAPEPLVVNLNGLDFMGCCAVAVLAHKAERC
MT Rv2638        DGRNEHLWRQLVTEAAAGVTAPGPLIVDVTGLDFMGCCAFAALADEAQRC
MB BCG2665       DGRNEHLWRQLVTEAAAGVTAPGPLIVDVTGLDFMGCCAFAALADEAQRC
MU MUL3282       DAANEHTWRQLIAETAASAPSPGLFVVDVSGLDFMGCCAFEVLAEQAGRC
MM MMAR2061      DAANEHTWRQLIAETAASAPSPGLFIVDVSGLDFMGCCAFEVLAEQAGRC
MU MUL2925       DASNEVVWQRLVNRSAAIAIAPGPFVIDIRDLDFIGSCAYAVLAQESVRC
MM MMAR2800      DASNEVVWQRLVNRSAAIAIAPGPFVIDIRDLDFIGSCAYAVLAQESVRC
MB BCG1943       DASNEVAWQRLVSKSAAIAIAPGPFVIDIRDLDFMGSCAYAVLAQESVRC
MT Rv1904        DASNEVAWQRLVSKSAAIAIAPGPFVIDIRDLDFMGSCAYAVLAQESVRC
MB BCG3746c      ---------------MADN--PTALVIDLSAVEFLG-SVGLKILAATSEK
MT RsfB          DLITAAALEEAIGEVVADN--PTALVIDLSAVEFLG-SVGLKILAATSEK
MM MAR5181       DLVTAPALEEAIGAVVADS--PTALVIDLSGVEFLG-SVGLKILAATYEK
MU MUL4427       DSTNCAALRDAIIKATLDE--PSAVVVNVSALQVPDEASWSIFVSARWQV
                                      *  .:::: ::. . .    :     .

                                                3
MT RsfA          RRRGVDVRLVSRDR-AVARIIHACGYGDVLPVHPTTESALSAT-------
MB BCG1427c      RRRGVDVRLVSRDR-AVARIIHACGYGDVLPVHPTTESALSAT-------
MT Rv2638        RCRGIDLRLVSHQP-IVARIAEAGGLSRVLPIYPTVDTALGKGTAGPARC
MB BCG2665       RCRGIDLRLVSHQP-IVARIAEAGGLSRVLPIYPTVDTALGKGTAGPARC
MU MUL3282       RARGVELRIASAAHTRVTRIVAACGLSDALPVYRSVDAALALPAPSD---
MM MMAR2061      RARGVELRIASAAHTRVTRIVAACGLSDALPVYRSVDAALALPAPSD---
MU MUL2925       RRRGVNLRLVSNQP-IVARTIAACGLRRLLPMYSSVEAALSPVPNGH---
MM MAR2800       RRRGVNLRLVSNQP-IVARTIAACGLRRLLPMYSSVEAALSPVPNGH---
MB BCG1943       RRRGVNMRLVSNQP-IVARTIAACGLRRLIPLYAMVETALAPPPSAH---
MT Rv1904        RRRGVNMRLVSNQP-IVARTIAACGLRRLIPLYATVETALAPPPSAH---
MBBCG3746c       IGQSVKFGVVARGS-VTRRPIHLMGLDKTFRLFSTLHDALTGVRGGRIDR
MT RsfB          IGQSVKFGVVARGS-VTRRPIHLMGLDKTFRLFSTLHDALTGVRGGRIDR
MM MMAR518       LGTAAEFGVVAQGP-ATRRPIHLTGLDKTFPLYPALDDALTGVRDGKLNR
MU MUL4427       DTPQHVPILLVCASRAGRELITRTGVTRFMPVYPTEKRAIKAVGRLARRK
                               :        .  *      : :. . *:
```

Fig. 4.9: Multiple sequence alignment of anti-anti-sigma factor members from various mycobacterial sources.
MT: *Mycobacterium tuberculosis,* **MB:** *Mycobacterium bovis,* **MU:***Mycobacterium ulcerans* and **MM:** *Mycobacterium marinum.*

Phylogenetic analysis: The phylogenetic analysis involving the retrieved anti-anti-sigma factor sequences demonstrated that the mycobacterial proteins occur in segregated clusters with good bootstrap values. Proteins with CC-(11)-C-(22)-C and CC-(11)-C-(22)-G patterns formed one cluster whereas anti-anti-sigma factors with a SC-(11)-C-(22)-C pattern clustered separately at a significant distance from those with CC-(11)-C-(22)-C pattern (Fig 4.10). The significant bootstrap value of 66.57% for the node from which members with CC-(11)-C-(22)-C pattern delineate shows that these proteins possess common features and suggests a common origin for them. Thus a family of anti-anti sigma factors with discreet differences in the sequences from other known anti-anti sigma factors has diverged to perform a function in mycobacterial system not reported in other microbial systems. Furthermore, based on the work reported earlier and here, this set of proteins may be involved in transcription regulation under stress conditions. They might have evolved to perform specialized roles under the diverse environmental conditions faced by the pathogen during the course of the infection.

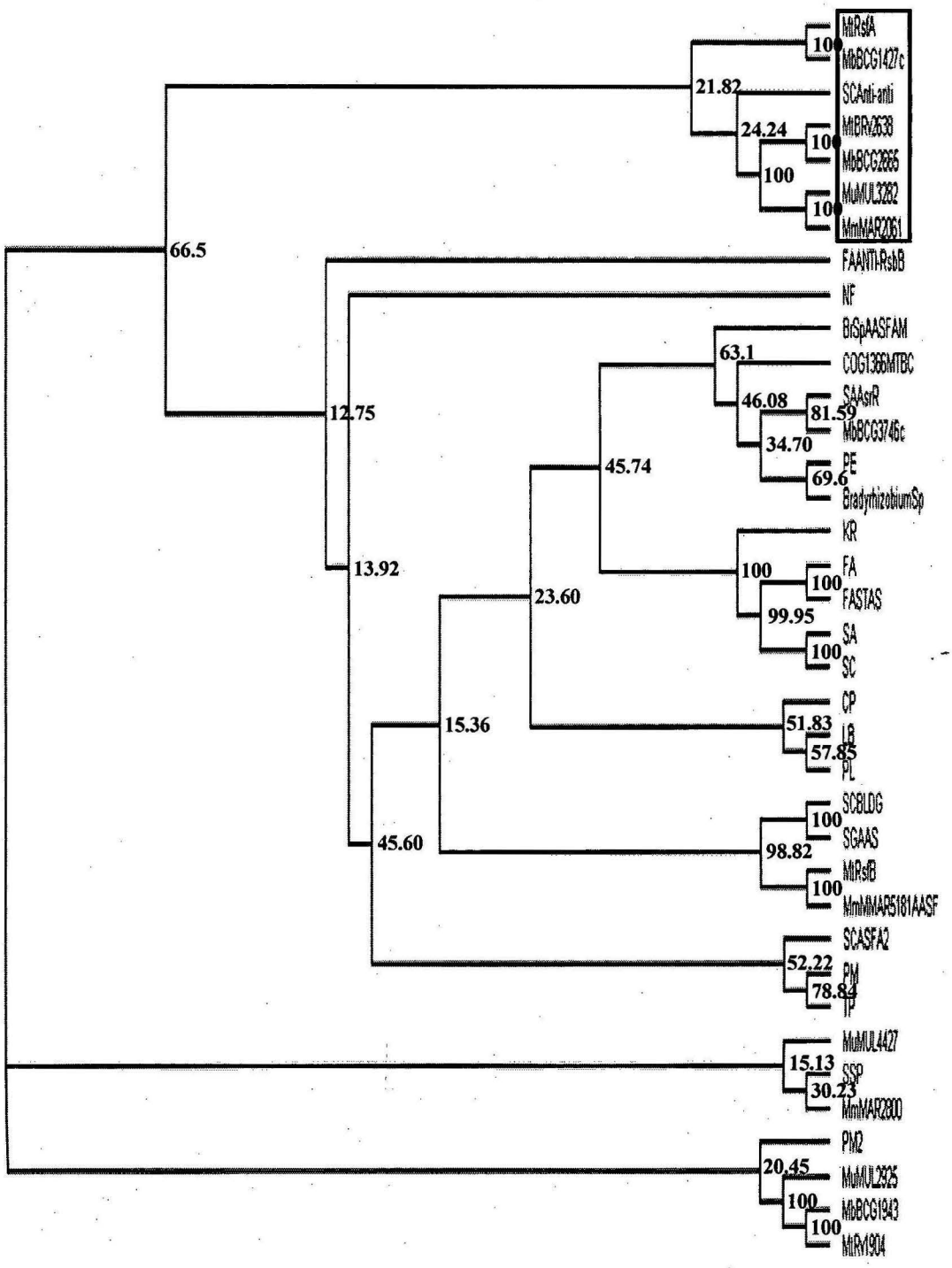


Fig 4.10: Phylogenetic analysis of anti-anti-sigma-factors from microbial sources.

Sequence co-variation in MtRsfA homologs: Proteins performing similar functions usually possess similar structural features like topologies and folds. The sequence perturbations in the primary sequence in one part of the structure mostly leads to changes in the other parts to accommodate/compensate for these changes. The co-variation in the residues of the protein sequence allows for the maintenance of the overall structural integrity (Campbell *et al.*, 2005). In the present instance, we calculated the positional entropies in the aligned sequences of anti-anti sigma factors from the whole set retrieved from the database searches and for those sequences retrieved from mycobacterial sources (Fig 4.11). The Positional entropy or informational entropy gives estimates about scores of multiple sequence alignment where the calculated values obtained from the method employed (in our case *The Scorecons Server*) are normalized for the Shannon's entropy so that conserved sequences *i.e*, those having low entropy score 1 and divergent sequences *i.e*, a high entropy score 0. The sequence numbers describing Cys containing residues are with respect to RsfA while those for Ser containing residues are with respect to RsfB. The analysis was carried out using the following criteria: (i) Overall positional entropy of 1 *i.e.*, a common residue in all samples. (ii) Positional entropy of 1 with respect to cysteine or serine containing proteins but overall entropy less than 1 *i.e.*, representative of the groups at that particular position or a primary co-variant signifying that the variation has come along with the change in the primary active residue. (iii) Positional entropy of 1 with respect to one group but less than 1 for the other group at that position *i.e.*, a secondary co-variant for that particular group. Fig 4.12A shows the positional entropy for the whole analyzed set while Fig 4.12B shows the analysis for mycobacterial sequences only. The group 1 describes the conserved residue in all the anti-anti sigma factor sequences selected for analysis. The group 2 fulfills the second criteria discussed above which can be described as the signature residues for their groups. Cys73 of MtRsfA was found to have replaced Ser61 of RsfB; similarly Ile51 of *M. tuberculosis* RsfB has been replaced by Val63. The positional entropy at these positions within the group is 1 *i.e.*, the position is highly conserved over the whole sequence set. The same entropy values at Ile/Val position points that the change in the primary sequence at this position is a direct consequence of introduction of Cys73 and these residues may be of primary importance for maintaining the overall stability of the structure. Group 3 describes the co-variations that have been brought by the introduction of cysteines in the sequence. The positions

do not follow a conservation pattern in serine containing anti-anti sigma factors but the positional entropy values of 1 at these positions in cysteine containing anti-sigma antagonists points towards the conserved nature of the residues at these positions. The formation of a disulphide bond between Cys73 and Cys109 must be introducing rigidity into the protein with additional structural constraints in the form of volumetric changes, charge imbalances etc. introduced into the protein. To compensate for these structural perturbations, the residues must have co-varied along with cysteines over the whole length of the protein to accommodate these changes in the proteins. The co-variation of the residues can also be for allowing the protein to perform specific function. One such example is Cys109 position which doesn't show conservation in serine containing antagonists but has to be essentially cysteine in redox sensor anti-anti sigma as it is the key to the formation of disulphide bond. Overall, the evolution of this class of anti-anti sigma factors has been accompanied by changes in the primary sequences which may be important for their functions under stress conditions and at the same time they help maintain the basic structural features of the anti-anti sigma factor proteins.

```
MT RsfA              ---------------MNPTQAGSFTTPVSNALKATIQHHDSAVIIHARGE
MB BCG 1427          ---------------MNPTQAGSFTTPVSNALKATIQHHDSAVIIHARGE
MU MUL 3282          -MSVDIAEPFTPHPMLSIRLLRELGDPHS-TLRAATDFRSTAVLIHAGGE
MM MMAR 2061         -MSVDIAEPFTPHPMLSIRLVRELGDPHS-TLRATTDFRSAAVLIHAGGE
MT Rv2638            MGLITTEPRSSPHP-LSPRLVHELGDPHS-TLRATTDGSGAALLIHAGGE
MB BCG2665           MGLITTEPRSSPHP-LSPRLVHELGDPHS-TLRATTDGSGAALLIHAGGE
MAP2742              MNSVIVEPMSATHLTLSTKLVYELGDPNS-TLRATTARSGSAVLIYAGGE
                                :.     .:  * * :*:*:    .:*::*:* **

                                                        1
MT RsfA              IDAANEHTWQDLVTKAAAATTAPEPLVVNLNGLDFMGCCAVAVLAHEAER
MB BCG1427c          IDAANEHTWQDLVTKAAAATTAPEPLVVNLNGLDFMGCCAVAVLAHKAER
MU MUL3282           VDAANEHTWRQLIAETAASAPSPGLFVVDVSGLDFMGCCAFEVLAEQAGR
MM MMAR2061          VDAANEHTWRQLIAETAASAPSPGLFIVDVSGLDFMGCCAFEVLAEQAGR
MT Rv2638            IDGRNEHLWRQLVTEAAAGVTAPGPLIVDVTGLDFMGCCAFAALADEAQR
MB BCG2665           IDGRNEHLWRQLVTEAAAGVTAPGPLIVDVTGLDFMGCCAFAALADEAQR
MAP2742              IDACNEDTWRHLVSEAAGVVTTPGPFVVDVTGLEFMGCCAFAVLTDEAKR
                     :*. **. *:.*::::*. ..:*  ::*::.**:******. .*:.:* *

                     2                  3
MT RsfA              CRRRGVDVRLVSRDR-AVARIIHACGYGDVLPVHPTTESALSAT-------
MB BCG1427c          CRRRGVDVRLVSRDR-AVARIIHACGYGDVLPVHPTTESALSAT-------
MU MUL3282           CRARGVELRIASAAHTRVTRIVAACGLSDALPVYRSVDAALALPAPSD---
MM MMAR2061          CRARGVELRIASAAHTRVTRIVAACGLSDALPVYRSVDAALALPAPSD---
MT Rv2638            CRCRGIDLRLVSHQP-IVARIAEAGGLSRVLPIYPTVDTALGKGTAGPARC
MB BCG2665           CRCRGIDLRLVSHQP-IVARIAEAGGLSRVLPIYPTVDTALGKGTAGPARC
MAP2742              CRQRGIDLRLVSCEP-IVGRIIDACGLSDILPIYPTVDSALSGADRW----
                     ** **:::*:.*     * **  * * .  **:: :.::**.
```

Fig 4.11: Multiple sequence alignment of *Mycobaterium tuberculosis* RsfA with putative redox sensor anti-anti sigma factors of other mycobacteria. **MT:** *Mycobacterium tuberculosis,* **MB:** *Mycobacterium bovis,* **MU:** *Mycobacterium ulcerans* and **MM:** *Mycobacterium marinum,* **MA:** (*Mycobacterium avium* subsp. Paratuberculosis: hypothetical protein).

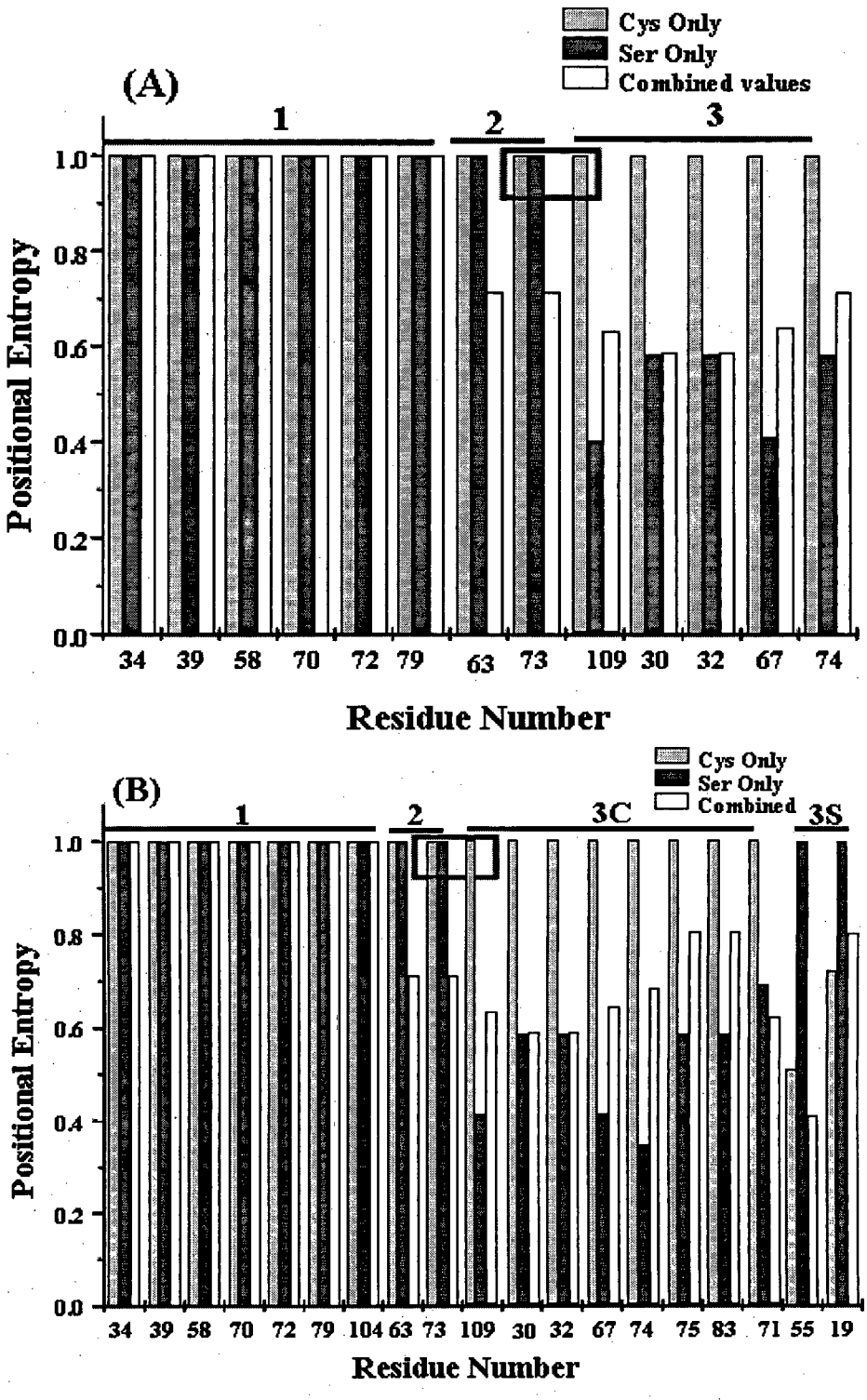


Fig 4.12: Positional entropies at the potentially conserved positions in cysteine and serine containing anti-anti sigma factors.

4.2.2.2 Cloning, over expression and purification of MtRsfA

The full-length *rsfA* gene 387 base pair in length was PCR amplified from genomic DNA of *M. tuberculosis*H37Rv and cloned into pET-21d at NcoI and HindIII to acquire a C-terminal 6x Histidine tag. The clones were confirmed through restriction digestion (Fig 4.13A) and the integrity of clone was confirmed by sequencing.

The expression plasmids were introduced into *E.coli* host strains like BL21 (DE3), Origami(DE3) and C41(DE3) that contain T7 RNA polymerase gene under the control of *lac*UV5 promoter. MtRsfA expressed both in Origami (DE3) and C41 (DE3) host cells. The conditions for growth were optimized at 37 °C by adding 0.3mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) to an OD_{600} of ~0.6 and growing for further twelve hours. The protein was insoluble in other cell variants and was soluble when expressed in the Origami(DE3) strain. The latter strain facilitates the formation of disulphide bonds in recombinant proteins. RsfA has a disulphide bond between Cys73-Cys109. In the Coomassie Brilliant Blue stained gel the over expressed protein band was evident with an apparent molecular mass of slightly less than 14kDa; this molecular mass compares with that of 13.5 kDa obtained from the amino acid sequence.

The His-tag present on the C-terminus of the expressed RsfA facilitated one step affinity purification from a NaCl Ni^{2+}-IDA column. The protein was purified using Imidazole gradient (10mM-500mM) in a buffer containing 50mM Tris-Cl, pH 8.5 and 50mM NaCl. The protein was highly pure with minor trace contaminants as observed on the SDS-PAGE (Fig 14.3B). The final purification step was performed by precipitating the protein with 60% ammonium sulphate and was dissolved in minimum volume of gel filtration running buffer (50mM Tris-Cl, pH8.5, 50mM NaCl and 5mM EDTA) utilizing the gel filtration chromatography column Superdex-75.The protein eluted at a volume of 13.07 ml in the Superdex-75 column which had been calibrated with proteins of known molecular weight. The elution volume of MtRsfA corresponds to a molecular weight of about 16kDa which is near to the molecular weight of 13.5kDa of MtRsfA. Thus it can be concluded that MtRsfA exists as a monomer in solution with an intra-chain disulphide linkage (Fig. 4.13C).

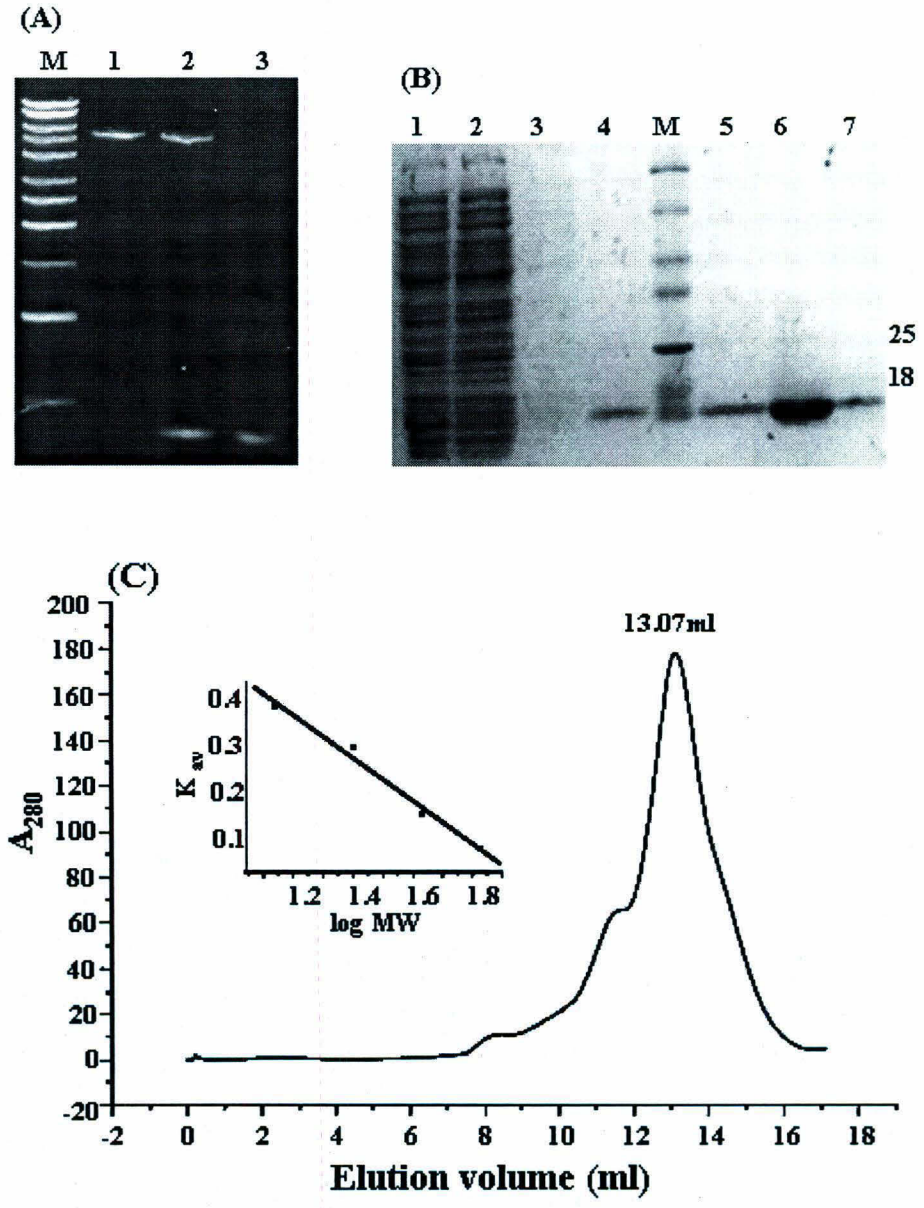


Fig 4.13: (A) Cloning of *M. tuberculosis* RsfA in pET21d.
LaneM: 1Kilobase DNA marker, Lane 1: NcoInd HindIII digested pET21d, Lane2: 387bp *rsfA* PCR amplified fragment digested with same enzymes, Lane3: Restriction digestion of *rsfA* clone in pET21d.
(B) Purification of MtRsfA. Lane1:Load, Lane2:Flowthrough, Lane3:Wash1, LaneM: Low Molecular Weight Protein Marker, Lane4:Wash2, Lanes5,6 and 7: Purified fractions From Ni^{2+}-IDA column.
(C) Size-exclusion chromatogram. The elution volume of 13.07ml points towards a molecular weight of ~ 16kDa for MtRsfA.

4.2.2.3 Effect of DTT on conformation of MtRsfA

The importance of Cys73-Cys109 disulphide bond in the RsfA function as anti-anti sigma factor has been earlier evaluated by mutational analysis (Beaucher *et al.*, 2002). MtRsfA binds to MtUsfX only when the disulphide bond is reduced. However, the actual significance of the disulphide bond in regulating its interactions with the cognate anti-sigma factor MtUsfX is not known. We utilized size exclusion chromatography to monitor conformational changes, if any, as a consequence of the reduction of the disulphide bridge. The gel-filtration elution profiles of MtRsfA from Superdex S-75 column were monitored under native and reduced conditions. The breaking of the Cys73-Cys109 disulphide bond under reducing conditions was achieved by incubating the protein with 5mM DTT. The loss of an intra-chain bond can be expected to bring small conformational or gross structural changes in the protein. Size exclusion chromatography has been used to calculate the Stokes radii of proteins in altered conformational states by following changes in the hydrodynamic properties (Ackers *et al.*, 1970, Kenneth *et al.*, 1995, Renu *et al.*, 2006). The reduced form of MtRsfA eluted earlier than its native form in the size exclusion chromatography column. Fig 4.14A represents gel filtration profiles of native and reduced MtRsfA. Elution volume of the native protein corresponds to a Stokes radius (R_S) of 18.5Å while that of the disulphide bond reduced form is 22.1Å (Fig 4.14B). Reduction of the disulphide bond might result in increased flexibility of the protein and allow for the enhanced hydrodynamic radius. The concomitant structural alterations may be necessary for the interactions with MtUsfX.

Circular Dichroism spectroscopy was used to detect any changes that may accompany the changes in the hydrodynamic radius of the protein. CD spectroscopy is used to identify secondary structural elements in proteins and the changes in these elements due to molecular interactions. MtRsfA is predicted to be a α/β- protein with a slightly higher helical content (31%) compared to the residues in the β-sheet regions (23%).The secondary structure predictions were obtained using the JPRED2 package and 3DPSSM Fold prediction server. Analysis of CD spectra with the K2D software provided estimates of the secondary structure content for RsfA (39% helix, 31% sheet).Similar results were obtained with the spectral data of reduced MtRsfA i.e., although reduction of the disulphide bond caused an increase in the radius of the protein, it didn't produce any gross secondary structural changes (Fig 4.14C). Thus

the reduction of disulphide bond may be essential for increasing the accessibility of certain residues that are important for the UsfX-RsfA interaction.

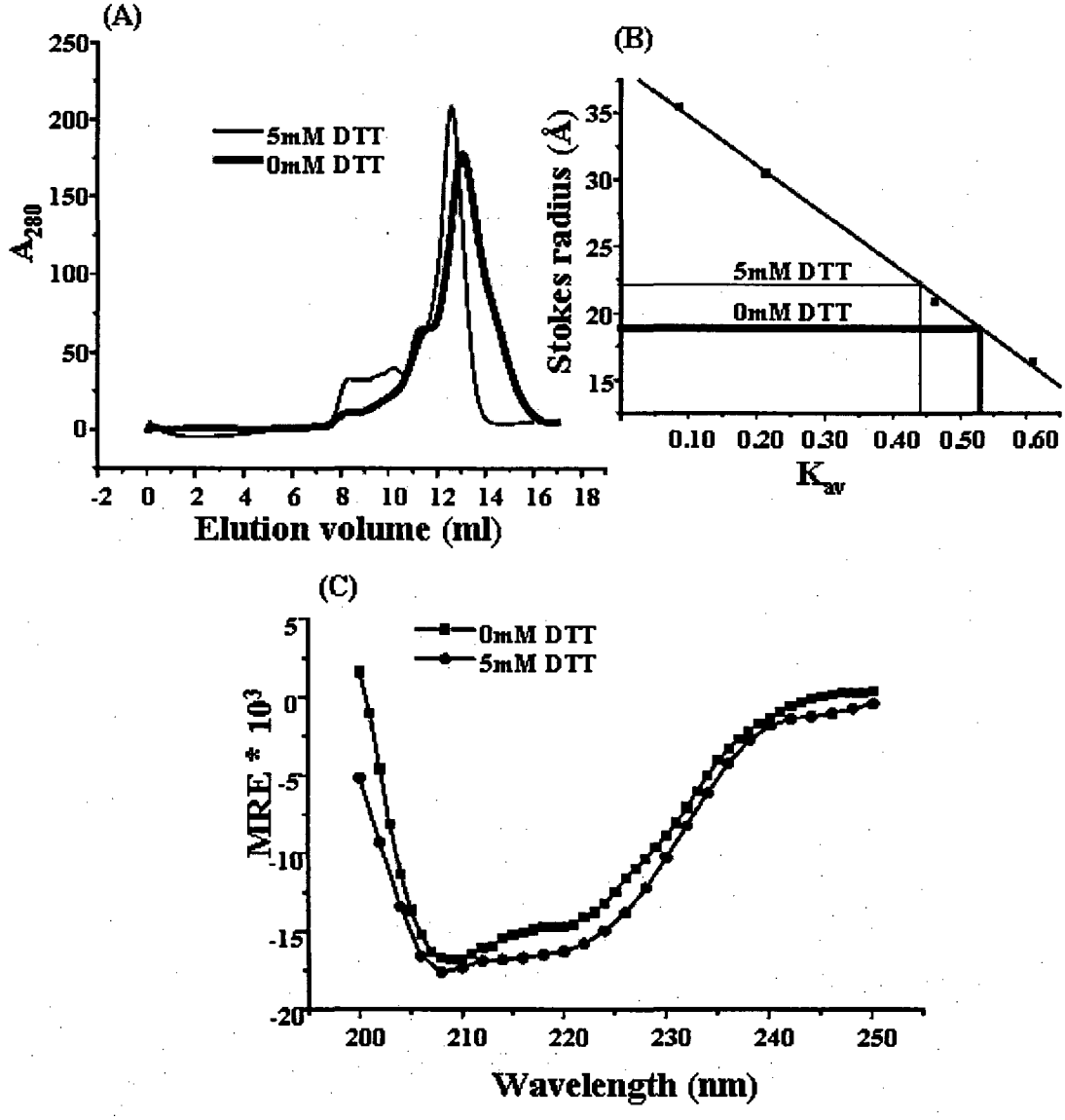


Fig 4.14: Change in structural properties of MtRsfA due to DTT.
(A) Gel filtration elution profile of MtRsfA in presence and absence of DTT.
(B) The Stokes radii of standard proteins were plotted against a function of their elution volume (K_{av}) on a Superdex 75 Column. The elution volumes of native and reduced (DTT treated) MtRsfA suggest Stokes radii of 18.5 and 22.1 Å respectively.
(C) Far UV circular dichroism spectra acquired from solutions of MtRsfA and MtRsfA in presence of 5mM DTT.

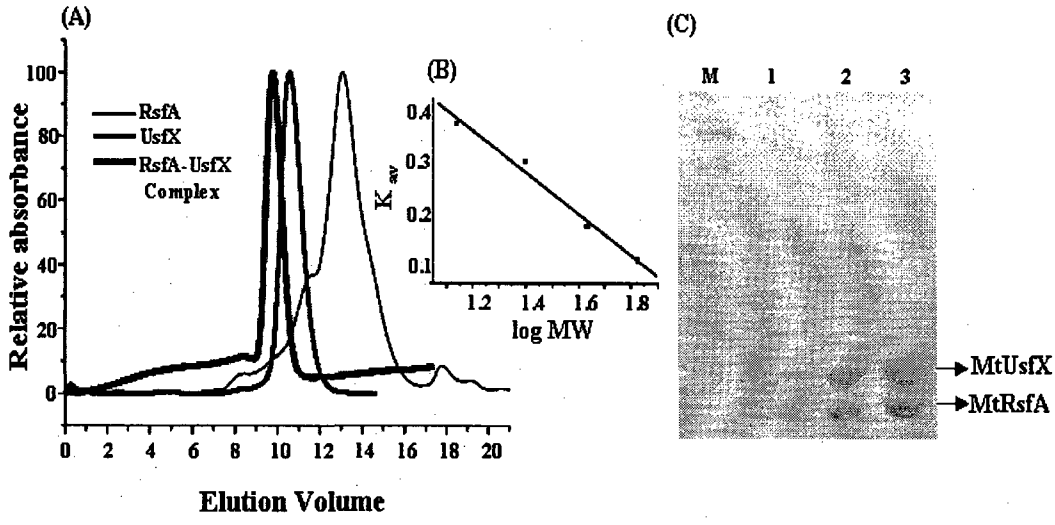
4.2.2.4 *In vitro* interaction between MtUsfX and MtRsfA

The formation of the UsfX-RsfA complex could be detected based on the elution profiles from size-exclusion chromatography experiments. MtRsfA and MtUsfX elute with volumes which correspond to molecular weights of 16kDa and 33kDa respectively and corresponds to quaternary associations of a monomer and dimer for the individual proteins. The UsfX-RsfA complex should elute as a higher molecular weight entity compared to the individual proteins. Fig 4.15A shows the size exclusion chromatogram of MtRsfA, MtUsfX and UsfX-RsfA complex. Clearly the two proteins have eluted as a complex and the interaction is stable in the solution. All the protein-protein interactions were conducted in reducing environment in presence of 5mM DTT and the size exclusion running buffer was also supplemented with 5mM DTT. In the absence of DTT there was a change in the elution pattern with the proteins eluting as a mixture from the column. The reduction of the disulphide bond (in the presence of DTT) apparently results in favorable conformation changes in MtRsfA to allow for the formation of a stable UsfX-RsfA complex. In the *in vivo* environment, in the event of a redox potential change, this would free the SigF protein for transcription.

4.2.2.5 Stoichiometry analysis

In physiological conditions, protein-protein interactions occur in a specific stoichiometric ratio inside the cells and these ratios have mechanistic implications related to the functioning of the interacting proteins inside the cell. In order to investigate the stoichiometry of the UsfX-RsfA complex we examined the results of gel filtration analysis The protein complex eluted at a volume of 9.7ml which corresponds to a molecular weight of about 52kDa.Using the molecular mass of MtRsfA (13.5kDa) and MtUsfX (15.55kDa) along with our early finding that MtUsfX exists as a dimer, it was concluded from the gel filtration analysis that UsfX: RsfA complex exists in a stoichiometric ratio of 2:1 with a molecular weight of 44.5kDa. Densitometry analysis also yielded a stoichiometry of UsfX:RsfA (1.6:1) Fig 4.15. The molecular mass of MtUsfX is almost equal to that of MtRsfA, while the intensity of the MtUsfX bands is almost twice than that of MtRsfA thus showing that MtUsfX is twice as that of MtRsfA. The partner-switching interactions involving sigma, anti-sigma and anti-anti sigma factors are governed by certain cellular signals and the nature of the signal is going to determine which of the interactions will prevail at that

time; sigma-anti sigma factor or an anti-anti sigma-anti-sigma factor interaction. Switching over from one partner to other involves various structural rearrangements. The binding stoichiometries determine these structural changes in particular steric clashes when the protein switches over from one partner to other. A stoichiometry of 2:2 has been described for anti-sigma and anti-anti sigma factor interaction (Shoko *et al.*, 2002) where one anti-anti sigma factor molecule dislodges the bound sigma factor from its cognate anti-sigma factor in an induced release mechanism and then another anti-anti sigma factor molecule binds to the assembled complex (Margeret *et al.*, 2003). There can be two reasons for the unusual stoichiometry. Either the complex that has eluted from the gel filtration column is an intermediate of induced release step or it may be the actual ratio present *in vivo* and may have a physiological relevance in terms of the redox potentials that are existent inside the cells during the formation of the complex. We are more convinced about 1:2 being the actual stoichiometry because this was existent even at higher concentrations of the reducing agent. Also the phosphorylation and dephosphorylation involved in serine containing factors which has been found to facilitate the binding of the second anti-anti-sigma factor molecule is not involved here.



(D)

Protein	Monomeric mass	Sample	Observed Intensity	Observed Intensity/mass	Calculated stoichiometry
MtRsfA	13.6kDa	Lane2 Lane3	15.7 16	1.16	
MtUsfX	15.6kDa	Lane2 Lane3	27.7 30	1.85	
Complex					**1/1.60**

Fig 4.15: Gel filtration and densitometry analysis of UsfX-RsfA complex.

(A) The UsfX-RsfA complex eluted at a volume of 9.7ml in Superdex-75 Column which corresponds to a molecular weight of about 52kDa.

(B) Standard molecular weight markers were used for calibrating the column.

(C) SDS-PAGE analysis of the peak fractions from gel filtration.

(D) Calculated stochiometry of RsfA:UsfX based on SDS-PAGE and densitometric scanning.

4.2.2.6 Structural studies on UsfX-RsfA interaction

Circular Dichroism spectroscopy was used to monitor changes in the secondary structure that may have occurred in the proteins due to complex formation. Fig 4.16A shows the CD spectra obtained for MtUsfX, MtRsfA and the UsfX-RsfA complex. No gross secondary structural changes were detected in either of the proteins and the CD spectra of the two proteins and UsfX-RsfA complex appear to be similar. Also no changes in MtRsfA structure (Section 4.3.3) were observed on reduction of the disulphide bond, presumably a prerequisite to complex formation so it can be concluded that the UsfX-RsfA interaction involves orientation of interacting residues without changing the secondary structure of the proteins. The thermal denaturation analysis using the two state model showed a higher thermal stability of the UsfX-RsfA complex as compared to RsfA (Fig 4.16B). This increased resistance to high temperatures can be attributed to the complex formation with MtUsfX. MtUsfX in interaction with its cognate sigma factor MtSigF was also found to be resistant to changes in secondary structure while changes in the secondary structure of MtSigF were observed there, also its interaction with MtSigF imparts greater thermal stability to UsfX-SigF complex than is found in MtSigF alone (Section 4.2.5). The conservation of secondary structure while switching partners from sigma factors to anti-anti sigma factors may be a way of maintaining the interacting surfaces for diverse proteins in the anti-sigma factors.

We also utilized fluorescence spectroscopy to monitor conformational changes in proteins due to complex formation. MtRsfA is characterized by a fluorescence maximum at about 340nm. The emission is blue shifted relative to that of free L-tryptophan which under the same conditions is observed at 354nm. As buried tryptophan residues in the folded protein show the fluorescence emission λmax at 330-340 nm, with an increase in the fluorescence yield, it can be concluded that the tryptophan molecule in MtRsfA is buried in the hydrophobic core of protein and not significantly exposed to the solvent. The fluorescence maximum observed for the UsfX-RsfA complex at 343nm is slightly red-shifted (by 3nm) when compared to that for MtRsfA and blue shifted by approximately 7nm when compared to MtUsfX which is characterized by a fluorescence maximum at ~349nm. Further the spectrum of the complex is characterized by an increase in the band width as compared to the spectra of the two individual proteins (Fig 4.16C). The fluorescence maxima of UsfX-RsfA complex suggests that there is a change in the environments of the tryptophans of

either one or both of proteins and from the fluorescence maxima of individual proteins it appears that it is the Trp-106 of MtUsfX which is getting shielded as a consequence of complex formation, the Trp-44 of MtRsfA is already in a hydrophobic environment so cant undergo much change. From these fluorescence spectra it can be concluded that the interaction between these two proteins although not involving changes in the structural elements but are accompanied by transitions in local environments of individual residues.

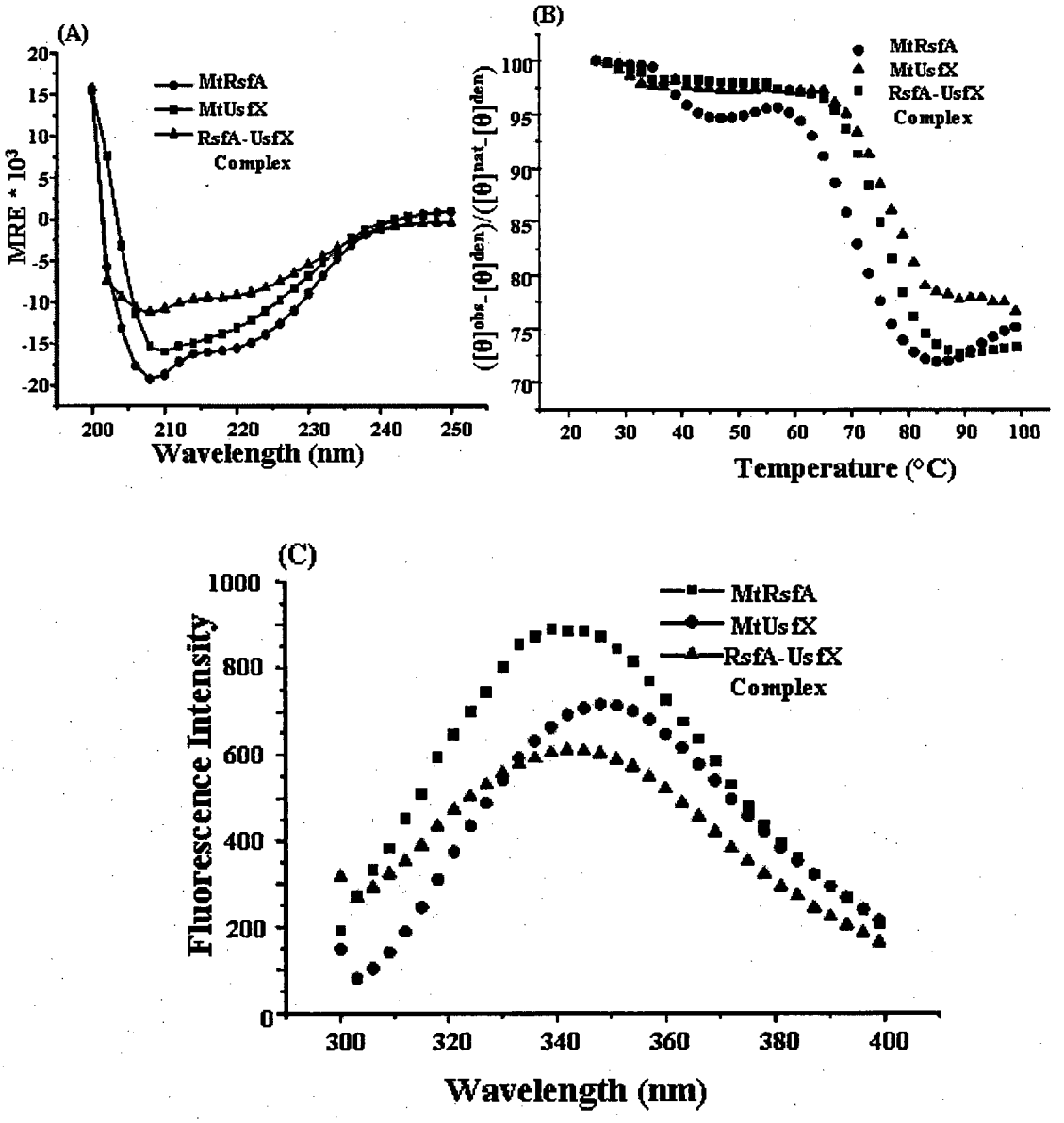


Fig 4.16: CD and fluorescence spectroscopy studies with MtUsfX, MtRsfA and UsfX-RsfA complex.

(A) Far UV circular dichroism spectra acquired from solutions of MtRsfA, MtUsfX, and the 2:1 UsfX: RsfA complex.

(B) Thermal denaturation profiles of the two proteins and the protein complex. Apparently stability is imparted in the complex from highly stable MtUsfX.

(C) Intrinsic fluorescence emission spectra obtained for MtRsfA, MtUsfX and UsfX-RsfA complex. The complex formation is accompanied by changes in the tryptophan fluorescence of the either of the proteins.

4.2.2.7 *In silico* analysis of UsfX-RsfA interaction

No success could be achieved either in the crystallization of MtRsfA or UsfX-RsfA complex. From our circular dichroism and fluorescence spectroscopy experiments we could identify changes at the amino acid conformational level as opposed to gross changes to the secondary structures of the proteins upon complex formation. We used molecular modeling and dynamics as a tool to probe the nature of the conformational changes that occur during UsfX-RsfA complex formation. The *in silico* models for the UsfX-RsfA complex were generated by comparative modeling techniques as described in section 4.2.1.5. The interacting surfaces were visualized using PYMOL and the major interacting regions were identified as the stretches encompassing residues 107-112 in MtRsfA and 100-111 in MtUsfX (Fig 4.17).Two models of the UsfX-RsfA complex were used as the starting point in the calculations. In the first one the Cys73-Cys109 bond of MtRsfA was maintained (model-I) while in the second model (model-II) the bond was broken as would be the case under reducing conditions. The latter condition as mentioned earlier is conducive for the complex formation. The molecular dynamics calculations were run on both models and the results were analyzed. The *r.m.s.d* values for the stretches involved in the UsfX-RsfA interfaces and also the stretch harboring Cys73 of MtRsfA are shown in Fig 4.18A.While there is little deviation in the region containing the Cys73 residue in both the models, significant *r.m.s.d* can be found in the 107-112 of MtRsfA and 100-111 of MtUsfX in model-II when compared to the corresponding values in model-I. A closer inspection shows that the *r.m.s.d* values for the residues in these streches can be attributed, amongst other things, to increased deviations of MtRsfA- Cys109, MtUsfX-His107 and MtUsfX-Phe104 (Table 4.3). The atomic distances between MtRsfA-Cys109 and MtUsfX-His107 were found to decrease during the course of the dynamics in model-II accompanied by a simultaneous increase in the atomic distances between MtRsfA-Cys109 and MtUsfX-Phe104. So during the rearrangement in model-II His107 of UsfX comes close to the Cys-109 of RsfA which exists as a thiol and this is accompanied by the movement of UsfX-Phe104 away from the thiol group (Fig.4.19A, B). Other increased interactions in model-II involve the MtRsfA-His81 and MtUsfX- Ala100 residue pair and also MtRsfA-Ile105 and MtUsfX-Val108 residues (Fig.4.19C). Some of these residues like Val108 and His107 of MtUsfX are conserved in mycobacteria and like pointed out earlier, the cysteines forming the

disulphide bridge in MtRsfA are also conserved. It is conceivable that the rearrangements observed here may be typical for members of the redox sensing anti-anti-sigma factor proteins from mycobacteria. The differences in the side chain rearrangements in the two models can be attributed to the local charge redistribution. In model-I the disulphide bridge is maintained and no major rearrangements of the sidechains of MtUsfX-His107 or Phe104 are observed. When the disulphide bridge breaks down with the corresponding formation of thiols (or thiolate), MtRsfA-Cys109 attracts the positively charged His107 residue closer and initiates this and other observed rearrangements in model-II.

Thus the Cys73-Cys109 disulphide bond in MtRsfA while acting as a sensor for changes in redox potential, appears to strength its interaction with MtUsfX in the reduced cellular atmosphere which is important for setting MtSigF free to drive the transcription in the altered atmosphere of the cell.

Table 4.4 *r.m.s.d*(Å) values for important residues in RsfA-UsfX interaction. The increased rmsd values in Cys73 and Cys109 are due to the loss of interaction among themselves. MtUsfX Phe104 and His107 have higher values when the disulphide bond is absent because of their proposed roles during complex formation.

Residue	Cys73- Cys109 Bond present	Cys73-Cys109 Bond absent
RsfA -Cys73	4.7	5.7
RsfA-Cys109	4.54	7.62
UsfX-Phe104	7.8	9.43
UsfX -His107	5.59	11.06
UsfX -Val108	9.1	7.27

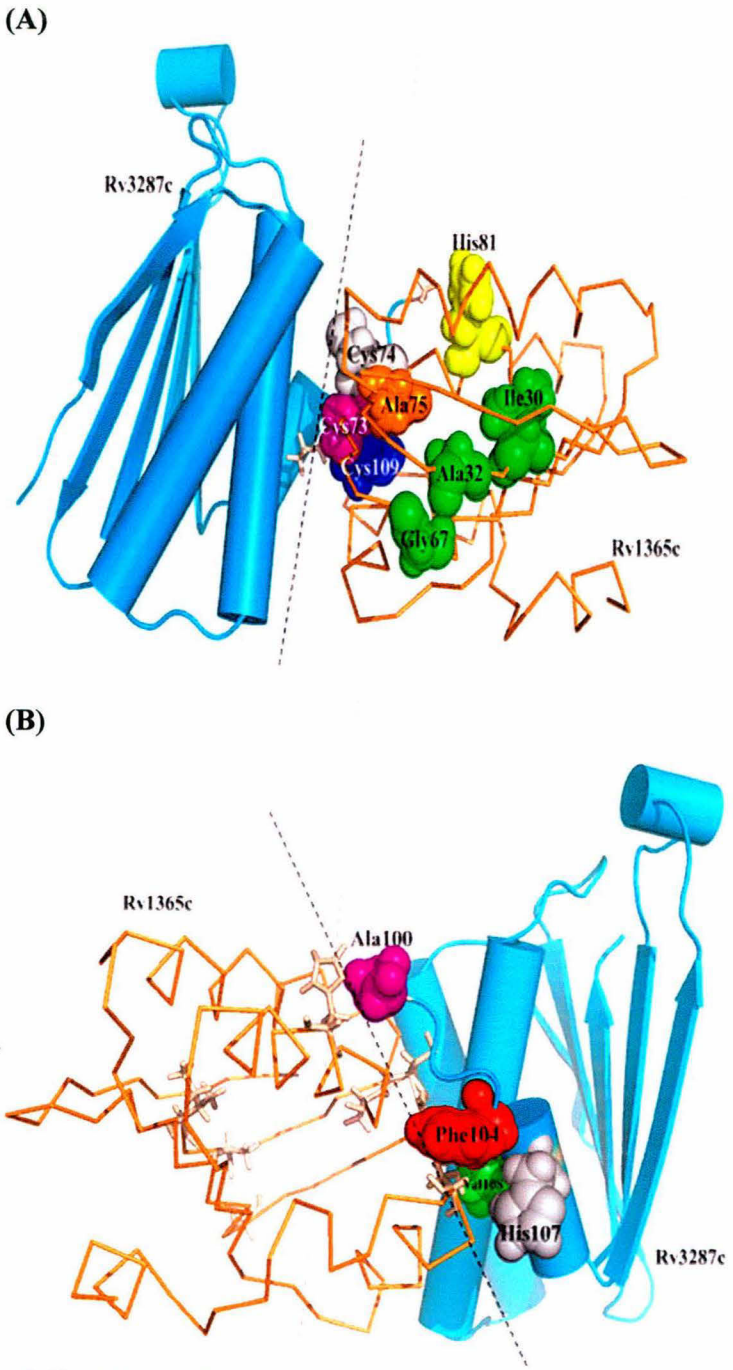


Fig 4.17: Modelled UsfX-RsfA complex.

(A) Interacting residues and the conserved residues of structural importance of MtRsfA (Rv1365c).

(B) The 180° rotated orientation of the above figure showing the interacting residues of MtUsfX (Rv3287c).

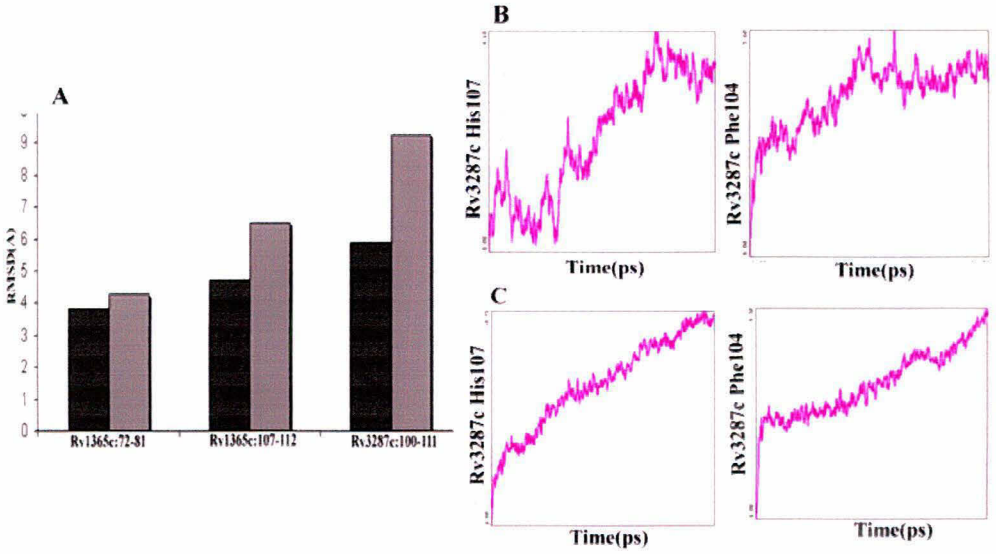


Fig 4.18: *r.m.s.d* (Å) and displacement graphs for residue stretches involved in UsfX-RsfA interaction. **(A)** Displacements in the functionally important stretches show that Cys109 of MtRsfA has a prominent role in interactions involving the region 100-111 of MtUsfX (Rv3287c). The black bars represent the case when theCys73-Cys109 bond is maintained while the grey bars represent values in the absence of the bond. **(B)** Displacement of MtUsfX His107 and Phe104 in the UsfX-RsfA complex when the Cys73-Cys109 bond is intact. **(C)** Increased displacement of MtUsfX His107 and Phe104 in the UsfX-RsfA complex in the absence of the Cys73-Cys109 bond.

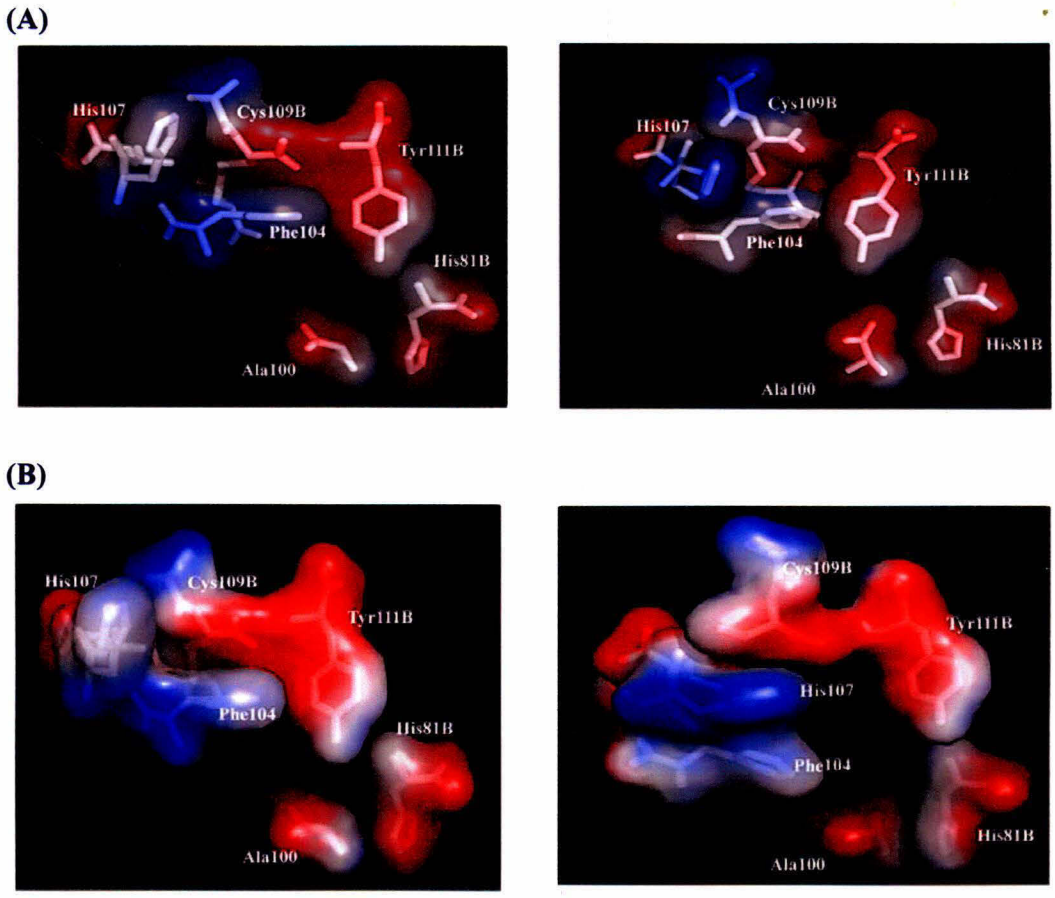


Fig 4.19: Interactions at the UsfX-RsfA interface.
(A) (*Left*): Stick representation of residues around Cys109 of MtRsfA in model-I where the Cys73-Cys109 disulphide bridge in MtRsfA is maintained. The van-der Waals surface area colored according to charge is also shown in this and in panel B. The snapshot is at the beginning of the MD simulations. (*Right*): Snapshot of the region at the end of the MD run. Clearly the residues do not exhibit large conformational changes in model-I. **(B)** (*Left*): Snapshot of the same region at the beginning of the MD simulations involving model-II where the Cys73-Cys109 disulphide bridge is broken. (*Right*): The snapshot of the region after the MD simulations. Clearly a rearrangement of the His107 and Phe014 side chains of MtUsfX takes place where the side chain of His107 comes close to Cys109 of MtRsfA. **(C)** Overlay of the UsfX-RsfA interface around Cys109 from model-I and model-II at the end of the respective MD simulations. Residues shown in blue are of model-II at the end of the MD simulations and are numbered with a * for clarity while those shown in green belong to model-I. Clearly increased interactions are observed between Val108 and Ile105 of MtUsfX and MtRsfA respectively in model-II.

4.3 Summary

MtUsfX like other anti-sigma factors was found to interact with its cognate sigma factor MtSigF and anti-anti sigma factor MtRsfA. A stable UsfX-SigF complex could be attained *in vivo* only by the co-expression of the two proteins. *In vitro* complex formation was difficult even in the presence of the nucleotides. Conclusively the interaction occurs in presence of some cellular factors and the presence of nucleotides only doesn't seem to be important for this interaction. A stoichiometry ratio of 2:1 was identified for the UsfX:SigF interaction with the interactions apparently stabilized by hydrophobic forces with a greater stability imparted to MtSigF upon complex formation. From nucleotide binding studies with UsfX-SigF complex it could be concluded that the protein interacting surface is not in the vicinity of the nucleotide binding site of MtUsfX. The structural rationalizing of the UsfX-SigF interaction with computational analysis confirmed the presence of hydrophobic residues at the interface and the distal nature of the nucleotide binding site and the protein interacting interface in MtUsfX.

Anti-sigma factor antagonists act as sensors for altering conditions inside the cell by sensing the changing chemical entities or chemical potentials inside the cells. MtRsfA from its sequence analysis was found to be a member of mycobacterium specific class of redox sensing anti-anti sigma factors which have evolved with a higher order conservation of structurally and functionally important residues. The reduction of the Cys73-Cys109 disulphide bond in MtRsfA with DTT results in an increase of the hydrodynamic radius and also leads to the formation of a stable UsfX-RsfA complex. The complex has a novel 1:2 stoichiometry compared to a 2:2 stoichiometry reported in other complexes involving anti-anti sigma factors. Circular dichroism studies ruled out gross secondary structural changes upon complex formation while fluorescence spectroscopy studies suggest smaller alterations. Using MD simulations analysis, the reduction of the disulphide bond of MtRsfA for the UsfX-RsfA interaction was found to be important to facilitate the specific conformational changes which lead to stabilization of the complex. The properties of the UsfX-RsfA interaction appear to be a common adaptation in redox sensing anti-anti-sigma factors as suggested by the conservation of important interacting residues across the sub-family of mycobacterial proteins.

REFERENCES

Ackers GL (1970). *Adv Protein Chem.* **24**, 343-466.

Agnieszka M, Yancho D, Daniel K, Kenton L, Zbigniew D, Jacek O and Zygmunt SD (2002). *Acta Cryst* **D58**, 1983-1991.

Alper S, Duncan L and Losick R (1994). *Cell* **77**,195-205.

Alper S, Dufour A, Garsin DA, Duncan L and Losick R (1996). *J Mol Biol* **260**,165–77.

Altschul SF, Thomas LM, Alejandro AS, Jinghui Z, Zheng Z, Miller W and David JL (1997). *Nucleic Acids Res* **25**, 3389-3402.

Altschul SF, Wootton C, Michael G, Agarwala JE, Morgulis RA, Schaffer AA and Yi-Kuo Y (2005). *FEBS J* **272**:5101-5109.

Barer MR and Harwood CR (1999). *Adv Microb Physiol.* **41**, 93-137.

Barne KA, Bown JA, Busby SJ and Minchin SD (1997). *EMBO J* **16**, 4034–4040.

Bashyam MD and Hasnain SE (2004). *Infect Genet Evol* **4**, 301–308.

Beaucher J, Rodrigue S, Jacques PE, Smith I, Brezinski R and Gaudreau L (2002). *Mol Microbiol* **45**,1527–40.

Betts JC, Lukey PT, Robb LC, McAdam RA and Duncan K (2002). *Mol Microbiol* **43**, 717–731.

Blaszczak A, Zylicz M, Georgopoulos C and Liberek K (1995). *EMBO J* **14**, 5085–93.

Blow DM, Chayen NE, Lloyd LF and Sardakis E (1994). *Protein Sci* **3**, 1638-43.

Bougie I, Parent A and Bisaillon M (2004). *Biochem J* **384**, 411-420.

Bradford M (1976). *Anal Biochem* **72**, 248-254.

Brody MS, Vijay K and Price CW (2001). *J Bacteriol* **183**, 6422–28.

Browning DF and Busby SJ (2004). *Nat Rev Microbiol* **2**, 57–65.

Ruck M, Gallegos MT, Studholme DJ, Guo Y and Gralla JD (2000). *J Bacteriol* **182**, 9–4136.

ski W and Klonowskas MM (1994). *J Biol Chem* **269**, 31359-31371.

B R, Travers AA, Dunn JJ and Bautz EK (1969).*Nature* **221**, 43–46.

Campbe. and Darst SA (2000). *J Mol Biol* **300**,17-28.

Campbell EA, Masuda S, Sun JL, Muzzin O, Olson CA, Wang S and Darst SA (2002). *Cell* **108,** 795–807.

Campbell FXV, Tarassov V, and Michnick SW (2005). *Proc Natl Acad Sci USA* **102,** 14988–14993.

Cappelli G, Volpe E, Grassi M, Liseo B, Colizzi V and Mariani F (2006). *Res Microbiol* **157,** 445–455.

Camus JC, Pryor MJ, Medigue C and Cole ST (2002). *Microbiology* **148,** 2967-73.

Chatterji D and Ojha AK (2001). *Curr Opin Microbiol* **4,** 160–165.

Chayan NE, Radcliffe JW and Blow DM (1993) *Protein Sci* **2,** 113-18.

Chien-Cheng C, Richard JL, Robin H, Michael DY and Olivier D (2003). *Mol Microbiol* **49,** 1657-1669.

Claudia LP, Gavin DM, Averell LG, Grant JJ, Wei-hau C, Jianhua Fu and Roger DK (1999). *Cell* **98,** 799–810.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV. Eiglmeier K, Gas S, Barry CE III, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin S, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S and Barrell BG (1998). *Nature* **393,** 537-544.

Cudney R, Patel S, Weisgraber K, Newhouse Y and McPherson A (1994). *Acta Cryst* **D50,** 414-23.

Cuff JA, Clamp ME, Siddiqui AS, Finlay M and Barton GJ (1998). *Bioinformatics* **14,** 892-893.

Daley, C. (2003) Tuberculosis latency in humans, in TUBERCULOSIS – 2nd ed, Ed. William N. Rom and Stuart Garay, Chapter 7, p. 85-99, Lippincott Williams & Wilkins, A Wolters Kluwer Company, New York

Dahl JL, Kraus CN, Boshoff HI, Doan B, Foley K, Avarbock D, Kaplan G, Mizrahi V, Rubin H and Barry CE III (2003) *Proc Natl Acad Sci USA* **100,** 10026–10031.

Dainese E, Rodrigue S, Delogu G *et al.* (2006) *Infect Immun* **74,** 2457–2461.

Delumeau O, Lewis RJ and Yudkin MD (2002). *J Bacteriol* **184,** 5583-89.

Delumeau O, Chen CC, Murray JW, Yudkin MD and Lewis RJ (2006). *J Bacteriol* **188,** 7885-92.

DeMaio J, Zhang Y, Ko C, Young DB and Bishai WR (1996). *Proc Natl Acad Sci USA* **93**,2790–2794.

DeMaio J, Zhang Y, Ko C and Bishai WR(1997). *Tuber Lung Dis* **78**, 3–12.

Deretic V and Fratti RA (1999). *Mol Microbiol* **31**,1603–1609.

Dieli F, Troye-Blomberg M, Ivanyi J, Fournie JJ, Krensky AM, Bonneville M, Peyrat MA, Caccamo N, Sireci G and Salerno A (2001).*J Infect Dis* **184**,1082–1085.

Dombroski AJ, Walter WA and Gross CA (1993). *Genes Dev* **7**, 2446–2455.

Dirk S, Sabine E, Martin IV, Yang L, Joseph AM, Irene MM, Gregory D, Brad E, Philip DB, Carl N and Gary KS (2003). *J Exp Med* **198**, 693–704.

Douglas FB and Stephen JWB (2004). *Nat Rev Microbl.***2**,1-9

Dove SL, Darst SA and Hochschild A (2003). *Mol Microbiol* **48**, 863–874.

Dye C, Scheele S, Dolin P, Pathania V and Raviglione MC (1999). *JAMA* **282**, 677-686.

Dufour A and Haldenwang WG (1994). *J Bacteriol* **176**, 1813–1820.

Duncan L and Losick R (1993). *Proc Natl Acad Sci USA* **90**,2325–29.

Duncan L, Alper S and Losick R (1996). *J Mol Biol* **260**, 147–164.

Ebright RH (2000). *J Mol Biol* **304**, 687–698.

Errington J (1996). *Trends Genet* **12**, 31–34.

Gaal T, Bartlett MS, Ross W, Turnbough CL Jr and Gourse RL (1997). *Science* **278**, 2092–2097.

Gardella T, Moyle H and Susskind MM (1989). *J Mol Biol* **206**, 579–590.

Geiman DE, Kaushal D, Ko C et al. (2004). *Infect Immun* **72**,1733–1745.

Geiduschek EP (1991). *Annu Rev Genet* **25**, 437–60.

George PR, Wright LL, Hopkins S and Furman PA (1991). *J Biol Chem* **266**, 19362-19368.

Gillen KL and Hughes KT (1991). *J Bacteriol* **173**, 2301–10.

Gillen KL and Hughes KT (1991). *J Bacteriol* **173**, 6453–59.

Gomez M, Doukhan L, Nair G and Smith I (1998). *Mol Microbiol* **29**, 617–628.

Gordon ST, Spencer RD, Leonard NJ, and WeberG (1970). *J Am Chem Soc* **92,** 687-695.

Gottesman S, Wickner S and Maurizi MR (1997). *Genes Dev* **11,** 815–23.

Graham RS, Lorenz W, Richard S, Joseph AM, Jason H, Ken GL, Douglas BY and Philip DB (2002). *Microbiology*, **148,** 3129–3138.

Greenstein AE, MacGurn JA, Baer CE, Falick AM, Cox JS, and Alber T (2007). *PLoS Pathog* **3(4),**e49. doi:10.1371/journal.ppat.0030049.

Gross C (1998). in *Transcription. Cold Spring HarborSymposia on Quantitative Biology, LXIII.* 141–155 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Grosset J, Truffot-Pernot C, Lacroix C and Ji B (1992). *Antimicrob Agents Chemother* **36,** 548-551.

Grossman AD, Erickson JW and Gross CA (1984). *Cell* **38,** 383–90.

Grossman AD, Straus DB, Walter WA and Gross CA (1987). *Genes Dev* **1,** 179– 84.

Gruber TM and Gross CA (2003). *Annu Rev Microbiol* **57,** 441–466.

Gruber TM, Markov D, Sharp MM, Young BA, Lu CZ, Zhong HJ, Artsimovitch I, Geszvain KM, Arthur TM and Burgess RR (2001). *Mol Cell* **8,** 21–31.

Guild N, Gayle M, Sweeney R, Hollingsworth T, Modeer T and Gold L (1988). *J Mol Biol* **199,** 241–58

Hahn MY, Raman S, Anaya M and Husson RN (2005). *J Bacteriol* **187,** 7062–7071.

Haldenwang WG (1995). *Microbiol Rev* **59,**1–30.

Hecker M and Volcker U (2001). *Adv Micob Physiol* **44,**35-91

Helmann JD (1999). *Curr Opin Microbiol* **2,** 135–141.

Hendrick JP and Hartl FU (1993). *Annu Rev Biochem* **62,** 349–84.

Hengge R and Bukau B (2003). *Mol Microbiol* **49,** 1451–1462.

Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994). Clustal W: *Nucleic Acids Res* **22,** 4673-4680.

Hinton DM, March-Amegadzie R, Gerber J and Sharma M (1996). *Methods Enzymol* **274,** 43–57.

Hong PC, Tsolis RM and Ficht TA (2000). *Infect Immun* **68,** 4102-4107.

Hu Y, Butcher PD, Sole KM, Mitchison DA and Coates AR (1998). *FEMS Microbiol Lett* **158,** 139-145.

Hu Y and Coates AR (1999). *J Bacteriol* **181,** 469–476.

Hu Y, Mangan JA, Dhillon J, Sole KM, Mitchison DA, Butcher PD and Coates AR (2000). *J Bacteriol* **182,** 6358-6365.

Hu Y and Coates AR (2001). *FEMS Microbiol Lett* **202,** 59–65.

Hu Y, Kendall S, Stoker NG and Coates AR (2004). *FEMS Microbiol Lett* **237,** 415–423.

Hughes KT, Gillen KL, Semon MJ and Karlinsey JE (1993). *Science* **262,**1277–28.

Hughes K and Mathee K (1998). *Annu Rev Microbiol* **52,** 231–286.

Issar S (2003). *Clinl Microbiol Rev* **16,** 463–496

Ishihama A (2000). *Annu Rev Microbiol* **54,** 499–518.

Jancarik J and Kim SH (1991). *J Appl Cryst* **24,** 409-11.

Jon Marles-Wright and Richard JL (2007). *Curr Opin in Str Biol* **17,** 755-760.

Jordi BJAM and Higgins CF (2000). *J Biol Chem* **275,** 12123–12128.

Jie-Oh L, Rieu P, Arnaout MA and Liddington R (1995). *Cell* **80,** 631-638.

Kaneko T, Tanaka N and Kumasaka T (2005). *Prot Sci* **14,**558-565.

Kang CM, Brody MS, Akbar S, Yang X and Price CW (1996). *J Bacteriol* **178,** 3846–53.

Karzai AW and McMacken R (1996). *J Biol Chem* **271,** 11236–46.

Kell DB, Kaprelyants AS, Weichart DH, Harwood CL and Barer MR (1998). *Antonie Van Leeuwenhoek.* **73,** 169-187.

Kell DB and Young M (2000). *Curr Opin Microbiol* **3,** 238-243.

Kelley LA, MacCallum RM and Sternberg MJE (2000). *J Mol Biol* **299,** 499-520.

Kenneth AB, Julie AM and Stephen CM (1995). *Anal Biochem* **231,** 301–308.

Kenton LL, Sarah MG, Peter JS and Zygmunt SD (2001). *Acta Cryst* **D57,** 679- 688.

Kutsukake K (1994). *Mol Gen Genet* **243,** 605–12.

Kutsukake K and Iino T (1994). *J Bacteriol* **176,** 3598–605

Lakowicz JR (1983). *Principle of Fluorescence Spectroscopy,* pp. 319-484, Plenum Press, New York

Laemmli UK (1976). *Nature* **227**, 680-85.

Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, Mcwilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007). *Bioinformatics* **23**, 2947-2948.

Lewis PJ, Magnin T and Errington J (1996). *Genes Cells* **1,** 881–94.

Liberek K, Wall D and Georgopoulos C (1995). *Proc Natl Acad Sci USA* **92,** 6224–28.

Liu X and Matsumura P (1995). *Gene* **164,** 81–84.

Lonetto M, Gribskov M and Gross CA (1992). *J Bacteriol* **174,** 3843–3849.

Lord M, Magnin T and Yudkin MD (1996). *J Bacteriol* **178,** 6730–35.

Maeda H, Fujita N and Ishihama A (2000). *Nucleic Acids Res* **28**, 3497–3503.

Magnin T, Lord M, Errington J and Yudkin MD (1996). *Mol Microbiol* **19,** 901–7.

Magnusson LU, Farewell A and Nystrom T (2005) *Trends Microbiol* **13,** 236–242.

Marchler-Bauer A, Anderson JB, Derbyshire MK, Deweese-Scott C, Gonzales NR, Gwadz M, *et al.,* (2007). *Nucleic Acids Res* **35,** d237-40.

Margaret S Ho, Karen C and Richard Losick (2003). *J Biol Chem* **278**, 20898–20905.

Margareta KS, Soumya SR and Seth AD (2004). *Mol Cell* **14,**127-138.

Margolis P, Driks A and Losick R (1991). *Science* **254,** 562–65

Mazeed S, Ofek G, Belachew A, Huang C, Zhou T and Kwong PD(2003) *Structure* **11**, 1061-70.

McCune R, Lee SH, Deuschle K and McDermott W (1957). *Am Rev Tuberc* **76,** 1106-1109.

McCune RM, Tompsett R and McDermott W (1956). *J Exp Med* **104**, 763-802.

McCune RM, Feldman FM, Lambert H and McDermott W (1966). *J Exp Med* **123,** 445-468.

McDermott W (1958). *The Yale J Biol Med* **30,** 257-291.

McKinney JD, Höner ZBK, Munoz- Elias EJ, Miczak BCA, Chan WT, Swenson D, Sacchettini JC, Jacobs WR Jr and Russell DG (2000). *Nature* **406,** 735-738.

Merelo JJ, Andrade MA, Prieto A and Morán F (1994). *Neurocomputing* **6,** 443-454.

Missiakas D and S Raina (1998). *Mol Microbiol* **28,**1059–1066.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, and Olson AJ (1998). *J Comp Chem* **19,** 1639-62.

Mooney RA, Darst SA and Landick R (2005). *Mol Cell* **20,**335-345.

Mukherjee R, Gomez M, Jayaraman N, Smith I and Chatterji D (2005). *Microbiology* **151,** 2385–2392.

Niranjala MDG, Gretta R, Jason H, Richard AS and Tanya P (2004). *Tuberculosis* **84,** 239–246.

Orme IM (2001). *Int J Tuberc Lung Dis.* **5,** 589-593.

Orsini G, Ouhammouch M, Le Caer JP and Brody EN (1993). *J Bacteriol* **175,** 85–93.

Ohnishi K, Kutsukake K, Suzuki H and Iino T (1990). *Mol Gen Genet* **221,**139–47.

Ohnishi K, Kutsukake K, Suzuki H and Iino T (1992). *Mol Microbiol* **6,** 3149–57

Packschies L, Theyssen H, Buchberger A, Bukau B, Goody RS and Reinstein J (1997). *Biochemistry* **36,** 3417– 22.

Paget MS and Helmann JD (2003). *Genome Biol* **4,** 203.

Panaghie G, Aiyar SE, Bobb KL, Hayward RS and de Haseth PL (2000). *J Mol Biol* **299,**
1217–1230.

Parida BK, Douglas T, Nino C and Dhandayuthapani S (2005). *Tuberculosis* **85,**347-355.

Pérez-Rueda E and Collado-Vides J (2000). *Nucleic Acids Res* **28,** 1838–1847.

Petersen C, Moller LB and Valentin-Hansen P (2002). *J Biol Chem* **277,** 31373–31380.

Picard-Jean F, Bougie I and Bisaillon M (2007). *Biochem J* **407,** 331–341.

Ping C, Rafael ER, Qing Li, Richard FS and William RB (2000). *Infect and Immun* **68,** 5575-5580.

Polissi A, Goffin L and Georgopoulos C (1995). *FEMS Microbiol Rev* **17,** 159–69.

Raivio TL and Silhavy TJ (2001). *Annu Rev Microbiol* **55,** 591–624.

Reid KL and Fink AL (1996). *Cell Stress Chaperon* **1,**127–37.

Rengarajan J, Barry RB and Eric JR (2005). *Proc Natl Acad Sci USA.* **102,** 8327–8332.

Renshaw PS, Panagiotidou P, Whelan A, Stephen VG, Glyn HR, Richard AW and Carr DM (2002). *J Biol Chem* **277,** 21598–21603.

Renu G, Ignacio JJ, Nandhini R, Ashish, Shobana KA, Venetta T, Mercedes R, Joanna KK, Erol F, Christopher MY and Juan A (2006). *The Journal of Immunology* **177,** 6579–6583.

Rhen M, Eriksson S, Clements M, Bergstrom S and Normark J (2003). *Trends Microbiol* **11,** 80-86.

Riccardo M, Roberta P, Sebastien R, Jocelyn B, Luc G and Issar Smith (2004). *J Bacteriol* **186,** 895–902

Salgado H, Santos-Zavelata A, Gama-Castro S, Millan-Zarate D, Diaz-Peredo E, Sanchez-Solano F, Perez-Reuda E, Bonavides-Martinez C and Collado-Vides J (2001). *Nucleic Acids Res* **29,** 72–74.

Sambrook J, Fritsch EF and Maniatis T (1989). Molecular Cloning, A laboratory Manual, *IInd Ed.* Cold Spring Harbor Laboratory Press, USA.

Sanderson A, Mitchell JE, Minchin SD and Busby SJW (2003). *FEBS Lett* **544,** 199–205.

Scanga CA, Mohan VP, Joseph H, Yu K, Chan J and Flynn FL (1999). *Infect Immun* **67,** 4531-4538.

Schnetz K (1995). *EMBO J* **14,** 2545–50.

Schneider DA, Gaal T and Gourse RL (2002). *Proc Natl Acad Sci USA* **99,** 8602–8607.

Sebastien R, Roberta P, Pierre-Etienne J, Luc G and Riccardo (2006). *FEMS Microbiol Rev* **30,** 926–941.

Sergei R and Sun PD (2002). *J Appl Cryst* **35,** 674-676.

Sergei R, Sean Li and Sun PD (2006). *Acta Cryst* **D62,** 605-612.

Sharp MM, Chan CL, Lu CZ, Marr MT, Nechaev S, Merritt EW, Severinov K, Roberts JW and Gross CA (1999). *Genes Dev* **13,** 3015–3026.

Shiping Wu, Susan TH, David LL, Andre K, Buka S, Hassan S, Veronica G, Benjamin W, Homayoun S, Randall JB, Ian MO and Peter FB (2004). *Mol Microbiol* **51,**1551–1562.

Shoko M, Katsuhiko SM, Sheng W, Anders OC, Jill D, Fred L, Darst SA and Campbell EA(2004). *J Mol Biol* **340,** 941–956.

Siegele DA, Hu JC, Walter WA and Gross CA (1989). *J Mol Biol* **206,** 591–603.

Song T, Dove SL, Lee KH and Husson RN (2003). *Mol Microbiol* **50,** 949–959.

Stern O and Volmer M (1919). *Physik.* **20,** 183-188.

Stevens A (1972). *Proc Natl Acad Sci USA* **69,** 603–7.

Stevens A (1973). *Biochem Biophys Res Commun* **54,** 488–93.

Stevens A (1976). In *RNA Polymerase*, ed. R Losick, MChamberlin. Cold Spring Harbor,NY: Cold Spring Harbor Lab.

Stock AM, Robinson VL and Goudreau PN (2000). *Annu Rev Biochem* **69,** 183–215.

Straus DB, Walter WA and Gross CA (1987). *Nature* **329,** 348–51.

Stragier P and Losick R (1996). *Annu Rev Genet* **30,** 297–341.

Sun R, Converse PJ, Ko C, Tyagi S, Morrison NE and Bishai WR (2004). *Mol Microbiol* **52,** 25–38.

Sung-Jin K and Ronald KC (1997). *Biochem and Biophy Res Comm* **234,** 681–685.

Sun Z and Zhang Y (1999). *J Bacteriol* **181,** 7626-7628.

Taeksun S, Simon LD, Kon Ho L and Robert NH (2003). *Mol Microbiol* **50,** 949–59.

Terese B (2003). *J Struc Biol* **142,** 66–76.

Thaller C, Weaver LH, Eichele E, Wilson E, Karlsson R and Jansonics JN (1981). *J Mol Biol* **147,** 465-69.

Thaller C, Weaver LH, Eichele E, Wilson E, Karlsson R and Jansonics JN (1985). *Meth in Enzymol* **114,** 132-135.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997). *Nucleic Acids Res* **24,** 4876-4882.

Theresa MM, Ko C and Bishai WR. (1999) *Antimicrob Agents and Chemo* 43, 218-225.

Tomsic M, Tsujikawa L, Panaghie G, Wang Y, Azok J and deHaseth PL (2001). *J Biol Chem* **276,** 31891–31896.

Tsujikawa L, Tsodikov OV and deHaseth PL (2002). *Proc Natl Acad Sci USA* **99,** 3493–3498.

Valdar WSJ (2002). *Proteins: Structure, Function, and Genetics* **43,** 227-241.

Voelker U, Dufour A and Haldenwang WG (1995). *J Bacteriol* **177**,114–22.

Voelker U,Voelker A and Haldenwang WG (1996). *J Bacteriol* **178,** 5456–63.

Voelker U,Voelker A and Haldenwang WG (1996). *J Bacteriol* **178,** 7020–23.

Wayne LG (1976). *Am Rev Respir Dis* **114,** 807- 811

Wayne LG and Lin Y (1982). *Infect Immun* **37**,1042-1049.

Wayne LG and Hayes LG (1996). *Infect Immun* **64,** 2062-2069.

Wayne LG and Hayes LG (1998). *Tuberc Lung Dis* **79,** 127-132.

Wayne LG and Sohaskey C (2001). *Annu Rev Microbiol* **55**,139–163.

Wise AA and Price CW (1995). *J Bacteriol* **177**,123–33.

Wosten MMSM (1998). *FEMS Microbiol Rev* **22,** 127–150.

Yamanaka K, Hwang J and Inouye M (2000). *J Bacteriol* **182,** 7078–7082.

Yang X, Kang CM, Brody MS, Price CW (1996). *Genes Dev* **10,** 2265–75.

Ying Z (2004). *Frontiers in Bioscience* **9,** 1136-1156.

Yukari CM and William RB (2000). *Nature Medicine* **6,** 1327-1329.

Zhang G, Campbell EA, Leonid M, Catherine R, Konstantin S and Darst SA (1999). *Cell* **98,** 811–824.

Zhang Y, Yang Y, Woods A, Cotter RJ and Sun Z (2001). *Biochem Biophys Res Commun* **284,** 542-547.

Zuber P, Healy J, Carter I, Luke H, Cutting S, Moran J, Charles P and Losick R (1989). *J Mol Biol* **206,** 605–614.

ANNEXURE

1. Chemicals & Strains

All the chemicals used in this work were of ultra pure grade and purchased from *Sigma-Aldrich, St. Louis, MO-63103, USA.* Growth media contents were from *HiMedia laboratories, Mumbai, India.* Ortho-phosphoric acid for Bradford reagent was purchased from *Qualigens India Ltd., Mumbai, India.*

Protein purification materials, columns, column materials and molecular biology kits were obtained from *Amersham Biosciences* (Part of *GE* healthcare), *UK.*

All the restriction enzymes were from *New England biolabs (UK) Ltd., Herts, SG4 0TY, UK.* T4 polynucleotide kinase was purchased from *Amersham Biosciences* (Part of *GE* healthcare), *UK.*

γ-^{32}P ATP was purchased routinely from *BRIT* (*Board of radiation and isotope technology*), *BARC, Mumbai, India.*

T7 based expression vectors *pET21d, pET23a,* bacterial strain BL21 (DE3) and other strains to aid cloning and expression were from *Novagen, EMD Biosciences Inc., Madison, WI-53719, USA* while T5 promoter based expression vector *pQE31* was from *Qiagen GmbH, Hilden, Germany.*

Centricons were purchased from *Amicon, Beverly, MA-01915, USA.* Hyperfilm-XP for autoradiograms was purchased from *Amersham Biosciences* (Part of *GE* healthcare), *UK.*

Primers were synthesized from *Sigma-Genosys, The woodlands, TX-77380, USA.* Lyophilized samples were dissolved in TE to make a 100 nM stock.

2. Reagents and Composition

All purification buffers, were properly filtered with 0.22 µm filters and degassed as per the guidelines of the column manufacturers.

A. Luria-Bertani (LB) media and YT Broth

LB: Tryptone 10g, Yeast extract 5g, NaCl 10g in 950 ml of deionized water. The contents are dissolved, made-up and sterilized by autoclaving for 20 min at 15 lb/in^2.

YT: Tryptone 16g, Yeast extract 10g, NaCl 5g in 950 ml of deionized water. The contents are dissolved, made-up and sterilized by autoclaving for 20 min at 15 lb/in^2.

For solid LB, 1.5 % agar is added before autoclaving.

B. Bradford Reagent

For 100 ml	
Coommassie Brilliant Blue G 250	10 mg
85 % Ortho phosphoric acid	10 ml
Absolute ethanol	5 ml

10 mg G-250 was dissolved in 5 ml ethanol and then thoroughly mixed with 10 ml ortho-phosphoric acid. Volume was made up to 100 ml with TDW and solution was filtered through whatman filter no.1 and kept in brown bottles.

C. Tris Acetic acid EDTA (50X)

For 1000 ml	
Tris	242 g
Acetic acid glacial	57.1ml
0.5 M EDTA	100 ml

D. SDS PAGE (12%)

Resolving gel	**5 ml**
H_2O	1.6 ml
30 % Acrylamide	2.0 ml
1.5 M Tris, pH 8.8	1.3 ml
10 % SDS	0.05 ml
10 % APS	0.05 ml
TEMED	0.002 ml

Stacking gel	**2 ml**
H_2O	1.4 ml
30 % Acrylamide	0.33 ml
1.0 M Tris, pH 6.8	0.25 ml
10 % SDS	0.02 ml
10 % APS	0.02 ml
TEMED	0.002 ml

F. PAGE running buffer (1X)

Tris	3.02 g
Glycine	18.8 g
10 % SDS	10.0 ml

G. Native gel loading dye

Sucrose	40 % (w/v)
Bromophenol blue	0.25 % (w/v)

H. Antibiotics Stock

100 mg/ml and 50 mg/ml stock solutions for ampicillin and kanamycin were prepared in TDW and and 5mg/ml tetracycline stock solution was prepared in ethanol and filtered with 0.22 μM syringe filters (*Millipore*, Massachusetts, USA).

I. IPTG & PMSF

1 M stock for isopropyl-beta-D-thiogalactopyranoside (IPTG) was prepared each time in TDW and filtered with 0.22 μM syringe filters and stored at -20°C. 200 mM stock of Phenyl methyl sulphonyl fluoride (PMSF) was made in absolute ethanol and stored at -20°C.